Geosystems Mathematics

Willi Freeden
M. Zuhair Nashed *Editors*

Handbook of Mathematical Geodesy

Functional Analytic and Potential Theoretic Methods

Birkhäuser

Geosystems Mathematics

Birkhäuser

This series provides an ideal frame and forum for the publication of mathematical key technologies and their applications to geo-scientific and geo-related problems. Current understanding of the highly complex system Earth with its interwoven subsystems and interacting physical, chemical, and biological processes is not only driven by scientific interest but also by the growing public concern about the future of our planet, its climate, its environment, and its resources. In this situation mathematics provides concepts, tools, methodology, and structures to characterize, model, and analyze this complexity at various scales. Modern high speed computers are increasingly entering all geo-disciplines. Terrestrial, airborne as well as spaceborne data of higher and higher quality become available. This fact has not only influenced the research in geosciences and geophysics, but also increased relevant mathematical approaches decisively as the quality of available data was improved.

Geosystems Mathematics showcases important contributions and helps to promote the collaboration between mathematics and geo-disciplines. The closely connected series *Lecture Notes in Geosystems Mathemactics and Computing* offers the opportunity to publish small books featuring concise summaries of cutting-edge research, new developments, emerging topics, and practical applications. Also PhD theses may be evaluated, provided that they represent a significant and original scientific advance.

Willi Freeden • M. Zuhair Nashed
Editors

Handbook of Mathematical Geodesy

Functional Analytic and Potential Theoretic Methods

Birkhäuser

Editors
Willi Freeden
Geomathematics Group
TU Kaiserslautern
Kaiserslautern, Germany

M. Zuhair Nashed
Department of Mathematics
University of Central Florida
Orlando, FL, USA

ISSN 2510-1544 ISSN 2510-1552 (electronic)
Geosystems Mathematics
ISBN 978-3-319-57179-9 ISBN 978-3-319-57181-2 (eBook)
https://doi.org/10.1007/978-3-319-57181-2

Library of Congress Control Number: 2018940865

Mathematics Subject Classification (2010): 86A30, 86A20, 31A25, 47A52, 65A20, 60J45

Cover illustration: designed by Prof. Dr. Schreiner, Buchs

Printed on acid-free paper

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Handbook of Mathematical Geodesy
Geosystems Mathematics, v–vi

Contents

Handbook of Mathematical Geodesy
Geosystems Mathematics, vii–viii

Preface

Geodesy, as most other disciplines, spans activities ranging from theoretical to applied border lines. In the twenty-first century, geodesy is strongly influenced by two scenarios: First, the technological progress, in particular, space observation has opened fundamentally new methods of measurements. Second, high speed computers have led to a strong "mathematization". As a consequence, geodesy is in great shape. However, the width and depth of new geodetic challenges will simultaneously require basic analysis and understanding of all technologically as well as mathematically driven components. These requirements are inextricably necessary to provide future improvements in diverse fields of geodetically involved public concern for our planet such as climate environment, expected shortage of natural resources, etc.

This "Handbook of Mathematical Geodesy" deals with mathematics as the key technology for modeling purposes and analysis of today's geodetic measurements and observations. It supplies deep modern and cutting-edge mathematical knowledge as transfer methodology from the reality space of measurements to the model space of mathematical structures and solutions, and vice versa. Essential interest is laid in studying the gravitational field usually in macroscopic sense, where the quantum behavior of gravitation may not be taken in account. Moreover, in geodetically reflected Earth's gravity work, velocities that are encountered are considerably smaller than the speed of the light. As a consequence, Newtonian physics can be safely used.

In detail, this Handbook is concerned with the following selection of topical areas:

- functional analysis and geodetic functional models
- constructive polynomial, spline and wavelet approximations
- mathematical treatment of geodetic observables and multiscale integrated concepts
- geodetic boundary value problems and oblique stochastic derivative problems
- Runge–Walsh mono-and multi-pole expansions on geodetic reference surfaces such as sphere, ellipsoid, telluroid, geoid, real Earth's surface
- regularization methods of ill-posed and inverse problems
- gravimetric and gradiometric (multiscale) modelling.

The objective of the handbook is twofold: on the one hand it serves as a self-consistent collection of newsworthy material at the graduate-student level for

all members of the mathematical community interested in any of the diverse problems relevant in today's geodesy. On the other hand, the book represents a valuable reference for all geodesists facing innovative modeling supplies involving recently measured datasets in their professional tasks. For both groups the Handbook provides important perspectives and challenges in crossing the traditional frontiers.

The Handbook consolidates the current knowledge by providing summaries and concepts as a guide for geodetic transfer from reality space ("measurements") to virtuality space ("models"). All in all, the work is an authoritative forum offering appropriate mathematical means of assimilating, assessing, and reducing to comprehensible form the flow of measured data and providing the methological basis for scientific interpretation, classification testing of concepts, modeling, and solution of problems.

The editors wish to express their particular gratitude to the people who not only made this handbook possible, but also made it extremely satisfactory:

- The *contributors to the handbook*, who dedicated much time, effort, and creative energy to the project. The handbook evolved continuously throughout the recruitment period, as more and more facets became apparent, many aspects were entirely new at the time of recruitment.
- The *folks at Birkhäuser*, particularly Clemens Heine, who initiated the whole work and gave a lot of encouragement and advice.
- *Helga Nutz*, Geomathematics Group of the University of Kaiserslautern, for reading most of the proofs and giving valuable comments.

Thank you very much for all exceptional efforts and support in creating a work offering exciting discoveries and impressive progress. We hope that the "Handbook of Mathematical Geodesy" will stimulate and inspire new research achievements in geodesy as well as mathematics.

February 2017

Willi Freeden, Kaiserslautern
M. Zuhair Nashed, Orlando

Handbook of Mathematical Geodesy
Geosystems Mathematics, ix–xiv

Introduction

Willi Freeden

In natural extension to the classical definition due to F.R. Helmert [2], *geodesy is the science that deals with the measurement and modeling of the Earth, including its gravity field*. So, the basis of geodetic science is its measurements, i.e., scalar numbers, vectors, tensors such as distances, angles, directions, velocities, accelerations. In this respect, the relevance of the gravity field manifests itself in twofold sense: from the need to handle heights and from the determination of the Earth's shape. Consequently, geodesy realizes a physical rather than a geometrical understanding of height by observing that a point is higher than another if water flows from the first to the second. In other words, "geometric" obligations do not allow to be separated from physical ones. The gravity field is still present, as the driving force.

Nowadays, geodesy as a measuring discipline is in great shape. In fact, computer facilities as well as measurement and observation methods open new research areas and opportunities. However, it is geodetic trademark to present measured values always together with a suitable modeling procedure for interpretation and an appropriate knowledge and estimation about reliability and accuracy. Following R. Rummel [6], this diligence demonstrates the geodesists role as notary of the Earth. As an evident consequence, however, this notarial role explains that geodesy is more than a discipline concerned only with measurements. Inherently, mathematics is implied as key technology bridging the real world of measurements and the virtual world of handling datasets, modeling geodetic quantities and processes, and providing illustrations and interpretations. Once more, the result of measurements are numbers, vectors, tensors, i.e., raw material. Mathematical handling and approximation of datasets as well as modeling techniques are necessary to connect the "reality space" with the "virtuality space". In this sense, a model represents the result of the transfer, it intends to be an image of the reality, expressed in mathematical language, so that an interaction between abstraction and concretization is involved. The mathematic's world of numbers and structures contains efficient tokens by which we are able to describe the rule-like aspect of a real problem. This description includes a simplification by abstraction: essential properties of, e.g., a certain geodetic problem are separated from unimportant ones and a solution scheme is set up. The "eye for similarities" enables mathematicians to recognize a

posteriori that resulting solutions become applicable to multiple cases not only in geodesy but also in other scientific disciplines after an appropriate adaptation.

Summarizing we are led to the following conclusion: *Mathematical Geodesy* is characterized by a twofold interaction. An input problem from reality space ("measurements") reduced by abstraction and transferred into virtuality space results in a mathematical output model which following a "circuit" (cf. Figure 1) becomes a new "concrete" input problem in reality space (usually in geodesy, but possibly also in other sciences).

As a consequence, the ideal process (circuit) for the solution process of geodetic problems (as proposed recently by R. Rummel [6]) canonically shows the following steps to be handled (see also the approach sketched in [1]):

- *Transfer from Reality to Virtuality Space:* Measurements and observational events in reality space lead to mathematical tokens and quantities as "row material" for modeling and processing in virtuality space.
- *Mathematical Modeling:* The observational input is translated into the language of the virtuality space, i.e., mathematics, requiring close cooperation between application-oriented and mathematical geodesists.
- *Development of a Mathematical Solution Method:* Appropriate analytic, algebraic, statistic, stochastic, and/or numeric methods must be taken into account; if necessary, new solution techniques must be proposed.
- *Retransfer from Virtuality to Reality Space:* The mathematical model is validated, the aim is a good accordance of model and measurement. If necessary, on the basis of new measurements, the model must be improved by use of modified "raw material".

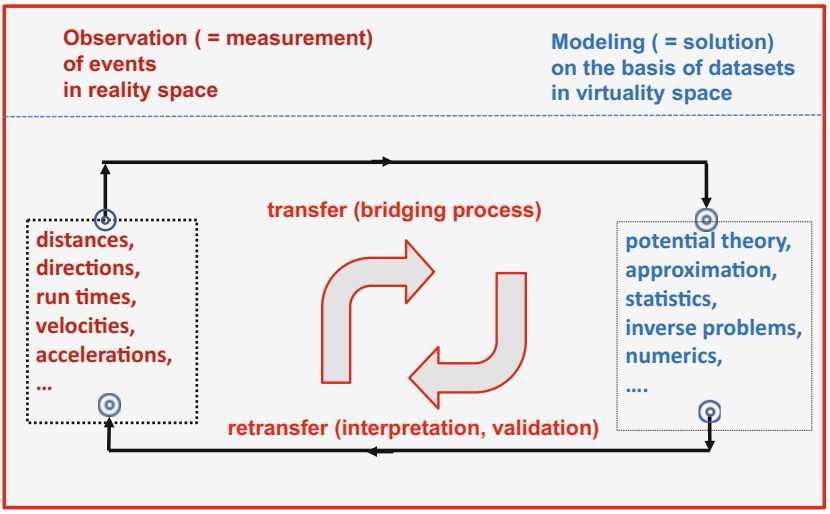


FIGURE 1. The circuit.

Usually, the circuit must be applied several times in an iterative way in order to get sufficient insight into the geodetic system. Obviously, the benefit of a circuit is a better, faster, cheaper, and more secure problem solution on the basis of the mentioned processes of modeling, simulation, visualization, and reduction, decorrelation and denoising of large amounts of data. The more measurements are available, the more one recognizes the causality between abstraction by mathematical concepts and their impact and cross-sectional importance to reality.

Evidently, the circuit in its ideal manifestation (as illustrated in Figure 1) has to follow an obligatory line, namely to provide an appropriate platform within which mathematically/geodetically interrelated features can be better motivated and understood, thereby canonically leading to an interdisciplinary palette of solution procedures in geodetic areas of application. In accordance with this intention, criteria must be found relative to which the scope and limitations of the various methods can be assessed. This is important both in theory and practice since there generally is no cure-all method for most of geodetic problems.

The interaction between abstraction and concretization characterizes the history of geodesy and its efforts as an independent science.

The questions, however, are why

- today's geodesists commonly restrict themselves to the reality space ("measurements") with a necessity to accept some "service fundamentals" of the virtuality space,
- today's mathematicians are interested only in rare exceptions in appropriate handling of geodetically relevant obligations including specific model developments.

Following an article about the interconnecting roles of geodesy and mathematics presented by H. Moritz [5], a prominent member of today's geodesy, the actual interrelationship shows a twofold appraisal from history:

- First, H. Moritz [5] states that the old days are gone when Carl Friedrich Gauss himself developed his epoch-making theories inspired by his geodetic concerns. Gone also are the days when Felix Klein (1849–1925), one of the leading mathematicians of his time, called geodesy "that geometrical discipline in which the idea of approximation mathematics has found its clearest and most consequent expression"(see [4], p. 128). Gone are the times when Henri Poincaré (1854–1912) investigated problems of astronomy and geodesy and actively participated in geodetic life. So, we are led to the conclusion that it apparently is the fault of today's mathematicians that they provide mathematics in an increasingly abstract way, without any regard to possible geodetic applications and, so to say in the scheme of Figure 1, out of touch with reality? Moritz' opinion is as follows: "In part, certainly, they are out of reality."
- Second, H. Moritz [5] is deeply convinced that an increasing abstraction is necessary to achieve progress, not only in mathematics, but also in to-

> day's geodesy. What is frequently overlooked by potential geodetic users of mathematical theory is that the modern abstract methods of mathematics, if properly understood, provide an extremely powerful tool for the solution of applied problems which could not be solved otherwise: the more abstract a method is, the more it is sometimes suitable for a concrete problem. Thus, we may also conclude that it apparently is the fault of modern geodesists to be restricted to measurement tasks, without any regard to virtuality space providing valuable mathematical concepts and, so to say, also out of touch with virtuality?

As a consequence, in the sense of Moritz' explications, today's circuits should follow the ideal way – at least to a considerable extent – that was initiated by Gauss as one of history's most influential mathematician and geodesist for an extremely fruitful interdisciplinary exchange. The heritage of Gauss's work has much to offer even these days to build a strong scientific bridge between mathematics and geodesy by the consequent continuation of the interplay between abstraction and concretization.

However, it must be confessed that today's circuits (in the sense as depicted in Figure 1) turn out to be too complex in their transfer demands from reality to virtuality space, and vice versa, as to be handled by only one ingenious geoscientist. In addition, geodetic changes have been accelerated dramatically. A last "tour de force" for a consolidation of a circuit by a single scientist in the aforementioned classical sense probably was the work by L. Hörmander [3] on the "Geodetic Boundary Value Problem", but only a few years later geodetic space observation by GPS made Hörmander's deep model approach unrealistic, since he started from the traditional assumption that the actual Earth's surface was unknown. Nowadays, the appalling résumé is that a large number of geodetic problems in their specific changes and modifications over the last years must be solved simultaneously. Interdisciplinary solutions are urgently required as answer to an increasingly complex geodetic world. In the opinion of the author, the scientific challenge is a "geodetic consortium", in which mathematics should not stay for geodesists in unloved external partnership and geodesy is of high quality problem-attractiveness for applied mathematicians.

Indeed, the leading role of mathematics for obligations in virtuality space must be acknowledged (again) within today's geodesy, so that mathematicians will become more enthusiastic about working on geodetic programs. A "geodetic consortium" reflecting the cross-sectional demands in reality as well as virtuality space is absolutely essential for a sustainable development in the future. No doubt, as this handbook will show, mathematicians can and should be integrated smoothly into the geodetic phalanx instead of restricting geodesists exclusively to measurements and mathematicians to mere service functions. Only a "geodetic consortium" consisting of scientists with equal standing, rights, and research position will be able to promote the significance of geodesy in its responsibility even for society similarly to the Gaussian epoch.

This *"Handbook of Mathematical Geodesy"* (HbMG) aims at providing innovative mathematical instruments in virtuality space in concrete adaptation to recent demands of gravity field reflected geodesy. It presents geodetically relevant tools and techniques from functional analysis, potential theory, constructive approximation, inverse theory, and numerics. A selected list of topics includes geodetically oriented functional analysis, inverse problem strategies, Gauss's understanding of least squares minimization and Nashed's concept of generalized inverse, harmonic reproducing kernel Hilbert space theory, the uncertainty principle in constructive approximation and its consequences for modeling measured datasets, Slepian function calculus, wavelet-based Meissl schemata of geodetic observables, fast spline multi-pole approximation, regularized functional matching pursuit and its variants, mono-pole and spherically oriented multi-pole Runge–Walsh approximation for use of, e.g., spherical instead of ellipsoidal harmonics even on ellipsoids or more complex geodetically relevant surfaces, stochastic geodetic boundary value problems, spectral tree regularization of "downward continuation problems" such as gradiometry, mollifier techniques in gravimetry, etc.

The handbook is meant as a mathematical addendum to the foundations in use within today's virtuality space (cf. Figure 1). The goal of the handbook is twofold:

- to make mathematicians aware of the particular mathematical developments and calamities occurring in modern geodetic concretizations,
- to make geodesists conscious of new tools, means, structures, methods, and procedures for handling recent measurements and observations by mathematical abstraction.

All in all, the handbook is understood as an essential step towards modern manifestations of "geodetic consortia" realizing the cross-sectional demands and requirements of today's circuits in well-balanced interdisciplinary way.

References

[1] Freeden, W.: Geomathematics: Its Role, its Aim, and its Potential. Handbook of Geomathematics, W. Freeden, M.Z. Nashed, and T. Sonar (Eds.), Vol. 1, 2nd Edition, 3–79, Springer, New York, Berlin, Heidelberg, 2015.

[2] Helmert, F.R.: Die Mathematischen und Physikalischen Theorien der Höheren Geodäsie 1. Vol. 1, Teubner, Leipzig, 1880.

[3] Hörmander, L.: The Boundary Problems of Physical Geodesy. The Royal Institute of Technology, Division of Geodesy, Report 9, Stockholm, 1975.

[4] Klein, F.: Elementarmathematik III. Die Grundlagen der Mathematischen Wissenschaften, Band 16, Springer, Berlin 1928.

[5] Moritz, H.: Geodesy and Mathematics. Zeszyty Naukowe Akademii Górniczo-Hutniezej I.M. Stanislawa Staszica, No. 780, Geodezja, 63: 38–43, Krakow, 1981.

[6] Rummel R.: Geodäsie in Zeiten des Wandels – Versuch einer Standortbestimmung. Zeitschrift für Geodäsie, Geoinformation und Landmanagement (ZfV), 139: 211–216, 2014.

Willi Freeden
Geomathematics Group
University of Kaiserslautern
MPI-Gebäude, Paul-Ehrlich-Str. 26
D-67663 Kaiserslautern, Germany
e-mail: freeden@rhrk.uni-kl.de

Handbook of Mathematical Geodesy
Geosystems Mathematics, 1–163

Gauss as Scientific Mediator Between Mathematics and Geodesy from the Past to the Present

Willi Freeden, Thomas Sonar, and Bertold Witte

Abstract. The objective of the paper is to document the pioneer dimension of Gauss's ideas, concepts, and methods in a twofold direction based on selected case examples, namely to demonstrate his mediation function between mathematics and geodesy to explain the historic development over the past centuries from the initial Gaussian ignition to modern characteristics and tendencies.

Keywords. Gauss, geometric number theory, numerical integration, integral theorems and boundary value problems, least squares adjustment.

1. Introduction

1.1. C.F. Gauss (1777–1855): A brief look at his life

Carl Friedrich Gauss, born on 30 April 1777 in Brunswick (Braunschweig), came from a humble background whose parents were only able to enroll him in a basic writing and counting school. His mathematics teacher discovered his exceptional arithmetic skills and became an advocate for the talented student to facilitate the placement in a grammar school. At the age of fourteen the young Gauss was introduced to the Duke of Brunswick who pledged to finance the education of the extraordinarily talented boy. At the age of 15, Gauss had the genial idea to transfer the principles of logarithm tables to the prime number theory. It was the first time in the history of prime number research that prime number probability became object of research for increasing number ranges. However, he was not able to prove his prime number assumption that the number $\pi(x)$ smaller than x behaves asymptomatically as the quotient from x and $\log(x)$. Still this became the starting point for a variety of number theoretical examinations of renowned mathematicians until present. In the year 1795 Gauss started his studies at the University of Göttingen. After a brief period of time he decided to study mathematics. In

1796 Gauss discovered the solution to an old geometric problem, namely the answer to the question "which straight lines, regular polygons can be exclusively constructed by compass and straightedge". The evidence for the constructability of the regular hexadecagon led Gauss to continue groundbreaking research work, the results of which he compiled to a significant number theoretical work already during his study period which, alas, was only published in 1801 because it had to be translated into Latin as it was customary at that time. This opus published under the title "Disquisitiones arithmeticae" (cf. Figure 2.1) assured Gauss the recognition, in particular, amongst the leading French mathematicians of the era (Cauchy, Laplace, d'Alembert, Laguerre). He was attributed to the best mathematicians of his time due to his number theory. The "Disquisitiones arithmeticae" contain many significant research results such as, e.g., the celebrated Fundamental Theorem of Gauss, or the Law of Quadratic Reciprocity of Legendre. The crowning result of his contributions was the complete solution of binomial equations, and a most unexpected achievement in placing the imaginary unit on a firm basis. He actually was the first to use the imaginary symbol "i", giving it the interpretation of a geometric mean (see [27] for more details). After having graduated from Göttingen – upon the Duke's request who continued to grant him financial support – he switched to the University of Helmstedt where he studied under Prof. Pfaff and finished his thesis in 1799 in which he provided the exact evidence of the "Fundamental Theorem of Algebra" (Latin title: "Demonstratio nova theorematis, omnem functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi posse"). In the following time Gauss almost exclusively dealt with strictly mathematic or geometric questions. Applications were secondary during this epoch although he had already been working on the least squares method since 1794 with a multitude of practical studies. The deeper he penetrated into mathematics, the more fully he was persuaded that its true meaning lies in its application to practical life and natural science (cf. [27]). On January 1st, 1801 the astronomer Piazzi from Palermo discovered a small planet named Ceres. He could only track it for a time period of 41 days. Due to the low number of measuring values, the known methods at that time did not allow to calculate the precise orbit of Ceres in order to locate this planetoid again. The interest in the new planet caused Gauss to temporarily lay aside his purely mathematical researches. Now he created applicable methods for orbit determinations. He sought the orbit which would fit the observations as good as possible using an ellipse by applying his method of least squares. In doing so, his way resulted in the solution of two different problems: first, to find an approximate orbit; second, to correct this orbit in such a way that it "satisfies" the observations as well as possible (for more details the reader is referred to [27]). At the turn of the year 1801/1802 von Zach in Gotha and Olbers in Bremen rediscovered the planetoid, its location agreeing exactly with the ephemeris computed by Gauss. The discovery of Ceres introduced him to the world as an astronomer of the highest order (cf. [27]).

Gauss received a lot of praise and recognition for his scientific achievements; in January 1802 the Petersburg Academy of Sciences nominated him Correspond-

FIGURE 1.1. Medal (headside (left), backside (right)) of the 150th anniversary of the death of C.F. Gauss, 2005 (from [122]).

ing Member and in September 1802 he was invited to teach at the Academy which he declined as other offers later. In 1807 he accepted the call to Göttingen to become a professor for astronomy and the director of the observatory. In the years to follow Gauss mainly worked on astronomic topics but still on the theory of numbers. In the meantime he refocused his research works on application-related topics of mathematics which he found in astronomy as well as in geodesy. From 1818 until approximately 1830 geodesy became the center of his activities. Gauss made many geodetic observations during the arc measurements and the land surveying (1821–1844) in the Kingdom of Hanover in which he personally participated. As a matter of fact, Gauss alone managed the comprehensive calculations for this surveying project. Geodesy derived large direct and indirect profits from this practical activity which, for instance, was lamented by Bessel because of the entailing large time commitment and physical strain. During these works of arc measurements and geodetic surveying Gauss also proved to be a gifted practitioner: Particular mention should be made to the heliotrope (cf. Figure 6.5) invented by him which significantly facilitated and accelerated surveying works. The fundamental studies in classical differential geometry, in particular on the theory of surfaces, and his contributions to potential theory and the further development of the least squares method can be attributed to his practical activity as geodesist and astronomer. The fundamental surface studies, the so-called "Disquisitiones generales circa superficies curvas" (Gauss's Works, Vol. IV), due to their content, cannot be allocated to the actual geodetic works (see [141]) but more to the mathematical fundus. However, they are closely related to geodesy. The significance for geodesy constitutes in the detailed elaboration on the theory of geodetic lines for which preparatory studies had been done by Bernoulli, Euler, and Clairaut, but not with the strong comprehensiveness and thoroughness Gauss applied. The "Disquisitiones" laid the grounds for suitable arithmetic formulas and coordinate systems on the geodetically relevant rotation ellipsoid. Gauss published practical applications for geodetic questions in his "Studies on Topics of Higher Geodesy" in

1843 and 1846 (see Gauss's Works, Vol. IV). Their importance for geodesy can be concisely expressed by the following words: Gauss is the founder of higher geodesy.

Since the collaboration with the Göttingen physicist Weber in the year 1831, Gauss published contributions to classical mechanics, Earth's magnetism, geometric optics, and electrodynamics. In mechanics, for example, he developed the principle of least constraint which was named after him and which has the advantage in comparison to other approaches that it can be demonstratively construed. The theory of optical systems of Gauss for the first time shows the exact principles of the passage of one light beam through a coaxial lens system when it is slightly tilted to the axis. The Gauss ocular is still used for autocollimation nowadays, e.g., for specific tasks in engineering surveying. Gauss's magnetic studies are recognized as his most significant contributions to physics. Gauss and Weber jointly developed the first electromagnetic telegraph.

The three areas of geomagnetism in which Gauss made great contributions were those related to the absolute measurement of the field, the analysis in terms of spherical harmonics, and the organization and equipping of magnetic observatories. Because of the insufficiency of observations over the globe at the time he worked, many of the investigations which he proposed on the basis of the spherical harmonic analysis had in spite of their tremendous significance to await later workers. These include the quantitative separation of internal and external sources, the effect of the Earth's ellipsoidal shape, the possible non-vanishing of the constant term in the expression for the potential and the possible existence of a non-potential portion of the field (see, e.g., [70, 179] and the references therein).

Following W.K. Bühler [17], Gauss did not venture deeply into electrodynamics. In this field he is an outsider fascinated by interesting ideas and phenomena. His legacy contains several interesting recordings, however, no reasonably rounded theory. In fragments regarding the nature of the electromagnetic field, Gauss attempts to describe the theory of long-distance effects – a theory which was then further developed by Weber and Neumann and finally superseded by Maxwell's theory of electromagnetism.

Measured by his rich mathematic knowledge, Gauss published a relatively small number of papers. After his death on 23 February 1855, large amount of unpublished mathematic ideas was detected in his legacy, among others, recordings on non-Euclidean geometry. Gauss supposedly asked himself the question in view of the different geometries, which geometry correctly depicts the physical reality. The experience taught that if the dimensions are sufficiently small, the Euclidean geometry applied regarding the measuring precision, based on which Gauss assumed that the Euclidean geometry also applies to the "infinite small", but that deviations occur with larger dimensions. Gauss wished to know whether our space is curved, i.e., non-Euclidean. For this purpose he measured the inner angles in a large triangle (cf. Figure 6.1). More concretely, Gauss took measurements from three mountains in Germany, Hohenhagen, near Göttingen, Brocken in the Harz Mountains and Inselsberg in the Thüringer Wald to the south. The three lines joining these locations form a great triangle, the angle at Hohenhagen

is close to a right angle, so the area of the triangle is close to half the product of the two short sides. Gauss assumed that light propagates along geodesics. During his high-precision measurement campaign he was not able to detect deviations from the Euclidean internal angle amount, implying that the Euclidean geometry can be applied for relatively large distances in the physical space. This result was negative for Gauss which may have been the reason for his decision not to publish his studies on non-Euclidean geometry. It is doubtful whether this interpretation is correct; Gauss presumably knew that potential deviations would only become obvious in triangles of astronomic dimensions (see [240] for more details). At a later time, Riemann, who was inspired by Gauss, created the Riemann geometry, the basis on which Einstein was able to build up his relativity theory which also finds its application, e.g., in satellite geodesy (see also [224]).

In conclusion, it has to be stated that it was extremely fortunate for geodesy that the world-renowned mathematician (*"Princeps Mathematicorum"*), Carl Friedrich Gauss (cf. Figure 1.2), had been so much fascinated by geodesy from early on in his mathematical career to which he dedicated a significant part of this work time over his life. In a letter to the astronomer Olbers in January 1802 he expressed it as follows: *"The most refined geometer and the perfect astronomer these are two separate titles which I highly esteem with all my heart, and which I worship with passionate warmth whenever they are united"* (see [27, 141] for more details).

1.2. Scientific bridge between mathematics and geodesy

Evidently, a contribution concerned with Gauss as scientific mediator between mathematics and geodesy (as intended by this publication) has to follow an obligatory line, namely to provide an appropriate platform within which mathematically/geodetically interrelated features can be better motivated and understood, thereby canonically leading to an interdisciplinary palette of solution procedures in diverse areas of application. In accordance with this intention, criteria must be found relative to which the scope and limitations of the various methods can be assessed. This is important both in theory and practice since there generally is no cure-all method for most of the problems in reality; it is imperative to be able to clarify why a certain method works in some context as well as when not to use that method.

The authors of this work are convinced that Gauss's suggestions and ideas as exemplary work is still helpful to understand the intuitive principles and interrelations that underlie the various methods and procedures to be needed for the solution of problems and desiderata. So, in summary, the present article may be regarded as an attempt to justify this basic assertion.

1.3. Mathematical circuit: abstraction and concretization

What is it exactly that enables mathematicians to build a bridge between geodesy and their discipline? What is exactly that enables the mathematicians to provide the transfer from concrete geodetic measurements and observables to abstract

FIGURE 1.2. Johann Carl Friedrich Gauss, born, 30 April 1777 Brunswick, Duchy of Brunswick-Wolfenbüttel, Holy Roman Empire, died, 23 February 1855 (aged 77) Göttingen, Kingdom of Hanover.

mathematical formalisms and models? What is exactly that enables the mathematicians to bridge the time gap from historic nomenclature to modern notation? Some answers should be given already at this early stage: The mathematics world of numbers and structures contains efficient tokens by which the rule-like aspect of geodetic problems can be described appropriately. In fact, this description includes as essential step a simplification by abstraction. Essential properties of the problem are separated from unimportant ones, further specified, and afterwards included into a solution scheme. The "eye for similarities" often enables mathematicians to recognize a posteriori that an adequately reduced problem may also arise from very different situations in various application areas, so that the resulting solutions may be applicable to multiple cases after an adequate adaptation or concretization. Without this ingredient, the abstraction remains essentially useless. The interaction between abstraction and concretization characterizes the history

of mathematics and its current development as a common language and independent standards in a unified setup. A problem reduced by abstraction has to be considered as a new "concrete" problem to be solved within a general framework, that determines the validity of a possible solution, not only in geodesy, but also in quite different sciences. The more examples and facets one knows, the more one recognizes the causality between the abstractness of mathematical concepts and their impact and cross-sectional significance.

As an immediate consequence, two important questions arise in a work making the attempt to pursue geodetic ideas and concepts from Gaussian time up to resulting mathematical settings in our days.

(i) What can be specified as mathematical structures and settings by a process of abstraction, i.e., a reduction to essential features?
(ii) What can be achieved from the reduced context obtained by abstraction as a new field of scientific interest, e.g., in geodesy or other sciences?

The purpose of this contribution is to discuss these questions for a selected collection of case studies. The interplay between abstraction and concretization will be explained for each of the problems, thereby offering new perspectives for future challenges. Dilemmas and methodologies will be indicated during the resolution process.

1.4. Specific strategies imposed on our work

The purpose of our work is not to describe the entire opus of the "science titan" Gauss in all its facets from the past until today. Such an approach would be doomed to fail in view of the enormous significance and the immense amount of ideas and works of Gauss. The objective of the authors is rather to document the pioneer dimension of Gauss's ideas, concepts, and methods in a twofold direction based on selected case examples, to demonstrate his mediation function between mathematics and geodesy firstly and secondly the historic development over the past centuries from the initial ignition by Gauss to the modern characteristics and tendencies in mathematics and/or geodesy. The authors consciously limited themselves to special topics in which they feel competent to focus on based on their scientific formation. It includes lattice point number theory with its basics for the development of modern sampling methods, Gauss integration with its impacts on modern fast algorithms for global modeling of geodetically and geophysically relevant quantities, the potential theory with the Gauss law as the initial point for geodetic boundary value problems to determine the physically defined Earth's figure and finally the least squares method in its canonic continuation into the issues of present ill-posed and inverse problems, e.g., in terms of pseudodifferential equations. Due to the reference of this contribution to mathematical geodesy, Gauss's contributions to geomagnetics were not reflected here (in this conjunction, reference is made to, e.g., publications of G.D. Garland [70], K. Reich, E. Roussanova [179] and their cross-references). In addition, a difficulty arose with translating Gauss terms into modern mindsets. The mathematic/geodetic language and forms

of expression have developed and changed over the past two centuries. The authors, alas, were not able to convey and suitably illustrate the development process of the notation. For reasons of legibility, a standardized unified notation was selected (except quotes) adjusted to the requirements of modern times as the last link of the term-defining process.

2. From Gaussian circle problem to geosampling

There is no doubt that the theory of numbers was Gauss's favourite subject. In a much quoted dictum, he asserted that

> *"Mathematics is the Queen of the Sciences and the Theory of Numbers is the Queen of Mathematics".*

Moreover, in the introduction to "Eisenstein's Mathematische Abhandlungen" Gauss wrote:

> *"The Higher Arithmetic presents us with an inexhaustible storehouse of interesting truths – of truths, too, which are not isolated but stand in the closest relation to one another, and between which, with each successive advance of the science, we continually discover new and sometimes wholly unexpected points of contact. A great part of the theories of Arithmetic derive an additional charm from the peculiarity that we easily arrive by induction at important propositions which have the stamp of simplicity upon them but the demonstration of which lies so deep as not to be discovered until after many fruitless efforts; and even then it is obtained by some tedious and artificial process while the simpler methods of proof long remain hidden from us."*

All this is well illustrated by what is perhaps Gauss's most profound publication, namely his "Disquisitiones arithmeticae"(cf. Figure 2.1). It has been described, quite justifiably, as the "Magna Carta of Number Theory", and the depth and originality of thoughts to be manifested in this work are particularly remarkable considering that they were written when Gauss was only about eighteen years of age. In view of the great impact Gauss had on large areas of modern number theory, anything even approaching a comprehensive representation of their influence seems untenable. It is not surprising that there is a huge amount of literature concerned with Gauss's number theoretical results, and his influence on modern mathematics is enormous. However, the obvious problem for our purposes here is the question if the "Queen of Mathematics" is actually able to show impacts to modern mathematical geodesy.

2.1. Lattice points inside circles

We start our bridge from Gaussian concepts of number theory to modern geodetically relevant sampling with a recapitulation of some results on the number of lattice points inside circles

$$\mathbb{S}^1_N = \{x \in \mathbb{R}^2 : |x| = N\} \tag{2.1}$$

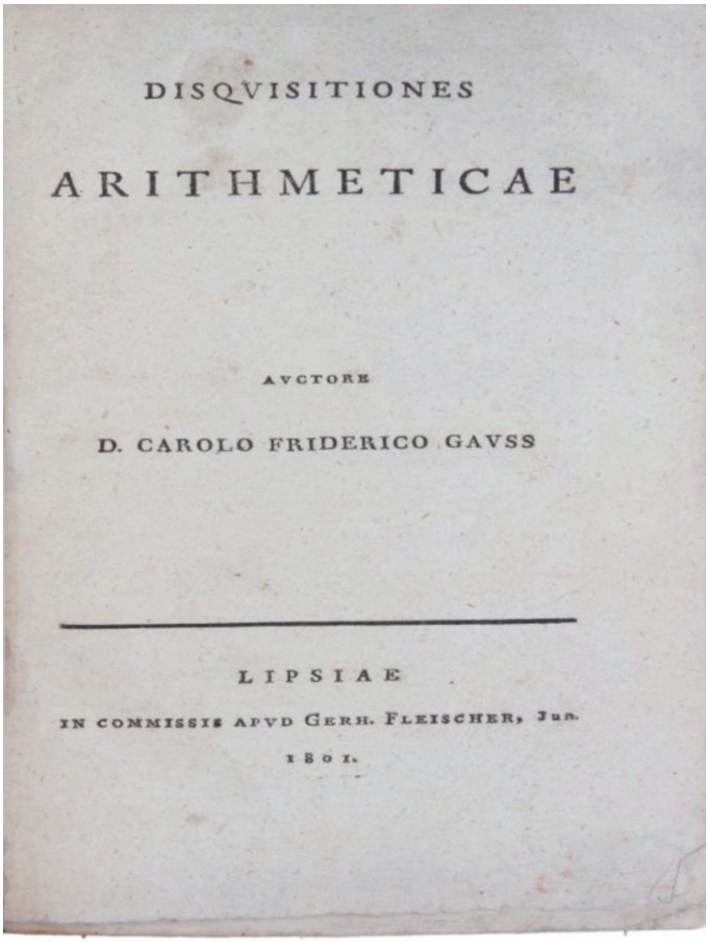


FIGURE 2.1. Title-page of **"Disquisitiones arithmeticae"**, original latin edition by Gerhard Fleischer, Lipsiae (Leipzig) 1801 (668 pages), first reprint as first issue of "Gesamtausgabe Carl Friedrich Gauss: Werke. Band 1, Dieterich, Göttingen 1863", reprint by Springer-Verlag, New York Heidelberg 1986, ISBN 0-387 96254-9 (English translation by Arthur A. Clarke, 1986, in revised form by William C. Waterhouse).

of radii $N > \frac{\sqrt{2}}{2}$ around the origin 0; more accurately, we deal with closed disks

$$\overline{\mathbb{B}_N^2} = \{x \in \mathbb{R}^2 : |x| \leq N\} \tag{2.2}$$

of radii $N > \frac{\sqrt{2}}{2}$ (for more background material and deeper number theoretical concepts the reader is referred, e.g., to the monographs [47, 66]).

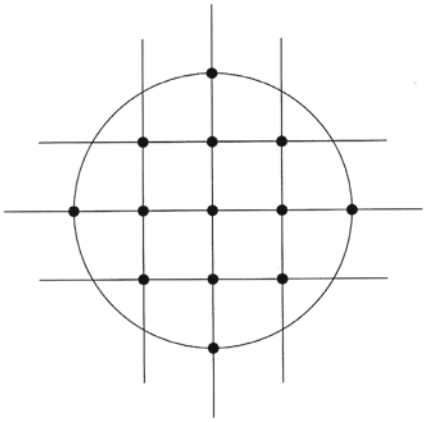


FIGURE 2.2. Lattice points inside a circle.

The problem of determining the total number of lattice points of $\mathbb{Z}^2$ inside and on a circle with radius N, i.e., the determination of the quantity

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}_N^2}\right) = \sharp\left\{(n_1,n_2)^T \in \mathbb{Z}^2 : n_1^2+n_2^2 \leq N^2\right\} \tag{2.3}$$

reaches back to L. Euler [33]. In today's nomenclature it can be equivalently expressed as a sum in the form

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}_N^2}\right) = \sum_{\substack{n_1^2+n_2^2\leq N^2;\\ (n_1,n_2)^T\in\mathbb{Z}^2}} 1. \tag{2.4}$$

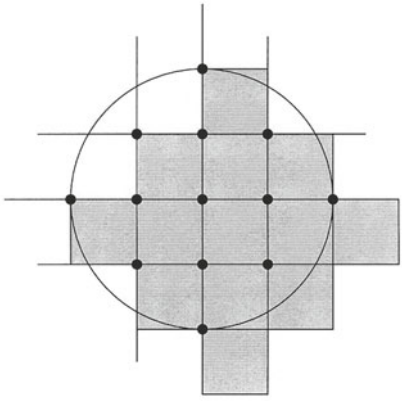


FIGURE 2.3. The polyhedral set $\mathbb{P}_N^2$.

Gauss [1801] found a simple, but efficient method for its estimation (cf. Figures 2.2 and 2.3): associate to every square the Northwest edge as lattice point. The union of all squares with lattice points inside $\overline{\mathbb{B}_N^2}$ defines a polyhedral set $\mathbb{P}_N^2$ with area

$$\|\mathbb{P}_N^2\| = \sharp_{\mathbb{Z}^2}(\overline{\mathbb{B}_N^2}) \tag{2.5}$$

(cf. Figure 2.3). Since the diagonal of each square is $\sqrt{2}$, the geometry of Figure 2.3 tells us that

$$\pi\left(N - \frac{\sqrt{2}}{2}\right)^2 \leq \sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}_N^2}\right) \leq \pi\left(N + \frac{\sqrt{2}}{2}\right)^2. \tag{2.6}$$

Therefore, $\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}_N^2}\right) - \pi N^2$ after division by N is bounded for $N \to \infty$, which is usually written with Landau's O-symbol as

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}_N^2}\right) = \pi N^2 + O(N). \tag{2.7}$$

In other words, the number of lattice points in $\overline{\mathbb{B}_N^2}$ is equal to the area of that circle plus a remainder of the order of the boundary. In particular,

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}_N^2}\right) \sim \pi N^2 \tag{2.8}$$

so that a method of determining the irrational, transcendent number π becomes obvious (for alternative approaches to π within the history of analysis the reader is referred to [204]):

$$\lim_{N\to\infty} \frac{\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}_N^2}\right)}{N^2} = \pi. \tag{2.9}$$

C.F. Gauss [71] illustrated his result by taking $N^2 = 100\,000$. In this case he calculated

$$\sum_{\substack{|g|^2 \leq 100\,000;\\ g\in\mathbb{Z}^2}} 1 = 314\ 197. \tag{2.10}$$

This calculation determines the number π up to three decimals after the comma.

2.2. Circle problem and Hardy's conjecture

The formula (2.7) due to C.F. Gauss [71] allows the following representation in the nomenclature of Landau's O-symbols:

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}_N^2}\right) = \pi N^2 + O(N). \tag{2.11}$$

The so-called *circle problem* is concerned with the question of determining the bound

$$\alpha_2 = \inf\left\{\gamma : \sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}^2_{\sqrt{N}}}\right) = \pi N + O(N^\gamma)\right\}. \tag{2.12}$$

Until now, we knew from (2.11) that $\alpha_2 \leq \frac{1}{2}$. An improvement of the Gaussian result, however, turns out to be very laborious, in fact, requiring a great effort.

A first remarkable result is due to W. Sierpinski [201], who proved by use of a method of his teacher G. Voronoi [220] that

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}^2_{\sqrt{N}}}\right) = \pi N + O\left(N^{\frac{1}{3}}\right), \tag{2.13}$$

i.e., $\alpha_2 \leq \frac{1}{3}$. The proof of Sierpinski is elementary (see, e.g., [47, 66] for more details); it is a link between geometry and number theory.

By use of advanced methods on exponential sums (based on the work by, e.g., H. Weyl [228], H.R. Chen [18], and many others) the estimate $\frac{1}{3}$ could be strengthened to some extent. It culminated in the publication by G. Kolesnik [128], who had as his sharpest result with these techniques

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}^2_{\sqrt{N}}}\right) - \pi N = O\left(N^{\frac{139}{429}}\right). \tag{2.14}$$

M.N. Huxley [117] devised a substantially new approach (not discussed here); his strongest result was the estimate

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}^2_{\sqrt{N}}}\right) - \pi N = O\left(N^{\frac{131}{416}}\right). \tag{2.15}$$

(note that $\frac{139}{429} = 0.324009\ldots$, while $\frac{131}{416} = 0.315068\ldots$).

Hardy's conjecture claims

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}^2_{\sqrt{N}}}\right) - \pi N = O\left(N^{\frac{1}{4}+\varepsilon}\right) \tag{2.16}$$

for every $\varepsilon > 0$. This conjecture seems to be still a challenge for future work. However, in the year 2007, S. Cappell and J. Shaneson deposited a paper entitled "Some Problems in Number Theory I: The Circle Problem" in the arXiv:math/0702613 claiming to prove the bound of $O(N^{\frac{1}{4}+\varepsilon})$ for $\varepsilon > 0$.

0.250000	Gauss (1801)
0.083333...	G. Voronoi (1903), Sierpinski (1906)
0.080357...	J.E. Littlewood, A. Walfisz (1924)
0.079268...	J.G. van der Corput (1928)
0.074324...	J.-R. Chen (1963)
0.074009...	G. Kolesnik (1985)
0.064903...	M.N. Huxley (2003)

TABLE 1. Incremental improvements for the value ε_2 in the estimate (2.17).

Table 1 lists incomplete incremental improvements for the quantity ε_2 of the upper limit for the circle problem

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}^2_{\sqrt{N}}}\right) - \pi N = O\left(N^{\frac{1}{4}+\varepsilon_2}\right). \tag{2.17}$$

For all recent improvements, the proofs became rather long and made use of some of the more heavy machinery in hard analysis.

Summarizing our results about lattice points inside circles (cf. [47]) we are confronted with the following situation:

$$\frac{1}{4} \;\le\; \alpha_2 \;\le\; \frac{1}{4} + \varepsilon_2 \tag{2.18}$$

and

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}^2_{\sqrt{N}}}\right) - \pi N \neq O\left(N^{\frac{1}{4}}\right), \tag{2.19}$$

$$\sharp_{\mathbb{Z}^2}\left(\overline{\mathbb{B}^2_{\sqrt{N}}}\right) - \pi N = O\left(N^{\frac{1}{4}+\varepsilon_2}\right), \tag{2.20}$$

where $0 < \varepsilon_2 \le \frac{1}{4}$ (for example, Huxley's bound $\varepsilon_2 = 0.064903\ldots$).

2.3. Variants of the circle problem

There are many perspectives to formulate variants of the Gaussian lattice point problem for the circle. It already was the merit of E. Landau [135] to point out particularly interesting areas, such as

- *General two-dimensional lattices*

  $$\Lambda = \{g = ng_1 + mg_2 \;:\; n, m \in \mathbb{Z}\} \tag{2.21}$$

  with $g_1, g_2 \in \mathbb{R}^2$ linearly independent (see Figure 2.4) can be used instead of the unit lattice $\mathbb{Z}^2$.

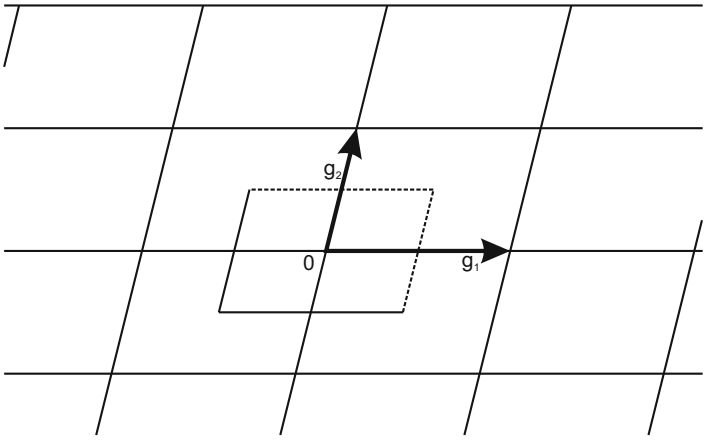


FIGURE 2.4. Two-dimensional lattice Λ generated by $g_1, g_2 \in \mathbb{R}^2$.

- The *remainder term* can be represented as alternating series, called *Hardy–Landau series* in terms of the Bessel function J_1 of order 1 (for the different facets of the proof see [103, 134, 160])

  $$\sum_{\substack{|g|\le N \\ g\in\Lambda}}{}' 1 = \frac{\pi N^2}{\|\mathcal{F}_\Lambda\|} + \lim_{R\to\infty} \frac{\pi N^2}{\|\mathcal{F}_\Lambda\|} \sum_{\substack{0<|h|\le R \\ h\in\Lambda^{-1}}} \frac{J_1(2\pi|h|N)}{\pi|h|N}, \tag{2.22}$$

where Λ is an arbitrary lattice in $\mathbb{R}^2$ and

$$\mathcal{F}_\Lambda = \left\{ x = x_1 g_1 + x_2 g_2 \in \mathbb{R}^2 : -\frac{1}{2} \le x_i < \frac{1}{2},\ i = 1,2 \right\} \tag{2.23}$$

is the *fundamental cell of* $\Lambda \subset \mathbb{R}^2$ with

$$\|\mathcal{F}_\Lambda\| = \sqrt{\det\left((g_i \cdot g_j)_{i,j=1,2}\right)} \tag{2.24}$$

as the area $\|\mathcal{F}_\Lambda\|$ *of* $\mathcal{F}_\Lambda$. Moreover, the following convention

$$\sum_{\substack{|g|\le N \\ g\in\Lambda}}{}' 1 = \sum_{\substack{|g|< N \\ g\in\Lambda}} 1 + \frac{1}{2} \sum_{\substack{|g|= N \\ g\in\Lambda}} 1 \tag{2.25}$$

is used in lattice point theory (note that the last sum only occurs if there is a lattice point $g \in \Lambda$ with $|g| = N$).

- Lattice points can be affected by non-constant weights (see [35])

$$\begin{aligned} &\sum_{\substack{|a+g|\le N \\ g\in\Lambda}}{}' e^{2\pi i y\cdot(a+g)} F(a+g) \\ &= \lim_{R\to\infty} \frac{1}{\|\mathcal{F}_\Lambda\|} \sum_{\substack{|h-y|\le R \\ h\in\Lambda^{-1}}} e^{2\pi i a\cdot h} \int_{\substack{|x|\le N \\ x\in\mathbb{R}^2}} F(x)\ e^{-2\pi i x\cdot(h-y)}\ dx, \end{aligned} \tag{2.26}$$

where dx is the volume element, $a, y \in \mathbb{R}^2$, F is twice continuously differentiable in $\overline{\mathbb{B}_N^2}$, $N > 0$, and the following convention has been used analogously to (2.25)

$$\sum_{\substack{|a+g|\le N \\ g\in\Lambda}}{}' \cdots = \sum_{\substack{|a+g|< N \\ g\in\Lambda}} \cdots + \frac{1}{2} \sum_{\substack{|a+g|= N \\ g\in\Lambda}} \cdots \ . \tag{2.27}$$

Note that, for $F = 1$, this formula leads back to

$$e^{2\pi i a\cdot y} \sum_{\substack{|g+a|\le N \\ g\in\Lambda}}{}' e^{2\pi i g\cdot y} = \lim_{R\to\infty} \frac{\pi N^2}{\|\mathcal{F}_\Lambda\|} \sum_{\substack{|h-y|\le R \\ h\in\Lambda^{-1}}} e^{2\pi i a\cdot h}\ \frac{J_1(2\pi|h-y|N)}{\pi|h-y|N}. \tag{2.28}$$

For $a = y = 0$ we obtain the *classical Hardy–Landau identity*, i.e., the identity

$$\sum_{\substack{|g|\le N \\ g\in\Lambda}}{}' 1 = \lim_{R\to\infty} \frac{\pi N^2}{\|\mathcal{F}_\Lambda\|} \sum_{\substack{|h|\le R \\ h\in\Lambda^{-1}}} \frac{J_1(2\pi|h|N)}{\pi|h|N}. \tag{2.29}$$

holds true. Observe that J_1 satisfies the asymptotic relation $J_1(r) = \frac{r}{2} + \cdots$, so that

$$\sum_{\substack{|g|\le N \\ g\in\Lambda}}{}' 1 = \frac{\pi N^2}{\|\mathcal{F}_\Lambda\|} + \lim_{R\to\infty} \frac{\pi N^2}{\|\mathcal{F}_\Lambda\|} \sum_{\substack{0<|h|\le R \\ h\in\Lambda^{-1}}} \frac{J_1(2\pi|h|N)}{\pi|h|N}. \tag{2.30}$$

- Generalizations to lattices $\Lambda \subset \mathbb{R}^q$ and regular regions $\mathcal{G} \subset \mathbb{R}^q, q \geq 2$, and continuous functions on $\overline{\mathcal{G}} = \mathcal{G} \cup \partial\mathcal{G}$ can be formulated in Gaussian summability (see [47] for the proof and a more detailed study)

$$\begin{aligned}
&{\sum_{\substack{a+g\in\overline{\mathcal{G}}\\ g\in\Lambda}}}' e^{2\pi i y\cdot(a+g)} F(a+g)\\
&= \lim_{\substack{\tau\to 0\\ \tau>0}} \frac{1}{\|\mathcal{F}_\Lambda\|} \sum_{h\in\Lambda^{-1}} e^{-\tau\pi^2 h^2} e^{2\pi i h\cdot a} \int_{\mathcal{G}} F(x) e^{-2\pi i x\cdot(h-y)}\, dx,\ a, y \in \mathbb{R}^q, \quad (2.31)
\end{aligned}$$

where a *regular region* $\mathcal{G}$ *in* $\mathbb{R}^q$ is understood to be an open and connected set $\mathcal{G} \subset \mathbb{R}^q$, $q \geq 2$, for which

(i) its boundary $\partial\mathcal{G}$ constitutes an orientable, piecewise smooth Lipschitzian manifold of dimension $q-1$,
(ii) the origin is contained in $\mathcal{G}$,
(iii) $\mathcal{G}$ divides $\mathbb{R}^q$ into the "inner space" $\mathcal{G}$ and the "outer space" $\mathbb{R}^q\backslash\overline{\mathcal{G}}$, $\overline{\mathcal{G}} = \mathcal{G} \cup \partial\mathcal{G}$.

Clearly,

$$\begin{aligned}
&{\sum_{\substack{a+g\in\overline{\mathcal{G}}\\ g\in\Lambda}}}' e^{2\pi i y\cdot(a+g)} F(a+g)\\
&\quad = \frac{1}{\|\mathcal{F}_\Lambda\|} \int_{\mathcal{G}} F(x) e^{2\pi i x\cdot y}\, dx\\
&\qquad + \lim_{\substack{\tau\to 0\\ \tau>0}} \frac{1}{\|\mathcal{F}_\Lambda\|} \sum_{\substack{0<|h|\leq R\\ h\in\Lambda^{-1}}} e^{-\tau\pi^2 h^2} e^{2\pi i h\cdot a} \int_{\mathcal{G}} F(x) e^{-2\pi i x\cdot(h-y)}\, dx,\ a, y \in \mathbb{R}^q.
\end{aligned} \quad (2.32)$$

The following abbreviation has been used consistently

$${\sum_{\substack{a+g\in\overline{\mathcal{G}}\\ g\in\Lambda}}}' \cdots = \sum_{\substack{a+g\in\mathcal{G}\\ g\in\Lambda}} \cdots + \sum_{\substack{a+g\in\partial\mathcal{G}\\ g\in\Lambda}} \alpha(a+g)\cdots \quad (2.33)$$

with $\alpha(a+g)$ denoting the *solid angle* subtended by $\partial\mathcal{G}$ at $a+g$ (note that, as geoscientifically relevant regular regions, we may choose the interior of the (actual) Earth's body or parts of it, the interior of geoscientifically relevant surfaces such as the geoid, telluroid, etc., but also ball, ellipsoid, cube, polyhedral bodies, etc. are included in accordance with the above definition; in potential-theoretic jargon, the solid angle will be explained by Definition 4.1 in a more detailed way).

2.4. Multivariate Shannon sampling

Let us continue with the observation that, for every $y \in \mathbb{R}^q$, the (for dimensions $q \geq 3$ formally understood) series

$$a \;\mapsto\; \lim_{N\to\infty} \frac{1}{\|\mathcal{F}_\Lambda\|} \sum_{\substack{|h-y|\le N\\ h\in\Lambda^{-1}}} e^{2\pi i a\cdot h} \underbrace{\int_{\mathcal{G}} F(x) e^{-2\pi i x\cdot(h-y)}\, dx}_{F^\wedge_{\mathcal{G}}(h-y)},\ a\in\mathbb{R}^q, \tag{2.34}$$

as well as the finite sum

$$a \;\mapsto\; {\sum_{\substack{a+g\in\overline{\mathcal{G}}\\ g\in\Lambda}}}' e^{2\pi i y\cdot(a+g)} F(a+g),\ a\in\mathbb{R}^q, \tag{2.35}$$

show Λ-periodicity, i.e., as functions of the variable $a\in\mathbb{R}^q$ they are periodic with respect to the lattice $\Lambda\subset\mathbb{R}^q$. As a consequence (see [57] for the details), Shannon-type sampling procedures can be obtained by formal integration of the lattice point identity (2.31) over a regular region $\mathcal{H}$ that is not necessarily equal to $\mathcal{G}$

$$\begin{aligned}
&\int_{\mathcal{H}} {\sum_{\substack{a+g\in\overline{\mathcal{G}}\\ g\in\Lambda}}}' e^{-2\pi i y\cdot(a+g)} F(a+g)\, da\\
&\quad= \sum_{\substack{(\mathcal{F}_\Lambda+\{g'\})\cap\overline{\mathcal{H}}\neq\emptyset\\ g'\in\Lambda}} \underbrace{\int_{\mathcal{G}\cap\bigcup\limits_{g\in\Lambda}(((\overline{\mathcal{H}}\cap(\mathcal{F}_\Lambda+\{g'\}))-\{g'\})+\{g\})} F(x) e^{-2\pi i y\cdot x}\, dx}_{=\,F^\wedge_{\mathcal{G}\cap\bigcup\limits_{g\in\Lambda}(((\overline{\mathcal{H}}\cap(\mathcal{F}_\Lambda+\{g'\}))-\{g'\})+\{g\})}(y)}\\
&\quad= \lim_{\substack{\tau\to 0\\ \tau>0}} \frac{1}{\|\mathcal{F}_\Lambda\|} \sum_{h\in\Lambda^{-1}} e^{-\tau\pi^2 h^2} \underbrace{\int_{\mathcal{G}} F(x) e^{-2\pi i h\cdot x}\, dx}_{=\,F^\wedge_{\mathcal{G}}(h)} \underbrace{\int_{\mathcal{H}} e^{2\pi i a\cdot(h-y)}\, da.}_{=\,K_{\mathcal{H}}(h-y)}
\end{aligned} \tag{2.36}$$

The identity (2.36) has many interesting properties. For example, by virtue of the Gaussian summability, the convergence of the cardinal-type series on the right-hand side of (2.36) may be exponentially accelerated. All manifestations of over- and undersampling can be explicitly analyzed by the finite sum of Fourier transforms on the left side of the identity (see [57]), dependent on the geometric configurations of the chosen regular regions $\mathcal{G},\mathcal{H}$ (note that the identity (2.36) also seems to be unknown for the uni-variate case in this generality).

The Gaussian summability of the cardinal series on the right-hand side of (2.36) is of great importance from numerical point of view; it enables a fast computation of the series. Nonetheless, W. Freeden, M.Z. Nashed [57]) show that the identity (2.36) additionally holds true in ordinary sense, i.e., we have

$$\begin{aligned}
&\int_{\mathcal{H}} {\sum_{\substack{a+g\in\overline{\mathcal{G}}\\ g\in\Lambda}}}' e^{-2\pi i y\cdot(a+g)} F(a+g)\, da\\
&\quad= \sum_{\substack{(\mathcal{F}_\Lambda+\{g'\})\cap\overline{\mathcal{H}}\neq\emptyset\\ g'\in\Lambda}} \underbrace{\int_{\mathcal{G}\cap\bigcup\limits_{g\in\Lambda}(((\overline{\mathcal{H}}\cap(\mathcal{F}_\Lambda+\{g'\}))-\{g'\})+\{g\})} F(x) e^{-2\pi i y\cdot x}\, dx}_{=\,F^\wedge_{\mathcal{G}\cap\bigcup\limits_{g\in\Lambda}(((\overline{\mathcal{H}}\cap(\mathcal{F}_\Lambda+\{g'\}))-\{g'\})+\{g\})}(y)}
\end{aligned}$$

$$= \frac{1}{\|\mathcal{F}_\Lambda\|} \sum_{h\in\Lambda^{-1}} \underbrace{\int_{\mathcal{G}} F(x)e^{-2\pi i h\cdot x}\, dx}_{=\, F^\wedge_{\mathcal{G}}(h)} \underbrace{\int_{\mathcal{H}} e^{2\pi i a\cdot(h-y)}\, da}_{=\, K_{\mathcal{H}}(h-y)}. \tag{2.37}$$

A simple, but significant case of (2.37) in Euclidean space $\mathbb{R}^q$ showing no phenomenon of aliasing is obtained under the special choice $\overline{\mathcal{G}} = \overline{\mathcal{H}} \subset \overline{\mathcal{F}_\Lambda}$ leading to the Shannon-type identity

$$F^\wedge_{\mathcal{G}}(y) \;= \frac{1}{\|\mathcal{F}_\Lambda\|} \sum_{h\in\Lambda^{-1}} F^\wedge_{\mathcal{G}}(h)\;\; K_{\mathcal{G}}(h-y). \tag{2.38}$$

In fact, the identity (2.38) is a multi-variate variant of the Shannon sampling theorem (cf. [198]), but now for (geoscientifically relevant) regions $\mathcal{G}$. The principal impact of Shannon sampling on information theory is that it allows the replacement of a bandlimited signal $F^\wedge_{\mathcal{G}}$ related to $\mathcal{G}$ by a discrete sequence of its samples without loss of any information. Also it specifies the lowest rate, i.e., the Nyquist rate (cf. [57]), that it enables to reproduce the original signal. In other words, Shannon sampling provides the bridge between continuous and discrete versions of a bandlimited function.

The Shannon sampling theorem has many applications in engineering and physics, for example, in signal processing, data transmission, cryptography, constructive approximation such as spectral analysis by Slepian functions (see, e.g., [202, 203]), partial differential equations such as Boltzmann equation (see, e.g., [171] and the references therein), and inverse problems such as the multi-variate discussion of a Fourier transform over a (geodetically relevant) regular region (see, e.g., [164] and the references therein for the uni-variate study of the antenna problem).

Finally, it should be remarked that sampling theory (in the sense of the representation of an analog signal in terms of its samples) has attracted considerable interest in the past three decades. Major advances in the theory and applications of sampling expansion in different function spaces have been made by mathematicians and engineers. There are several journals on signal processing and computational harmonic analysis that publish papers on sampling theory and its applications. The field also sports its own journal "Sampling Theory in Signal and Image Processing" and has its own "professional society". We quote from Wikipedia:

> *"SampTA* (*Sampling Theory and Applications*) *is a biennial interdisciplinary conference for mathematicians, engineers, and applied scientists. The main purpose of SampTA is to exchange recent advances in sampling theory and to explore new trends and directions in related areas of applications. The SampTA conference series began as a small workshop in* 1995 *in Riga, Latvia, but the meetings grew into full-fledged conferences attracting an even mix of mathematicians and engineers as the interest in sampling theory blossomed."*

2.5. Paley–Wiener spline interpolation

Under the aforementioned assumption that $\mathcal{G}$ is a regular region with $\overline{\mathcal{G}} \subset \overline{\mathcal{F}_\Lambda}$, standard Fourier inversion (see, e.g., [161]) guarantees that

$$\begin{aligned} F_{\mathcal{G}}^{\wedge}(y) &= \int_{\mathbb{R}^q} F_{\mathcal{G}}^{\wedge}(x) \left(\int_{\mathcal{G}} e^{2\pi i a\cdot(x-y)}\, da \right) dx \\ &= \int_{\mathbb{R}^q} F_{\mathcal{G}}^{\wedge}(x)\; K_{\overline{\mathcal{G}}}(x-y)\; dx \end{aligned} \tag{2.39}$$

holds true for all $y \in \mathbb{R}^q$, where $\int_{\mathbb{R}^q} \dots$ is understood in the following sense:

$$\int_{\mathbb{R}^q} \dots = \lim_{N\to\infty} \int_{\substack{|x|\le N \\ x\in\mathbb{R}^q}} \dots \quad . \tag{2.40}$$

Hence, the *Paley–Wiener space* $B_{\overline{\mathcal{G}}}$

$$B_{\overline{\mathcal{G}}} = \left\{ y \mapsto \int_{\mathcal{G}} e^{-2\pi i a\cdot y} F(a)\; da,\; y \in \mathbb{R}^q : F \in L^2(\mathcal{G}) \right\}. \tag{2.41}$$

is a reproducing kernel Hilbert space with the uniquely determined kernel

$$K_{\mathcal{G}}(x-y) = \int_{\mathcal{G}} e^{2\pi i a\cdot(x-y)}\; da. \tag{2.42}$$

The reproducing kernel framework enables us to realize minimum-norm-interpolation, i.e., *spline interpolation*, within the Paley–Wiener space $B_{\overline{\mathcal{G}}}$. More concretely, we are interested in finding a "smooth" spline interpolant to $F_{\mathcal{G}}^{\wedge} \in B_{\overline{\mathcal{G}}}$ from a given finite dataset

$$\{F_{\mathcal{G}}^{\wedge}(h) : h \in \Xi,\; \Xi \subset \Lambda^{-1}\}. \tag{2.43}$$

For that purpose, we consider the finite-dimensional space $Spline^{\Xi}_{B_{\overline{\mathcal{G}}}}$ consisting of all *Paley–Wiener spline functions* $\hat{S}$ of the form

$$\hat{S}(y) = \sum_{h'\in\,\Xi} \hat{a}_{h'} \int_{\mathcal{G}} e^{2\pi i a\cdot(h'-y)}\; da, \quad y \in \mathbb{R}^q, \quad \hat{a}_{h'} \in \mathbb{C}. \tag{2.44}$$

It is easy to deduce from the theory of Gramian determinants (see, e.g., [23]) that there is one and only one spline in $Spline^{\Xi}_{B_{\overline{\mathcal{G}}}}$, denoted by $\hat{S}_F$, with coefficients $\hat{a}_h^F$ satisfying the linear equations

$$\sum_{h'\in\,\Xi} \hat{a}_{h'}^F \int_{\mathcal{G}} e^{2\pi i a\cdot(h'-h)} da \;=\; F_{\mathcal{G}}^{\wedge}(h), \quad h \in \Xi. \tag{2.45}$$

$\hat{S}_F$ fulfills the *minimum-norm-property*

$$\left(\int_{\mathcal{G}} \left| \hat{S}_F(y) \right|^2 dy \right)^{1/2} = \min_{\substack{\hat{S}\in B_{\overline{\mathcal{G}}} \\ \hat{S}(h)=F_{\mathcal{G}}^{\wedge}(h) \\ h\in\Xi}} \left(\int_{\mathcal{G}} |\hat{S}(y)|^2\; dy \right)^{1/2}.$$

As a consequence, suppose that the dataset

$$\{F_{\mathcal{G}}^{\wedge}(h) = \int_{\mathcal{G}} e^{-2\pi i a\cdot h} F(a)\ da : h \in \Xi,\ \Xi \subset \Lambda^{-1}\} \tag{2.46}$$

is given such that $\hat{S}_F$ is "close" to $F_{\mathcal{G}}^{\wedge}$, for example, if the nodal width of Ξ is "small enough" (see the convergence theorems presented in [44]). Then, the Fourier inversion formula (see, e.g., [49]) yields the identity

$$\begin{aligned} \alpha(x)\ F(x) &\simeq \int_{\mathbb{R}^q} e^{2\pi i x\cdot y} \hat{S}_F(y) dy \\ &= \sum_{h'\in\Xi} \hat{a}_{h'}^F \int_{\mathbb{R}^q} e^{2\pi i x\cdot y} \int_{\mathcal{G}} e^{2\pi i a\cdot(h'-y)}\ da\ dy \\ &= \sum_{h'\in\Xi} \hat{a}_{h'}^F e^{2\pi i x\cdot h'},\ x \in \mathbb{R}^q, \end{aligned} \tag{2.47}$$

with coefficients $\hat{a}_h^F$ determined from the already known linear (spline) equations (2.45). In other words, our spline interpolation technique approximately solves the multi-variate inversion problem of determining the function $\alpha F \in L^2(\mathcal{G})$ from a finite set of discrete values of $F_{\mathcal{G}}^{\wedge}$ (note that combined spline interpolation/smoothing (as proposed by W. Freeden, B. Witte [62]) can be used instead of spline interpolation if the data are (partially) noisy).

Finally, a generalization of the *multi-variate antenna problem* should be explained in more detail: In antenna theory, the identification problem is the one in which the far-field radiation pattern is known, either exactly or approximately, by means of physical measurements. The problem is to find the aperture distribution producing this given pattern. The synthesis problems are ones in which we are given a desired far-field pattern, specified exactly or partially (e.g., through samples of the far field at a finite number of points or through measurements contaminated by noise), and we wish to determine a source (a constrained or unconstrained aperture distribution) whose far-field radiation pattern approximates the desired pattern in some acceptable manner. In terms of the operator equation,

$$A_{\mathcal{G}} F(y) = \int_{\mathcal{G}} e^{-2\pi i a\cdot y} F(a)\ da = F_{\mathcal{G}}^{\wedge}(y), \tag{2.48}$$

F represents the aperture distribution, G represents the far-field pattern, and $A_{\mathcal{G}}$ the operator which relates these two functions, characterizes the antenna structure.

For the convenience of the reader, we recast the operator-theoretic setting of a finite area source antenna: Consider the problem, appropriately normalized, of a linear aperture on $\mathcal{G}$. Then, the aperture distribution F is related to the far field $F_{\mathcal{G}}^{\wedge} \in B_{\overline{\mathcal{G}}}$ by means of an integral equation (2.48) of the form the pattern being limited to a visible range containing the lattice points $h \in \Xi$, $\Xi \subset \Lambda^{-1}$. In fact, assuming that $F_{\mathcal{G}}^{\wedge}(h)$ is known for the lattice points $h \in \Xi$, $\Xi \subset \Lambda^{-1}$ we are led back to the aforementioned spline problem, and an approximation to the aperture

distribution is known from formula (2.47) to handle, for example, the multi-variate antenna problem.

Summarizing our number theoretical excursion starting from the Gaussian circle problem via the Hardy–Landau lattice point identities we resulted in new geoscientifically relevant Shannon sampling procedures of high practical applicability enabling Paley–Wiener spline interpolation of Fourier transforms from discrete data, regularly located in lattice points.

3. From Gaussian integration to geocubature

Gauss's famous method of approximate integration, almost immediately after its discovery and throughout the whole 19th century, attracted the attention of the leading mathematicians of the time. The story of Gaussian quadrature began with Newton and Cotes. Following W. Gautschi [85], Newton, in 1676, was the first to suggest a rule of approximate integration. Cotes, independently, arrived at similar methods, and brought them into workable form after learning of Newton's ideas. In 1814, Gauss took the work of Newton and Cotes as a point of departure. He combined it with his own knowledge on the hypergeometric series to develop his famous new method of integration. In more detail, if the set of (distinct) nodes in the Newton–Cotes rules vary freely and the weights are computed in accordance with the Newton–Cotes formula, Gauss's problem was, what is the maximum degree of polynomial exactness that can be achieved? How are the nodes to be selected in order to realize this optimal situation? These were questions raised by C.F. Gauss [76], and answered most elegantly by means of his theory of continued fractions associated with hypergeometric series. The continued fraction was well known to Gauss, being a special case of his general continued fraction for ratios of hypergeometric functions (see [74]). Gauss's breakthrough discovery first inspired Jacobi to provide an elegant alternative derivation. Christoffel then significantly generalized the method and subsequently extended it to arbitrary measures of integration. Stieltjes established the legitimacy of the method, by proving its convergence, while Markov endowed it with an error term. Thus, by the end of the 19th century, the Gauss integration method became firmly entrenched in the repertoire of numerical methods of approximation.

In spite of the huge literature about Gauss's rules and its enormous applications there are still new aspects in the context of approximate integration. In this work, we embed the Gaussian numerical integration in the theory of Legendre operators and its inversion by means of mathematical physics, i.e., by use of the theory of Green's functions. This procedure (cf. [36]) leads us to new error terms in adaptation to the integrand under consideration which finally contribute to the convergence of geocubatures in terms of spherical harmonics.

3.1. Gaussian integration revisited

In order to derive our approach to Gaussian quadrature formulas it is necessary to have some basic knowledge on classical Legendre polynomials. We start with

its explicit representation. Then we list some integral formulas involving Legendre operators which turn out to be useful in Gaussian quadrature.

Legendre Polynomials. We begin with the explicit definition of the Legendre polynomial (for more details see, e.g., [50] and the references therein).

Definition 3.1. The function $P_n : [-1,1] \to \mathbb{R},\ n = 0, 1, \ldots$, defined by

$$P_n(t) \;=\; \sum_{s=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^s \frac{(2n-2s)!}{2^n (n-2s)!(n-s)!\, s!}\, t^{n-2s}, \quad t \in [-1,1] \tag{3.1}$$

is called the *Legendre polynomial*.

$P_n : [-1,1] \to \mathbb{R}$ is uniquely determined by the properties:

(i) P_n is a polynomial of degree n on the interval $[-1,1]$,

(ii) $\displaystyle\int_{-1}^{1} P_n(t) P_m(t)\, dt = 0$ for $n \neq m$,

(iii) $P_n(1) = 1$.

This is easily seen from the usual process of orthogonalization. In particular, we have, for $n = 0, \ldots, 4$,

$$P_0(t) = 1, \qquad P_1(t) = t, \qquad P_2(t) = \frac{3}{2}t^2 - \frac{1}{2}, \tag{3.2}$$

$$P_3(t) = \frac{5}{2}t^3 - \frac{3}{2}t, \quad P_4(t) = \frac{35}{8}t^4 - \frac{15}{4}t^2 + \frac{3}{8}. \tag{3.3}$$

A graphical impression of the first Legendre polynomials can be found in Figure 3.1. Furthermore, an easy calculation shows that

$$\int_{-1}^{1} P_n(t) P_m(t)\, dt = \frac{2}{2n+1}\, \delta_{nm}. \tag{3.4}$$

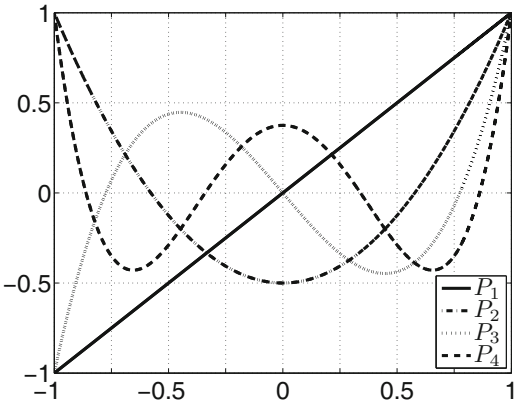


FIGURE 3.1. Legendre polynomials $t \mapsto P_n(t)$, $t \in [-1,1]$, $n = 1, \ldots, 4$.

The Legendre polynomial satisfies the estimate

$$|P_n^{(k)}(t)| \;\le\; |P_n^{(k)}(1)|, \tag{3.5}$$

where

$$P_n^{(k)}(1) = \left(\frac{1}{2}\right)^k \frac{1}{k!}\, n(n+1)\left((n(n+1)-1\cdot 2)\ldots(n(n+1)-k(k-1))\right). \tag{3.6}$$

A straightforward calculation yields

$$\Bigg(\underbrace{(1-t^2)\left(\frac{d}{dt}\right)^2 - 2t\frac{d}{dt}}_{=\mathrm{L}_t} + \underbrace{n(n+1)}_{=\mathrm{L}^\wedge(n)}\Bigg) P_n(t) = 0, \quad t\in[-1,1], \tag{3.7}$$

where L_t is called the *Legendre operator*. We therefore obtain the following lemma.

Lemma 3.1. *The Legendre polynomial P_n is the only twice differentiable eigenfunction of the "Legendre operator" L as defined in (3.7) on the interval $[-1,1]$, corresponding to the eigenvalues $\mathrm{L}^\wedge(n) = n(n+1)$, $n = 0,1,\ldots$, and bounded on $[-1,1]$ with $P_n(1) = 1$.*

The differential equation (3.7) shows that P_n and P_n' cannot vanish simultaneously such that P_n has no multiple zeros. Moreover, the Legendre polynomial P_n has n different zeros in the interval $(-1,1)$.

Lemma 3.2 (Rodriguez Formula). *For $n = 0,1,\ldots$,*

$$P_n(t) = \frac{1}{2^n n!}\left(\frac{d}{dt}\right)^n (t^2-1)^n, \quad t\in[-1,1]. \tag{3.8}$$

Integrating by parts we obtain the *Rodriguez rule*

$$\int_{-1}^{1} F(t)P_n(t)\,dt = \frac{1}{2^n n!}\int_{-1}^{1} F^{(n)}(t)(1-t^2)^n\,dt \tag{3.9}$$

for every $F\in C^{(n)}[-1,1]$. It is not hard to show that

$$P_{n+1}'(t) - tP_n'(t) = (n+1)P_n(t), \tag{3.10}$$

$$(t^2-1)P_n'(t) = ntP_n(t) - nP_{n-1}(t), \tag{3.11}$$

$$(n+1)P_{n+1}(t) + nP_{n-1}(t) - (2n+1)tP_n(t) = 0. \tag{3.12}$$

The formulas (3.10)–(3.12) are known as the *recurrence formulas for the Legendre polynomials*.

From [209] we borrow the following estimate of the Legendre polynomial.

Lemma 3.3. *For $n = 1,2,\ldots$ and $t\in(-1,1)$,*

$$|P_n(t)| \le \frac{1}{\sqrt{\pi}}\left(\frac{4}{n(1-t^2)}\right)^{1/2}. \tag{3.13}$$

L^2-Orthonormalized Legendre Polynomials. The system $\{P_n^*\}_{n=0,1,\ldots}$ given by

$$P_n^*(t) = \sqrt{\frac{2n+1}{2}} P_n(t), \quad t \in [-1,1] \tag{3.14}$$

is orthonormal in the sense

$$\int_{-1}^{1} P_n^*(t) P_m^*(t)\ dt = \delta_{n,m}. \tag{3.15}$$

The system $\{P_n^*\}_{n=0,1,\ldots}$ of orthonormal Legendre polynomials is closed in the Hilbert space $(L^2([-1,1]), \langle\cdot,\cdot\rangle_{L^2([-1,1])})$ and is closed in the Banach space

$$(C^{(0)}([-1,1]), \|\cdot\|_{C^{(0)}([-1,1])}).$$

In consequence, any function $F \in L^2([-1,1])$ can be represented as orthogonal expansion by means of $\{P_n^*\}_{n=0,1,\ldots}$

$$\lim_{N\to\infty} \left(\int_{-1}^{1} \left| F(t) - \sum_{n=0}^{N} \int_{-1}^{1} F(x) P_n^*(x)\ dx\ P_n^*(t) \right|^2 dt \right)^{1/2} = 0. \tag{3.16}$$

Our formulas arising orthonormal Legendre polynomials also give a different and useful intuition to *Gauss–Legendre sampling*: The kernel $K_n(\cdot,\cdot) : [-1,1] \times [-1,1] \to \mathbb{R}$ given by

$$K_n(t,x) = \sum_{k=0}^{n} P_k^*(t) P_k^*(x), \quad x,t \in [-1,1] \tag{3.17}$$

is a *reproducing kernel of the space* $Pol_{0,\ldots,n}$ *of* (*algebraic*) *polynomials of degree* $\leq n$, i.e., we have

(i) for each $t \in [-1,1]$, $K_n(t,\cdot)$ is a member of $Pol_{0,\ldots,n}$,
(ii) $F(t) = \langle K_n(t,\cdot), F\rangle_{L^2([-1,1])}$ for all $F \in Pol_{0,\ldots,n}$.

The uniquely determined reproducing kernel (3.17) can be expressed in explicit way (see, e.g., [209]): From the recursion relation (3.12) we easily obtain by elementary manipulation

$$\begin{aligned}(k+1)(P_{k+1}(x)P_k(t) - P_{k+1}(t)P_k(x)) - k(P_k(x)P_{k-1}(t) - P_k(t)P_{k-1}(t)) \\ = (2k+1)(x-t)P_k(x)P_k(t).\end{aligned} \tag{3.18}$$

Summation over k from 1 to n yields the identity

$$\begin{aligned}(x-t)\sum_{k=1}^{n}(2k+1)P_k(x)P_k(t) \\ = (n+1)(P_{n+1}(x)P_n(t) - P_{n+1}(t)P_n(x)) - (x-t),\end{aligned} \tag{3.19}$$

where we have used that $P_0(x) = 1$, $P_1(x) = x$. Therefore, for $x \neq t$, we find the so-called *Christoffel–Darboux formula*

$$K(x,t) = \frac{n+1}{2} \frac{P_{n+1}(x)P_n(t) - P_{n+1}(t)P_n(x)}{x-t}. \tag{3.20}$$

Equivalently, we have

$$K(x,t) = \frac{n+1}{\sqrt{2n+1}\sqrt{2n+3}} \frac{P^*_{n+1}(x)P^*_n(t) - P^*_{n+1}(t)P^*_n(x)}{x-t}. \tag{3.21}$$

We take $x_{i,n+1}, i = 1, \ldots, n+1$, to be the (distinct) zeros of P_{n+1}. Then, $K(x_{j,n+1}, x_{i,n+1}) = 0$ for $i \neq j$ and $K(x_{i,n+1}, x_{i,n+1}) = \sum_{k=0}^{n} P_k^2(x_{i,n+1}) > 0$ since the zeros are interlaced. As a consequence (cf. [167]), we have

$$\begin{aligned} F(t) &= \sum_{i=1}^{n+1} F(x_{i,n+1}) \frac{K(x_{i,n+1}, t)}{K(x_{i,n+1}, x_{i,n+1})} \\ &= \sum_{i=1}^{n+1} F(x_{i,n+1}) \frac{P_{n+1}(t)}{P'_{n+1}(x_{i,n+1})(t - x_{i,n+1})}. \end{aligned} \tag{3.22}$$

Green's Function for the Legendre Operator. As already known, the Legendre operator L has a half-bounded and discrete eigenspectrum $\{\mathrm{L}^\wedge(n)\}_{n=0,1,\ldots}$ such that

$$(\mathrm{L}_t + \mathrm{L}^\wedge(n))P_n(t) = 0, \quad t \in [-1,1], \tag{3.23}$$

where

$$\mathrm{L}^\wedge(n) = n(n+1), \quad n = 0, 1, \ldots. \tag{3.24}$$

Thus, $\mathrm{Spect}_{\mathrm{L}} = \{\mathrm{L}^\wedge(n) \in \mathbb{R} : \mathrm{L}^\wedge(n) = n(n+1),\ n = 0, 1, \ldots\}$ is the *eigenspectrum of the Legendre operator*.

The Hilbert theory of Green functions (cf. [111]) leads to the following definition.

Definition 3.2. A function $G(\mathrm{L} + \lambda; \cdot, \cdot)$: $[-1,1] \times [-1,1] \to \mathbb{R}, (x,t) \mapsto G(\mathrm{L} + \lambda; x, t), \lambda \in \mathbb{R}$, is called *Legendre* (*Green*) *function with respect to the operator* $\mathrm{L} + \lambda$, $\lambda \in \mathbb{R}$, if it satisfies the following properties:

(i) (*Boundedness*) For each fixed $x \in [-1,1]$, $G(L + \lambda; x, \cdot)$ is a continuous function on $(-1,1)$ satisfying the conditions

$$|G(\mathrm{L} + \lambda; x, 1)| < \infty, \tag{3.25}$$

$$|G(\mathrm{L} + \lambda; x, -1)| < \infty. \tag{3.26}$$

(ii) (*Differential equations*) For each fixed $x \in [-1,1]$, $G(\mathrm{L} + \lambda; x, \cdot)$ is twice continuously differentiable in $[-1,1]\backslash\{x\}$. For $\lambda \notin \mathrm{Spect}_{\mathrm{L}}$ and $t \in [-1,1]\backslash\{x\}$ we have

$$(\mathrm{L}_t + \lambda)G(\mathrm{L} + \lambda; x, t) = 0, \quad t \in [-1,1]\backslash\{x\}, \tag{3.27}$$

while, for $\lambda = \mathrm{L}^\wedge(n) \in \mathrm{Spect}_{\mathrm{L}}$, we have

$$(\mathrm{L}_t + \lambda)G(\mathrm{L} + \lambda; x, t) = -P^*_n(x)P^*_n(t), \quad t \in [-1,1]\backslash\{x\}. \tag{3.28}$$

(iii) (*Characteristic singularity*)

$$(1-x^2)\frac{d}{dt}G(\mathrm{L}+\lambda;x,t)\Big|_{t=x-0}^{t=x+0}=1. \tag{3.29}$$

(iv) (*Normalization*) For each $x\in[-1,1]$ and $\lambda=\mathrm{L}^\wedge(n)\in\mathrm{Sepct}_\mathrm{L}$,

$$\int_{-1}^{+1}G(\mathrm{L}+\mathrm{L}^\wedge(n);x,t)P_n^*(t)\;dt=0. \tag{3.30}$$

The function P_n^* is an eigenfunction of the Green function with respect to the operator L in the sense of the linear integral equation

$$(1-\delta_{\lambda,\mathrm{L}^\wedge(n)})P_n^*(t)=(\lambda-\mathrm{L}^\wedge(n))\int_{-1}^{1}G(\mathrm{L}+\lambda;t,u)P_n^*(u)\;du. \tag{3.31}$$

The bilinear expansion of $G(\mathrm{L}+\lambda;\cdot,\cdot)$ therefore reads

$$G(\mathrm{L}+\lambda;x,t)=\sum_{\substack{k=0\\ \mathrm{L}^\wedge(k)\neq\lambda}}^{\infty}\frac{1}{\lambda-\mathrm{L}^\wedge(k)}P_k^*(x)P_k^*(t), \tag{3.32}$$

where $\Sigma_{\mathrm{L}^\wedge(k)\neq\lambda}$ means that the summation is to be extended over all $k\in\mathbb{N}_0$ with $\mathrm{L}^\wedge(k)\neq\lambda$. Obviously, because of $P_n(t)=O((n(1-t^2))^{-1/2})$, $t\in(-1,1)$, the bilinear expansion (3.32) is absolutely and uniformly convergent both in x and t for every compact subinterval of $(-1,1)$.

From the completeness of the system $\{P_n^*\}_{n=0,1,\ldots}$ we easily obtain that $G(\mathrm{L}+\lambda;\cdot,\cdot)$ is uniquely determined by its defining properties (i)–(iv).

A particular role is played by the Legendre (Green) function with respect to the operator L (i.e., $\lambda=0$). It is explicitly available as elementary function (cf. [36])

$$G(\mathrm{L};x,t)=\begin{cases}\frac{1}{2}\ln((1+t)(1-x))+\ln 2-\frac{1}{2}, & x\le t\\ \frac{1}{2}\ln((1-t)(1+x))+\ln 2-\frac{1}{2}, & t\le x.\end{cases} \tag{3.33}$$

In the sense of the Fredholm–Hilbert theory of linear integral equations we may interpret the Green function $G(\mathrm{L}+\lambda;\cdot,\cdot)$ for $\lambda\neq 0$ as resolvent of the kernel

$$\begin{aligned}G(\mathrm{L}+\lambda;x,t)=G(\mathrm{L};x,t)+\lambda\int_{-1}^{1}G(\mathrm{L}+\lambda;x,u)G(\mathrm{L};t,u)\;du\\ -\frac{1}{2\lambda}-\frac{1}{\lambda}P_n^*(x)P_n^*(t)\delta_{\lambda,\mathrm{L}^\wedge(n)}.\end{aligned} \tag{3.34}$$

Inserting the bilinear expansions we therefore obtain, for each $\lambda\neq 0$,

$$\begin{aligned}G(\mathrm{L}+\lambda;x,t)=G(\mathrm{L};x,t)-\lambda\sum_{k=0}^{\infty}{}^{*}\;\frac{1}{(\lambda-\mathrm{L}^\wedge(k))\mathrm{L}^\wedge(k)}P_k^*(x)P_k^*(t)\\ -\;\frac{1}{2\lambda}-\frac{1}{\lambda}P_n^*(x)P_n^*(t)\delta_{\lambda,\mathrm{L}^\wedge(n)},\end{aligned} \tag{3.35}$$

where the series on the right-hand side converges uniformly and absolutely both in x and t on the interval $[-1,1]$, and the symbol Σ^* means, that the summation is extended over all $k \in \mathbb{N}_0$ satisfying $(\lambda - \mathrm{L}^\wedge(k))\mathrm{L}^\wedge(k) \neq 0$.

Next we introduce Legendre (Green) functions with respect to the iterated operator $(\mathrm{L}+\lambda)^p$, $p \in \mathbb{N}$.

Definition 3.3. For $\lambda \in \mathbb{R}$, the function $G((\mathrm{L}+\lambda)^p;\cdot,\cdot)$, $p = 2,3,\ldots$, defined recursively by

$$G((\mathrm{L}+\lambda)^p;x,t) = \int_{-1}^{1} G((\mathrm{L}+\lambda)^{p-1};x,u)G(\mathrm{L}+\lambda;t,u)\,du, \tag{3.36}$$

is called *Legendre (Green) function with respect to the operator* $(\mathrm{L}+\lambda)^p$.

$G((\mathrm{L}+\lambda)^p;\cdot,\cdot)$ is symmetric in its arguments, i.e., for all $x,t \in [-1,1]$

$$G((\mathrm{L}+\lambda)^p;x,t) = G((\mathrm{L}+\lambda)^p;t,x). \tag{3.37}$$

The bilinear expansion of $G((\mathrm{L}+\lambda)^p;\cdot,\cdot)$ reads as follows

$$G((\mathrm{L}+\lambda)^p;x,t) = \sum_{\substack{k=0 \\ \lambda \neq \mathrm{L}^\wedge(k)}}^{\infty} \frac{1}{(\lambda - \mathrm{L}^\wedge(k))^p} P_k^*(x)P_k^*(t). \tag{3.38}$$

For $p = 2,3,\ldots$, the series on the right-hand side is absolutely and uniformly convergent both in $x \in [-1,+1]$ and $t \in [-1,1]$.

If F is continuously differentiable at the point $t \in (-1,1)$, then

$$\begin{aligned}(\mathrm{L}_t+\lambda)&\int_{-1}^{1} G(\mathrm{L}+\lambda;t,u)F(u)\,du \\ &= F(t) - \delta_{\lambda,\mathrm{L}^\wedge(n)}P_n^*(t)\int_{-1}^{+1} P_n^*(u)F(u)\,du.\end{aligned} \tag{3.39}$$

In particular, we have

Lemma 3.4. *For $p = 2,3,\ldots$ and $t \neq x$*

$$(\mathrm{L}_t+\lambda)^{p-1}G((\mathrm{L}+\lambda)^p;x,t) = G(\mathrm{L}+\lambda;x,t). \tag{3.40}$$

Integral Formulas for the Legendre Operator. Suppose that $F:[-1,1]\to\mathbb{R}$ is of the class $C^{(2)}([-1,1])$. Assume that $\lambda \in \mathbb{R}$, $x \in (-1,1)$, and $\varepsilon > 0$ (sufficiently small). Then partial integration, i.e., the so-called Green–Lagrange formula yields

$$\begin{aligned}&\int_{-1}^{x-\varepsilon} \{F(t)(\mathrm{L}_t+\lambda)G(\mathrm{L}+\lambda;x,t) - G(\mathrm{L}+\lambda;x,t)(\mathrm{L}_t+\lambda)F(t)\}\,dt \\ &\quad + \int_{x+\varepsilon}^{1} \{F(x)(\mathrm{L}_t+\lambda)G(\mathrm{L}+\lambda;x,t) - G(\mathrm{L}+\lambda;x,t)(\mathrm{L}_t+\lambda)F(t)\}\,dt \\ &= -(1-t^2)\left\{F(t)\frac{d}{dt}G(\mathrm{L}+\lambda;x,t) - G(\mathrm{L}+\lambda;x,t)\frac{d}{dt}F(t)\right\}\Big|_{t=x-\varepsilon}^{t=x+\varepsilon}.\end{aligned} \tag{3.41}$$

Observing the differential equation and the characteristic singularity of the Green function with respect to $\mathrm{L}+\lambda$ we obtain by letting $\varepsilon \to 0$ the following integral formulas.

Theorem 3.1 (Integral Formula for $\mathrm{L}+\lambda$). *Let x be a point in $(-1,1)$. Suppose that F is of class $C^{(2)}([-1,1])$. Then, for $\lambda \notin \mathrm{Spect}_{\mathrm{L}}$,*

$$F(x) = \int_{-1}^{1} G(\mathrm{L}+\lambda; x, t)(\mathrm{L}_t + \lambda)F(t)\ dt. \tag{3.42}$$

Furthermore, for $\lambda \in \mathrm{Spect}_{\mathrm{L}}$, i.e., $\lambda = \mathrm{L}^{\wedge}(n) = n(n+1)$,

$$\begin{aligned} F(x) = P_n^*(x) \int_{-1}^{1} F(t)P_n^*(t)\ dt \\ + \int_{-1}^{+1} G(\mathrm{L}+\lambda; x, t)(\mathrm{L}_t+\lambda)F(t)\ dt. \end{aligned} \tag{3.43}$$

In particular, for $\lambda = 0$, we have

$$F(x) = \frac{1}{2}\int_{-1}^{+1} F(t)\ dt + \int_{-1}^{+1} G(\mathrm{L}; x, t)\ \mathrm{L}_t F(t)\ dt. \tag{3.44}$$

Repeated application of the Green–Lagrange formula yields

$$\begin{aligned} &\int_{-1}^{1} G(\mathrm{L}+\lambda; x, t)(\mathrm{L}_t+\lambda)F(t)\ dt \\ &\quad = \int_{-1}^{1} G((\mathrm{L}+\lambda)^p; x, t)(\mathrm{L}_t+\lambda)^p F(t)\ dt \end{aligned} \tag{3.45}$$

$p = 2, 3, \ldots$, provided that F is sufficiently often differentiable. This leads us to the following extension of Theorem 3.1.

Theorem 3.2 (Integral Formula for $(\mathrm{L}+\lambda)^p$). *Suppose that $p \in \mathbb{N}$, $\lambda \in \mathbb{R}$, $x \in (-1,1)$, and $F \in C^{(2p)}([-1,1])$.*

Then, for $\lambda \notin \mathrm{Spect}_{\mathrm{L}}$

$$F(x) = \int_{-1}^{1} G((\mathrm{L}+\lambda)^p; x, t)(\mathrm{L}_t+\lambda)^p F(t)\ dt. \tag{3.46}$$

Moreover, for $\lambda \in \mathrm{Spect}_{\mathrm{L}}$, i.e., $\lambda = \mathrm{L}^{\wedge}(n) = n(n+1)$,

$$\begin{aligned} F(x) = P_n^*(x) \int_{-1}^{1} F(t)P_n^*(t)\ dt \\ + \int_{-1}^{1} G((\mathrm{L}+\lambda)^p; x, t)(\mathrm{L}_t+\lambda)^p F(t)\ dt. \end{aligned} \tag{3.47}$$

Finally, for $m \in \mathbb{N}_0$ and $\lambda \in \mathbb{R}$ with $\lambda \neq \mathrm{L}^{\wedge}(m+k)$, $k \in \mathbb{N}$, we let

$$G^{\perp}_{0,\ldots,m}((\mathrm{L}+\lambda)^p; x, t) = G((\mathrm{L}+\lambda)^p; x, t) - G_{0,\ldots,m}((\mathrm{L}+\lambda)^p; x, t), \tag{3.48}$$

where

$$G_{0,\dots,m}((\mathrm{L}+\lambda)^p;x,t)=\sum_{\substack{k=0\\ \lambda\neq \mathrm{L}^\wedge(k)}}^{m}\frac{1}{(\lambda-\mathrm{L}^\wedge(k))^p}P_k^*(x)P_k^*(t). \tag{3.49}$$

$G^{\perp}_{0,\dots,m}((\mathrm{L}+\lambda)^p;\cdot,\cdot)$ is called *mth truncated Legendre (Green) function with respect to the operator* $(\mathrm{L}+\lambda)^p$. These functions admit an easy reformulation of Theorem 3.2 after application of partial integration.

Theorem 3.3. *Suppose that* $p\in\mathbb{N}$, $m\in\mathbb{N}_0$. *Assume that* $\lambda\in\mathbb{R}$ *with*

$$\lambda\neq \mathrm{L}^\wedge(m+1),\mathrm{L}^\wedge(m+2),\dots.$$

Then, for all $x\in(-1,1)$ *and* $F\in C^{(2p)}([-1,1])$,

$$\begin{aligned}F(x)=&\sum_{n=0}^{m}\int_{-1}^{+1}F(t)P_n^*(t)\ dt\ P_n^*(x)\\ &+\int_{-1}^{1}G^{\perp}_{0,\dots,m}((\mathrm{L}+\lambda)^p;x,t)\ (\mathrm{L}_t+\lambda)^pF^{\perp}_{0,\dots,m}(t)\ dt,\end{aligned} \tag{3.50}$$

where we have used the abbreviation

$$F^{\perp}_{0,\dots,m}=F-\underbrace{\sum_{n=0}^{m}\int_{-1}^{1}F(t)P_n^*(t)\ dt\ P_n^*}_{=F_{0,\dots,m}}. \tag{3.51}$$

Theorem 3.3 enables a comparison between the functional value $F(x)$ at the point x and the mth truncated orthogonal expansion of F in terms of Legendre polynomials at x with explicit representation of the remainder term in integral form.

Later on, this formula will be used to formulate adaptive estimates of the remainder terms in Gaussian numerical integration.

Gaussian Quadrature. Let $x_{1,n},\dots,x_{n,n}$ with $x_{1,n}<\dots<x_{n,n}$ be the zeros of the Legendre polynomial P_n of degree n. Then it is well known that there is precisely one zero of the polynomial P_{n+1} in each interval $(-1,x_{1,n}),\dots,(x_{n,n},1)$.

For later use we want to prove the following preparatory result.

Lemma 3.5. *The matrix*

$$A=\left(P_k^*(x_{j,n})\right)_{\substack{k=0,\dots,n-1\\ j=1,\dots,n}} \tag{3.52}$$

is non-singular.

Proof. Assume that the matrix is singular. Then there exist coefficients C_0, …, C_{n-1} with $c=(C_0,\dots,C_{n-1})^T\neq(0,\dots,0)^T$ such that $c^TA=0$, i.e., the following polynomial of degree $\leq n-1$

$$Q(x)=\sum_{l=0}^{n-1}C_lP_l^*(x) \tag{3.53}$$

has as zeros $x_{1,n}, \ldots, x_{n,n}$. Hence, $Q = 0$. Now, for $k = 0, \ldots, n-1$,

$$0 = \int_{-1}^{1} Q(x) P_k^*(x)\ dx = C_k. \tag{3.54}$$

But this is a contradiction. Thus, Lemma 3.5 must be true. □

We are now interested in the *"Gaussian integration formula"* involving the "*Gaussian n-point formula*" $\mathcal{G}_n$

$$\mathcal{I}(F) = \int_{-1}^{1} F(t)\ dt \simeq \mathcal{G}_n(F) = \sum_{i=1}^{n} w_{i,n} F(x_{i,n}), \tag{3.55}$$

where the knots $x_{i,n}$, $i = 1, \ldots, n$, are the zeros of the Legendre polynomial P_n and the weights $w_{i,n}, i = 1, \ldots, n$, are determined by the linear system

$$\begin{aligned}
\sum_{i=1}^{n} w_{i,n} P_0^*(x_{i,n}) &= \int_{-1}^{1} P_0(t) P_0^*(t)\ dt \quad &= \sqrt{2},\\
\sum_{i=1}^{n} w_{i,n} P_1^*(x_{i,n}) &= \int_{-1}^{1} P_1(t) P_0^*(t) dt \quad &= 0,\\
&\vdots &\vdots\\
\sum_{i=1}^{n} w_{i,n} P_{n-1}^*(x_{i,n}) &= \int_{-1}^{1} P_{n-1}(t) P_0^*(t)\ dt &= 0.
\end{aligned} \tag{3.56}$$

From Lemma 3.5 it is clear that the system (3.56) is uniquely solvable in the unknown weights $w_{1,n}, \ldots, w_{n,n}$.

Central for our considerations is the following well-known theorem (see, e.g., [206, 226]).

Theorem 3.4. *Let $x_{1,n}, \ldots, x_{n,n}$ be the zeros of the Legendre polynomial P_n. Furthermore, let $w_{1,n}, \ldots, w_{n,n}$ be the (unique) solution of the linear system* (3.56).

(i) *Then, the weights are positive, i.e., $w_{i,n} > 0$ for $i = 1, \ldots, n$, and we have*

$$\int_{-1}^{+1} P(t)\ dt = \sum_{i=1}^{n} w_{i,n} P(x_{i,n}) \tag{3.57}$$

for all $P \in \mathrm{Pol}_{0,\ldots,2n-1}$, i.e., for all algebraic polynomials of degrees $\leq 2n-1$,

(ii) *Conversely, if* (3.57) *is valid for real numbers $w_{i,n}, x_{i,n}$, $i = 1, \ldots, n$, then the knots $x_{i,n}$, $i = 1, \ldots, n$, are the zeros of the Legendre polynomial P_n and the weights $w_{i,n}, i = 1, \ldots, n$, satisfy the linear equations* (3.56).

(iii) *Finally, there exist no real numbers $x_{i,n}, w_{i,n}, i = 1, \ldots, n$, such that* (3.57) *holds true for all $P \in \mathrm{Pol}_{2n}$.*

Proof. Consider a polynomial $P \in \mathrm{Pol}_{0,\ldots,2n-1}$. Then there exist $Q, R \in Pol_{n-1}$ with

$$Q = \sum_{r=0}^{n-1} A_r P_r^* \tag{3.58}$$

and

$$R = \sum_{r=0}^{n-1} B_r P_r^* \tag{3.59}$$

such that

$$P = P_n^* Q + R. \tag{3.60}$$

Because of the orthogonality of the Legendre polynomials it is clear that

$$\begin{aligned}\int_{-1}^{+1} P(t)\ dt &= \int_{-1}^{1} P(t)\underbrace{P_0(t)}_{=1}\ dt \\ &= \int_{-1}^{1} Q(t)P_n^*(t)\ dt + \int_{-1}^{1} R(t)P_0(t)\ dt \\ &= \int_{-1}^{1} R(t)P_0(t)\ dt. \end{aligned} \tag{3.61}$$

In connection with (3.59) this implies

$$\int_{-1}^{1} P(t)\ dt = \sqrt{2}\ B_0. \tag{3.62}$$

Moreover, because of $P_n^*(x_{i,n}) = 0$, $i = 1, \dots, n$, we obtain

$$\begin{aligned}\sum_{i=1}^{n} w_{i,n} P(x_{i,n}) &= \sum_{i=1}^{n} w_{i,n} \underbrace{P_n^*(x_{i,n})}_{=0} Q(x_{i,n}) + \sum_{i=1}^{n} w_{i,n} R(x_{i,n}) \\ &= \sum_{i=1}^{n} w_{i,n} \sum_{k=0}^{n-1} B_r P_r^*(x_{i,n}) \\ &= \sqrt{2}\ B_0. \end{aligned} \tag{3.63}$$

This shows us that (3.57) holds true.

Inserting the polynomials $\tilde{T}_k \in \mathrm{Pol}_{0,\dots,2n-1}$, $k = 1, \dots, n$, given by

$$\tilde{T}_k(x) = \prod_{\substack{j=1 \\ j\neq k}}^{n} (x - x_{j,n})^2, \quad x \in [-1, 1], \tag{3.64}$$

into (3.57) we get for $k = 1, \dots, n$

$$0 < \int_{-1}^{+1} \tilde{T}_k(t)\ dt = \sum_{i=1}^{n} w_{i,n} \tilde{T}_k(x_{i,n}) = \sum_{i=1}^{n} w_{i,n} \prod_{\substack{j=1 \\ j\neq k}}^{n} (x_{i,n} - x_{j,n})^2, \tag{3.65}$$

hence, it is clear that

$$w_{i,n} > 0, \quad i = 1, \dots, n. \tag{3.66}$$

This proves (i).

Next we verify (iii). Suppose that $w_{i,n}, x_{i,n}$, $i = 1, \ldots, n$, are given in such a way that (3.57) is valid for all $P \in \mathrm{Pol}_{0,\ldots,2n}$. Choose $H \in \mathrm{Pol}_{0,\ldots,2n}$ of the form

$$H(x) = \prod_{j=1}^{n} (x - x_{j,n})^2. \tag{3.67}$$

Then

$$0 < \int_{-1}^{1} H(x)\ dx = \sum_{i=1}^{n} w_{i,n} H(x_{i,n}) = 0. \tag{3.68}$$

This is a contradiction.

In order to guarantee (ii) we apply the Gauss rule (3.57) especially to the Legendre polynomials P_k^*, $k = 0, \ldots, n-1$, such that

$$\sum_{i=1}^{n} w_{i,n} P_k^*(x_{i,n}) = \int_{-1}^{1} P_k^*(x)\ dx = \int_{-1}^{1} P_k^*(x) \underbrace{P_0(x)}_{=1}\ dx. \tag{3.69}$$

In other words, the weights have to satisfy (3.56), i.e.,

$$\sum_{i=1}^{n} w_{i,n} P_k^*(x_{i,n}) = \begin{cases} \sqrt{2}, & k = 0, \\ 0, & k = 1, \ldots, n-1. \end{cases} \tag{3.70}$$

We now consider $P = P_k^* P_n^*$ for $k = 0, \ldots, n-1$, and apply (3.57). This gives

$$0 = \int_{-1}^{1} P_k^*(t) P_n^*(t)\ dt = \sum_{i=1}^{n} w_{i,n}\ P_k^*(x_{i,n})\ P_n^*(x_{i,n}) \tag{3.71}$$

for $k = 0, \ldots, n-1$. Consequently, the vector $a = (A_1, \ldots, A_n)^T$ given by

$$(A_1, \ldots, A_n)^T = (w_{1,n} P_n^*(x_{1,n}), \ldots, w_{n,n} P_n^*(x_{n,n}))^T \tag{3.72}$$

satisfies the linear system

$$0 = \sum_{i=1}^{n} A_i P_k^*(x_{i,n}), \quad k = 0, \ldots, n-1, \tag{3.73}$$

where the coefficient matrix is non-singular. But this means that $A_i = 0$, $i = 1, \ldots, n$, hence, $P_n(x_{i,n}) = 0$, $i = 1, \ldots, n$, as required. □

Remainder Terms Involving Green's Function. Now, the integral formula (Theorem 3.3) comes into play. For all values $\lambda \in \mathbb{R} \backslash \{\mathrm{L}^\wedge(2n), \mathrm{L}^\wedge(2n+1), \ldots\}$ and for functions $F \in C^{(2p)}([-1, +1])$, $p \in \mathbb{N}$, we have

$$\begin{aligned} \sum_{i=1}^{n} w_{i,n} F(x_{i,n}) = & \int_{-1}^{1} F(t)\ dt \\ & + \sum_{i=1}^{n} w_{i,n} \int_{-1}^{1} G_{0,\ldots,2n}^{\perp}((\mathrm{L} + \lambda)^p; x_{i,n}, t)\ (\mathrm{L}_t + \lambda)^p F(t)\ dt. \end{aligned} \tag{3.74}$$

In other words, the remainder

$$\mathcal{E}_n(F) = \mathcal{I}(F) - \mathcal{G}_n(F), \tag{3.75}$$

when the integral

$$\mathcal{I}(F) = \int_{-1}^{1} F(t)\ dt, \quad F \in C^{(2p)}([-1,1]), \tag{3.76}$$

is replaced by the "*Gaussian n-point formula*"

$$\mathcal{G}_n(F) = \sum_{i=1}^{n} w_{i,n} F(x_{i,n}), \tag{3.77}$$

is expressible by means of the $2n$th truncated Legendre (Green) function in integral form

$$\mathcal{E}_n(F) = -\sum_{i=1}^{n} w_{i,n} \int_{-1}^{1} G^{\perp}_{0,\dots,2n}((\mathrm{L}+\lambda)^p; x_{i,n}, t)\ (\mathrm{L}_t+\lambda)^p F(t)\ dt \tag{3.78}$$

for all $\lambda \in \mathbb{R}\backslash\{\mathrm{L}^{\wedge}(2n), \mathrm{L}^{\wedge}(2n+1), \dots\}$.

Applying the Cauchy–Schwarz inequality to (3.78) we obtain

$$|\mathcal{E}_n(F)| \le \sqrt{A^{(p)}_{\lambda}(n)} \sqrt{\int_{-1}^{1} |(\mathrm{L}_t+\lambda)^p F(t)|^2\ dt}, \tag{3.79}$$

for $\lambda \in \mathbb{R}\backslash\{\mathrm{L}^{\wedge}(2n), \mathrm{L}^{\wedge}(2n+1), \dots\}$, where we have used the abbreviation

$$A^{(p)}_{\lambda}(n) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,n} w_{j,n} G^{\perp}_{0,\dots,2n}((\mathrm{L}+\lambda)^{2p}; x_{i,n}, x_{j,n}) \tag{3.80}$$

with

$$\begin{aligned} &G^{\perp}_{0,\dots,2n}((\mathrm{L}+\lambda)^{2p}; x_{i,n}, x_{j,n}) \\ &\quad = \int_{-1}^{1} G^{\perp}_{0,\dots,2n}((\mathrm{L}+\lambda)^p; x_{i,n}, t) G^{\perp}_{0,\dots,2n}((\mathrm{L}+\lambda)^p; t, x_{j,n})\ dt. \end{aligned} \tag{3.81}$$

Summarizing our results we obtain the following remainder estimate.

Theorem 3.5. *Let $x_{i,n}, i = 1, \dots, n$, be the zeros of the Legendre polynomial P_n, $n \ge 1$, and let $w_{i,n}, i = 1, \dots, n$, be the unique solution of the linear system*

$$\sum_{i=1}^{n} w_{i,n} P_k(x_{i,n}) = \int_{-1}^{1} P_0(t) P_k(t)\ dt, \tag{3.82}$$

$k = 0, \dots, n-1$.

Then, for values $\lambda \in \mathbb{R}\backslash\{\mathrm{L}^{\wedge}(2n), \mathrm{L}^{\wedge}(2n+1), \dots\}$ and for functions $F \in C^{(2p)}[-1,1]$, we have

$$\left| \int_{-1}^{1} F(t)\ dt - \sum_{i=1}^{n} w_{i,n} F(x_{i,n}) \right| \le \sqrt{A^{(p)}_{\lambda}(n)} \sqrt{\int_{-1}^{1} |(\mathrm{L}_t+\lambda)^p F(t)|^2\ dt}, \tag{3.83}$$

where

$$A_\lambda^{(p)}(n) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}w_{j,n} \sum_{k=2n}^{\infty} \frac{1}{(\lambda - \mathrm{L}^\wedge(k))^{2p}} P_k^*(x_{i,n})P_k^*(x_{j,n}). \tag{3.84}$$

Remainder Term Estimation. From (3.78), we get the estimate

$$|\mathcal{E}_n(F)| \le \sqrt{A_\lambda^{(p)}(n)} \sqrt{\int_{-1}^{1} \left|(\mathrm{L}_t + \lambda)F_{0,\dots,2n-1}^{\perp}(t)\right|^2 \, dt}, \tag{3.85}$$

where

$$F_{0,\dots,2n-1}^{\perp}(t) = F(t) - \underbrace{\sum_{n=0}^{2n-1} \int_{-1}^{1} F(u)P_n^*(u) \, du \, P_n^*(t)}_{=F_{0,\dots,2n-1}}, \tag{3.86}$$

$t \in [-1, 1]$.

Observing the estimate $|P_n(t)| \le 1$ for all $n \in \mathbb{N}_0$ and $t \in [-1, 1]$ and the properties of the Gaussian weights we get

$$|A_\lambda^p(n)| \le \frac{1}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}w_{j,n} \sum_{k=2n}^{\infty} \frac{2k+1}{(\lambda - \mathrm{L}^\wedge(k))^{2p}} = 2 \sum_{k=2n}^{\infty} \frac{2k+1}{(\lambda - \mathrm{L}^\wedge(k))^{2p}} \tag{3.87}$$

for all $\lambda \in \mathbb{R}\backslash\{\mathrm{L}^\wedge(2n), \mathrm{L}^\wedge(2n+1), \dots\}$.

Theorem 3.6. *For $F \in C^{(2n)}([-1, 1])$,*

$$\begin{aligned} &\int_{-1}^{1} F(t) \, dt - \sum_{i=1}^{n} w_{i,n}F(x_{i,n}) \\ &\quad = \sum_{i=1}^{n} w_{i,n} \int_{-1}^{1} G_{2n}^{\perp}\big((L + \lambda)^n; x_{i,n}, t\big)(\mathrm{L}_t + \lambda)^n F_{0,\dots,2n-1}^{\perp}(t) \, dt, \end{aligned} \tag{3.88}$$

where $F_{0,\dots,2n-1}^{\perp}(t)$ is defined by (3.86).

Convergence of Gaussian Quadrature. From Theorem 3.5 we obtain

$$\left| \int_{-1}^{1} F(t) \, dt - \sum_{i=1}^{n} w_{i,n}F(x_{i,n}) \right| \le \sqrt{A_0^{(p)}(n)} \sqrt{\int_{-1}^{1} (\mathrm{L}_t F(t))^2 \, dt}, \tag{3.89}$$

where

$$A_0^{(p)}(n) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}w_{j,n} \sum_{k=2n}^{\infty} \frac{1}{(k(k+1))^{2p}} P_k^*(x_{i,n})P_k^*(x_{j,n}). \tag{3.90}$$

The sequence

$$\left\{ 2 \sum_{k=2n}^{\infty} \frac{2k+1}{(k(k+1))^{2p}} \right\}_{n\in\mathbb{N}} \tag{3.91}$$

is (strictly) monotonously decreasing and positive. Furthermore, we have the rough estimate

$$0 \leq A_0^{(p)}(n) \leq 2 \sum_{k=2n}^{\infty} \frac{2k+1}{(k(k+1))^{2p}}. \tag{3.92}$$

Therefore, we come to the conclusion.

Theorem 3.7. *The sequence* $\{A_o^{(p)}(n)\}_{n\in\mathbb{N}}$ *has the limit*

$$\lim_{n\to\infty} A_0^p(n) = 0 \tag{3.93}$$

such that, for all $F \in C^{(2p)}([-1,1])$,

$$\lim_{n\to\infty} \mathcal{E}_n(F) = 0, \tag{3.94}$$

i.e., the Gaussian quadrature rule is convergent:

$$\lim_{n\to\infty} \sum_{i=1}^{n} w_i F(x_{i,n}) = \int_{-1}^{1} F(t)\ dt. \tag{3.95}$$

Comparing different numerical integration techniques with the Gaussian method we see that the last method assures the highest degree of polynomial precision. The critical point of the estimate (3.85), however, is to find an estimate for $(\mathrm{L}+\lambda)^{2p}F(t)$ with $(L+\lambda)^{2p}$ being appropriately adapted to the specific properties of the function such that the bound of $|\mathcal{E}_n(F)|$ becomes small. Moreover, it should be critically mentioned that the pointsets $\{x_{1,n},\ldots,x_{n,n}\}_{n\in\mathbb{N}}$ are not hierarchically distributed.

Gauss's work was simplified by Jacobi and further developed by F.G. Mehler [149], E.B. Christoffel [19], and others through much of the 19th century. Eventually, there emerged a coherent theory which received its first systematic expositions by E.B. Christoffel [19], R. Radau [177], and E. Heine [106] in his book on spherical functions. An outline of the developments in numerical quadrature that took place in a period of approximately 200 years from Newton via Gauss to early developments in the 20th century can be found in [188], and a German edition of the four principal memoirs (of Newton, Cotes, Gauss, and Jacobi) is due to A. Kowalewski [129]. A more detailed survey on developments up to the second half of the 20th century is given by W. Gautschi [85], culminating in the following comment:

> *"Gauss's discovery must be rated as one of the most significant events of the* 19*th century in the field of numerical integration and perhaps in all of numerical analysis. The result not only has great beauty and power, but also influenced many later developments in computing and approximation. It soon inspired contemporaries, such as Jacobi and Christoffel, to perfect Gauss's method and to develop it into new directions. Towards the end of the century, it inspired K. Heun* [109] *to generalize Gauss's idea to ordinary differential equations, which in turn led to significant developments in the numerical solution of differential equations, notably the discovery of the Runge–Kutta method* (*cf.* [133]). *Gauss's influence continues into the* 20*th century and is still felt today, ..."*

Whether or not the Gauss method had actually been widely used in practice up to the seventieth of the last century is a matter of some doubt, since the method requires the evaluation of functions at irrational arguments, hence, tedious interpolation. All this changed when powerful digital computers entered the scene, which generated a phase of renewed interest in Gaussian quadrature. The formulas began to be routinely applied, and increased usage, in turn, led to important new theoretical developments.

3.2. Periodic approximate integration

Next we are interested in trapezoidal sums based on Euler-type summation formulas for approximate integration of one-dimensional periodic functions.

To this end we first list some obvious results on *"τ-dilated lattices"*

$$\tau\mathbb{Z} = \{\tau g : g \in \mathbb{Z}, \tau > 0\}. \tag{3.96}$$

The *fundamental cell* $\mathcal{F}_{\tau\mathbb{Z}}$ of the lattice $\tau\mathbb{Z}$ is given by

$$\mathcal{F}_{\tau\mathbb{Z}} = \left\{ x \in \mathbb{R} : -\frac{1}{2}\tau \le x < \frac{1}{2}\tau \right\}. \tag{3.97}$$

A function $F : \mathbb{R} \to \mathbb{C}$ is called *$\tau\mathbb{Z}$-periodical* if $F(x+g) = F(x)$ holds for all $x \in \mathcal{F}_{\tau\mathbb{Z}}$ and $g \in \tau\mathbb{Z}$. The function $\Phi_h : \mathbb{R} \to \mathbb{C},\ h \in \frac{1}{\tau}\mathbb{Z}$, given by

$$x \mapsto \Phi_h(x) = \frac{1}{\sqrt{\tau}} e^{2\pi i h x} \tag{3.98}$$

is $\tau\mathbb{Z}$-periodical and satisfies the identity

$$\int_{\mathcal{F}_{\tau\mathbb{Z}}} \Phi_h(x)\Phi_{h'}(x)\, dx = \begin{cases} 1, & h = h' \\ 0, & h \neq h'. \end{cases} \tag{3.99}$$

The space of all $F \in C^{(m)}(\mathbb{R})$ that are $\tau\mathbb{Z}$-periodical is denoted by $C^{(m)}_{\tau\mathbb{Z}}(\mathbb{R})$, $0 \le m \le \infty$. Clearly, the space $L^2_{\tau\mathbb{Z}}(\mathbb{R})$ is the completion of $C^{(0)}_{\tau\mathbb{Z}}(\mathbb{R})$ with respect to the norm $\|\cdot\|_{L^2_{\tau\mathbb{Z}}(\mathbb{R})}$. The system $\{\Phi_h\}_{h\in\frac{1}{\tau}\mathbb{Z}}$ is orthonormal with respect to the $L^2_{\tau\mathbb{Z}}(\mathbb{R})$-inner product. By convention, we say that λ is an *eigenvalue of the lattice $\tau\mathbb{Z}$ with respect to the operator* Δ of the second-order derivative (i.e., the one-dimensional Laplace operator), if there is a non-trivial solution U of the differential equation $(\Delta + \lambda)U = 0$ satisfying the "boundary condition" of periodicity $U(x+g) = U(x)$ for all $x \in \mathcal{F}_{\tau\mathbb{Z}}$ and $g \in \tau\mathbb{Z}$. From classical Fourier analysis we know that the operator Δ has a half-bounded and discrete eigenspectrum $\{\Delta^\wedge(h)\}_{h\in\frac{1}{\tau}\mathbb{Z}} \subset \mathbb{R}$ such that $(\Delta_x + \Delta^\wedge(h))\,\Phi_h(x) = 0, x \in \mathcal{F}_{\tau\mathbb{Z}}$, with eigenvalues $\Delta^\wedge(h)$ given by $\Delta^\wedge(h) = 4\pi^2h^2, h \in \frac{1}{\tau}\mathbb{Z}$, and eigenfunctions $\Phi_h(x) = \tau^{-\frac{1}{2}}e^{2\pi i h x}, h \in \frac{1}{\tau}\mathbb{Z}, x \in \mathcal{F}_{\tau\mathbb{Z}}$.

Consequently, the *eigenspectrum of the operator* Δ (with respect to $\tau\mathbb{Z}$) is given by

$$\mathrm{Spect}_\Delta(\tau\mathbb{Z}) = \left\{ \Delta^\wedge(h) : \Delta^\wedge(h) = 4\pi^2h^2,\ h \in \frac{1}{\tau}\mathbb{Z} \right\}. \tag{3.100}$$

Trapezoidal Integration Rule. For purposes of *numerical integration of one-dimensional periodic functions* we introduce the *definition of the $\tau\mathbb{Z}$-lattice Green*

function with respect to the operator $\Delta + \lambda$, $\lambda \in \mathbb{R}$. It can be given in canonical way in the sense of the classical Hilbert approach (cf. [111]):

Definition 3.4. A function $G(\tau\mathbb{Z}; \Delta + \lambda; \cdot) : \mathbb{R} \to \mathbb{R}$ is called the *Green function for the Helmholtz operator* $\Delta + \lambda$, $\lambda \in \mathbb{R}$, *with respect to the lattice* $\tau\mathbb{Z}$ (in brief, $\tau\mathbb{Z}$-lattice Green function for $\Delta + \lambda$), if it fulfills the following properties:

(i) (*Periodicity*) $G(\tau\mathbb{Z}; \Delta + \lambda; \cdot)$ is continuous in $\mathbb{R}$, and

$$G(\tau\mathbb{Z}; \Delta + \lambda; x + g) = G(\tau\mathbb{Z}; \Delta + \lambda; x) \tag{3.101}$$

for all $x \in \mathbb{R}$ and $g \in \tau\mathbb{Z}$.

(ii) (*Differential Equation*) $G(\tau\mathbb{Z}; \Delta + \lambda; \cdot)$ is twice continuously differentiable for all $x \notin \tau\mathbb{Z}$ with

$$(\Delta_x + \lambda) G(\tau\mathbb{Z}; \Delta + \lambda; x) = 0 \tag{3.102}$$

provided that $\lambda \notin \mathrm{Spect}_\Delta(\tau\mathbb{Z})$,
$G(\tau\mathbb{Z}; \Delta + \lambda; \cdot)$ is twice continuously differentiable for all $x \notin \tau\mathbb{Z}$ with

$$(\Delta_x + \lambda) G(\tau\mathbb{Z}; \Delta + \lambda; x) = -\frac{1}{\tau} \sum_{\substack{\lambda - (\Delta)^\wedge(h) = 0 \\ h \in \frac{1}{\tau}\mathbb{Z}}} e^{2\pi i h x} \tag{3.103}$$

provided that $\lambda \in \mathrm{Spect}_\Delta(\tau\mathbb{Z})$ (note that the summation on the right side of (3.103) is to be taken over all lattice points $h \in \frac{1}{\tau}\mathbb{Z}$ satisfying $\lambda - (\Delta)^\wedge(h)) = 0$, i.e., $4\pi^2 h^2 = \lambda$).

(iii) (*Characteristic Singularity*)

$$x \mapsto G(\tau\mathbb{Z}; \Delta + \lambda; \cdot) - \frac{1}{2} x \,\mathrm{sign}(x) \tag{3.104}$$

is continuously differentiable for all $x \in \mathcal{F}_{\tau\mathbb{Z}}$.

(iv) (*Normalization*) For all $h \in \frac{1}{\tau}\mathbb{Z}$ with $(\Delta + \lambda)^\wedge(h) = 0$,

$$\int_{\mathcal{F}_{\tau\mathbb{Z}}} G(\tau\mathbb{Z}; \Delta + \lambda; x) \overline{e^{2\pi i h x}}\, dx = 0. \tag{3.105}$$

The Fourier expansion reads

$$G(\tau\mathbb{Z}; \Delta + \lambda; x) \;= \frac{1}{\tau} \sum_{\substack{\lambda - (\Delta)^\wedge(h) \neq 0 \\ h \in \frac{1}{\tau}\mathbb{Z}}} \frac{1}{\lambda - (\Delta)^\wedge(h)} e^{2\pi i h x}, \;\; x \in \mathbb{R}. \tag{3.106}$$

By ordinary partial integration we obtain

Theorem 3.8 ($\tau\mathbb{Z}$-Euler Summation Formula for the Helmholtz Operator $\Delta + \lambda$, $\lambda \in \mathbb{R}$). *Let* F *be of class* $\mathrm{C}^{(2)}([a, b])$, $a < b$. *Suppose that* x *is a point of* $\mathcal{F}_{\tau\mathbb{Z}}$.

Then, the following identity holds true:

$$\begin{aligned}\sum_{\substack{g+x\in[a,b]\\ g\in\tau\mathbb{Z}}}{}'F(g+x) &= \sum_{\substack{\Delta^\wedge(h)=\lambda\\ h\in\frac{1}{\tau}\mathbb{Z}}}\int_a^b F(y)\overline{e^{2\pi ihy}}\;dy\;e^{2\pi ihx}\\ &+\int_a^b G(\tau\mathbb{Z};\Delta+\lambda;x-y)(\Delta_y+\lambda)F(y)\;dy \qquad (3.107)\\ &+\{F(y)(\nabla_y G(\tau\mathbb{Z};\Delta+\lambda;x-y))-(\nabla_y F(y))G(\tau\mathbb{Z};\Delta+\lambda;x-y)\}\,|_a^b,\end{aligned}$$

where the sum on the right side is to be taken over all points $h\in\frac{1}{\tau}\mathbb{Z}$ for which $\lambda-(\Delta)^\wedge(h)=0$. In case of $\lambda-(\Delta)^\wedge(h)\neq 0$ for all $h\in\frac{1}{\tau}\mathbb{Z}$, this sum is understood to be zero.

The case $\lambda=0$ leads back to the concept for the Laplace operator, i.e., the operator of the second derivative

$$\begin{aligned}\sum_{\substack{g+x\in[a,b]\\ g\in\tau\mathbb{Z}}}{}'F(g+x) &= \frac{1}{\tau}\int_a^b F(y)\;dy\\ &+\int_a^b G(\tau\mathbb{Z};\Delta;x-y)(\Delta_y)F(y)\;dy\\ &+\{F(y)(\nabla_y G(\tau\mathbb{Z};\Delta;x-y))-(\nabla_y F(y))G(\tau\mathbb{Z};\Delta;x-y)\}\,|_a^b, \qquad (3.108)\end{aligned}$$

(note that $\tau=2\pi$ provides the Euler summation formula (see [33]) in its original form).

As an immediate consequence, a (generalized) trapezoidal rule of the following kind

$$\begin{aligned}&\tau\sum_{\substack{g+x\in[a,b]\\ g\in\tau\mathbb{Z}}}{}'F(g+x)-\tau\,\{F(y)(\nabla_y G(\tau\mathbb{Z};\Delta;x-y))-(\nabla_y F(y))\;G(\tau\mathbb{Z};\Delta;x-y)\}\,|_a^b\\ &\simeq\int_a^b F(y)\;dy \qquad (3.109)\end{aligned}$$

holds true, where the remainder term

$$\tau\int_a^b G(\tau\mathbb{Z};\Delta;x-y)\Delta_y F(y)\;dy \qquad (3.110)$$

is of the order $O(\tau^2)$. Note that the formula requires the knowledge of $F(y)$ and $\nabla_y F(y)$ at the endpoints a and b of the interval $[a,b]$. Especially, for $x=0$, we have the classical rule

$$\begin{aligned}&\tau\sum_{\substack{g\in[a,b]\\ g\in\tau\mathbb{Z}}}{}'F(g)-\tau\,\{F(y)(\nabla_y G(\tau\mathbb{Z};\Delta;y))-(\nabla_y F(y))\;G(\tau\mathbb{Z};\Delta;y)\}\,|_a^b\\ &\simeq\int_a^b F(y)\;dy. \qquad (3.111)\end{aligned}$$

Under the assumption $\lambda \in \mathbb{R}\backslash\{(\Delta)^\wedge(m), (\Delta)^\wedge(m+1), \ldots, m \in \mathbb{N}\}$ we introduce

$$G^{\perp}_{0,\ldots,m-1}(\tau\mathbb{Z}; \Delta + \lambda; x) = G(\tau\mathbb{Z}; \Delta + \lambda; x) - \frac{1}{\tau} \sum_{\substack{\lambda - (\Delta)^\wedge(h) \neq 0 \\ h \leq m-1 \\ h \in \frac{1}{\tau}\mathbb{Z}}} \frac{1}{\lambda - (\Delta)^\wedge(h)} e^{2\pi i h x}. \tag{3.112}$$

The Fourier series of $G^{\perp}_{0,\ldots,m-1}(\tau\mathbb{Z}; \Delta + \lambda; x),\ x \in \mathbb{R}$, is given by

$$G^{\perp}_{0,\ldots,m-1}(\tau\mathbb{Z}; \Delta + \lambda; x) \ = \ \frac{1}{\tau} \sum_{\substack{(\Delta+\lambda)^\wedge(h) \neq 0 \\ h \geq m \\ h \in \frac{1}{\tau}\mathbb{Z}}} \frac{1}{\lambda - (\Delta)^\wedge(h)} e^{2\pi i h x},\ x \in \mathbb{R}. \tag{3.113}$$

By partial integration we therefore obtain from Theorem 3.8

Corollary 3.1 (Extended Trapezoidal Rule). *Let x be an arbitrary point of $\mathbb{R}$. Suppose that F is of class $\mathrm{C}^{(2)}([a,b])$. Then, for values*

$$\lambda \in \mathbb{R}\backslash\{(\Delta)^\wedge(m), (\Delta)^\wedge(m+1), \ldots\},\ m \in \mathbb{N},$$

$$\begin{aligned}
{\sum_{\substack{g+x \in [a,b] \\ g \in \tau\mathbb{Z}}}}' F(g+x) = {} & \frac{1}{\tau} \int_a^b F(y)\ dy \\
& + \frac{1}{\tau} \sum_{\substack{0 \leq h \leq m-1 \\ \Delta^\wedge(h) = \lambda \\ h \in \frac{1}{\tau}\mathbb{Z}}} \int_a^b F(y) \overline{e^{2\pi i h y}}\ dy\ e^{2\pi i h x} \\
& + \left\{F(y) \nabla_y G^{\perp}_{0,\ldots,m-1}(\tau\mathbb{Z}; \Delta + \lambda; x - y)\right\}|_a^b \\
& - \left\{\nabla_y F(y)\ G^{\perp}_{0,\ldots,m-1}(\tau\mathbb{Z}; \Delta + \lambda; x - y)\right\}|_a^b \\
& + \int_a^b G^{\perp}_{0,\ldots,m-1}(\tau\mathbb{Z}; \Delta + \lambda; x - y)(\Delta_y + \lambda) F(y)\ dy.
\end{aligned} \tag{3.114}$$

Quadrature of Periodic Functions. Next we turn to the well-known approximate integration of 2π-periodic functions. More explicitly, we choose $a = 0$ and $b = 2\pi$. Moreover, we let F be of class $C^{(2)}_{2\pi}(\mathbb{R})$. Then we are able to deduce from Corollary 3.1 that, for $\tau = \frac{2\pi}{m}$, $m \in \mathbb{N}$,

$$\begin{aligned}
\frac{2\pi}{m} {\sum_{\substack{g \in [0,2\pi] \\ g \in \frac{2\pi}{m}\mathbb{Z}}}}' F(g) = {} & \int_0^{2\pi} F(y)\ dy \\
& + \sum_{\substack{0 \leq h \leq m-1 \\ 4\pi^2 h^2 \neq \lambda \\ h \in \frac{m}{2\pi}\mathbb{Z}}} \int_0^{2\pi} F(y) \overline{e^{2\pi i h y}}\ dy\ e^{2\pi i h x} \\
& + \int_0^{2\pi} G^{\perp}_{0,\ldots,m-1}\left(\frac{2\pi}{m}\mathbb{Z}; \Delta + \lambda; y\right)(\Delta_y + \lambda) F(y)\ dy
\end{aligned} \tag{3.115}$$

holds for all $\lambda \in \mathbb{R}\backslash\{\Delta\hat{}(m), \Delta\hat{}(m+1), \ldots\}$. Because of the 2π-periodicity of F we are able to write

$$\frac{2\pi}{m} \sum_{\substack{g\in[0,2\pi] \\ g\in\frac{2\pi}{m}\mathbb{Z}}}{}' F(g) = \frac{2\pi}{m} \sum_{k=1}^{m} F\left((k-1)\frac{2\pi}{m}\right). \tag{3.116}$$

Furthermore, we notice that the integration rule

$$\int_0^{2\pi} F(y)\ dy = \frac{2\pi}{m} \sum_{k=1}^{m} F\left((k-1)\frac{2\pi}{m}\right) = \frac{2\pi}{m} \sum_{k=0}^{m-1} F\left(k\frac{2\pi}{m}\right) \tag{3.117}$$

holds true for all trigonometric polynomials F up to degree $m-1$ in the interval $[0, 2\pi]$ (see, e.g., [206]). In order to guarantee this it is sufficient to verify that (3.117) will be exact for the functions $x \mapsto e^{ilx}$, $x \in \mathbb{R}$, $l = 1, \ldots, m-1$.

For $m = 0$, the assertion is evidently true. Choosing $l \in \{1, \ldots m-1\} \subset \mathbb{N}$ we obtain

$$\int_0^{2\pi} e^{ily} dy = \frac{1}{il}\left(e^{2\pi il} - 1\right) = 0 \tag{3.118}$$

and

$$\sum_{k=1}^{m} e^{il((k-1)\frac{2\pi}{2m})} = \frac{e^{ilm\frac{2\pi}{2m}} - 1}{e^{il\frac{2\pi}{2m}} - 1} = \frac{e^{2\pi il} - 1}{e^{il\frac{2\pi}{2m}} - 1} = 0. \tag{3.119}$$

As an immediate consequence of (3.115) we therefore find in combination with our integral formulas that

$$\begin{aligned} \int_0^{2\pi} F(y)\ dy = & \sum_{k=0}^{m-1} \frac{2\pi}{m} F\left(\frac{2\pi k}{m}\right) \\ & - \int_0^{2\pi} G^{\perp}_{0,\ldots,m-1}\left(\frac{2\pi}{m}\mathbb{Z}; \Delta + \lambda; y\right)(\Delta_y + \lambda)F(y)\ dy \end{aligned} \tag{3.120}$$

is valid for all $F \in C^{(2)}_{2\pi}(\mathbb{R})$ and $\lambda \in \mathbb{R}\backslash\{m^2l^2, m^2(l+1)^2, \ldots, l \in \mathbb{N}\}$.

The remainder term

$$\mathcal{E}_{m-1}(F) = -\int_0^{2\pi} G^{\perp}_{0,\ldots,m}\left(\frac{2\pi}{m}\mathbb{Z}; \Delta + \lambda; y\right)(\Delta_y + \lambda)F(y)\ dy \tag{3.121}$$

of the quadrature depends on the properties of the integrand F. It provides an appropriate representation if the parameter λ can be chosen such that $(\Delta + \lambda)F$ is small (with respect to a certain topology).

An estimate in the sense of A. Sard [190] is obtainable via the $L^2_{2\pi}([0, 2\pi])$-topology

$$|\mathcal{E}_{m-1}(F)| \leq \sqrt{(G^{\perp}_{0,\ldots,m})^{(2)}\left(\frac{2\pi}{m}\mathbb{Z}; \Delta + \lambda, 0\right)} \sqrt{\int_0^{2\pi} |(\Delta_y + \lambda)F(y)|^2\ dy}, \tag{3.122}$$

where

$$(G_{0,\dots,m-1}^{\perp})^{(2)}\left(\frac{2\pi}{m}\mathbb{Z};\Delta+\lambda;0\right)=\frac{m}{2\pi}\sum_{\substack{m^2l^2\neq\lambda\\ l\geq m}}\left(\frac{1}{\lambda-m^2l^2}\right)^2 e^{imlx},\ x\in\mathbb{R}. \quad (3.123)$$

3.3. Latitude-longitude constituted spherical harmonics

We begin our considerations with some basic background on the conventionally used *associated Legendre spherical harmonic system* in physical geodesy (see, e.g., [107], these and other types of spherical harmonic systems can be found in [46, 161]): Let H_n be a homogeneous, harmonic polynomial of degree n with the following properties:

(i) $H_n(\mathbf{t}x)=H_n(x)$ for all orthogonal transformations $\mathbf{t}$ leaving $\varepsilon^3=(0,0,1)^T$ fixed, i.e., $\mathbf{t}\varepsilon^3=\varepsilon^3$,
(ii) $H_n(\varepsilon^3)=1$.

Then it is known that H_n is uniquely determined by

$$H_n(x)=r^nP_n(t), \quad (3.124)$$

with

$$x=r\xi,\quad \xi\in\Omega=\{\xi\in\mathbb{R}^3:|\xi|=1\}, \quad (3.125)$$

(note that we use Ω instead of $\mathbb{S}^2$ for the unit sphere in $\mathbb{R}^3$)

$$\xi=t\varepsilon^3+\sqrt{1-t^2}\,(\cos\lambda\varepsilon^1+\sin\lambda\varepsilon^2), \quad (3.126)$$

$$t=\cos\theta,\ t\in[-1,1],\ \lambda\in[0,2\pi), \quad (3.127)$$

where

$$P_n(t)=\sum_{k=0}^{n}C_{\frac{n-k}{2}}(1-t^2)^{\frac{n-k}{2}}t^k \quad (3.128)$$

and

$$C_{\frac{n-k}{2}}=\begin{cases}0 & ,\quad n-k \text{ odd},\\ (-\frac{1}{4})^{\frac{n-k}{2}}\frac{n!}{((\frac{n-k}{2})!)^2k!}, & n-k \text{ even}.\end{cases} \quad (3.129)$$

Equivalently, we have

$$P_n(t)=n!\sum_{l=0}^{\lfloor\frac{n}{2}\rfloor}\left(-\frac{1}{4}\right)^l\frac{(1-t^2)^lt^{n-2l}}{(l!)^2(n-2l)!}. \quad (3.130)$$

By a straightforward calculation we are able to introduce from

$$P_{n,m}(t)=(1-t^2)^{m/2}\left(\frac{d}{dt}\right)^mP_n(t),\ \ t\in[-1,1], \quad (3.131)$$

the so-called *associated Legendre function of degree n and order m*

$$P_{n,m}(t) = (1-t^2)^{m/2} \sum_{k=0}^{\lfloor \frac{n-m}{2} \rfloor} (-1)^k \frac{(2n-2k)!}{2^n k!(n-k)!(n-m-2k)!} t^{n-m-2k} \quad (3.132)$$

(note that P_n is equal to $P_{n,0}$). In connection with $P_{n,m} = 0$ for $m > n$, the preceding result leads to the following statement (see, e.g., [46]).

Lemma 3.6 (Associated Legendre Polynomial of Degree n and Order m). *For $n = 0, 1, \ldots, m = 0, 1, \ldots$, and $t \in [-1, 1]$ we have*

$$P_{n,m}(t) = (1-t^2)^{\frac{m}{2}} \sum_{k=0}^{n} C^m_{\frac{n-m-k}{2}} (1-t^2)^{\frac{n-m-k}{2}} t^k, \quad (3.133)$$

where the generating coefficients $C^m_{\frac{n-m-k}{2}}$ of the associated Legendre polynomial of degree n and order m are given by

$$C^m_{\frac{n-m-k}{2}} = \begin{cases} (\frac{1}{2})^m \frac{(n+m)!}{(n-m)!m!} C_{\frac{n-m-k}{2}}, & n-m-k \text{ even}, \ 0 \le k \le n-m \\ 0, & \text{otherwise.} \end{cases} \quad (3.134)$$

The associated Legendre functions $P_{n,m}$, $n = 0, 1, \ldots$, $m = 1, \ldots, n$, satisfy the differential equation

$$(1-t^2)\left(\frac{d}{dt}\right)^2 P_{n,m}(t) - 2t\frac{d}{dt}P_{n,m}(t) + \left(n(n+1) - \frac{m^2}{1-t^2}\right) P_{n,m}(t) = 0. \quad (3.135)$$

For every $m = 1, 2, \ldots$, the system

$$\left\{ \left(\frac{2n+1}{2} \frac{(n-m)!}{(n+m)!} \right)^{1/2} P_{n,m} \right\}_{n=m,m+1,\ldots}$$

is a complete orthonormal system in $L^2([-1, 1])$.

The functions G, H defined by

$$G: \ t \mapsto G(t) \ = P_{n,j}(t), \qquad t \in (-1, 1), \quad (3.136)$$

$$H: \lambda \mapsto H(\lambda) = \begin{cases} \cos(j\lambda) \\ \sin(j\lambda) \end{cases}, \quad \lambda \in [0, 2\pi), \quad (3.137)$$

respectively, satisfy the differential equations

$$(1-t^2)G''(t) - 2tG'(t) + \left(n(n+1) - \frac{j^2}{1-t^2}\right) G(t) = 0, \quad (3.138)$$

$$H''(\lambda) + j^2 H(\lambda) = 0. \quad (3.139)$$

Therefore, the functions $L^{\mathbb{R}}_{n,1}, \ldots, L^{\mathbb{R}}_{n,2n+1} \in C^{(\infty)}(\mathbb{S}^2)$ given by

$$L^{\mathbb{R}}_{n,j}(\xi) = \begin{cases} P_{n,|j|}(t)\cos(j\lambda), & j = -n, \ldots, 0 \\ P_{n,j}(t) \ \sin(j\lambda), & j = \ \ 1, \ldots, n \end{cases} \quad (3.140)$$

satisfy the differential equation (for graphical illustrations see Figures 3.2, 3.3, and 3.4),

$$(\Delta_\xi^* + (\Delta^*)^\wedge(n))\ L_{n,j}^{\mathbb{R}}(\xi) = 0, \quad \xi \in \mathbb{S}^2,\ (\Delta^*)^\wedge(n) = n(n+1), \tag{3.141}$$

$j = -n, \ldots, n$. In addition, the functions $L_{n,j}^{\mathbb{R}} \in C^{(\infty)}(\mathbb{R}^3), j = -n, \ldots, n$, given by

$$L_{n,j}^{\mathbb{R}}(x) = |x|^n L_{n,j}^{\mathbb{R}}(\xi), \quad x = |x|\xi, \quad \xi \in \Omega, \tag{3.142}$$

form homogeneous harmonic polynomials of degree n in $\mathbb{R}^3$.

Definition 3.5. Let $L_{n,j}^{\mathbb{R}}, j = -n, \ldots, n$, be defined by (3.140). Then, $L_{n,j}^{\mathbb{R}}$ is called *associated Legendre* (*spherical*) *harmonic of degree n and order j*. Correspondingly, the system $\{Y_{n,j}\}_{j=-n,\ldots,n}$ given by

$$Y_{n,j} = C_{n,j} L_{n,j}^{\mathbb{R}}, \quad j = -n, \ldots, n, \tag{3.143}$$

with

$$C_{n,j} = \sqrt{(2 - \delta_{j,0}) \frac{2n+1}{4\pi} \frac{(n-|j|)!}{(n+|j|)!}} \tag{3.144}$$

is called (*fully*) $L_{\mathbb{R}}^2(\mathbb{S}^2)$*-orthonormal system of associated Legendre* (*spherical*) *harmonics*.

In terms of associated Legendre harmonics, the *addition theorem* allows the following reformulation that is standard in all geosciences (see, e.g., [46, 159]).

Remark 3.1 (Addition Theorem for the system $\{Y_{n,j}\}$). Suppose that $\xi, \eta \in \Omega$ are given by

$$\begin{aligned} \xi = &\sqrt{1-t_\xi^2}\cos\lambda_\xi\ \varepsilon^1 + \sqrt{1-t_\xi^2}\sin\lambda_\xi\ \varepsilon^2 + t_\xi\ \varepsilon^3 \\ &-1 \le t_\xi \le 1, t_\xi = \cos\theta_\xi,\ 0 \le \lambda_\xi < 2\pi, \end{aligned} \tag{3.145}$$

$$\begin{aligned} \eta = &\sqrt{1-t_\eta^2}\cos\lambda_\eta\ \varepsilon^1 + \sqrt{1-t_\eta^2}\sin\lambda_\eta\ \varepsilon^2 + t_\eta\varepsilon^3 \\ &-1 \le t_\eta \le 1, t_\eta = \cos\theta_\eta,\ 0 \le \lambda_\eta < 2\pi, \end{aligned} \tag{3.146}$$

respectively, so that

$$\begin{aligned} \xi \cdot \eta &= t_\xi t_\eta + \sqrt{1-t_\xi^2}\sqrt{1-t_\eta^2}(\cos\lambda_\xi\cos\lambda_\eta + \sin\varphi_\xi\sin\lambda_\eta) \\ &= t_\xi t_\eta + \sqrt{1-t_\xi^2}\sqrt{1-t_\eta^2}\cos(\lambda_\xi - \lambda_\eta). \end{aligned} \tag{3.147}$$

Then we have

$$\begin{aligned}
&\frac{2n+1}{4\pi} P_n(t_\xi t_\eta + \sqrt{1-t_\xi^2}\sqrt{1-t_\eta^2}\cos(\lambda_\xi - \lambda_\eta)) \\
&\quad = \frac{1}{4\pi} P_n(t_\xi) P_n(t_\eta) \\
&\qquad + \frac{2n+1}{2\pi} \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_{n,m}(t_\xi) P_{n,m}(t_\eta) \cos(m(\lambda_\xi - \lambda_\eta)) \\
&\quad = \sum_{j=-n}^{n} Y_{n,j}(\xi) Y_{n,j}(\eta). \qquad (3.148)
\end{aligned}$$

In other words, summing up all spherical harmonics involving associated Legendre functions via the addition theorem leads (apart from a multiplicative factor) to the orthogonal invariant Legendre (kernel) functions.

FIGURE 3.2. Zonal ($j = 0$) spherical harmonics of different degrees $1, \ldots, 6$ (from left to right). The black and white color indicate the zones of different signs of the function, respectively.

FIGURE 3.3. Tesseral ($k \neq \pm n$) spherical harmonics of degree $n = 4$. The black and white colors indicate the zones of different signs of the function, respectively.

FIGURE 3.4. Sectorial ($k = \pm n$) spherical harmonics of different degrees. The black and white colors indicate the zones of different signs of the function, respectively.

Definition 3.6 (Complex-Valued Spherical Harmonics). Let $n \in \mathbb{N}_0$, $j \in \mathbb{Z}$, with $-n \le j \le n$. The function

$$\xi \mapsto Y_{n,j}^{\mathbb{C}}(\xi) = (-1)^j \sqrt{\frac{2n+1}{4\pi}\frac{(n-j)!}{(n+j)!}}\, P_{n,j}(\cos(\theta))\, e^{ij\lambda}$$

is called (*complex*) *associated Legendre spherical harmonic of degree n and order j, where* θ, λ *are the spherical* coordinates *of* ξ (*note that i denotes the imaginary unit with* $i^2 = -1$).

These spherical harmonics are orthonormal with respect to the canonical scalar product of the space $L^2_{\mathbb{C}}(\Omega)$ of complex-valued square-integrable functions on the unit sphere Ω. Their addition theorem may be rewritten as follows

$$\sum_{j=-n}^{n} \overline{Y_{n,j}^{\mathbb{C}}(\xi)} Y_{n,j}^{\mathbb{C}}(\eta) = \frac{2n+1}{4\pi} P_n(\xi \cdot \eta). \tag{3.149}$$

For further details on this representation of spherical harmonics the reader is referred to, e.g., [50] and [161].

For $n \in \mathbb{N}_0$, $j \in \mathbb{Z}$ with $j = -n, \ldots, n$,

$$\xi \mapsto Y_{n,j}(\xi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-|j|)!}{(n+|j|)!}} P_{n,|j|}(\cos(\theta)) \begin{cases} \sqrt{2}\cos(j\varphi)\,, & j<0 \\ 1\,, & j=0 \\ \sqrt{2}\sin(j\varphi)\,, & j>0 \end{cases} \tag{3.150}$$

can be represented in the form

$$\xi \mapsto Y_{n,j}(\xi) = \begin{cases} \frac{\sqrt{2-\delta_{0,j}}}{2}\left(Y_{n,j}^{\mathbb{C}}(\xi) + \overline{Y_{n,j}^{\mathbb{C}}(\xi)}\right), & j \le 0, \\ \frac{(-1)^j\sqrt{2}}{2i}\left(Y_{n,j}^{\mathbb{C}}(\xi) - \overline{Y_{n,j}^{\mathbb{C}}(\xi)}\right), & j > 0, \end{cases} \tag{3.151}$$

for all $\xi \in \Omega$, $n \in \mathbb{N}_0$, and $j \in \mathbb{Z}$ with $j = -n, \ldots, n$ (note that in case of the real-valued spherical harmonics the indexing with negative integers is just one possibility to distinguish the two types with sine and cosine).

3.4. Latitude-longitude integration

As is well known longitude-latitude point sets on the unit sphere $\Omega \subset \mathbb{R}^3$ enable approximate integration weights being available in a comparatively easy and fast way, thereby establishing spherical harmonics exact cubature formulas which are of tremendous importance, e.g., in mathematical geodesy.

In what follows we are first concerned with the arrangement of equiangular point sets. Systems of this type can be gained by suitably dividing $[0, \pi) \times [0, 2\pi]$ into longitude-latitude grids. Clearly, there are numerous possibilities for a choice of locations concerning the type and number of latitude as well as longitude distributions. Nonetheless, these choices are decisive for the quality of the resulting integration formulas.

For application of longitude-latitude grid integration, suppose that the spherical harmonics are represented in the following (complex) way

$$Y_{n,j}^{\mathbb{C}}(\xi) = \underbrace{(-1)^j \sqrt{\frac{2n+1}{4\pi}\frac{(n-j)!}{(n+j)!}}}_{=C_n^j} \; P_{n,j}(\cos(\theta)) \; e^{ij\lambda}, \tag{3.152}$$

$n = 0, 1, \ldots,\ j = -n, \ldots, n$ and $\xi = t\varepsilon^3 + \sqrt{1-t^2}(\cos\lambda\varepsilon^1 + \sin\lambda\varepsilon^2), t = \cos\theta,\ t \in [-1, 1],\ \lambda \in [0, 2\pi)$.

On the one hand, taking into account that $Y_{0,0}^{\mathbb{C}} = \frac{1}{\sqrt{4\pi}}$, we obtain

$$\int_{\mathbb{S}^2} Y_{k,l}^{\mathbb{C}}(\xi) \; dS(\xi) = \begin{cases} \sqrt{4\pi}\,, \text{ if } k = l = 0, \\ 0\,, \text{ else} \end{cases} \tag{3.153}$$

(dS denotes the surface element). On the other hand, it can be readily seen that

$$\int_{\mathbb{S}^2} Y_{k,l}^{\mathbb{C}}(\xi) \; dS(\xi) = \int_0^\pi C_k^l P_{k,|l|}(\cos(\theta))\sin(\theta) \; d\theta \int_0^{2\pi} (\cos(l\lambda) + i\sin(l\lambda)) \; d\lambda. \tag{3.154}$$

The combination of (3.153) and (3.154) leads us to the equation

$$\begin{aligned} &\int_0^\pi C_k^l P_{k,|l|}(\cos(\theta))\sin(\theta) \; d\theta \left(\int_0^{2\pi} \cos(l\lambda) \; d\lambda + i \int_0^{2\pi} \sin(l\lambda) \; d\lambda \right) \\ &\quad = \begin{cases} \sqrt{4\pi}\,, \text{ if } k = l = 0, \\ 0\,, \text{ else.} \end{cases} \end{aligned} \tag{3.155}$$

This identity is the point of departure for the derivation of exact integration formulas. Together with $C_0^0 = \frac{1}{\sqrt{4\pi}}$ we easily see that

$$\begin{aligned} &\int_0^\pi P_{k,|l|}(\cos(\theta))\sin(\theta) \; d\theta \left(\int_0^{2\pi} \cos(l\lambda) \; d\lambda + i \int_0^{2\pi} \sin(l\lambda) \; d\lambda \right) \\ &\quad = \begin{cases} \sqrt{4\pi}\,, \text{ if } k = l = 0, \\ 0\,, \text{ else.} \end{cases} \end{aligned} \tag{3.156}$$

3.5. Cubature based on periodic integration

In accordance with (3.154) we are able to replace the integral on the whole unit sphere by latitude-longitude integration. As a consequence, the desired cubature rule on the sphere can be based on already known approaches to numerical integration from one dimension. In order to establish exact integration formulas we pursue the following three-step strategy:

(i) Determine a number $N_\lambda \in \mathbb{N}$, the nodes $\lambda_0, \ldots, \lambda_{N_\lambda - 1} \in [0, 2\pi)$, and the weights $a_0^\lambda, \ldots, a_{N_\lambda-1}^\lambda \in \mathbb{R}$, so that the integration formula

$$\int_0^{2\pi} \cos(l\lambda) d\tau + i \int_0^{2\pi} \sin(l\lambda)\, d\lambda = \sum_{j=0}^{N_\lambda - 1} a_j^\lambda \cos(l\lambda_j) + i \sum_{j=0}^{N_\lambda - 1} a_j^\lambda \sin(l\lambda_j)$$
$$= \begin{cases} 2\pi\,, \text{ if } l = 0, \\ 0\,, \text{ else} \end{cases} \tag{3.157}$$

is exact for $l = 0, \ldots, m$. Evidently, (3.157) is fulfilled for $l \neq 0$ if the longitude λ is discretized according to (i).This is why the integral

$$\int_0^\pi P_{k,|l|}(\cos(\theta)) \sin(\theta) d\theta$$

is only significant for $l = 0$ and $k = 0, \ldots, m$.

(ii) Concerning the discretization of the latitude θ it is sufficient to consider the following integration rule:
Determine a number $N_\theta \in \mathbb{N}$, the nodes $\theta_0, \ldots, \theta_{N_\theta - 1} \in [0, \pi]$, and the weights $a_0^\theta, \ldots, a_{N_\theta-1}^\theta \in \mathbb{R}$, so that

$$\int_0^\pi \underbrace{P_{k,0}(\cos(\theta))}_{=P_k(\cos(\theta))} \sin(\vartheta) d\theta = \sum_{j=0}^{N_\theta - 1} a_j^\theta P_k(\cos(\theta_j)) \tag{3.158}$$

is exact for $k = 0, \ldots, m$.

(iii) Combine the latitude-longitude nodal systems to the product set

$$(\theta_k, \lambda_j) \in [0, \pi) \times [0, 2\pi), \quad k = 0, \ldots, N_\theta - 1, \quad j = 0, \ldots, N_\lambda - 1, \tag{3.159}$$

and collect the integration weights $a_0^\lambda, \ldots, a_{N_\lambda-1}^\lambda$ and $a_0^\theta, \ldots a_{N_\theta-1}^\theta$.

We begin with the discretization with respect to the longitude that is well-known from (3.120).

Lemma 3.7 (Longitude Integration). *For $N_\lambda \in \mathbb{N}$, assume that the nodes are given by $\lambda_j = j\frac{2\pi}{N_\lambda} \in [0, 2\pi)$, $j = 0, \ldots, N_\lambda - 1$. Then, the integration rule*

$$\int_0^{2\pi} T(\lambda)\, d\lambda = \sum_{j=0}^{N_\lambda - 1} \frac{2\pi}{N_\lambda} T\left(\frac{2\pi j}{N_\lambda}\right) \tag{3.160}$$

holds true for all trigonometric functions T of degree $\leq N_\lambda - 1$.

In order to discretize the latitude θ, however, it is evident that the integration weights cannot be expected to be equal for all samples. In fact, there are several possibilities to determine point sets and weights which fulfill (3.157). The methods presented here reflect straightforward numerical application.

In accordance with (3.156) we are interested in constructing an integral formula of the form

$$\int_0^\pi P_k(\cos(\theta))\sin(\theta)\ d\theta = \sum_{j=0}^{N_\theta-1} a_j^\theta P_k(\cos(\theta_j)) \tag{3.161}$$

for $k = 0, \ldots, m$. Since we already know that the set $\{P_0, \ldots, P_{N_\theta}\}, N_\theta \in \mathbb{N}$, represents a *Tschebyscheff system* in the uni-variate sense (see also [23]), integration formulas of the type (3.157) exist if the angles $\theta_i, i = 0, \ldots, N_\theta - 1$, are chosen to be pairwise distinct. In particular, we can think of an equiangular distribution of the samples on latitudes. As a matter of fact, we briefly explain an integration technique whose origin dates back to J.R. Driscoll, R.M. Healy [26]. For that purpose we notice that, for an even $N_\theta \in \mathbb{N}$,

$$\int_0^\pi P_k(\cos(\theta))\sin(\theta)\ d\theta = \frac{1}{2}\int_{-\pi}^\pi P_k(\cos(\theta))\sin(\theta)\mathrm{sgn}(\theta)\ d\theta,\ k = 0, \ldots, N_\theta - 1. \tag{3.162}$$

Remembering the well-known sign expansion (see, e.g., [144])

$$\mathrm{sgn}(\theta) = \frac{4}{\pi}\sum_{l=0}^\infty \frac{1}{2l+1}\sin((2l+1)\theta), \quad \theta \in [-\pi, \pi]. \tag{3.163}$$

we obtain for $k = 0, \ldots, N_\theta - 1$

$$\int_0^\pi P_k(\cos(\theta))\sin(\theta)\ d\theta = \frac{2}{\pi}\sum_{s=0}^\infty \int_{-\pi}^\pi P_k(\cos(\theta))\sin(\theta)\frac{1}{2s+1}\sin((2s+1)\theta)\ d\theta. \tag{3.164}$$

It is clear that, for $k = 0, \ldots, N_\theta - 1$, the term $P_k(\cos(\theta))\sin(\theta)$ represents a trigonometric function of degree $\leq N_\theta$. Due to the orthogonality relations we readily find that

$$\int_{-\pi}^\pi P_k(\cos(\theta))\sin\theta\sin((2s+1)\theta)\ d\theta = 0, \quad s > \frac{N_\theta}{2} - 1. \tag{3.165}$$

Hence, the series (3.164) reduces to

$$\int_0^\pi P_k(\cos(\theta))\sin(\theta)\ d\theta = \frac{2}{\pi}\sum_{s=0}^{\frac{N_\theta}{2}-1} \int_{-\pi}^\pi P_k(\cos(\theta))\sin(\theta)\frac{1}{2s+1}\sin((2s+1)\theta)\ d\theta. \tag{3.166}$$

The integrand on the right side is a trigonometric function of order $2N_\theta - 1$, hence, it is not difficult to show that the identity

$$\frac{2}{\pi}\sum_{s=0}^{\frac{N_\theta}{2}-1}\int_{-\pi}^{\pi} P_k(\cos(\theta))\sin(\theta)\frac{1}{2s+1}\sin((2s+1)\theta)\ d\theta \tag{3.167}$$

$$= \frac{2}{N_\theta}\sum_{j=-N_\theta}^{N_\theta-1} P_k\left(\cos\left(j\frac{\pi}{N_\theta}\right)\right)\sin\left(j\frac{\pi}{N_\theta}\right)\sum_{s=0}^{\frac{N_\theta}{2}-1}\frac{1}{2s+1}\sin\left((2s+1)\,j\frac{\pi}{N_\theta}\right)$$

holds true for $k = 0, \ldots, N_\theta - 1$. Moreover, the symmetry of the trigonometric polynomials enables us to rewrite the first sum on the right-hand side of (3.167) as follows

$$\int_0^\pi P_k(\cos(\theta))\sin(\theta)\, d\theta \tag{3.168}$$
$$= \frac{2}{N_\theta} \sum_{j=0}^{N_\theta-1} P_k\left(\cos\left(j\frac{\pi}{N-\theta}\right)\right) \sin\left(j\frac{\pi}{N-\theta}\right) \sum_{s=0}^{\frac{N_\theta}{2}-1} \frac{1}{2s-1} \sin\left((2s+1)j\frac{\pi}{N_\theta}\right)$$

for $k = 0, \ldots, N_\theta - 1$. Finally, taking into account that the system of Legendre polynomials P_k, for $k = 0, \ldots, N_\theta - 1$, forms a basis of the space of all polynomials of degree less or equal to $N_\theta - 1$ we arrive at following result.

Theorem 3.9 (Latitude Integration). *Let $N_\theta \in \mathbb{N}$ be an even number. Suppose that $\theta_j = j\frac{\pi}{N_\theta}$, $j = 0, \ldots, N_\theta - 1$. Then, the integration rule*

$$\int_0^\pi P(\cos(\theta))\sin(\theta)\, d\theta = \int_{-1}^{+1} P(t)\, dt = \sum_{j=0}^{N_\theta-1} a_j^\theta P(\cos(\theta_j)) \tag{3.169}$$

with

$$a_j^\theta = \frac{4}{N_\theta} \sin\left(j\frac{\pi}{N_\theta}\right) \sum_{s=0}^{\frac{N_\theta}{2}-1} \frac{1}{2s+1} \sin\left((2s+1)j\frac{\pi}{N_\theta}\right) \tag{3.170}$$

holds true for all polynomials of degrees $\leq N_\theta - 1$.

Now, especially for $N_\theta = N_\lambda = m + 1$, $m \in \mathbb{N}$ odd, the combination of the above schemes for longitude-latitude distribution enables us to arrange an *equiangular longitude-latitude distributed grid* (ELLG) *integration rule*, where the nodes and weights, respectively, are given by

$$(\theta_j, \lambda_k) \in [0, \pi) \times [0, 2\pi),\ \theta_j = j\frac{\pi}{m+1},\ j = 0, \ldots, m,\ \lambda_k = k\frac{2\pi}{m+1},\ k = 0, \ldots, m, \tag{3.171}$$

and

$$a_j^\theta = \frac{4}{m+1} \sin\left(j\frac{\pi}{m+1}\right) \sum_{s=0}^{\frac{m+1}{2}-1} \frac{1}{2s+1} \sin\left((2s+1)j\frac{\pi}{m+1}\right),\ j = 0, \ldots, m,$$
$$a_k^\lambda = \frac{2\pi}{m+1}, \quad k = 0, \ldots, m. \tag{3.172}$$

Figure 3.5 gives a graphical illustration of the equiangular longitude-latitude grid (ELLG) in the (θ, λ)-plane. As a consequence, any spherical harmonic Y of degree $\leq m$, i.e., $Y \in Harm_{0,\ldots,m}$ can be integrated exactly in the form

$$\int_\Omega Y(\eta)\, dS(\eta) = \sum_{j=0}^{m} \sum_{k=0}^{m} a_j^\theta\, a_k^\lambda\, Y(\theta_j, \lambda_k). \tag{3.173}$$

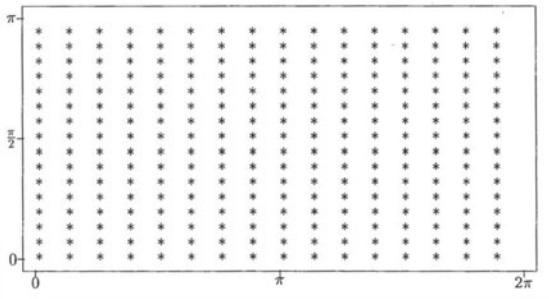


FIGURE 3.5. The ELLG longitude-latitude grid with $m = 15$.

In particular, if a product FY is of class $Harm_{0,\ldots,m}$, then the inner product $\langle F, Y\rangle_{\mathrm{L}^2(\Omega)}$ can be evaluated exactly by means of a proposed latitude-longitude grid based on $(m+1)^2$ points.

3.6. Cubature based on Gaussian integration

From the identity (3.156) we know that integration involving spherical harmonics along spherical longitudes and latitudes, respectively, can be treated independently from each other. Thus, classical Gaussian quadrature can also be used to construct an efficient cubature formula for latitude approximation.

We first choose the longitude $\lambda \in [0, 2\pi)$ to be discretized in an equiangular way. Our purpose is to reduce the number N_θ of the required latitudes in relation to the already discussed equiangular latitude-longitude grid (ELLG) by use of Gaussian quadrature.

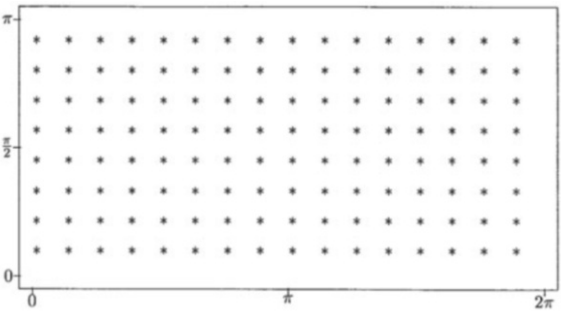


FIGURE 3.6. Gaussian Grid (GG) with $m = 15$.

As a consequence, let N_θ and N_λ, as usual, denote the number of spherical latitudes and longitudes, respectively, to be involved. Assume, especially, that $m \in \mathbb{N}$ is odd, and N_θ and N_λ, are chosen (see Figure 3.6 for an example of a Gaussian grid (GG))

as follows:

$$N_\lambda = m+1, \tag{3.174}$$

$$N_\theta = \frac{1}{2}(m+1). \tag{3.175}$$

Let $x^\theta_{1,N_\theta} < \cdots < x^\theta_{N_\theta,N_\theta}$ be the zeros of the Legendre polynomial of degree N_θ, and suppose that the grid (GG) is given by

$$(\theta_j, \lambda_k) \in [0,\pi] \times [0, 2\pi), \tag{3.176}$$

with

$$\theta_j = \arccos(x^\theta_{j,N_\theta}), \quad j = 1, \ldots, N_\theta = \frac{1}{2}(m+1), \tag{3.177}$$

$$\lambda_k = k\frac{2\pi}{m+1}, \quad k = 0, \ldots, m. \tag{3.178}$$

Then, for $Y \in Harm_{0,\ldots,m} = \bigoplus_{k=0}^m Harm_k$, we have

$$\begin{aligned}\int_\Omega Y(\eta)\, dS(\eta) &= \sum_{j=0}^{\frac{1}{2}(m+1)} \sum_{k=0}^{m} w^\theta_{j,N_\theta} a^\lambda_k Y(\theta_j, \lambda_k) \\ &= \frac{2\pi}{m+1} \sum_{j=0}^{\frac{1}{2}(m+1)} w^\theta_{j,N_\theta} \sum_{k=0}^{m} Y(\theta_j, \lambda_k).\end{aligned} \tag{3.179}$$

Thus, for $m \in \mathbb{N}$ odd and $Y, F \in Harm_{0,\ldots,\frac{m-1}{2}}$, the product YF is of class $Harm_{0,\ldots,m}$, hence, the inner product of Y and F

$$\langle Y, F\rangle_{L^2(\Omega)} = \int_\Omega Y(\eta)F(\eta)\, dS(\eta) \tag{3.180}$$

can be calculated in exact way by use of the Gaussian grid (GG) with $\frac{1}{2}(m+1)^2$ points.

In particular, this result can be applied to the determination of the Fourier coefficients of an arbitrary function $F \in L^2(\Omega)$. In more detail, if the Fourier series expansion of a function $F \in L^2(\Omega)$

$$F \sim \sum_{k=0}^{\infty} \sum_{l=-n}^{n} \langle F, Y_{k,l}\rangle_{\mathrm{L}^2(\mathbb{S}^2)}\, Y_{k,l} \tag{3.181}$$

with the Fourier coefficients given by the usual projection

$$\langle F, Y_{k,l}\rangle_{L^2(\Omega)} = \int_\Omega F(\eta)Y_{k,l}(\eta)\, dS(\eta). \tag{3.182}$$

is replaced by its bandlimited version of degree m (for sufficiently large odd m)

$$F_{0,\ldots,m} = \sum_{k=0}^{m} \sum_{l=-n}^{n} \langle Y_{k,l}, F\rangle_{\mathrm{L}^2(\Omega)}\, Y_{k,l}, \tag{3.183}$$

we are allowed to conclude that

$$\langle F, Y_{k,l}\rangle_{L^2(\Omega)} = \langle F_{0,\ldots,m}, Y_{k,l}\rangle_{L^2(\Omega)},\ k = 0,\ldots,m,\ l = -n,\ldots,n. \tag{3.184}$$

In connection with (3.179) we then obtain with $N_\theta = \frac{1}{2}(m+1)$ the following discrete version

$$F_{0,\ldots,m} = \sum_{k=0}^{m}\sum_{l=-n}^{n}\left(\frac{2\pi}{m+1}\sum_{j=0}^{\frac{1}{2}(m+1)} w_{j,N_\theta}^{\theta}\sum_{k=0}^{m} Y_{k,l}(\theta_j,\lambda_k)\ F(\theta_j,\lambda_k)\right) Y_k^l. \tag{3.185}$$

Recently, a novel cubature formula based on the relation of spherical harmonics to Wigner functions has been proposed by McEwen and Wiaux (for more details see [148]), that achieves the same degree of accuracy as the Gaussian formula with (approximately) half as many points as the aforementioned Driscoll–Healy approach (the details will not be discussed here).

3.7. Latitude-Longitude grids and Weyl's equidistributions

Both the equiangular longitude-latitude grid (ELLG) and the Gauss grid (GG) are constructed by a uniform distribution of points in the two-dimensional interval $[0,\pi]\times[0,2\pi)$. Unfortunately, this means that these pointsets are not equidistributed on the sphere in the sense of Weyl (see [64] for more details). In fact, the "density of points around the poles" on $\mathbb{S}^2$ is comparatively high, while larger distances between the nodes can be observed around the equator. In the case of (ELLG), even a certain number of points coincides with the poles. The Gaussian grid (GG) achieves the same degree of accuracy with (approximately) half as many points as the system (ELLG). However, the zeros of the Legendre polynomials have to be included which is not as simple as for the grid (ELLG). Besides the prescription of the location of nodal points where the measurements must be done, the observation that the integrand is usually not sampled uniformly on the sphere $\mathbb{S}^2$ is of disadvantage except, e.g., for satellite measurements, where the data are acquired in similar way. No doubt, the facts that the approximate integration rules are explicitly known, easily implementable, and economical even for a higher degree m, make latitude-longitude sampling attractive.

The cubature method presented here is only senseful for regular grids generated by partitioning of the two-dimensional interval $[0,\pi)\times[0,2\pi)$. For scattered data distributions this integration technique is not applicable. In this case, other cubature formulas (for example, spline integration as proposed by W. Freeden [38, 39]) have to come into play.

4. From Gaussian theorem to geoidal determination

In 19th century, the forces in physics were believed to be derived from potentials which satisfy Laplace's equation. This explains the term "potential theory". Consequently, in mathematical sense, potential theory was the study of functions that could serve as (approximants of) potentials. Nowadays, we know that the equations that describe forces are systems of non-linear partial differential equations, such as Einstein equations (see, e.g., [1, 146, 225]) and that the Laplace equation is only valid as a limiting case. Nevertheless, the term potential theory has remained as a convenient term for describing the study of functions satisfying the Laplace equation and its generalizations.

4.1. Gauss's role and influence to potential theory

In classical physical geodesy following the law proposed by I. Newton (1643–1727) in 1687 the only forces considered were the forces of mutual attraction acting upon two material particles of small size or two material points. These forces are directly proportional to the product of the masses of these particles and inversely proportional to the square of the distance between them. Thus, the first and the most important problem from the point of view of physical geodesy was to study the forces of attraction of a material point by a finite (regular) material body – a spheroid and, in particular, an ellipsoid (since many celestial bodies have this shape). After first achievements by Newton and others, studies carried out by J.L. Lagrange (1736–1813) in 1773, A. Legendre (1752–1833) between 1784–1794 and by P.S. Laplace (1749–1827) continued in 1782–1799 became of major importance. Lagrange established that the field of gravitational forces, as it is called now, is a (gradient) potential field. He introduced a function which was called in 1828 by G. Green (1793–1841) a potential function and later in 1840 by C.F. Gauss just a potential.

Already Gauss, Green, and their contemporaries discovered that the method of potentials can be applied not only to solve problems in the theory of gravitation but, in general, to solve a wide range of problems in mathematical physics, in particular, in electrostatics and magnetism. In this connection, potentials became to be considered not only for the physically realistic problems concerning mutual attraction of positive masses, but also for problems with "masses of arbitrary sign", or charges. Representative boundary value problems were defined in potential theory, such as the Dirichlet problem and the Neumann problem, the electrostatic problem of the static distribution of charges on conductors or the Robin problem, and the problem of sweeping-out mass (balayage method). To solve the aforementioned problems in the case of domains with sufficiently smooth boundaries certain types of potentials turned out to be efficient, i.e., special classes of parameter-dependent integrals such as volume potentials of distributed mass, single – and double layer potentials, Green potentials, etc. Results obtained by A.M Lyapunov (1857–1918) and V.A. Steklov (1864–1926) at the end of 19th century were fundamental for the creation of strong methods of the solution of the

main boundary value problems. Studies in potential theory concerning properties of different potentials have acquired an independent significance. In the first half of the 20th century, a great stimulus for the generalization of the principal problems and the completion of the existing formulations in potential theory was made on the basis of the general notions of a Radon measure, capacity, and generalized functions. Nowadays, potential theory is closely related in its development to the theory of analytic functions and to some extend to the probability theory (see, e.g., [1, 225]). Together with further studies of classical boundary value problems and inverse problems, the modern period of the development of potential theory is usually characterized by the application of methods and notions of topology and functional analysis, and the use of abstract axiomatic methods.

4.2. Vector analytic and potential theoretic tools

Newton's law of universal gravitation, first published in his Principia in 1687, asserts that the force v exerted on a point mass Q at $x \in \mathbb{R}^3$ by a system of finitely many point masses q_i at $y_i \in \mathbb{R}^3, i = 1, \ldots, N$, is equal to

$$x \mapsto v(x) = \sum_{i=1}^{N} \frac{Cq_iQ}{|x-y_i|^2} \frac{x-y_i}{|x-y_i|}, \quad x \neq y_i,\ i = 1, \ldots, N, \tag{4.1}$$

with a constant $C < 0$ (like masses attract). The same law of interaction between point charges was discovered experimentally by C.A. de Coulomb (1736–1806) and announced in 1785, now with $C > 0$ (like charges repel). Note that the numerical value of the constant C depends on the unit system one is using to measure force, mass (or charge), and distance. After the introduction of the function

$$x \mapsto V(x) = \sum_{i=1}^{N} \frac{Cq_i}{|x-y_i|}, \quad x \neq y_i,\ i = 1, \ldots, N, \tag{4.2}$$

into the theory of gravitation by D. Bernoulli in 1748, J.-L. Lagrange noticed in 1773 that

$$v(x) = Q\ \nabla V(x), \quad x \neq y_i,\ i = 1, \ldots, N. \tag{4.3}$$

Hence the function V completely describes the gravitational (or electrostatic) field.

For a continuous distribution of charges with density ρ, vanishing outside $\overline{\mathcal{G}}$, the potential becomes

$$V(x) = \frac{C}{4\pi} \int_{\mathcal{G}} \rho(y)\ \frac{1}{|x-y|}\ dy, \quad x \in \mathcal{G}^c, \tag{4.4}$$

where dy is the volume element and $\mathcal{G}^c = \mathbb{R}^3 \backslash \overline{\mathcal{G}}$ is the outer space of $\mathcal{G}$. As observed by P.S. Laplace in 1782, the function $G(\Delta; |\cdot - y|)$, $y \in \mathbb{R}^3$, given by

$$G(\Delta; |x-y|) = \frac{1}{4\pi} \frac{1}{|x-y|}, \quad x \neq y, \tag{4.5}$$

in today's jargon called the *fundamental solution of the Laplace equation* (4.6) in $\mathbb{R}^3 \backslash \{y\}$, satisfies

$$\Delta_x G(\Delta; |x-y|) = 0,\ x \in \mathbb{R}^3 \backslash \{y\}. \tag{4.6}$$

Later, the solutions of the Laplace equation came to be known as harmonic functions. It should, however, be remarked that the Laplace equation had been also considered by Lagrange in 1760 in connection with his study of fluid flow problems. Laplace's result was completed by his student S.D. Poisson (1781–1840) in 1813, when he showed that $\Delta V = -C\rho$ for smooth enough densities ρ.

We summarize these classical results in today's mathematical formalism:

Theorem 4.1. *Suppose that $\mathcal{G}$ is a regular region in $\mathbb{R}^3$ (as introduced earlier in Section 2).*

(1) *Let $F : \overline{\mathcal{G}} \to \mathbb{R}$ be an integrable bounded function. Then*

$$V(x) = C \int_{\mathcal{G}} \rho(y)\ G(\Delta; |x-y|)\ dy, \quad x \in \mathcal{G}^c, \tag{4.7}$$

satisfies

$$\Delta_x C \int_{\mathcal{G}} \rho(y)\ G(\Delta; |x-y|)\ dy = 0 \tag{4.8}$$

for all $x \in \mathcal{G}^c$, i.e., V is harmonic in $\mathcal{G}^c$.

(2) *Let $\rho : \overline{\mathcal{G}} \to \mathbb{R}$ be of class $C^{(0)}(\overline{\mathcal{G}})$. Then V as defined by the volume integral in (4.7) is of class $C^{(0)}(\overline{\mathcal{G}})$. Furthermore, we have*

$$\nabla_x V(x) = C \int_{\mathcal{G}} \rho(y)\ \nabla_x G(\Delta; |x-y|)\ dy, \quad x \in \overline{\mathcal{G}}. \tag{4.9}$$

Moreover, the so-called Poisson equation under the assumption of μ-Hölder continuity, $\mu \in (0, 1]$, can be formulated as follows:

Theorem 4.2. *If ρ is of class $C^{(0,\mu)}(\overline{\mathcal{G}})$, $\mu \in (0, 1]$, then the Poisson differential equation*

$$-\Delta_x C \int_{\mathcal{G}} \rho(y)\ G(\Delta; |x-y|)\ dy = C\rho(x) \tag{4.10}$$

holds true for all $x \in \mathcal{G}$.

In Theorem 4.2 the assumption of μ-Hölder continuity of ρ, $\mu \in (0, 1]$, is needed for its proof. As a matter of fact, H. Petrini [173] showed that the μ-Hölder continuity of ρ, $\mu \in (0, 1]$, is necessary to imply the second continuous differentiability of the Newton volume potential.

The Gauss integral theorem (from 1813) and the related Green formulas (cf. [91]) are among the basic tools of potential theory. They are also indispensable for a variety of problems in physical geodesy (cf. [107, 112, 159]).

Theorem 4.3 (Gauss's Integral Theorem). *Let $\mathcal{G}$ be a regular region. Let $F : \overline{\mathcal{G}} \to \mathbb{R}$ be a scalar field, $f : \overline{\mathcal{G}} \to \mathbb{R}^3$ a vector field, that is continuous on $\overline{\mathcal{G}}$ and differen-*

tiable in $\mathcal{G}$, respectively. Then

$$\int_{\partial\mathcal{G}} F(y)\nu(y)\ dS(y) = \int_{\mathcal{G}} \nabla_y F(y)\ dy, \tag{4.11}$$

$$\int_{\partial\mathcal{G}} f(y)\cdot\nu(y)\ dS(y) = \int_{\mathcal{G}} \nabla_y \cdot f(y)\ dy, \tag{4.12}$$

$$\int_{\partial\mathcal{G}} \nu(y)\wedge f(y)\ dS(y) = \int_{\mathcal{G}} \nabla_y \wedge f(y)\ dy, \tag{4.13}$$

provided that the integrand on the right-hand side is Lebesgue-integrable on $\mathcal{G}$. The vector field $\nu : \partial\mathcal{G} \to \mathbb{R}^3$ is the (unit) normal field pointing into the exterior of $\mathcal{G}$ (dS is the surface element).

The identities (4.12) and (4.13) are valid for all vector fields, whatever their physical meaning is. Of special interest is the case (4.12) in which f may be understood to be the velocity vector of an incompressible fluid. Inside the surface $\partial\mathcal{G}$ there may be sources in which the fluid is generated or sinks in which the fluid is annihilated. The divergence $\nabla\cdot f$ measures the strength of the sources and sinks. The volume integral $\int_{\mathcal{G}} \nabla\cdot f(y)\ dy$ is the total amount of the fluid generated in unit time. The surface integral $\int_{\partial\mathcal{G}} f(y)\cdot\nu(y)\ dS(y)$ is the total amount of fluid flowing in unit time across the surface $\partial\mathcal{G}$. Therefore, the Gauss formula expresses a balance equation, namely the evident fact that both integrals in (4.12) are equal.

Gravitational Interpretation. In the case where f is the vector of the gravitational force, i.e., we especially choose instead of f the field $v = \nabla V$, the intuitive interpretation of the Gauss integral theorem is not so obvious, but the analogy to the balance equation of fluid flow is often helpful. In gravitation we can take advantage of the Poisson equation

$$\nabla\cdot v = \Delta V = -C\rho. \tag{4.14}$$

This equation (cf. [107]) can be interpreted to mean that the masses are the sources of the gravitational field; the strength or the sources, $\nabla\cdot v$, is proportional to the mass density ρ. The right-hand side of (4.14) is called the flux of force, in our case gravitational flux, also in analogy to the fluid flow.

Next we come to the *interior Green formulas* for regular regions $\mathcal{G} \subset \mathbb{R}^3$. Suppose that $f = \nabla F$, where $F \in C^{(1)}(\overline{\mathcal{G}})\cap C^{(2)}(\mathcal{G})$, i.e., $F : \overline{\mathcal{G}} \to \mathbb{R}$ is continuously differentiable in $\overline{\mathcal{G}}$ and $F|\mathcal{G}$ is twice continuously differentiable in $\mathcal{G}$. Let ΔF be Lebesgue-integrable in $\mathcal{G}$. Then we obtain from the Gauss theorem (4.12)

$$\int_{\partial\mathcal{G}} \frac{\partial F}{\partial\nu}(y)\ dS(y) = \int_{\mathcal{G}} \Delta F(y)\ dy, \tag{4.15}$$

where, as always, $\frac{\partial}{\partial\nu} = \nu\cdot\nabla$ denotes the derivative in the direction of the outer (unit) normal field ν.

Under the special choice $f = F\ \nabla G$ the Gauss Theorem yields

Theorem 4.4 (Interior First Green Theorem). *Suppose that $\mathcal{G} \subset \mathbb{R}^3$ is a regular region. For $F \in C^{(1)}(\overline{\mathcal{G}})$, $G \in C^{(1)}(\overline{\mathcal{G}}) \cap C^{(2)}(\mathcal{G})$ with ΔG Lebesgue-integrable on $\mathcal{G}$ we have*

$$\int_{\mathcal{G}} (F(y)\Delta G(y) + \nabla F(y) \cdot \nabla G(y))\ dy = \int_{\partial\mathcal{G}} F(y) \frac{\partial G}{\partial \nu}(y)\ dS(y). \tag{4.16}$$

Taking $f = F\ \nabla G - G\ \nabla F$ we finally obtain

Theorem 4.5 (Interior Second Green Theorem). *Suppose that $\mathcal{G} \in \mathbb{R}^3$ is a regular region. For $F, G \in C^{(1)}(\overline{\mathcal{G}}) \cap C^{(2)}(\mathcal{G})$ with $\Delta F, \Delta G$ Lebesgue-integrable on $\mathcal{G}$ we have*

$$\begin{aligned} &\int_{\mathcal{G}} (G(y)\Delta F(y) - F(y)\Delta G(y))\ dV(y) \\ &\quad = \int_{\partial\mathcal{G}} \left(G(y) \frac{\partial F}{\partial \nu}(y) - F(y) \frac{\partial G}{\partial \nu}(y) \right)\ dS(y). \end{aligned} \tag{4.17}$$

In what follows, we collect some basic material well-known from classical potential theory in the Euclidean space $\mathbb{R}^3$. First we have a closer look at the fundamental solution of the Laplace equation. Observing its specific properties we are able to formulate the third interior Green formula. Mean value theorems and maximum/minimum principle are the canonical consequences. Harmonic functions are recognized to be analytic in their harmonicity domain. The Kelvin transform enables us to study harmonic functions which are regular at infinity. Keeping the regularity at infinity in mind we are finally led to exterior Green formulas. The third exterior Green formula is formulated in analogy to its interior counterpart (for the proof the reader is referred to, e.g., [49]). Essential tools are the limit relations

$$\lim_{r \to 0+} \int_{\Omega_r(x)} F(y)\ \frac{x-y}{|x-y|} \cdot \nabla_y G(\Delta; |x-y|)\ dS(y) = F(x), \tag{4.18}$$

$$\lim_{r \to 0+} \int_{\Omega_r(x)} F(y)\ G(\Delta; |x-y|)\ dS(y) = 0 \tag{4.19}$$

for a continuous functions F on $\mathbb{B}_\rho(x)$, $\rho > 0, x \in \mathbb{R}^3$. In fact, from (4.18), (4.19), we easily obtain in combination with the Second Green Theorem.

Theorem 4.6 (Interior Third Green Theorem). *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region with continuously differentiable boundary $\partial\mathcal{G}$. Suppose that $F : \overline{\mathcal{G}} \to \mathbb{R}$ is of class $C^{(1)}(\overline{\mathcal{G}}) \cap C^{(2)}(\mathcal{G})$ with ΔF Lebesgue-integrable on $\mathcal{G}$. Then*

$$\begin{aligned} &\int_{\partial\mathcal{G}} \left(G(\Delta; |x-y|) \frac{\partial}{\partial \nu(y)} F(y) - F(y) \frac{\partial}{\partial \nu(y)} G(\Delta; |x-y|) \right)\ dS(y) \\ &\qquad - \int_{\mathcal{G}} G(\Delta; |x-y|) \Delta F(y)\ dy) = \alpha(x) F(x), \end{aligned} \tag{4.20}$$

where $\alpha(x)$ id solid angle $\alpha(x)$ subtended by the boundary $\partial\mathcal{G}$ at the point $x \in \mathbb{R}^3$

Especially, for a continuously differentiable boundary $\partial\mathcal{G}$, we obtain

Corollary 4.1 (Interior Third Green Theorem). *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region with continuously differentiable boundary $\partial\mathcal{G}$. Suppose that $F : \overline{\mathcal{G}} \to \mathbb{R}$ is of class $C^{(1)}(\overline{\mathcal{G}}) \cap C^{(2)}(\mathcal{G})$ with ΔF Lebesgue-integrable on $\mathcal{G}$. Then*

$$\int_{\partial\mathcal{G}} \left(G(\Delta; |x-y|) \frac{\partial}{\partial\nu(y)} F(y) - F(y) \frac{\partial}{\partial\nu(y)} G(\Delta; |x-y|) \right) dS(y)$$
$$- \int_{\mathcal{G}} G(\Delta; |x-y|) \Delta F(y)\, dy) = \begin{cases} F(x), & x \in \mathcal{G}, \\ \frac{1}{2} F(x), & x \in \partial\mathcal{G}, \\ 0, & x \in \mathcal{G}^c. \end{cases} \tag{4.21}$$

As special case we obtain for continuously differentiable functions F in $\overline{\mathcal{G}}$ which are harmonic in $\mathcal{G}$ the so-called *Interior Fundamental Theorem.*

Corollary 4.2. *Suppose that $F : \overline{\mathcal{G}} \to \mathbb{R}$ is of class $C^{(1)}(\overline{\mathcal{G}}) \cap C^{(2)}(\mathcal{G})$ with $\Delta F = 0$ on $\mathcal{G}$. Then*

$$\int_{\partial\mathcal{G}} \left(G(\Delta; |x-y|) \frac{\partial F}{\partial\nu(y)}(y) - F(y) \frac{\partial}{\partial\nu(y)} G(\Delta; |x-y|) \right) dS(y)$$
$$= \begin{cases} F(x), & x \in \mathcal{G}, \\ \dfrac{1}{2} F(x), & x \in \partial\mathcal{G}, \\ 0, & x \in \mathcal{G}^c. \end{cases} \tag{4.22}$$

These formulas, which turn out to be the point of departure for the limit and jump relations in potential theory (see, e.g., [49, 121, 191]), are also due to Gauss.

Letting $F = 1$ in $\overline{\mathcal{G}}$, we obviously find in connection with (4.5) and Corollary 4.2 the following definition (cf. (2.33)).

Definition 4.1 (Solid Angle). Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region. Then the solid angle $\alpha(x)$ subtended by the boundary $\partial\mathcal{G}$ at the point $x \in \mathbb{R}^3$ is given by

$$\alpha(x) = - \int_{\partial\mathcal{G}} \frac{\partial}{\partial\nu(y)} G(\Delta; |x-y|)\, dS(y). \tag{4.23}$$

Note that we have

$$a(x) = \begin{cases} 1 & , \quad x \in \mathcal{G} \\ \frac{1}{2} & , \quad x \in \partial\mathcal{G} \\ 0 & , \quad x \notin \overline{\mathcal{G}}. \end{cases} \tag{4.24}$$

provided that $\mathcal{G} \subset \mathbb{R}^q$ is a regular region with continuously differentiable boundary $\partial\mathcal{G}$ (cf. Figure 4.1). In the case of the cube $\mathcal{G} = (-1, 1)^3 \subset \mathbb{R}^3$ we especially have (i) $\alpha(x) = 1$ if x is located in the open cube $\mathcal{G}$, (ii) $\alpha(x) = \frac{1}{2}$ if x is located on one of the six faces of the boundary $\partial\mathcal{G}$ of the cube $\mathcal{G}$ but not on an edge or in a vertex, (iii) $\alpha(x) = \frac{1}{4}$ if x is located on one of the eight edges of $\partial\mathcal{G}$ but not

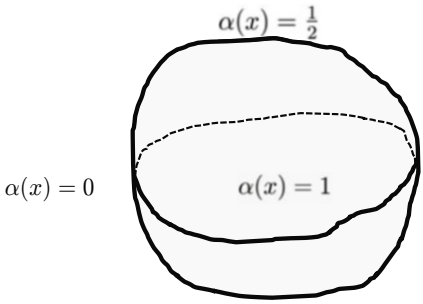


FIGURE 4.1. Solid angle subtended at $x \in \mathbb{R}^3$ by the surface $\partial\mathcal{G}$ of a regular region $\mathcal{G}$ with "smooth boundary".

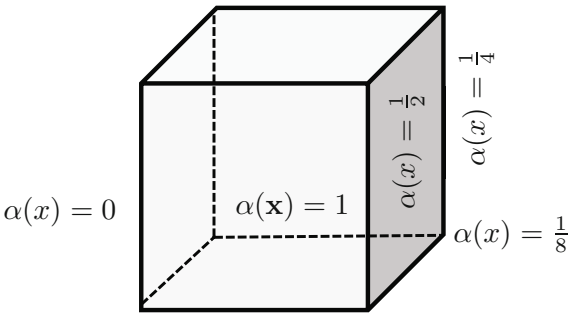


FIGURE 4.2. Solid angle subtended at $x \in \mathbb{R}^3$ by the surface $\partial\mathcal{G}$ of the "non-smooth" cube $\mathcal{G} = (-1, 1)^3$.

in a vertex, (iv) $\alpha(x) = \frac{1}{8}$ if x is located in one of the eight vertices of $\partial\mathcal{G}$ (cf. Figure 4.2).

It should be remarked that the divergence theorem first appeared in Lagrange's 1860 posthumous work, and it was proved in a special case already by Gauss in 1813. The general three-dimensional case was treated by M.V. Ostrogradsky in 1826. In a preliminary section of his groundbreaking 1828 essay, George Green proved several reductions of three-dimensional volume integrals to surface integrals, similar in spirit to the divergence theorem, and independently of M.V. Ostrogradsky. Nowadays, those are called Green's identities and best viewed as consequences of the Gauss integral theorem.

Next we are concerned with the *mean value theorem* (in today's mathematical language) that dates back to C.F. Gauss (1840).

Theorem 4.7. *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region. Then the following statements are equivalent:*

(a) *$U : \mathcal{G} \to \mathbb{R}$ is harmonic in $\mathcal{G}$, i.e., $U \in C^{(2)}(\mathcal{G})$ and $\Delta U = 0$ in $\mathcal{G}$,*

(b) $U : \mathcal{G} \to \mathbb{R}$ *possesses the* Mean Value Property *on* $\mathcal{G}$*, i.e.,* U *is of class* $C^{(0)}(\mathcal{G})$ *and for all* $x \in \mathcal{G}$ *and all* $r > 0$ *with* $\mathbb{B}_r(x) = \{z \in \mathbb{R}^3 : |z| < r\} \Subset \mathcal{G}$

$$U(x) = \frac{1}{4\pi r^2} \int_{|x|=r} U(y)\, dS(y), \tag{4.25}$$

(c) U *is of class* $C^{(0)}(\overline{\mathcal{G}})$ *and for all* $r > 0$ *with* $\mathbb{B}_r(x) \Subset \mathcal{G}$

$$\int_{|x|\leq r} (U(x) - U(y))\, dy = 0. \tag{4.26}$$

A central result in the theory of harmonic functions is the Maximum/Minimum Principle. Essential tool is the Mean Value Property.

Theorem 4.8 (Maximum/Minimum Principle). *Let* $\mathcal{G} \subset \mathbb{R}^3$ *be a regular region. Suppose that* U *is harmonic in* $\mathcal{G}$ *and non-constant. Then* U *does not reach its minimum or maximum in* $\mathcal{G}$*. If, in addition,* U *is of class* $C^{(0)}(\overline{\mathcal{G}})$*, then* U *reaches its minimum and maximum in* $\overline{\mathcal{G}}$*, and the extremal points are lying on* $\partial\mathcal{G}$*. More precisely,*

$$\sup_{x\in\overline{\mathcal{G}}} |U(x)| \leq \sup_{x\in\partial\mathcal{G}} |U(x)|. \tag{4.27}$$

A direct consequence of the Maximum/Minimum Principle is the following *stability theorem.*

Theorem 4.9. *Let* $\mathcal{G} \subset \mathbb{R}^3$ *be a regular region. Suppose that* U *and* V *are of class* $C^{(0)}(\overline{\mathcal{G}}) \cap C^{(2)}(\mathcal{G})$*, and harmonic in* $\mathcal{G}$*. Let* ε *be an arbitrary positive number. If*

$$\sup_{x\in\partial\mathcal{G}} |U(x) - V(x)| \leq \varepsilon, \tag{4.28}$$

then

$$\sup_{x\in\overline{\mathcal{G}}} |U(x) - V(x)| \leq \varepsilon. \tag{4.29}$$

Now we are prepared to establish the (real) analyticity of harmonic functions.

Theorem 4.10 (Analyticity). *Let* $\mathcal{G} \subset \mathbb{R}^3$ *be a regular region. Suppose that* U *is harmonic on* $\mathcal{G}$*. Then* U *is (real) analytic, i.e., for* $x_0 \in \mathcal{G}$ *there exists* $\rho > 0$ *such that*

$$U(x_0 + h) = \sum_{j=0}^{\infty} \frac{1}{j!} ((h \cdot \nabla)^j U)(x_0) \tag{4.30}$$

for all $h \in \mathbb{R}^3$ *with* $|h| < \rho$*.*

The Newton (volume) potential extended over $\mathcal{G}$ is harmonic in the exterior $\mathcal{G}^c = \mathbb{R}^3 \backslash \overline{\mathcal{G}}$. This is the reason why potential theory under geoscientifically relevant aspects essentially aims at concepts in the outer space of a regular region. The treatment of the outer space in the Euclidean space $\mathbb{R}^3$, however, includes the discussion at infinity. As a consequence, Green's integral theorems must be formulated under geophysically relevant conditions imposed on harmonic functions

at infinity. Mathematically (see, e.g., [49]), the "regularity at infinity" can be deduced via the Kelvin transform by a transition from functions harmonic in the inner space to their counterparts in outer space, and vice versa.

Theorem 4.11. *If U is harmonic in $\mathcal{G}^c$ and U converges to zero for $|x| \to \infty$ uniformly with respect to all directions, then $|x||U(x)|$ and $|x|^2|\nabla U(x)|$ are bounded for $|x| \to \infty$.*

Theorem 4.11 leads us to the definition of the *"regularity at infinity"*.

Definition 4.2. A function $U : \mathcal{G}^c \to \mathbb{R}$ is called regular at infinity, if U satisfies the asymptotic relation $|U(x)| = O(|x|^{-1})$ and $|\nabla U(x)| = O(|x|^{-2})$, $|x| \to \infty$, uniformly with respect to all directions $x/|x|$.

Now we are prepared to discuss exterior versions of the Green identities involving harmonic functions being regular at infinity. All these identities can be obtained by first considering the auxiliary set $\mathcal{G}_R^c(0) = \mathcal{G}^c \cap \mathbb{B}_R(0)$ (with R sufficiently large such that $\mathcal{G} \Subset \mathbb{B}_R(0)$, i.e., $\overline{\mathcal{G}} \subset \mathbb{B}_R(0)$ and $\operatorname{dist}(\partial\mathcal{G}, \partial\mathbb{B}_R(0)) > 0$) and afterwards letting R tend to infinity (note that $\mathcal{G}_R^c(0)$ as the difference of the two regular regions $\mathbb{B}_R(0)$ and $\mathcal{G}$ allows the application of the interior Green formulas).

Theorem 4.12 (Exterior First Green Theorem). *Let F be a function of class $C^{(2)}(\mathcal{G}^c) \cap C^{(1)}(\overline{\mathcal{G}^c})$ such that F is harmonic in $\mathcal{G}^c$ and regular at infinity. Suppose that the function $H \in C^{(1)}(\overline{\mathcal{G}^c})$ satisfies the asymptotic relations*

$$|y|^2 \, |F(y)\nabla H(y)| = o(1) \tag{4.31}$$

and

$$|\nabla F(y) \cdot \nabla H(y)| = O\left(\frac{1}{|y|^{3+\varepsilon}}\right), \ \varepsilon > 0. \tag{4.32}$$

Then

$$\int_{\mathcal{G}^c} \nabla F(y) \cdot \nabla H(y) \, dy = \int_{\partial\mathcal{G}} F(y) \frac{\partial H}{\partial \nu}(y) \, dS(y), \tag{4.33}$$

where ν is the outer unit normal field to $\mathcal{G}^c$, i.e., the inner unit normal field to $\mathcal{G}$.

Theorem 4.13 (Exterior Second Green Theorem). *Let the function*

$$F, G \in C^{(1)}(\overline{\mathcal{G}^c}) \cap C^{(2)}(\mathcal{G}^c)$$

be harmonic in $\mathcal{G}^c$ and regular at infinity. Then

$$\int_{\partial\mathcal{G}} \left(F(y) \frac{\partial}{\partial \nu} H(y) - H(y) \frac{\partial}{\partial \nu} F(y) \right) \, dS(y) = 0. \tag{4.34}$$

Theorem 4.14 (Exterior Third Green Theorem). *Suppose that $\mathcal{G}$ is a regular region with continuously differentiable boundary $\partial\mathcal{G}$. Let U be of class $C^{(1)}(\overline{\mathcal{G}^c}) \cap C^{(2)}(\mathcal{G}^c)$*

such that U is harmonic in $\mathcal{G}^c$ and regular at infinity. Then

$$\int_{\partial\mathcal{G}} \left(G(\Delta; |x-y|) \frac{\partial}{\partial \nu(y)} U(y) - U(y) \frac{\partial}{\partial \nu(y)} G(\Delta; |x-y|) \right) dS(y) = \begin{cases} U(x) & , \quad x \in \mathcal{G}^c \\ \frac{1}{2} U(x) & , \quad x \in \partial\mathcal{G} \\ 0 & , \quad x \in \mathcal{G}, \end{cases} \tag{4.35}$$

where ν is the outer unit normal field to $\mathcal{G}^c$, i.e., the inner unit normal field to $\mathcal{G}$, and $\alpha(x)$ is the solid angle subtended by the boundary $\partial\mathcal{G}$ at $x \in \mathbb{R}^3$.

4.3. Gravitational potential

Since the time of Newton, the theories for the precise determination of the attraction of the Earth, exerting on a point of the Earth's surface, did have a special significance for geodesy. These investigations are especially connected to the names of Huygens, MacLaurin, d'Alembert, Lagrange, Legendre, and Laplace as well as to their scholars Poisson and Plana (cf. [216]). In his treatise about his investigations on potential theory concerning homogeneous ellipsoids (Theoria attractonis corporum sphaeroidorum ellipticorum homogeneorum methodo nova tractata, Vol. II, Göttingen, [1813]),, presented to the "Königliche Gesellschaft der Wissenschaften zu Göttingen" (Royal Society of Sciences at Göttingen), in which Gauss derives the well-known his name bearing "Gauss's Integral Theorem" (divergence theorem) about the transformation of a volume integral in a surface integral, he writes:

> *"The solution of Laplace deserves because of its elegance and its sagacity common admiration; but just the fact that it was necessary to use special finesses and artifices in order to solve the problem, may among the geometricians awake the desire for a simpler, less complicated and more direct solution. This desire did not completely come true by a new proof of the main sentence given by Legendre (Hist. de l'acad. roy. des sc.* 1788*, Sur les intégrales doubles), though the here shown extraordinary analytical skilfulness was by all geometricians recognized. Later Biot and Plana have tried to simplify the solution (Mm. de l'institut T. VI; Memorie di matematica e di fisica della societ. italiana T. XV). But also these two solutions belong to one of the most complicated applications of analysis. We hope that a new solution will not be unwelcome to mathematicians and astronomers, which will use a quite different way."*

As a consequence of the historic prework, the determination of the Earth's shape and its *gravity potential* are considered to be main tasks of physical geodesy. Both problems relate to the work of G.G. Stokes [207]. Nowadays, they can be tackled simultaneously by the so-called *Molodensky problem* (cf. [151]). In order to understand Molodensky's setup, we have to start once more with the *gravitational potential* V of the Earth, related to its mass by the Newton volume integral (4.4). As a consequence of the Gauss Theorem we already know that V is harmonic in the outer space, i.e., $\Delta V = 0$ in $\mathcal{G}^c$. Moreover, V is regular at infinity. As the Earth

is rotating, we can also assign a *centrifugal potential* Φ to the rotating body. This results in the so-called *gravity potential* W given by

$$W = V + \Phi. \tag{4.36}$$

The gradient of the gravity potential is the *gravity vector*

$$w = \nabla W. \tag{4.37}$$

The magnitude of w, in geodesy usually denoted by $g = |w|$ and simply called the *gravity* (*intensity*), can be obtained on the surface $\partial\mathcal{G}$ of the Earth from gravimetric measurements. The direction of w gives the direction of the *plumb line* and can be obtained from astronomic observations and today also from satellite measurements. Plumb lines are not straight, but intersect each equipotential surface of W normally, such that the gravity vector w at any given point is tangential to the plumb line at this point.

Combining leveling with the gravimetric and astronomic measurements which determine w allows us to get W on the surface up to an additional constant which can be determined from additionally knowing at least one distance. All data sets are assumed to be corrected for influences like gravitational potentials of other celestial bodies or the Earth's precession and atmosphere.

4.4. Geodetic boundary value problems

L. Hörmander [113] in his seminal work stated the situation related to Molodensky's problem as the following idealized setting (where we follow almost literally the approach described in [3]):

(i) The Earth $\overline{\mathcal{G}}$ is a rigid body rotating with a known constant angular speed ω around a fixed axis, which we choose as the x_3-axis. The centrifugal potential is given by

$$\Phi(x) = \frac{1}{2}\omega^2(x_1^2 + x_2^2), \quad x = (x_1, x_2, x_3)^T. \tag{4.38}$$

(ii) The center of gravity is the origin 0 of our (Earth's fixed) system of coordinates.
(iii) The gravity vector w is known at every point P of $\partial\mathcal{G}$.
(iv) The gravity potential W is known at every point P of $\partial\mathcal{G}$.
(v) $\partial\mathcal{G}$ can be mapped to the unit sphere $\Omega = \{x \in \mathbb{R}^3 : |\boldsymbol{x}|^2 = 1\}$, i.e., there is a differentiable embedding $\varsigma : \Omega \to \mathbb{R}^3$ such that $\varsigma(\Omega) = \partial\mathcal{G}$.

In the Molodensky problem, we are looking for the unknown gravity potential W in the whole space $\mathbb{R}^3$ and the unknown embedding ς. As the gravitational potential V is a harmonic function, i.e., $\Delta V = 0$ outside the Earth, we obtain

$$\Delta W = \Delta(V + \Phi) = \Delta V + \Delta\Phi = 2\omega^2 \tag{4.39}$$

outside $\partial\mathcal{G}$. The assumption of having the barycenter at the origin further on imposes on V that the asymptotic relation

$$V(x) = \frac{C}{|x|} + O(|x|^{-3}), \quad |x| \to \infty, \tag{4.40}$$

holds true, which can be seen by taking a multipole expansion into account.

All in all, Molodensky's problem is a non-linear free-boundary problem and, therefore, hard to solve mathematically. As a consequence, a variety of approximate methods exist. Usually, Molodensky's approach is modified via linearization. In fact, linearization and a sophisticated iterative process which avoids the loss of regularity are essential in treating Molodensky's problem (note that we do not cover here the iteration procedures as proposed by L. Hörmander [113]).

Linearizing Molodensky's problem amounts to the introduction of two ingredients:

(α) an approximate surface $\partial\mathcal{T}$ for the Earth, called *telluroid*,
(β) an approximate potential U, called *normal potential*.

Classically, in the geodetic context, the following steps are carried out:

(1) Choose a reference body $\mathcal{E}$, usually an ellipsoid, which lies completely inside the Earth and rotates with the same angular speed ω as the Earth around the x_3-axis.
(2) For every point P on $\partial\mathcal{G}$, find a point Q_0 on the surface $\partial\mathcal{E}$ such that $x_P - x_{Q_0}$ is normal to $\partial\mathcal{E}$. x_P and x_{Q_0} are the position vectors of P and Q.
(3) Determine the normal potential U such that
   – its gravitational potential part is caused by a mass identical to the Earth's mass,
   – its centrifugal potential part is identical to (4.38),
   – the reference surface $\partial\mathcal{E}$ is an equipotential surface of U.

   Note that U can be calculated explicitly as its centrifugal part is known and its gravitational part is the solution to an exterior Dirichlet problem with boundary values given on $\partial\mathcal{E}$ (see, e.g., [94, 107]).
(4) Compute the gradient ∇U of U, called *normal gravity vector field* and denoted by $u = \nabla U$ with the magnitude $\gamma = |u|$ called *normal gravity*. As $\partial\mathcal{E}$ is an equipotential surface of U, u is normal to $\partial\mathcal{E}$ for every point on $\partial\mathcal{E}$, i.e., $u(x_{Q_0})$ is parallel to $x_P - x_{Q_0}$.
(5) For every point P on the real Earth surface $\partial\mathcal{G}$ choose a point Q according to one of the following conditions:
   (A) Q lies on the line between P and its corresponding point Q_0 on the surface $\partial\mathcal{E}$ so that

$$W(x_P) = U(x_Q), \tag{4.41}$$

   (B)

$$\frac{w(x_P)}{|w(x_P)|} = \frac{u(x_Q)}{|u(x_Q)|} \tag{4.42}$$

   and

$$W(x_P) = U(x_Q), \tag{4.43}$$

(C) $$w(x_P) = u(x_Q). \tag{4.44}$$

All points Q chosen in this way make up the *telluroid* $\partial\mathcal{T}$. Each of the above conditions gives a slightly different telluroid. None of them is an equipotential surface, neither for W, nor for U. Condition (A) is the most popular one and the one originally used by Molodensky, whereas (B) is theoretically more correct and (C) seems to be better adapted to the so-called *Marussi condition* (cf. [113, 131]). The Marussi condition says that the Jacobi matrix of u, i.e. the Hessian of U, should be non-singular at every point Q, i.e.,

$$\det\left(\left(\frac{\partial^2 U(x)}{\partial x_i\,\partial x_j}\right)_{i,j=1}^{3}\right)\Bigg|_{x=x_Q} \neq 0 \tag{4.45}$$

for all points Q on the telluroid $\partial\mathcal{T}$. We shall see later why this condition is needed.

In principle, the telluroid can be chosen by any surface as long as there is a one-to-one mapping between $\partial\mathcal{G}$ and $\partial\mathcal{T}$ (in this respect it should be remarked that the geoid may be assumed to be known with an accuracy of about one centimeter or less). In order to be suitable as a point of departure in the context of linearization, $\partial\mathcal{T}$ should be close to $\partial\mathcal{G}$ (in some sense) and chosen in a way that brings advantages during the following process of linearization (note that a more correct notation is to write $Q_0(P)$ and $Q(P)$ as the points Q_0 and Q are dependent on P. We will do so whenever it may help to avoid any confusion).

The introduction of the normal gravity field u also suggests the definition of so-called *normal plumb lines*, i.e., lines which intersect each equipotential surface of u normally, such that the normal gravity vector u at any given point is tangential to the normal plumb line at this point.

The normal potential has been well established in geodesy long before Molodensky, whereas other surfaces have been in use for a long time before the telluroid, e.g., the *geoid* $\partial\mathcal{G}$. As mentioned above, $\partial\mathcal{E}$ is an equipotential surface of the normal potential U. Denoting the value of U on $\partial\mathcal{E}$ by U_0, the geoid $\partial\mathfrak{G}$ is defined as the equipotential surface of W for which we have $W = U_0$ on $\partial\mathfrak{G}$.

Points on the geoid $\partial\mathfrak{G}$ can be related to points on $\partial\mathcal{G}$. We denote by $P_0 \in \partial\mathcal{G}$ the point related to $P \in \partial\mathcal{G}$. We can determine P_0 from P by moving along the plumb line from P towards the center of the Earth until we reach the geoid. Another possibility would be to use the normal plumb line to reach the geoid or to choose P_0 as the point on the geoid that lies also on the line between $P \in \partial\mathcal{G}$ and the corresponding $Q_0 \in \partial\mathcal{E}$. With an appropriate choice of $\partial\mathcal{E}$ and the normal potential U, all of these methods yield almost the same point P_0. We define the *distance vector* between $P_0(P)$ and $Q_0(P)$ as

$$d(P) = x_{P_0}(P) - x_{Q_0}(P). \tag{4.46}$$

Its magnitude is the *geoidal undulation*, in the geodetic context denoted by

$$N(P) = |d(P)| = |x_{P_0}(P) - x_{Q_0}(P)|\,. \tag{4.47}$$

Linearization of Molodensky's Problem. In order to linearize Molodensky's problem, we start from the assumption (due to L. Hörmander [113]) that instead of one embedding $\varsigma : \Omega \to \mathbb{R}^3$ we have a family of smooth embeddings depending on a parameter $\chi \in [0,1]$, i.e.,

$$\mathfrak{S} : \Omega \times [0,1] \to \mathbb{R}^3, \text{ such that } \mathfrak{S}(\Omega, 0) = \partial\mathcal{T},\ \mathfrak{S}(\Omega, 1) = \partial\mathcal{G} \tag{4.48}$$

smooth with respect to χ, too. Moreover, we let

$$\mathfrak{W} : \mathbb{R}^3 \times [0,1] \to \mathbb{R} \tag{4.49}$$

be a family of potentials such that $\mathfrak{W}(\cdot, 0) = U$ and $\mathfrak{W}(\cdot, 1) = W$. We assume that

$$\mathfrak{W}(\cdot, \chi) = \mathfrak{V}(\cdot, \chi) + \Phi, \tag{4.50}$$

i.e., the gravitational part v of $\mathfrak{W}(\cdot, \chi)$ depends on χ and the centrifugal potential part Φ is independent of χ (this is in line with the assumptions on U and W). Corresponding to $\mathfrak{W}$, we arrive at a family of gravity vectors

$$\mathfrak{g} = \nabla\mathfrak{W} \text{ such that } \mathfrak{g}(\cdot, 0) = u \text{ and } \mathfrak{g}(\cdot, 1) = w. \tag{4.51}$$

We are now able to discuss the composition

$$\overline{\mathfrak{W}} = \mathfrak{W} \circ \mathfrak{S} : \Omega \times [0,1] \to \mathbb{R},\ (\xi, \chi) \mapsto \overline{\mathfrak{W}}(\xi, \chi) = \mathfrak{W}(\mathfrak{S}(\xi, \chi), \chi). \tag{4.52}$$

For $\chi = 0$, we find $\mathfrak{W}(\mathfrak{S}(\xi, 0), 0) = U(x_Q)$ for $Q \in \partial\mathcal{T}$, i.e., boundary values of U on the telluroid. For $\chi = 1$, we obtain $\mathfrak{W}(\mathfrak{S}(\xi, 1), 1) = W(x_P)$ for $P \in \partial\mathcal{G}$, i.e., boundary values of W on the real Earth's surface. Analogously, we introduce

$$\overline{\mathfrak{g}} = \mathfrak{g} \circ \mathfrak{S} : \Omega \times [0,1] \to \mathbb{R},\ (\xi, \chi) \mapsto \overline{\mathfrak{g}}(\xi, \chi) = \mathfrak{g}(\mathfrak{S}(\xi, \chi), \chi). \tag{4.53}$$

Note that W and w are supposed to be known on $\partial\mathcal{G}$ whereas U and u can be calculated analytically outside $\partial\mathcal{E}$. Now, the linearization can be performed by differentiation with respect to χ, denoted by a dot. We first notice that

$$\dot{\mathfrak{W}} = \frac{\partial\mathfrak{W}}{\partial\chi} = \frac{\partial\mathfrak{V}}{\partial\chi} = \dot{\mathfrak{V}}. \tag{4.54}$$

Thus, $\dot{\mathfrak{W}}$ is harmonic and satisfies (4.40). For boundary values we obtain

$$\begin{aligned} \dot{\overline{\mathfrak{W}}}(\xi, \chi) &= \dot{\mathfrak{W}}(\mathfrak{S}(\xi, \chi), \chi) + \nabla\mathfrak{W}(\mathfrak{S}(\xi, \chi), \chi)^T \dot{\mathfrak{S}}(\xi, \chi) \\ &= \dot{\mathfrak{W}}(\mathfrak{S}(\xi, \chi), \chi) + \mathfrak{g}(\mathfrak{S}(\xi, \chi), \chi)^T \dot{\mathfrak{S}}(\xi, \chi), \end{aligned} \tag{4.55}$$

In the same way we arrive at

$$\begin{aligned} \dot{\overline{\mathfrak{g}}}(\xi, \chi) &= \dot{\mathfrak{g}}(\mathfrak{S}(\xi, \chi), \chi) + J(\mathfrak{g})(\mathfrak{S}(\xi, \chi), \chi)^T \dot{\mathfrak{S}}(\xi, \chi) \\ &= \dot{\mathfrak{g}}(\mathfrak{S}(\xi, \chi), \chi) + \mathrm{Hess}(\mathfrak{W})(\mathfrak{S}(\xi, \chi), \chi)\dot{\mathfrak{S}}(\xi, \chi), \end{aligned} \tag{4.56}$$

with $J(\mathfrak{g})$ being the Jacobian of $\mathfrak{g}$ and $\mathrm{Hess}(\mathfrak{W})$ the Hessian of $\mathfrak{W}$. As immediate results we are led to the aforementioned *Marussi condition*

$$\det\left(\left(\frac{\partial^2\mathfrak{W}(\mathfrak{S}(\xi, \chi), \chi)}{\partial x_i\, \partial x_j}\right)_{i,j=1}^{3}\right) \neq 0 \tag{4.57}$$

for all $(\xi,\chi) \in \Omega \times [0,1]$. Rewriting the equations (4.55), (4.56) in composition form, we see that

$$\dot{\overline{\mathfrak{W}}} = \dot{\mathfrak{W}} \circ \mathfrak{S} + (\mathfrak{g} \circ \mathfrak{S})^T \dot{\mathfrak{S}}, \tag{4.58}$$

$$\dot{\bar{\mathfrak{g}}} = \dot{\mathfrak{g}} \circ \mathfrak{S} + (\mathrm{Hess}(\mathfrak{W}) \circ \mathfrak{S}) \dot{\mathfrak{S}}. \tag{4.59}$$

From (4.59), we are able to deduce that

$$\dot{\mathfrak{S}} = (\mathrm{Hess}(\mathfrak{W}) \circ \mathfrak{S})^{-1} \left(\dot{\bar{\mathfrak{g}}} - \dot{\mathfrak{g}} \circ \mathfrak{S}\right) \tag{4.60}$$

In connection with (4.58) we therefore obtain from (4.60)

$$\begin{aligned}
\dot{\overline{\mathfrak{W}}} &= \dot{\mathfrak{W}} \circ \mathfrak{S} + (\mathfrak{g} \circ \mathfrak{S})^T (\mathrm{Hess}(\mathfrak{W}) \circ \mathfrak{S})^{-1} \left(\dot{\bar{\mathfrak{g}}} - \dot{\mathfrak{g}} \circ \mathfrak{S}\right) \\
&= \dot{\mathfrak{W}} \circ \mathfrak{S} + \left(\dot{\bar{\mathfrak{g}}} - \dot{\mathfrak{g}} \circ \mathfrak{S}\right)^T (\mathrm{Hess}(\mathfrak{W}) \circ \mathfrak{S})^{-1} (\mathfrak{g} \circ \mathfrak{S}) \\
&= \dot{\mathfrak{W}} \circ \mathfrak{S} + \left(\dot{\bar{\mathfrak{g}}} - \dot{\mathfrak{g}} \circ \mathfrak{S}\right)^T \left((\mathrm{Hess}(\mathfrak{W}))^{-1} \mathfrak{g}\right) \circ \mathfrak{S} \\
&= \dot{\mathfrak{W}} \circ \mathfrak{S} + \left(\dot{\bar{\mathfrak{g}}} - \dot{\mathfrak{g}} \circ \mathfrak{S}\right)^T \left((J(\mathfrak{g}))^{-1} \mathfrak{g}\right) \circ \mathfrak{S},
\end{aligned} \tag{4.61}$$

where we have used the fact that the Hessian is symmetric. $\left((J(\mathfrak{g}))^{-1} \mathfrak{g}\right)$ is the tangent of the curve along which the gravity field has a fixed direction. Such lines are called *isozenithals* in geodesy. We introduce

$$\mathfrak{M} = \left((J(\mathfrak{g}))^{-1} \mathfrak{g}\right) \tag{4.62}$$

to gain the more compact notation. As a matter of fact, we are able to detect the following equivalencies

$$\dot{\overline{\mathfrak{W}}} = \dot{\mathfrak{W}} \circ \mathfrak{S} + (\mathfrak{M} \circ \mathfrak{S})^T \left(\dot{\bar{\mathfrak{g}}} - \dot{\mathfrak{g}} \circ \mathfrak{S}\right) \tag{4.63}$$

$$\Leftrightarrow \dot{\overline{\mathfrak{W}}} - (\mathfrak{M} \circ \mathfrak{S})^T \dot{\bar{\mathfrak{g}}} = \dot{\mathfrak{W}} \circ \mathfrak{S} - (\dot{\mathfrak{g}} \circ \mathfrak{S})^T (\mathfrak{M} \circ \mathfrak{S}) \tag{4.64}$$

$$\Leftrightarrow \dot{\overline{\mathfrak{W}}} - (\mathfrak{M} \circ \mathfrak{S})^T \dot{\bar{\mathfrak{g}}} = \left(\dot{\mathfrak{W}} - \dot{\mathfrak{g}}^T \mathfrak{M}\right) \circ \mathfrak{S}, \tag{4.65}$$

such that

$$\dot{\overline{\mathfrak{W}}} = \dot{\mathfrak{W}} \circ \mathfrak{S} + (\mathfrak{M} \circ \mathfrak{S})^T \left(\dot{\bar{\mathfrak{g}}} - \dot{\mathfrak{g}} \circ \mathfrak{S}\right) \tag{4.66}$$

$$\Leftrightarrow \dot{\overline{\mathfrak{W}}} - (\mathfrak{M} \circ \mathfrak{S})^T \dot{\bar{\mathfrak{g}}} = \left(\dot{\mathfrak{W}} - \left(\nabla \dot{\mathfrak{W}}\right)^T \mathfrak{M}\right) \circ \mathfrak{S}. \tag{4.67}$$

Looking at our considerations in more detail we notice that we have specified a boundary condition for $\dot{\mathfrak{W}}$. Since the values on the left-hand side of (4.67) are supposed to be known. The boundary condition is of Robin-type with an oblique derivative due to the occurrence of the vector $\mathfrak{M}$ which is in general not normal to the boundary surface $\mathfrak{S}(\Omega,\cdot)$.

Geodetic Quantities. Although our derivation as presented above is mathematically sound, it looks rather unfamiliar to the reader with a geodetic background. However, the quantities above can be easily related to well-known and often used quantities from geodesy.

We formally approximate derivatives with respect to χ by differences. As results we are able to specify the following quantities:

(1) the *potential anomaly* by

$$\dot{\overline{\mathfrak{W}}}(x,\chi) \approx \frac{\overline{\mathfrak{W}}(\mathfrak{S}(\xi,1),1) - \overline{\mathfrak{W}}(\mathfrak{S}(\xi,0),0)}{1} = W(x_P) - U(x_Q) = \Delta W(x_P), \quad (4.68)$$

(2) the *gravity anomaly* by

$$\dot{\overline{\mathfrak{g}}}(x,\chi) \approx \frac{\overline{\mathfrak{g}}(\mathfrak{S}(\xi,1),1) - \overline{\mathfrak{g}}(\mathfrak{S}(\xi,0),0)}{1} = w(x_P) - u(x_Q) = \Delta w(x_P), \quad (4.69)$$

(3) the *disturbing potential* by

$$\dot{\mathfrak{W}}(x,\chi) \approx \frac{\mathfrak{W}(x,1) - \mathfrak{W}(x,0)}{1} = W(x) - U(x) = T(x), \quad (4.70)$$

(4) the *gravity disturbance vector* δ_g

$$\dot{\mathfrak{g}}(x,\chi) \approx \frac{\mathfrak{g}(x,1) - \mathfrak{g}(x,0)}{1} = w(x) - u(x) = \delta w(x). \quad (4.71)$$

By comparison we are immediately led to $\nabla T = \delta w$ and the harmonicity of T outside $\partial\mathcal{G}$. The settings (4.68)–(4.71) also show, that T and δw can be understood for all $x \in \mathbb{R}^3$, whereas the anomalies ΔW and Δw can only be defined for pairs of corresponding points on certain surfaces.

It is common practice in geodesy to work on the geoid instead of the real surface of the Earth and the ellipsoid instead of the telluroid, although there is no need mathematically. In this case, the anomalies are defined with respect to $P_0 \in \partial\mathcal{G}$ and $Q_0 \in \partial\mathcal{E}$. Moreover, a remove-compute-restore or gravity reduction technique is required that removes all masses outside the geoid by condensing the topography such that the disturbing potential can be treated as being harmonic outside the geoid, computes this disturbing potential and then restores the removed masses while adapting the potential (see, e.g., [112, 159] for a more detailed discussion of remove-restore procedures).

In order to approximate the vector $\mathfrak{M}$, we go back to u in (4.62). Introducing the (local) unit vector in (local) direction of the isozenithal as ε_τ and writing derivatives in direction of the isozenithal as $\frac{\partial}{\partial\tau}$, it can be shown (see, e.g., [131, 158]) and the references therein) that

$$\mathfrak{M} \approx -\left(\frac{1}{\gamma}\frac{\partial\gamma}{\partial\tau}\right)^{-1}\varepsilon_\tau. \quad (4.72)$$

Introducing $-\Delta g'(x) = -\Delta w^T(x)\varepsilon_\tau$ as the component of the gravity vector in the downward direction of the isozenithal, the identity (4.67) shows that

$$\left[\frac{1}{\gamma(x)}\frac{\partial\gamma(x)}{\partial\tau}\Delta W(x) - \Delta g'(x) = \frac{1}{\gamma(x)}\frac{\partial\gamma(x)}{\partial\tau}T(x) - \frac{\partial T(x)}{\partial\tau}\right]\Bigg|_{x=x_P}. \quad (4.73)$$

In the case of a non-rotating sphere (see, e.g., [158]), we get with $|x| = r$

$$u(x) = -\frac{GM}{|x|^3}x = -\frac{GM}{r^2}\varepsilon_r, \tag{4.74}$$

$$\frac{1}{\gamma(x)}\frac{\partial\gamma(x)}{\partial\tau} = \frac{1}{\gamma(x)}\frac{\partial\gamma(x)}{\partial r} = -\frac{2}{r}. \tag{4.75}$$

where G is the gravitational constant and M is the (mean) Earth's mass.

Given the *reference ellipsoid with the semi-principal axes a, a and b, $a > b$*, consider a point P with the ellipsoidal (or geodetic) coordinates (h, θ, λ), whereas h is the height above the ellipsoid, $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$ the latitude and $0 \le \lambda < 2\pi$ the longitude. x_P can be written (see, e.g., [89] and the references therein) as

$$x_P = \begin{pmatrix} \left(\frac{a^2}{\sqrt{a^2\cos^2(\lambda)+b^2\sin^2(\lambda)}} + h\right)\cos(\theta)\cos(\lambda) \\ \left(\frac{a^2}{\sqrt{a^2\cos^2(\lambda)+b^2\sin^2(\lambda)}} + h\right)\cos(\theta)\sin(\lambda) \\ \left(\frac{b^2}{\sqrt{a^2\cos^2(\lambda)+b^2\sin^2(\lambda)}} + h\right)\sin(\theta) \end{pmatrix}. \tag{4.76}$$

In *spherical approximation*, P is mapped on a point $\overline{P}$ with the spherical coordinates (r, θ, λ), i.e.,

$$x_{\overline{P}} = \begin{pmatrix} r\cos(\theta)\cos(\lambda) \\ r\cos(\theta)\sin(\lambda) \\ r\sin(\theta) \end{pmatrix} \tag{4.77}$$

whereas $r = \sqrt[3]{a^2 b} + h$. Hence, (4.73) reduces to

$$\left[\frac{\partial T(x)}{\partial r} + \frac{2}{r}T(x) = -\Delta w^T(x)e_r + \frac{2}{r}\Delta W(x)\right]\Bigg|_{x=x_{\overline{P}}}. \tag{4.78}$$

The boundary value problem consisting of the Laplace equation

$$\Delta_x T(x) = 0, \quad x \in \mathcal{G}^c, \tag{4.79}$$

and (4.78) is called*simple Molodensky problem.*

If the telluroid is chosen according to condition (a), ΔW vanishes on the Earth's surface. Let us further assume that r can be chosen to take a value R identical for all points. The direction of the isozenithal is then identical to the direction of the radius vector e_r. We also approximate

$$\Delta w^T(x_P)e_r \approx \Delta g(x_P) = g(x_P) - \gamma(x_Q). \tag{4.80}$$

Δg is called *gravity anomaly.* It is approximately, but not exactly, the magnitude of the gravity anomaly vector Δw. The gravity anomaly Δg can be computed directly from measurements of g as γ can be calculated due to the definition of U.

With these assumptions, we arrive at the classical (exterior) *Stokes boundary value problem* (see [207])

$$\Delta_x T(x) = 0, \qquad |x| > R, \tag{4.81}$$

$$\frac{\partial T}{\partial r}(x) + \frac{2}{R} T(x) = -\Delta g(x), \qquad |x| = R, \tag{4.82}$$

$$T(x) = \frac{C}{|x|} + O(|x|^{-3}), |x| \to \infty. \tag{4.83}$$

As the derivative with respect to r coincides on the sphere with the normal derivative, this is no longer an oblique-derivative problem.

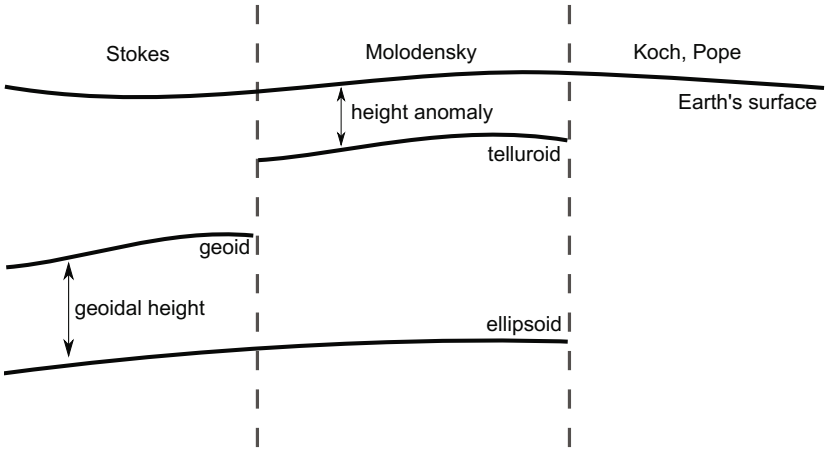


FIGURE 4.3. Different realizations of the geodetic boundary value problem (modified illustration following R. Rummel [185], see also [3]).

The Stokes problem links the disturbing potential T to gravity anomalies Δw. On the other hand, we have seen that $\nabla T = \delta w$. On the sphere $\mathbb{S}^2_R$ around the origin with radius R, the normal component of this relation is given by

$$\frac{\partial T}{\partial r} = \delta w^T \varepsilon_r. \tag{4.84}$$

Using an approximation of the form

$$\delta w^T(x)\varepsilon_r \approx \delta g(x) = g(x) - \gamma(x), \tag{4.85}$$

thus introducing the *gravity disturbance* δg, we obtain the *exterior Neumann boundary value problem* (see, e.g., [112])

$$\Delta_x T(x) = 0, \qquad |x| > R, \tag{4.86}$$

$$\frac{\partial T}{\partial r}(x) = \delta g(x) \qquad |x| = R, \tag{4.87}$$

$$T(x) = \frac{c}{|x|} + O(|x|^{-3}), \quad |x| \to \infty. \tag{4.88}$$

In the same way as before, δg is approximately the magnitude of the gravity disturbance vector δw. If the approximation (4.85) is not admissible, which is also the case if we do not use spherical approximation, the relation evaluation $\nabla T = \delta w$ on a boundary surface yields an oblique derivative problem.

A historical remark involving Neumann's problem is due to M.I. Yurkina (cf. [112]): The Neumann problem is a classical problem of potential theory, with a long history. Neumann's problem is named after Carl Neumann, who edited his father's (Franz Neumann) lectures from 1850s. The external spherical Neumann problem also occurs in [121]. It is again found in [114]. In future, because of GNSS (Global Navigation Satellite System such as GPS, GLONASS, BEIDOU, GALILEO), gravity disturbances may be expected to be much more available than gravity anomalies. This observation implies a renewed importance of the Neumann problem.

Another quantity that can be used to determine the disturbing potential is the deflection of the vertical. The *deflection of the vertical* is the difference between the direction of the reference normal vector $\nu'(x) = u(x)/|u(x)|^{-1}$, associated with the reference potential U and the reference gravity vector w, and the normal vector $\nu(x) = -w(x)/|w(x|$, associated with the (actual) gravity potential W and the (actual) gravity vector w. There are different definitions based on whether ν is evaluated at the real surface of the Earth or at the geoid and whether ν' is evaluated at the real surface of the Earth, the telluroid, or the reference ellipsoid.

Considering a point P_0 on the geoid $\partial\mathcal{G}$, we can decompose ∇T into a normal part and a surface part, defining the latter as

$$\nabla^{\mathrm{S}} T(x_{P_0}) = \nabla T(x_{P_0}) - (\nabla T(x_{P_0}) \cdot \nu(x_{P_0}))\, \nu(x_{P_0}) \tag{4.89}$$

as $\partial\mathcal{G}$ is an equipotential surface of W and, thus, $\nu(x_{P_0})$ is normal to $\partial\mathcal{G}$. The reference gravity vector u can be decomposed similarly by defining

$$\begin{aligned} u^{\mathrm{S}}(x_{P_0}) &= u(x_{P_0}) - (u(x_{P_0}) \cdot \nu(x_{P_0}))\, \nu(x_{P_0}) \\ &= -u(x_{P_0})\nu'(x_{P_0}) + u(x_{P_0})\, (\nu'(x_{P_0}) \cdot\ \nu(x_{P_0}))\, \nu(x_{P_0}) \\ &= -ux_{P_0})\, [\nu'(x_{P_0}) - (\nu'(x_{P_0}) \cdot \nu(x_{P_0}))\, \nu(x_{P_0})]\,. \end{aligned} \tag{4.90}$$

Note that $w(x) = -g(x)\nu(x)$ by definition, i.e., there are no surface components of $w(x_{P_0})$ at the geoid. Therefore, we get for the surface components

$$\begin{aligned} u^{\mathrm{S}}T(x_{P_0}) &= -u^{\mathrm{S}}(x_{P_0}) \\ &= -|u(x_{P_0})|\,[\nu'(x_{P_0}) - (\nu'(x_{P_0}) \cdot \nu(x_{P_0}))\,\nu(x_{P_0})] \\ &= -\gamma u(x_{P_0})\Theta(x_{P_0}) \end{aligned} \tag{4.91}$$

with the (vectorial) deflection of the vertical defined as

$$\Theta(x_{P_0}) = \nu'(x_{P_0}) - (\nu'(x_{P_0}) \cdot \nu(x_{P_0}))\,\nu(x_{P_0}). \tag{4.92}$$

We can also use Bruns's formula (cf. [15]) in the form

$$T(x_{P_0}) = \gamma(x_{Q_0})N(x_{P_0}) \tag{4.93}$$

with Q_0 being the point on the reference ellipsoid associated with P_0 and N the geoidal undulation (leading the approach given by F.A. Vening Meinesz [217].

In spherical approximation (see, e.g., [159]), we can estimate $u(\mathbf{x}_{Q_0})$ by the constant value u_0 to arrive at

$$\nabla^{\mathrm{S}}N(x_{P_0}) = -\frac{\gamma(x_{P_0})}{\gamma_0}\Theta(x_{P_0}). \tag{4.94}$$

Traditionally, it is also assumed that $\frac{\gamma(x_{P_0})}{\gamma_0} \approx 1$ and that $\nu'(x_{P_0})$ can be approximated by the ellipsoidal normal, i.e., $\nu'(x_{Q_0})$.

In spherical notation, we have

$$\xi = \sin\theta\varepsilon^3 + \cos\theta(\cos\lambda\varepsilon^1 + \sin\lambda\varepsilon^2),\ \ 0 \le \lambda < 2\pi,\ 0 \le \theta \le \pi \tag{4.95}$$

(λ: spherical longitude, θ: spherical latitude), where $\varepsilon^1, \varepsilon^2, \varepsilon^3$, respectively, form the (canonical) orthonormal basis in $\mathbb{R}^3$ and a moving orthonormal triad on the unit sphere Ω is given in the form

$$\varepsilon^r = \begin{pmatrix} \cos\lambda\cos\theta \\ \sin\lambda\cos\theta \\ \sin\theta \end{pmatrix}, \quad \varepsilon^\lambda = \begin{pmatrix} -\sin\lambda \\ \cos\lambda \\ 0 \end{pmatrix}, \quad \varepsilon^\theta = \begin{pmatrix} -\cos\lambda\sin\theta \\ \sin\lambda\sin\theta \\ \cos\theta \end{pmatrix}, \tag{4.96}$$

so that (4.91) and (4.94) lead to

$$\frac{1}{R}\nabla_\xi^* T(R\xi) = -\frac{GM}{R^2}\Theta(R\xi) \tag{4.97}$$

$$\iff \quad \frac{1}{R}\nabla_\xi^* N(R\xi) = -\Theta(R\xi). \tag{4.98}$$

The *surface gradient* ∇_ξ^* in local spherical coordinates is given by

$$\nabla_\xi^* = \varepsilon^\lambda \frac{1}{\cos\theta}\frac{\partial}{\partial\lambda} + \varepsilon^\theta \frac{\partial}{\partial\theta}, \tag{4.99}$$

Note that G is the gravitational constant and R is the (mean) Earth's radius used in spherical approximation. The particular representation of ∇_ξ^* yields the scalar

equations

$$-\frac{1}{R}\frac{\partial N}{\partial \lambda}(\lambda,\theta) = \mathrm{NSC}(\lambda,\theta), \tag{4.100}$$

$$-\frac{1}{R}\frac{1}{\cos\lambda}\frac{\partial N}{\partial \lambda}(\lambda,\theta) = \mathrm{EWC}(\lambda,\theta) \tag{4.101}$$

with $\mathrm{NSC}(\lambda,\vartheta)$, the north-south component, and $\mathrm{EWC}(\lambda,\vartheta)$, the east-west component of Θ. The difference between (4.98) on the one hand and (4.100) and (4.101) on the other hand is, that the former is an isotropic vectorial differential equation whereas the latter is an anisotropic system of two scalar differential equations. The solution of the isotropic vectorial differential equation (and its multi-scale approximation) can be found in [61], while the solution of the scalar anisotropic differential equations is due to F.A. Vening Meinesz [217].

For Dirichlet, Neumann, and Stokes problems, there exist solution procedures based on an integral representation of T (see, e.g., [159]). However, evaluating these integrals can be cumbersome. Therefore, it is necessary to consider suitable integration and approximation formulas to derive a numerical procedure that allows a fast and precise determination of the disturbing potential T from given boundary data (numerical realizations by locally supported wavelets have been recently proposed by W. Freeden, M. Schreiner [46], W. Freeden, K. Wolf [63], W. Freeden et al. [65], W. Freeden, C. Gerhards [49]). This leads us to the consideration of finite pointset methods on spheres and low-pass filtering involving truncated series of spherical harmonics. Even harmonic spline Runge methods can be performed, e.g., after a suitable Kelvin transform (see [3, 20, 49, 55, 56]).

Simplified Geodetic Models. If the normal potential U is determined as above, it has the same monopole component as W, as the mass of the reference ellipsoid is assumed to be equal to the mass of the Earth. As a consequence, the monopole component of T has to vanish, such that we actually have $T(x) = O(|x|^{-3})$ as $|x| \to \infty$. Even more, the traditional approach to be realized in standard textbooks of geodesy (see, e.g., [107, 112]) is based on the *Pizzetti oriented concept* (see, e.g., [174]) such that the following assumptions may be supposed to be valid:

(i) The mass within the reference ellipsoid for establishing the disturbing potential F is equal to the mass of the Earth.

(ii) The center of the reference ellipsoid coincides with the center of the Earth.

(iii) The value of the potential on the geoidal surface and the value of the normal potential on the reference ellipsoidal surface are the same.

(iv) There are no masses outside the geoid (remove-restore-principle from masses outside the geoid).

(v) The constructive approximation is simplified for reasons of computational economy from an ellipsoidal to a spherical framework.

In the Pizzetti oriented concept we finally arrive at the following manifestations

$$\int_{|x|=R} T(x)\ dS(x) = 0, \qquad \int_{|x|=R} T(x)(\varepsilon^k \cdot x)\ dS(x) = 0, \quad k = 1,2,3. \tag{4.102}$$

resulting in the identities

$$\int_{|x|=R} \Delta g(x)\ dS(x) = 0, \quad \int_{|x|=R} \Delta g(x)(\varepsilon^k \cdot x)\ dS(x) = 0, \quad k = 1,2,3, \tag{4.103}$$

$$\int_{|x|=R} \delta g(x)\ dS(x) = 0, \quad \int_{|x|=R} \delta g(x)(\varepsilon^k \cdot x)\ dS(x) = 0, \quad k = 1,2,3. \tag{4.104}$$

To some extent, the conceptional vagueness of disturbing potential and geoid via boundary problems based on heterogeneously distributed boundary data can be overcome by combining terrestrial and satellite data. With precise positioning by satellites, for example, intercontinental height links can be established between local geopotential realizations. Moreover, nowadays, GNSS can be used to determine the real Earth's surface. This fact led K.R. Koch, A.J. Pope [125] to reformulate the geodetic boundary value problem as an exterior oblique derivative problem without any need to introduce an auxiliary surface such as telluroid or ellipsoid. Conventionally, a solution of an oblique derivative problem corresponding to the boundary values

$$\frac{\partial V}{\partial \lambda} = F, \qquad l = \frac{\nabla V}{|\nabla V|} \tag{4.105}$$

and

$$\inf_{x \in \partial \mathcal{G}} (l(x) \cdot \nu(x)) > 0, \tag{4.106}$$

is represented by a strongly singular (Fredholm) integral equation based on a potential of a single layer as ansatz function that causes difficulties for numerical realizations. W. Freeden, H. Kersten [51, 52], W. Freeden, V. Michel [56], W. Freeden, C. Gerhards [49] show that the *geodetic boundary value problem using the gravitational vector* (4.105) *on the known Earth's surface* can be deduced from discrete data sets by a constructive Runge theorem in terms of, e.g., mono- and/or multipoles situated on an arbitrary closed surface lying completely inside the Earth.

4.5. Concluding remarks about the Earth's figure

In the 18th century it was well established that deviations between measurement and ellipsoidal theory were existent, which could not been explained by measurement errors (i.e., inaccurate measurements). The geodetic arc measurements as well as the pendulum measurements pointed to the problem that local deflections of the vertical occur because of the different composition of the masses of the Earth. The actual composition of the Earth's masses could not correspond to the presumed assumption. Already Ch. Maire, R.J. Boscović [145] discussed the gravity changes as local distortions, which they traced back to the influence of mountains, a simple conclusion of Newton's potential theory. Also the strongly different values for the flattening of the ellipsoid coming from different geodetic

arc measurements questioned the regular curvature of the meridians. P.S. Laplace [136] concluded in his "Traité de Mécanique Céleste" that the Earth is noticeably different from an elliptic shape.

On the basis of the well-known scientific findings at that time and his own knowledge about the significance of the deflections of the vertical, which he traced back to the visibility of the irregularity of the masses as well as to the different densities below the Earth's surface, C.F. Gauss came as a first scientist to a comprehensive definition of the mathematical figure of the Earth which was fundamental in physical geodesy. This definition has been published in 1828: "Bestimmung des Breitenunterschiedes zwischen den Sternwarten von Göttingen und Altona" (Determination of the latitude difference between the observatories of Göttingen and Altona), (Gauss Werke vol. IX, p. 49). In this treatise one can find the following sentences:

> *"In our opinion the topic is viewed from a wrong angle, if such phenomena are always considered as local deviations of the plumb line, and these deviations are regarded as it were only an exception. What we call in a geometric sense the surface of the Earth is nothing else as the same surface, which everywhere intersects the direction of gravity at right angles and part of it coincides with the surface of the world's oceans. The direction of the gravity is determined at every point by the shape of the part of the solid Earth and its unequal density. At the exterior rind of the Earth, from which we alone know anything, its shape and density appear as extreme irregular; the irregularity of the density may extend fairly deep under the outer rind and cannot be computed, because there is nearly no data available. The geometrical surface is the product of the total effect of these unequal distributed elements. In consideration of this situation nothing can prevent us to view the Earth as a whole as a spheroid of revolution, from which its real (geometrical) surface will overall deflect almost by stronger, weaker, shorter or longer undulations."*

For this definition of the geometrical surface of the Earth, Gauss's scholar J.B. Listing [143] chose the term "geoid", which is up to date used, and is the reference surface for all heights above sea level. Gauss explains the term "height" in a letter to Olbers (Gauss works, vol. IX, p. 375):

> *"This surface (the surface of the oceans) is called a horizontal surface (couche de niveau); equal heights are given to the points of this surface without caring by no means about whether or how much these points deviate from an ellipsoidal spheroid. The heights above this surface can be measured by a barometer as well as trigonometrically so that both must correspond to each other."*

According to H. Moritz [155] this surface of the Earth, in principle defined by Gauss, can in its importance for geodesy be considered as a change like that of Copernicus, because the simple model of the surface of an ellipsoid as an ideal was replaced by the physically interpreted Earth. Since that time the ellipsoid

serves in geodesy mainly as surface for computations in geodetic surveying and as a reference surface for the geoid and the Earth's gravity field. The Gaussian definition solves the problems with complicated correction models, which occur by reducing measurements for the determination of deflections of the vertical. The surface of an Earth representing ellipsoid of revolution cannot be exactly horizontal because of gravitational pull of the mountains, that is vertical to the plumb line. The Gaussian definition does not imply a simple geometrical surface, but it is defined by specifying all locations in Euclidean space $\mathbb{R}^3$ satisfying the simple equation $W = W_0 = \text{const}$. In other words, W is equated with the constant W_0, with that the geoid is explained as a level (equipotential) surface of the Earth's gravity field. Gauss introduced with his definition a principally different model, which is defined through the observations and represents a reversal of perspective (Gauss says that this definition is a change of the point of view).

A generalization of the Gaussian definition was achieved by H. Bruns [15]:

> *"The problem of scientific geodesy is the investigation of the potential function of the Earth."*

Instead of looking at a special equipotential surface, namely the geoid, all possible equipotential surfaces are considered in this definition as well as the gravity potential. This definition by Bruns can especially be used in satellite geodesy, because a satellite is moving in the gravity field of the Earth so that its orbit is determined by the potential W.

In today's geodesy concerned with Earth's figure, a result first motivated by C. Runge [187] in one-dimensional complex analysis and later generalized, e.g., by J.L. Walsh [222], I.N. Vekua [218], and L. Hörmander [113] to potential theory in three-dimensional Euclidean space $\mathbb{R}^3$ is of basic interest. For geodetically relevant application (see, e.g., [130, 155, 159, 189] and the references therein), the *Runge–Walsh theorem* may be formulated as follows (cf. [156]): Let the Earth's interior $\mathcal{G} \subset \mathbb{R}^3$ be a regular region, i.e., a bounded region $\mathcal{G} \subset \mathbb{R}^3$ dividing $\mathbb{R}^3$ uniquely in $\mathcal{G}$ and the outer space $\mathcal{G}^c = \mathbb{R}^3 \backslash \overline{\mathcal{G}}, \overline{\mathcal{G}} = \mathcal{G} \cup \partial\mathcal{G}$, such that $\mathcal{G}$ contains the origin and the boundary $\partial\mathcal{G}$ is an orientable smooth Lipschitzian manifold of dimension 2. The Earth's gravitational potential in $\mathcal{G}^c$ that is harmonic in $\mathcal{G}^c$ and regular at infinity can be approximated by a function that is harmonic outside an arbitrarily given Runge (in geodesy called Bjerhammar) ball $\mathcal{A} \Subset \mathcal{G}$, i.e., $\mathcal{A} \subset \mathcal{G}$ with $\text{dist}(\overline{\mathcal{A}}, \partial\mathcal{G}) > 0$ in the sense that, for any given $\varepsilon > 0$, the absolute error between the two functions is smaller than ε for all points outside and on any closed surface completely surrounding $\partial\mathcal{G}$ in its outer space. The value ε may be arbitrarily small, and the surrounding surface may be arbitrarily close to the surface.

Obviously, the Runge–Walsh theorem in the preceding formulation represents a pure existence theorem. It guarantees only the existence of an approximating function and does not provide a constructive method to find it.

The situation, however, is completely different if spherical geometrics are exclusively involved in the Runge concept. Assuming that both $\mathcal{A}, \mathcal{G}$ are concentric balls around the origin with $\mathcal{A} \Subset \mathcal{G}$, a constructive approximation of a potential

in the outer space $\mathcal{G}^c$ is available, e.g., by outer harmonic (orthogonal) expansions (see, e.g., [81, 121, 168, 223]). More concretely, within the classical context of a twofold spherical configuration, a constructive version of the Runge–Walsh theorem can be guaranteed by finite truncations of Fourier expansions in terms of outer harmonics, where the $L^2(\partial\mathcal{G})$-convergence of the Fourier series implies uniform converges on any pointset $\mathcal{K} \Subset \mathcal{G}^c$. The Fourier coefficients are obtained by integration over the sphere $\partial\mathcal{G}$. The gravitational potential is available (in spectral sense) by tables of the Fourier coefficients. Nowadays, in fact, outer harmonic expansions constitute the conventional geodetic tools in globally reflected approximation of the Earth's gravitational potential and its observables.

From a superficial point of view, one could suggest that approximation by truncated series expansions in terms of outer harmonics is closely related to spherical geometries $\partial\mathcal{A}, \partial\mathcal{G}$. W. Freeden [37], however, showed that the essential steps to a constructive Fourier approach can be extended to any regular (i.e., not-necessarily spherical) region $\mathcal{G}$ and to any regular (i.e., not-necessarily spherical) Runge domain $\mathcal{A} \Subset \mathcal{G}$. The Runge–Walsh theorem in this formulation avoids any difficulty with the convergence to the gravitational potential by the generalized Fourier series for arbitrary sets $\mathcal{K} \Subset \mathcal{G}^c$. In analogy to the spherical case, however, it likewise does not help to specify convergence inside $\mathcal{A}^c \backslash \overline{\mathcal{G}^c}$, so that any attempts (see [10]) to reduce gravitational information via infinite Fourier series downward from $\partial\mathcal{G}$ to the Runge surface $\partial\mathcal{A}$ are not justifiable by the Runge–Walsh framework.

In summary, the Runge–Walsh concept reflects constructive approximation capabilities of the Earth's gravitational (and not gravity) potential even if geoscientifically realistic (i.e., not necessarily spherical) geometries come into play. For numerical computations, the Runge concept may be regarded as the justification why boundary value problems in geodesy have to be solved without any need to use specific telluroidal reflected trial function systems such as numerically more difficult to handle ellipsoidal harmonics, Lamé functions, etc. Instead outer spherical harmonics (i.e., multi-poles) and/or mono-poles showing a larger harmonicity domain than the exterior of a telluroid (in linear Molodensky problem) or the actual Earth (in oblique derivative problem) can be taken into account.

5. From Gaussian least squares adjustment to inverse multi-scale regularization

5.1. Gauss's historic role and influence

Mathematicians interested in geodesy and astronomy tried to use the probability theory since the middle of the eighteenth century for the evaluation of redundant measurements. Like in the theory of games, values of measurements can be regarded as random samples of a random process. Already in the year 1755, Simpson pointed out that the mean value of a number of measurements can be advantageously used (see [7] for more details). Especially in the works of the French mathematicians Lagrange and Laplace, important theoretical foundations were treated. The actual reasoning for the adjustment theory with its core, *the method of least squares*, was done by Gauss and Legendre around 1800. Gauss probably hit on the fundamental idea in the autumn 1794 reading a publication of Lambert concerning applications of mathematics (see [69]), but published his method not before 1809 (cf. "Theoria motus corporum coelestium in sectionibus conicis solem ambientium"). He closed this publication by giving an explanation of his method ("Determinatio orbitae observationibus quotcumque quam proximae satisfacientis"). Further works followed in the year 1810 ("Disquisitio de elementis ellipticis Palladis") in 1816 ("Bestimmung der Genauigkeit der Beobachtungen" 1880) and finally, in 1821 and 1823 in a systematic presentation ("Theoria combinationis observationum erroribus minimis obnoxiae, pars prior et pars posterior"). Three years later these works were completed by the note "Supplementum theoriae combinationis observationum erroribus minimis obnoxiae".

Actually, in the year 1806, Legendre first published this adjustment method in a supplement of his treatise "Nouvelles méthodes pour la détermination des orbites comètes" and termed this principle "Sur la méthode des moindres carrés" (cf. [139]). This term was also taken by Gauss. Legendre explains in his paper that there is among all principles no simpler one for the evaluation of observations than to minimize the sum of the squared errors in order to estimate an unknown quantity.

Already about 1798 Gauss succeeded in giving his approach a foundation based on the probability theory. It is possible that the notice in his diary from June 17th of that year refers to this achievement: "Calculus probabilitatis contra Laplace defensus" (Gauss Werke X, p. 533). Since 1801 Gauss applied this method nearly daily for his astronomic computations (Gauss Werke, Vol. IV, p. 98). A well-known application during this time was the computation of the orbit of the Planetoid Ceres (cf. [27]), which could only be observed by Piazzi over 90° of its orbit. In spite of the small number of observations Gauss was able to compute the ephemeris of Ceres so precisely that Ceres could be redetected at the predetermined place. The discovery of the planetoid Ceres introduced Gauss to the world as a theoretical astronomer of the highest reputation ([27]). Why Gauss published his "Theoria motus corporum coelestium ..." so late can probably only be explained by the

endeavour of Gauss to give his works "an inner perfection". Gauss's motto was:

"pauca sed matura".

This is probably the reason that Legendre could publish before him in the year 1806 the method without giving a thorough reasoning. R. Adrain [2] (see also the comments by E. Hammer [97]) introduced this principle, too.

Legendre characterized this *method of least squares* as the most general, most accurate, and very easy to apply adjustment principle and proved that this method leads for one unknown to the arithmetic mean and for two or three unknowns to the center point of the system. According to Legendre a kind of equilibrium between the deviations of the measurements will be fixed, comparable to the determination of a space point by measurements, which lies in the center point of the system. Independently from Legendre, R. Adrain [2] derived in his publication "Research concerning the probabilities of the errors which happen in making observations" the function of the error probability and the method of least squares resulting out of it. The correspondence of great mathematicians of that time, e.g., between Gauss and Laplace, supports the impression that there was no dispute concerning the priority. Rather the correspondence proves that this simple principle of the method of least squares was highly regarded by the contemporaries (see also [87]).

According to V. Bialas [7], the method of least squares can be regarded as a characteristic example of the increasing importance of mathematics in geodesy in the first half of the 19th century. At the same time this method also is an expression of how the unity of observational practice, the evaluation of measuring results, and the corresponding theory is growing together.

Reasons for the Least Squares Method. The introduction of the "mean square error" and its definition as an independent measure of precision was an important assumption for the foundation of the least squares method, which can be written by the following objective function using today's representation (see, e.g., [238]): $v^T P v = \min$. Here P is a weight matrix, e.g., a diagonal matrix, in the Gaussian approach, and v is the vector of residuals. Besides the complete definition of this measure of precision Gauss also demonstrated in his "Theoria combinationis observationum erroribus minimis obnoxiae" how to get the mean square errors of the unknowns as well as the weight of a function of the adjusted unknowns in the course of an adjustment of observation equations (see [77] and the comments by G.W. Stewart [205]).

Approximate Reasoning. Following a publication by R. Dedekind [25], Gauss did argue in his lecture about the least squares method against the adjustment principle of minimizing the absolute sum of errors (deviations) and the algebraic sum being zero. This principle was first applied by Ch. Maire, R.J. Boscović [145] and later by P.S. Laplace [136]

$$\sum_{i=1}^{n} |v_i| = \min, \tag{5.1}$$

under the constraint

$$\sum_{i=1}^{n} v_i = 0. \tag{5.2}$$

Nowadays, geodesy calls the principle $\sum_{i=1}^{n} |v_i| = \min$ together with the constraint (5.2) *the Boscovič–Laplace method*. It seems that F.Y. Edgeworth [28] used this method primarily but without the constraint (5.2) for the estimation of many unknowns. A solution suitable for practical geodetic computations, which may be used to determine the most favorable weights in base extensional networks, was achieved by K. Friedrich [67]. About 50 years earlier, O. Schreiber [194] installed this method to solve similar problems. Mathematically, the Boscović-Laplace method (5.1) and (5.2) may be regarded as an ℓ_1-method under constraints. No doubt, ℓ_1-type approaches are important for geodesy. It is especially appropriate to detect gross errors (outliers) in observations. It may also be understood as a problem of linear programming. Using the ℓ_1-method a decomposition in consistent and non-consistent observation equations can be deduced so that an outlier test can be developed. Concerning this outlier test the question, if there exist a primal or dual degeneration of solutions, is of great numerical significance.

One of Gauss's arguments against the Boscović–Laplace method shall be mentioned: With a greater number of observations a bigger error would not exert a stronger impact on the results than many smaller errors, of which the absolute values have the same sum. An error, which occurs n-times, must have a stronger impact than n-single errors. Instead of the errors themselves the squares are therefore taken to estimate the usefulness of a hypothesis of smallness of their corresponding sum of squared errors. Gauss rejected errors with a higher power, which have always to be even, because the computations would than be extremely complicated so that the treatment would not be worth the effort (cf. [25]).

Today, for practical purposes, e.g., in engineering surveying, the estimation procedures for the least squares method and the ℓ_1-method are usually applied in simultaneous combination, because the corresponding programs are available. Specifically with the help of the ℓ_1-method outliers are detected, localized, and then eliminated. In some cases the ℓ_1-method turns out to be insufficient, to detect reliable outliers, so that other robust estimation methods come into play (see, e.g., [98, 115, 119, 123, 124, 142, 147, 192] and the references therein for more details).

Probabilistic Reasoning: Gauss's work concerning the theory of probability starts with a justification for the least squares method. In his "Theoria motus corporum coelestium" the problem was set in such a way that, with measurements of equal precision, the accidental errors exhibit a differentiable density of the probability distribution. The distribution is determined under the supposition that the most probable value for the measured quantity with any number of observations is equal to the arithmetic mean of the measured values (cf. [87]). Under this assumption Gauss was led to the well-known formula (cf. Figure 5.1) for the (Gaussian) probability distribution (i.e., the bell curve).

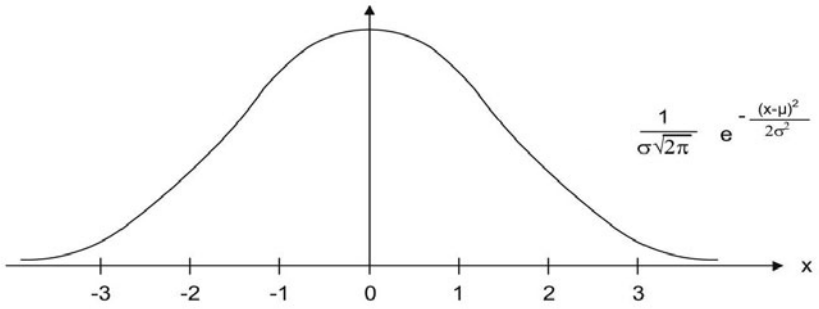


FIGURE 5.1. Gaussian distribution function with μ mean or expectation of the distribution (and also its median and mode), σ standard deviation, σ^2 variance.

The formula provides a measure for the precision of the observations. Gauss was aware of certain arbitrariness of this quantity. In fact, he wrote in Articles 178 and 179 of his publication "Theoria motus corporum coelestium: Functio modo eruta omni quidem igore errorum probilitatis exprimere certo non potest...".

"Hoc principium, quod in omnibus applicationibus mathesis ad philosophiam naturalem usum frequentissimum."

The function just found cannot, it is true, express rigorously the probabilities of the errors: For since the possible errors are in all cases confined within certain limits, the probability of errors exceeding those limits ought always to be zero, while our formula always gives some value. However, this defect, which every analytical function must, from its nature, labor under, is of no importance in practice, because the value of our function decreases so rapidly, when it has acquired a considerable magnitude, that it can safely be considered as vanishing. This principle, which promises to be of most frequent use in all applications of the mathematics to natural philosophy, must, everywhere, be considered an axiom with the same propriety as the arithmetical mean of several observed values of the same quantity is adopted as the most probable value (translated by P.J. Davis [23]).

After Gauss has determined the probability distribution for the measuring errors (errors of observations) in such a way, he concluded in one of the next paragraphs: The probability density of a given totality of observations will attain its maximum under the condition that the sum of the squares of the deviations of the observed quantities with respect to the true values of the quantities, which have to be measured, will become a minimum. This principle can also be applied to observations of different precision. If the mean square error is however unknown and the deviations of the approximate values from the true values have to be estimated, then the student distribution should be chosen according to the number of degrees of freedom. Especially with a small number of observations the results

will be different to the ones determined with the procedure given by Gauss (see, e.g., [87]).

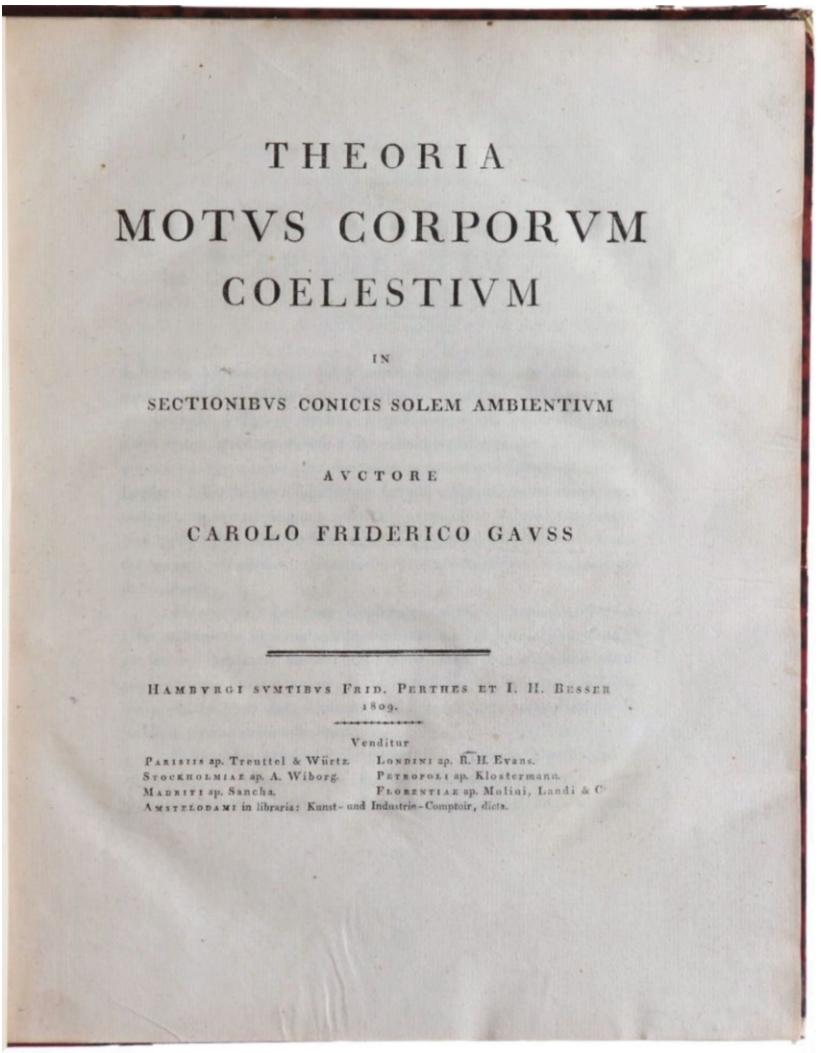
In case that the Gauss's error law and the axiom of the arithmetic mean underlying this law would not be valid, then the unknowns determined by the method of least squares are not any more the most probable values. Gauss came to a second justification presumably on the basis of an article by Laplace, who proved in his "Théorie analytique des Probabilités" ([138]) that the unknowns determined by the method of least squares have the smallest average errors. This proof is independent of the form of the error law, if for all observations the same error law is valid and if positive as well as negative errors of equal absolute value are equal probable. Laplace furnished this proof for two unknowns under the condition that the number of the observations is infinite. Gauss did not take this assumption into account and he proved this in his "Theoria observationum erroribus minimis obnoxiae". In the "Anzeigen" ([77]) the following sentence can be found:

> *"..., die Funktion für die Wahrscheinlichkeit der Fehler sei, welche sie wolle, und die Anzahl der Beobachtungen möge gross oder klein sein."* (*..., the function for the probability of the errors may be as it likes to be and the number of observations may be large or small*).

In consistency with G.W. Stewart [205] we present a list of what was new in Gauss's treatment of random errors:

(1) The careful distinction between systematic and random errors.
(2) The use of the first moment of a distribution to measure its center.
(3) The use of the second moment to measure precision.
(4) A Chebyshev-like inequality.
(5) The correct formula for the expectation of a function of a random variable.
(6) The rate of convergence of the sample mean and variance.
(7) The correct formula for estimating the precision of observations from the residual sum of squares.

Gaussian Adjustment Theory. For geodetic purposes, even in our days, the adjustment of observation equations (Gauss–Markoff model) is in principle applied in such a way as Gauss has proposed it. H. Wolf [236] explains: After linearization of the observation equations, usually by Taylorizing, the system of normal equations is set up and solved by use of the original Gaussian algorithm. Weight reciprocals and weighting coefficients as well as the partial derivatives of the observations with respect to the unknowns are computed. The error calculation consists of the computation of the mean square error of unit weight, the observations, the unknowns, and their functions. The well-known control formula for the computation of the sum of the squared residuals can also be dated back. Further, the problem is solved to specify the alterations in the unknowns, which are caused by an addition of a further observation equation or by a change of a single weight. Also the way of getting homogeneity by multiplying with the square root of the weight can be found in Gauss's work.



THEORIA
MOTVS CORPORVM
COELESTIVM
IN
SECTIONIBVS CONICIS SOLEM AMBIENTIVM
AVCTORE
CAROLO FRIDERICO GAVSS

HAMBVRGI SVMTIBVS FRID. PERTHES ET I. H. BESSER
1809.

Venditur
PARISIIS ap. Treuttel & Würtz. LONDINI ap. R. H. Evans.
STOCKHOLMIAE ap. A. Wiborg. PETROPOLI ap. Klostermann.
MADRITI ap. Sancha. FLORENTIAE ap. Molini, Landi & C.
AMSTELODAMI in libraria: Kunst- und Industrie-Comptoir, dicta.

FIGURE 5.2. Title-page of **"Theoria motus corporum coelestium in sectionibus conicis solem ambientium"**, original latin edition by Friedrich Perthes and I.H. Besser, Hamburg, 1809, also: "Carl Friedrich Gauss: Werke, herausgegeben von der (Königlichen) Gesellschaft der Wissenschaften zu Göttingen" (Göttinger Digitalisierungszentrum).

Examples of Application by Gauss and Others. Starting with the successful computation of the orbit of the planetoid Ceres (dwarf planet) up to the computation of the geodetic arc measurements in the Kingdom of Hanover Gauss validated the practical usefulness of his method. Besides the justification and development of his

least squares method Gauss proposed numerous approximate adjustments, which have naturally to be free of arbitrariness and contradiction as well as qualified for adjustments in successive steps and in groups. In order to adjust the angle observations taken on the single stations of the net for his arc measurements (station adjustments) Gauss used partly angle unknowns and partly bearing unknowns. To solve the adjustments he developed the successive approximation method (cf. [141]), which was applied by an adjustment of observation equations as well as of condition equations (cf. [86]). After getting the normal equation system with each step only the residual for one unknown is determined. For this step the residuals of the other unknowns are however set to zero. Gauss adjusted trigonometric nets partly with angle residuals and partly with bearing residuals using the condition equation method. The correlations coming from the station adjustments were not regarded (cf. [236]). However, C.F. Gauss [80] pointed out that, for this purpose, the observations to be adjusted must be independent of each other.

For the adjustment of condition equations it is not necessary to put up the normal equation system, if directly the condition equations are step by step fulfilled. This possibility for the application of the Gaussian successive approximation procedure with height nets was firstly pursued by C.A. Vogler [219]. For the adjustment of the geodetic arc measurement net in the kingdom of Hanover with condition equations Gauss used a successive approximation procedure, which he did not publish in detail, but it was later reconstructed by L. Krüger [132]. Subsequently, Krüger developed this procedure to the so-called Krüger method by two groups: C.F. Gauss [80] recommended only in Article 20 of his publication "Supplementum theoria combinationis ..." to divide the condition equations in two groups for the adjustment of large triangulation nets. More explicitly, Gauss writes in Article 20 (due to G.W. Stewart [205]):

> *"Quoties multitudo aequationum conditionalium permagna est, determinatio correlatorum ..."* (*"When the number of conditional equations is very large, the calculation of the correlates A, B, C, etc. by direct elimination becomes so laborious that the endurance of the calculator is not equal to the task. In such cases it is often better to use the theory of the preceding subsection to compute the complete adjustment by successive approximation ..."*).

A further important geodetic application example was the computation of the dimensions of the Earth's ellipsoid by the Gauss-scholar Schmidt, who at Gauss's suggestion adjusted the geodetic arc measurements of good quality known at that time. For the polar flattening he actually came to the result of $1 : 298{,}39$, an amazing result, if compared to the corresponding value of the World Geodetic System (WGS 84) of $1 : 298,257$.

The Gaussian Influence on More Recent Developments. It is possible to relate various settings of today directly to the Gaussian conceptions, e.g., the *covariance*. In fact, the essence inherent in the definition of the covariance can be deduced

from Articles 17 and 18 of Gauss's publication "Theoria Combinationis observationum..., Pars Prior"([77]). We present Article 17 here in the the original Latin form and Articles 17 and 18 in their English translation:

Article 17 of Gauss's Latin publication "Theoria Combinationis Observationum..., Pars Prior" in Latin language ([77]:

17.

Si valor quantitatis, quae ab alia quantitate incognita pendent, per observationem praecisione absoluta non gaudentem determinata est, valor incognitae hinc calculatus etiam errori obnoxius erit, sed nihil in hac determinatione arbitrio relinquiter. At si plures quantitates ab eadem incognita pendentes per observationes haud absolute exactas innotuerunt, valorum incognitae vel per quamlibet harum observationum eruere possumus, vel per aliquam plurium observationum combinationem, quod innitis modis diveris er potest. Quamquam vero valor incognitae tali modo prodiens errori semper obnoxius manet, tamen in alia combinatione maior, in alia minor error metuendus ert. Similiter res se habebit, si plures quantitates a pluribus incognitis simul pendentes sunt observatae: prout observationum multitudo multitudini incognitarum vel aequalis, vel hac minor, vel maior fuerit, problema vel determinatum, vel indeterminatum, vel plus quam determinatum erit (*generaliter saltem loquendo*)*, et in casu tertio ad incognitarum determinationem observationes innitis modis diversis combinari poterunt. E tali combinationum varietate eas eligere, quai maxime ad rem faciant, i.e., quae incognitarum valores erroribus minimis obnoxios suppeditent, problema sane est in applicatione matheseos ad philosophiam naturalem longe gravissimum.*

In Theoria motus corporum coelestium ostendimus, quomodo valores incognitarum maxime probabiles eruendi sint, si lex probabilitatis errorum observationum cognita sit; et quum haec lex natura sua in omnibus fere casibus hypothetica maneat, theorem illam ad legem maxime plausibilem applicavimus, ubi probabilitas erroris x quantiti exponentiale e mit Exponent proportionalis supponitur, unde methodus a nobis dudum in calculis praesertim astronomicis, et nunc quidem a plerisque calculatoribus sub nomine methodi quadratorum minimorum usitata demanavit.

Postea ill. Laplace, rem alio modo aggressus, idem principium omnibus aliis etiamnum praeferendum esse docuit, quaecumque fuerit lex probabilitatis errorum, si modo multitudo sit permagna. At pro multitudine observationum modica, res intacta mansit , ita ut si lex nostra hypothetica respuature, methodus quadratorum minimorum eo tantum nomine prae aliis commendabilis habenda sit, quod calculorum concinnitati maxime est adaptata.

Geometris itaque gratum fore speramus, si in hac nova argumenti tractatione docuerimus, methodum quadratorum minimorum exhibere combinationem ex omnibus optimam, non quidem proxime, sed absolute, quaecumque fuerit lex probabilitatis errorum, quaecumque observationum multitudo, si modo notionem erroris medii non ad menterm ill. Laplace set ita, ut in artt. 5 *et* 6 *a nobis factum est, stabiliamus.*

Ceterum expressis verbis hic praemonere convenit, in omnibus disquisitionibus sequentibus tantummodo de erroribus irregularibus atque a parte constante liberis sermonem esse, quum proprie ad perfectam artem observandi pertineat, omnes errorum constantium causeas summo studio amovere. Quaenam vero subsidia calculator tales observationes tractare suscipiens, quas ab erroribus constantibus non liberas esse iusta suspicio adest, ex ipso calculo probabilium petere possit, disquisitioni peculiari alia occasione promulgandae reservamus.

Articles 17 and 18 of Gauss's Latin publication "Theoria Combinationis Observationum..., Pars Prior" in English language [translated by G.W. Stewart [205]]:

17.

Suppose a quantity that depends on another unknown quantity is estimated by an observation that is not absolutely precise. If the unknown is calculated from this observation, it will also be subject to error, and there will be no freedom in this estimate of it. But if several quantities depending on the same unknown have been determined by inexact observations, we can recover the unknown either from one of the observations or from any of an infinite number of combinations of the observations. Although the value of an unknown determined in this way is always subject to error, there will be less error in some combinations than in others.

A similar situation occurs when we observe several quantities depending on several unknowns. The number of observations may be equal to, less than, or greater than the number of unknowns. In the first case the problem is well determined; in the second it is indeterminate. In the third case the problem is (generally speaking) overdetermined, and the observations can be combined in an infinite number of ways to estimate the unknowns. One of the most important problems in the application of mathematics to the natural sciences is to choose the best of these many combinations, i.e., the combination that yields values of the unknowns that are least subject to the errors.

In my Theory of the motion of heavenly bodies I showed how to calculate most probable values of the unknowns, provided the probability law of the observation errors is known. But in almost all cases this law can only be hypothetical, and for this reason I applied the theory to the most plausible law, in which the probability of an error x is proportional

e^{-hhxx}. *From this supposition came a method which I had already used for some time, especially in astronomical calculations. It is now used by many calculators under the name of the method of least squares.*

Later Laplace attacked the problem from a different angle and showed that if the number of observations is very large then the method of least squares is to be preferred, whatever the probability law of the errors. But for a modest number of observations, things are as they were, and if one rejects my hypothetical law, the only reason for recommending the method of least squares over other methods is that it lends itself to easy calculation.

I therefore hope that mathematicians will be grateful if in this new treatment of the subject I show that the method of least squares gives the best of all combinations – not approximately, but absolutely, whatever the probability law of the errors and whatever the number of observations – provided only that we take the notion of mean error not in the sense of Laplace but as in Arts. 5 *and* 6.

Here we should say that in the sequel we will be concerned only with random errors having no constant part, since the craft of taking observation requires that we take pains to remove all causes of constant errors. On another occasion I will give a special treatment about what help a calculator can expect from the calculus of probabilities when he undertakes to treat observations he suspects are not free of constant errors.

18.

Problem. Given a function U *of the unknown quantities* V, V', V'', *etc., find the mean error* M *to be feared in estimating* U *when, instead of the true values of* V, V', V'', *etc. one uses independently observed values having mean errors* m, m', m'', *etc.*

Solution. Let e, e', e'', *etc. denote the errors in the observed values of* V, V', V'', *etc., and let* $\lambda, \lambda', \lambda''$, *etc. be the differential quotients* e, e', e'' *etc. at the true values of* V, V', V'', *etc. Then the resulting error in* U *can be represented by the linear function*

$$\lambda e + \lambda' e' + \lambda'' e'' + etc. = E$$

provided the observations are precise enough so that we can neglect squares and products of the errors. From this it follows first that the mean value of E *is zero, since the observation errors are assumed to have no constant parts. Moreover, the mean error to be feared in this value of* U *is the square root of the of the mean value of* EE*; that is,* MM *is the mean value of the sum*

$$\lambda\lambda ee + \lambda'\lambda' e'e' + \lambda''\lambda'' e''e'' + \text{ect.} + 2\lambda\lambda' ee' + 2\lambda\lambda'' ee'' + 2\lambda'\lambda'' + \text{etc.}$$

Now the mean value of $\lambda\lambda ee$ *is* $\lambda\lambda mm$, *the mean value of.* $\lambda'\lambda' e'e'$ *is* $\lambda'\lambda' m'm'$, *etc. The mean values of the products* $2\lambda\lambda' ee'$, *etc. are all*

zero. Hence it follows that

$$M = \sqrt{\lambda\lambda mm + \lambda'\lambda' m'm' + \lambda''\lambda'' m''m'' + \text{etc.}}$$

It is appropriate to append some comments to this solution.

I. *Since we have taken the observation errors to be quantities of the first order and have neglected quantities of higher orders, we may use the values of the differential quotients* $\frac{dU}{dV}$*, etc. that come from the observed quantities* V, V', V''*, etc. to evaluate our formula instead of* $\lambda, \lambda', \lambda''$*, etc. Obviously this substitution makes no difference at all when* U *is a linear function.*

II. *Let* p, p', p''*, etc. be the weights of the observation errors with respect to an arbitrary unit, and let* P *be the weight of the estimate of* U *derived from the observed quantities* V, V', V''*, etc. If we prefer to work in terms of these quantities rather than the mean errors, then we have*

$$P = \frac{1}{\frac{\lambda\lambda}{p} + \frac{\lambda'\lambda'}{p'} + \frac{\lambda''\lambda''}{p''} + etc.}.$$

III. *Let* T *be another function of the quantities* V, V', V''*, etc., and for the true value of these quantities let*

$$\frac{dT}{dV} = \kappa, \quad \frac{dT'}{dV'} = \kappa', \quad \frac{dT''}{dV''} = \kappa'', \quad etc.$$

Then the error in the estimate for T *obtained from the observed values* V, V', V''*, etc. is* $E' = \kappa e + \kappa' e' + \kappa'' e'' + etc.$*, and the error to be feared in this estimate is* $\sqrt{\kappa\kappa mm + \kappa'\kappa' m'm' + \kappa''\kappa'' n''m'' + etc.}$ *The errors* E *and* E' *are clearly not independent, and, unlike the products* ee'*, the mean value of* EE' *is not zero but* $\kappa\lambda mm + \kappa'\lambda' m'm' + \kappa''\lambda'' m''m'' + \text{etc.}$

IV. *Our problem also extends to the case where the quantities* V, V', V''*, etc. are not obtained directly from observations but are derived from arbitrary combinations of observations. However, the individual quantities must be mutually independent, i.e., based on different observations. If this condition does not hold, then the formula for* M *will be in error. For example, if some observation that was involved in the calculation of* V *is also used in the calculation of* V'*, the errors* e *and* e' *will no longer be independent, and the mean value of the product* ee'*, will not be zero. However, if we can ascertain the relation of* V *and,* V' *with the simple observations from which they were derived, we can determine the mean value of the product* ee' *by the methods of comment III and hence give the correct formula for* M*.*

Summarizing, we are allowed to say that Gauss was, in principle, acquainted with the essence of covariance.

The Chebyshev Principle: The adjustment principle, established by Chebyshev in 1853, to make the biggest residual correction as small as possible, $|v_{\max}| = \min$, also follows from the Gaussian principle. This perception can be already found in Gauss's publication (see Figure 5.2 for its title-page): "Theoria motus corporum coelestium in sectionibus conicis solem ambientium" [1809]. Gauss writes in Book 2, Article 186 (in translation by P.J. Davis [23], p. 270):

> *"If we were to adopt a power with an infinite even exponent, we should be led to that system in which the greatest differences become less than in any other system".*

Hansen's Relationship to Pseudoinverses. According to H. Wolf [237], P.A. Hansen (1795–1874) used the so-called Bessel zero point correction (in an adjustment of a triangulation net) to go from the Bessel angle unknowns on to the bearing unknowns. The arising singularity of the station adjustment removed Hansen by an addition of a so-called restriction equation. Based on this procedure – already in the year 1867 – a method was developed, that is used today for the (numerical) computation of pseudoinverses.

After a watch-maker's apprenticeship Hansen because of his astronomical interests became a collaborator of Heinrich Christian Schumacher, the director of the Altona observatory (Altona is nowadays a suburb of Hamburg). Here, P.A. Hansen was able to gain know-how and experience in astronomical computations and observation techniques also in solving geodetic problems, which he could acquire during the Danish arc measurement of that time. In 1825 he was on recommendation of Gauss and Schumacher appointed as director of the Seeberg observatory near to Gotha, Duchy of Sachsen-Coburg-Gotha in Thuringia. Here, he was in charge of the surveying of the Duchy of Sachsen-Coburg-Gotha. The results of this triangulation served as a basis for the measurements to establish a cadastre. For this purpose the net, represented in Figure 5.3, had to be condensed by a low order triangulation and by traverses. By the following measurements of parcels a cadastral map was the final result. The triangulation net necessary for this task had to be adjusted by the method of least squares. During this work he hit on the problem to overcome the singularity of the station normal equations. His works of this period and the further development of Gauss's method of least squares were published in "Abhandlungen der Mathematisch-Physikalischen Klasse der Königlich-Sächsischen Gesellschaft der Wissenschaften, Leipzig", (see [99–102]).

Robinson's Thinking about the Origin of Pseudoinverses. The idea of a pseudo- (or generalized) inverse may be briefly motivated as follows: Suppose that

$$A \in \mathbb{R}^{m\times n}, \quad A = (A_{j,l})_{\substack{j=1,\dots,m \\ l=1,\dots,n}} \tag{5.3}$$

is understood as a linear mapping from $\mathbb{R}^n$ into $\mathbb{R}^m$. If $y = (y_1, \dots, y_m)^T$ in $\mathbb{R}^m$ is known, then the problem is to find $x = (x_1, \dots, x_n)^T$ in $\mathbb{R}^n$ such that

$$x \;\mapsto\; y \;=\; Ax \tag{5.4}$$

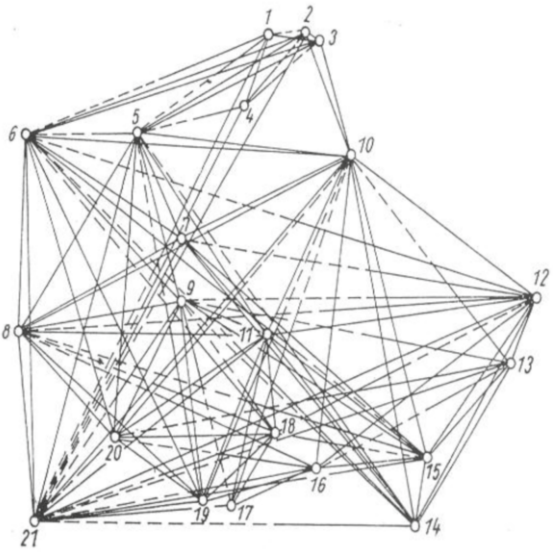


FIGURE 5.3. The Hansen triangulation net of the Duchy of Sachsen-Coburg-Gotha; point 21 (Inselsberg) is identical with the same point of the Gaussian arc measurement net (for comparison see Figure 6.1), point 11 is the location of the Seeberg observatory. In contrast to Gauss's measurement campaign Hansen's net served as a basis for the establishment of a cadastre.

holds true. If $n = m$ and $A \in \mathbb{R}^{m\times n}$ is invertible, with inverse A^{-1}, then the solution clearly is $x = A^{-1}y$. But, in case of $n \neq m$, the problem is a proper understanding of the "invertibility" of $Ax = y$, $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. The question is if there will be a mapping $A^\dagger$, called a *generalized inverse* (usually called, *pseudoinverse*) of A, such that $x = A^\dagger y$ (so that $A^\dagger = A^{-1}$ in case of $n = m$)?

During the last decades of the last century, the concept of a pseudo- (or generalized) inverse has rated considerable attention in the mathematical as well as geodetic literature (a bibliography, for example, listing over 1700 references on the subject is due to M.Z. Nashed [163]). Early interest in the first half of the last century in the subject of pseudoinverses was initiated by a paper on matrices by R. Penrose [172]. However, this concept had been considered somewhat earlier. For example, E.H. Moore [152] presented a development of the notion (see also [4, 9, 68, 153, 169, 178, 200]). Moreover, in the setting of integral and differential operators the concept was considered even earlier by I. Fredholm [34] and W.A. Hurwitz [116], and by D. Hilbert [111] (see [180] for a discussion of generalized inverses in classical analysis, and see also [5, 14, 163] for brief historical sketches of the subject).

The relation between Gauss's ideas and the concept of the pseudoinverse was discussed by D.W. Robinson [183]. His paper attempts to show that although Gauss did not formalize the notion of a pseudoinverse, he provided the essential ingredients to produce one. Next we follow this approach almost literally, however, formulated within a today's notational framework of linear algebra.

The point of departure for a mathematical concretization of Gauss's role is what is usually called today the *full-rank linear model*, which can be described as follows: Given a set $\mathcal{L}_1, \ldots, \mathcal{L}_n$ of linear functionals on $\mathbb{R}^n$, a set $y = (y_1, \ldots, y_m)^T$ of observations, a set $e = (e_1, \ldots, e_m)^T$ of errors, and a set $w = (w_1, \ldots, w_m)^T$ of positive numbers. The problem is to find $x = (x_1, \ldots, x_n)^T$ that makes the errors

$$e = (e_1, \ldots, e_m)^T = (\mathcal{L}_1 x - y_1, \ldots, \mathcal{L}_m x - y_m)^T$$

in a certain metric, as small as possible, subject to the condition that the weights constituting the metric influence the precision of the respective equations. It should be noted that Gauss considered the adoption of several possible principles to solve this problem, but finally argues for the minimization of "the sum of the squares of the differences between the observed and the computed values multiplied by numbers which measure the degree of precision" (see [72]). In more detail, Gauss sought to minimize the sum

$$\sum_{j=1}^{m} w_j e_j^2 = \sum_{j=1}^{m} w_j (\mathcal{L}_j x - y_j)^2 \tag{5.5}$$

of the *weighted squares of the errors*.

In terms of modern notation we are led to the following context: Let $\mathbb{R}^m$ be equipped with an inner product defined by

$$\langle d, z \rangle = \sum_{j=1}^{m} w_j \; d_j z_j, \; d \in \mathbb{R}^m, z \in \mathbb{R}^m, \tag{5.6}$$

with associated norm

$$\|d\| = \langle d, d \rangle^{1/2} = \left(\sum_{j=1}^{m} w_j d_j^2 \right)^{1/2}. \tag{5.7}$$

If $y = (y_1, \ldots, y_m)^T$ is a given element of $\mathbb{R}^m$, and if $\mathcal{L}_1, \ldots, \mathcal{L}_m$ is used to provide the linear mapping $A : \mathbb{R}^m \to \mathbb{R}^n$, $x \mapsto Ax = (\mathcal{L}_1 x, \ldots, \mathcal{L}_m x)^T$, by letting

$$\mathcal{L}_j x = \sum_{l=1}^{n} A_{j,l} x_l, \; j = 1, \ldots, m, \tag{5.8}$$

then the problem is to minimize

$$\|Ax - y\| = \langle Ax - y, Ax - y \rangle^{1/2} = \sum_{j=1}^{m} w_j \left(\sum_{l=1}^{n} A_{j,l} x_l - y_j \right)^2. \tag{5.9}$$

C.F. Gauss [72] provided what he termed a "very expeditious algorithm" for solving this problem. He argued that if $A^T \in \mathbb{R}^{n\times m}$ means what is now called the *adjoint* of A, i.e., $A^T = (A_{l,j})_{\substack{l=1,\ldots,n,\\ j=1,\ldots,m}}$, then the minimization problem is equivalent to the solution of the system

$$A^T Ax = A^T y \tag{5.10}$$

with $A^T A \in \mathbb{R}^{n\times n}$. These equations $A^T Ax = A^T y$ are known to be the *normal equations*, which Gauss solved by the process of elimination to obtain the unique solution denoted by $b = (b_1, \ldots, b_n)^T$. Thus, the component b_i was Gauss's choice for the best estimate of x_i, $i = 1, \ldots, n$. Although Gauss did not formalize the notion of linear rank, it is clear from his context that he wished to consider only the case, where the rank of A is so that $n \leq m$.

At this stage, D.W. Robinson [183] comes to the following conclusion: In the "Theoria motus"(1809), Gauss did not hint at the idea of a pseudoinverse. However, this work was only his first on the subject of leastsquares. In 1821 he presented to the Royal Academy of Sciences in Göttingen the first part of his "Theoria combinations", followed by the second part in 1823 and a supplement in 1826. His purpose in preparing this lengthy paper was to improve the foundations of the theory of least squares ([77], Art. 17). He accomplished this by introducing the dual inner product in the space $(\mathbb{R}^n)^*$ of linear functionals. As a consequence, the functional $\mathcal{P}_j$, defined by

$$\mathcal{P}_j : \mathbb{R}^m \ni y \mapsto \mathcal{P}_j y = y_j,\ j = 1, \ldots, m \tag{5.11}$$

has the norm

$$\|\mathcal{P}_j\| = \sqrt{\langle \mathcal{P}_j, \mathcal{P}_j \rangle} = \frac{1}{\sqrt{w_j}}. \tag{5.12}$$

Thus, by considering once more the full-rank linear model $n \leq m$, $y_j = \mathcal{P}_j y$ is an estimate of

$$\mathcal{L}_j x = \sum_{l=1}^{n} A_{j,l} x_l = y_j = \mathcal{P}_j y \tag{5.13}$$

having the "weight" $w_j = \|\mathcal{P}_j\|^{-2}$. In other words, the estimate of $\mathcal{L}_j x$ is given as the image of y under the linear functional $\mathcal{P}_j$, which satisfies $\mathcal{P}_j A = \mathcal{L}_j$ with the additional property that

$$w_j = \|\mathcal{P}_j\|^{-2},\ j = 1, \ldots, m. \tag{5.14}$$

The idea is to require that, for every linear functional J in $(\mathbb{R}^n)^*$, the best estimate of Jx be the image under y of the linear functional $\mathcal{K}$ in $(R^n)^*$ satisfying $\mathcal{K}A = J$ and having the maximum weight $\|\mathcal{K}\|^{-2}$.

More concretely, Gauss formulated what he termed the "problem", which may be rephrased in the language of inner product spaces as follows: Given J_i satisfying $J_i x = x_i$, find, among the linear functionals $\mathcal{K} : \mathbb{R}^n \to \mathbb{R}$ with $\mathcal{K}A = J_i$,

the one having minimum norm. Gauss successfully solved this problem. In fact, he obtained the explicit solution in the form

$$\mathcal{K}_i = J_i(B^TB)A^T, i = 1, \ldots, m, \tag{5.15}$$

where B^TB is the result of transforming by elimination $z = AA^Tx$ into $x = (B^TB)z$ (cf. [77], Art. 20 Eq. (4)). Thus, the best estimate of $x_i = \mathcal{K}_i y$ was taken to be $\mathcal{K}_i y = J_i(B^TB)A^T$, $i = 1, \ldots, m$. Furthermore, Gauss argued that this best estimate $\mathcal{K}_i y$ was equal to the value b_i obtained by the method of least squares ([77], Art. 21). By composing these functionals together, if

$$A^\dagger : \mathbb{R}^m \to \mathbb{R}^n, y \mapsto (\mathcal{K}_1 y, \ldots, \mathcal{K}_n y)^T, \tag{5.16}$$

then $b = A^\dagger y$ minimizes $\|Ax - y\|$. In this sense, $A^\dagger$ may be recognized as a pseudoinverse of A.

In conclusion, D.W. Robinson [183] was led to the following statement: Gauss did not formally display $A^\dagger$ in his 1821 paper. The ingredients for the construction of a pseudoinverse were essentially available to him, but he did not use them toward this end. Indeed, there appears to be no evidence that he was inclined to proceed in that direction. On the other hand, his approach to the problem of determining best estimates is certainly in the spirit of pseudoinverses. This is suggested by the diagrams (again a contemporary device), when A is associated to $(\mathcal{L}_1, \ldots, \mathcal{L}_m)$ and $A^\dagger$ with $(\mathcal{K}_1, \ldots, \mathcal{K}_n)$:

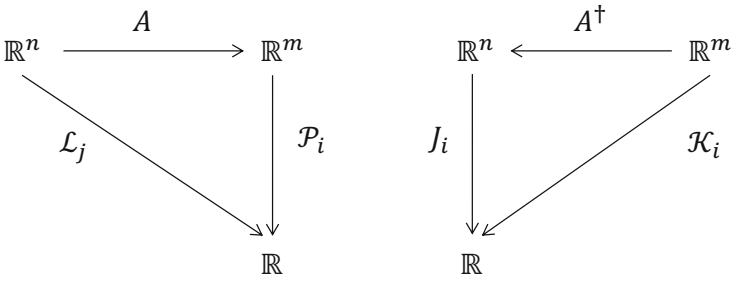


FIGURE 5.4. Diagram illustrating the roles of A and $A^\dagger$ in Gauss's approach. Specifically, the idea of "mapping back" is suggested in Figure 5.4 (cf. [183]).

Moreover, in the supplement to "Theoria combinationis", C.F. Gauss (1826) showed that the solution b which minimized $\|Ax-y\|$ was expressible as the image of y under a mapping depending only upon A.

In the opening paragraph of Article 8 of this supplement, Gauss stated his objective:

> *"When one wants to find the most likely values of several unknowns, depending on the same observations, or when one does not know which unknowns it is preferable to derive from the observations, it is convenient to proceed in a different way"* (*translation from D.W. Robinson* [183]).

The "different way" was to solve the normal equations $A^TAx = A^Td$ explicitly. He did so by letting $z = A^TAx$ and obtained $x = B^TBz$ by elimination. Although he did not use the formal "inverse", he appreciated that B^TB was a function of A^TA and noted, in particular, that A^TA was symmetric. He then expressed the solution of the normal equations in the form $b = (B^TB)A^Ty$ (see the supplement to [77], Art. 8). Consequently, his objective was satisfied: b was the image of y under (B^TB). In other words, interpreted in contemporary language, $(B^TB)A^T$ is a pseudoinverse of A. Thus Gauss's approach to this problem was consistent with the objectives of the theory of pseudoinverses, and his explicit solution is readily identified with the generalized inverse representation $b = (A^TA)^\dagger A^Ty$ used today (see, for example, [163]).

Finally, once more following D.W. Robinson [183], the two pseudoinverses identified above are the same; in fact, they are equal to what is called today the pseudoinverse (or Moore–Penrose) inverse. Indeed, by using the standard definition, for an inner product of two linear mappings, Gauss's problem may be viewed as a functional version of the following problem: Given the identity map $I : \mathbb{R}^n \to \mathbb{R}^n$, find among the linear mappings $\mathcal{K} : \mathbb{R}^m \to \mathbb{R}^m$, satisfying $\mathcal{K}A = I$, the one having minimum norm. Since $A^\dagger$ is the solution to this problem, it is what is called the best approximate solution of the equation $\mathcal{K}A = I$, and in this case it is the pseudoinverse (generalized or Moore–Penrose) inverse of A (cf. [172]). Moreover, it follows from Gauss's explicit solution $\mathcal{K}_i$ of the best linear estimate problem and the fact that $(J_1x, \ldots, J_nx) = x$, that, for every y in $\mathbb{R}^m$, $A^\dagger y = (\mathcal{K}_1y, \ldots, \mathcal{K}_ny)^T$. That is $A^\dagger = (B^TB)A^T$. Consequently, we are able to conclude that, under the conditions of the Gaussian full-rank linear model, $A^\dagger, (B^TB)A^T$, and the Moore–Penrose inverse of A are one and the same operator.

All in all, the conclusion of D.W. Robinson [183] is as follows:

> *"While generalized inverses were not a part of Gauss's vocabulary, equivalent expressions may be found in his writings. Specifically, in his solution of the problem of least squares, Gauss established explicit formulas which may readily be identified with generalized inverses. Indeed, by translating his analytical formulation of the problem into the more geometrical framework of vector spaces and linear mappings, the features of generalized inverses are easily recognized. In particular, Gauss's development of best linear estimates was in the spirit of generalized inverses. This observation suggests the possibility that Gauss's view of least squares in* 1826 *may have been more geometrical than the analytical form of his presentation could express. He may well have conceived best linear estimates in some mode of geometrical thought, but communicated the results in the accepted and more rigorous analytical mode of the day,*

since the geometrical tools of the early 19th century were limited, and the conceptual framework needed to develop a theory of generalized inverses was not available to him."

Hints for Further Developments. Next we list some areas of today's mathematics which can be essentially traced back to Gauss's work:

(a) The probability distribution function ("bell curve") introduced by Gauss has as normal distribution obtained a special significance in statistics, because theory and practice of confidence intervals and the resulting hypothesis tests were not possible without the Gaussian distribution (see [236]).
(b) According to W. Jordan, the least squares method can be used to solve optimization problems by an iterative process, if the observations are redundant, e.g., for the detection of optimal weight distributions, a problem, which can primarily be settled by use of the Boscović–Laplace method respectively, the simplex algorithm of linear programming (cf. [236]).
(c) Also the procedures of spherical and spherical-harmonic prediction and collocation, which made their headway by H. Moritz [154, 155] are based on the method of least squares (note that "least squares collocation" as it is called in geodesy is just named "minimum norm interpolation" in mathematical language, while "least squares adjustment" as understood in geodesy is nothing more than "smoothing", e.g., in the sense of C. Reinsch [181]). Furthermore, the prediction method can be considered as a kind of inter- or extrapolation to determine a trend function.

These approximation methods were transferred and widely extended to the spline context by W. Freeden and many others [39, 40, 42, 43, 62, 64, 95, 96, 182, 199, 221].

Clearly, the list of examples of Gaussian ideas as presented here is rather incomplete. Furthermore, our approach can be deepened by far, which will be done partly later in a functional analytic jargon.

5.2. Bridge between least squares solutions and pseudoinverses

Until now, our historic survey on least squares and pseudoinverse facets has been formulated within the context of the original framework. In what follows we would like to answer the question what exactly enabled mathematicians to build the bridge from Gaussian least squares theory to essential settings of today's theory of inverse problems. To this end, the following two question (Q1) and (Q2) have to be answered:

(Q1) What can be specified as mathematical structures and settings from least squares by a process of abstraction, i.e., a reduction to essential features?
(Q2) What can be achieved from the reduced context obtained by abstraction from least squares theory as a new field of scientific interest, e.g., in geodesy or other sciences?

The answers can be given in a three step (S1), (S2), (S3) procedure:

(S1) We specify the transfer from least squares solutions to the theory of the pseudoinverse in a *finite-dimensional matrix calculus*. The pseudoinverse is seen, comparably, but more generally to the Gaussian approach, as the best-approximate solution within the set as least squares solutions in finite-dimensional settings.

(S2) In turn, the reduction of the pseudoinverse to its specific properties involving functional analytic means based on operator theoretical background leads us to the concretization of the pseudoinverse in finite-dimensional spaces that can be extended to infinite-dimensional spaces in a straightforward way.

(S3) As a consequence, we shall be able by use of an *infinite-dimensional operator calculus* to solve ill-posed problems for (compact) operator equations by regularization. In particular, dilemmas and methodologies of resolution of ill-posed problems become obvious with particular reference to the problem of finding minimum norm least squares solutions of first kind integral equations (and, more generally, of linear operator equations with non-closed range).

5.3. Pseudoinverse for finite-dimensional matrix equations

Since any numerical approximation procedure usually leads to finite-dimensional problems involving a singular functional analytic context, we first consider finite systems of linear equations (see, e.g., M.Z. Nashed [163, 164] for more details). After having treated the finite-dimensional situation, we turn to the analogous theory in operator framework.

Spectral Matrix Representation. We start with a linear matrix equation of the form

$$Ax = y, \tag{5.17}$$

with $A \in \mathbb{K}^{n\times n}$ being a Hermitian positive definite matrix with n rows and n columns. From the spectral theory (see, e.g., [226]) it is well known that there exist eigenvalues $0 < \lambda_1 \leq \cdots \leq \lambda_n$ and a corresponding unitary matrix $U = (u_1, \ldots, u_n)$ of eigenvectors $u_i \in \mathbb{C}^n \backslash \{0\}$ (i.e., $u_i^H u_j = \delta_{ij}$) such that A has a representation of the form

$$U^H A U = \underbrace{\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ & & \ddots \\ 0 & & \lambda_n \end{pmatrix}}_{n\times n} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n). \tag{5.18}$$

The *condition number* of A is given by the quotient of the largest and smallest eigenvalue, i.e., $\kappa = \frac{\lambda_n}{\lambda_1}$ (note that $\lambda_1 > 0$). For the sake of simplicity and coherence with the analysis for the infinite-dimensional case below, we shall assume here that the scaling is such that $\lambda_n = 1$, so that $\kappa = \lambda_1^{-1}$. The condition number is a measure for the stable solvability of the problem (5.17).

Ill-Conditioned Matrix Equations. Assume that we have *noisy data* y^ε instead of y, which satisfy the Euclidean norm estimate

$$\|y^\varepsilon - y\| \leq \varepsilon \tag{5.19}$$

in the Euclidean norm on $\mathbb{K}^n$. Let x^ε denote the solution with right-hand side y^ε. Then it follows from the spectral representation that

$$x^\varepsilon - x \;=\; U \;\mathrm{diag}\left(\frac{1}{\lambda_1},\ldots,\frac{1}{\lambda_n}\right)\; U^H\;(y^\varepsilon - y). \tag{5.20}$$

Hence, observing the orthogonality of eigenvectors we are led to the estimate

$$\|x^\varepsilon - x\|^2 = \sum_{i=1}^{n} \lambda_i^{-2} |u_i^H (y^\varepsilon - y)|^2 \leq \lambda_i^{-2} \|y^\varepsilon - y\|^2. \tag{5.21}$$

In other words, we have

$$\|x^\varepsilon - x\| \leq \kappa \|y^\varepsilon - y\| \leq \kappa\, \varepsilon. \tag{5.22}$$

The sharpness of this estimate can be immediately seen for $y^\varepsilon - y = \varepsilon u_1$. It is clear that with increasing condition number of the matrix A, the noise amplification increases in the worst case. For large κ one therefore speaks of an *"ill-conditioned problem"* (IPP). But it should be remarked that a finite-dimensional linear problem is never ill-posed (in the sense that the third condition in Hadamard's classification is violated), but for κ large one certainly comes close to this case.

We also observe that errors in low frequencies (i.e., corresponding to eigenvectors with large eigenvalues) are amplified less. Following our nomenclature we see that an error in the lowest frequency, i.e., $y^\varepsilon - y = \varepsilon\, u_n$, is not amplified at all. In fact, we just obtain $\|x^\varepsilon - x\| = \varepsilon$ from the spectral representation. This is a typical effect for inverse problems. It means that not all possible versions of noise of the same size are equally bad, high-frequency noise corresponding to low eigenvalues is always worse than low-frequency noise. However, in practice, we are able to make any assumption on the noise only in rare exceptions, so that a regularization method has to deal with arbitrary noise.

Until now, we have assumed that the matrix A is Hermitian positive definite, i.e., the minimal eigenvalue is positive. If this is not the case, the matrix has a non-trivial null space. If λ_r denotes the minimal non-zero eigenvalue, then the solution formula becomes

$$x = \sum_{i=r}^{n} \lambda_i^{-1} u_i u_i^H y, \tag{5.23}$$

and the problem is solvable if and only if $u_i^H y = 0$ for $i < r$. If the data set is noisy, i.e, instead of y we have y^ε, we are led to use the projection Py^ε onto the range of A. In doing so we obtain for the corresponding solution x^ε with data Py^ε that

$$x^\varepsilon - x = \sum_{i=r}^{n} \lambda_i^{-1} u_i u_i^H (Py^\varepsilon - y). \tag{5.24}$$

Since $u_i^H P y^\varepsilon = u_i^H y^\varepsilon$ for $i \geq r$ we thus can estimate similarly as described above

$$\|x^\varepsilon - x\| \leq \lambda_r \varepsilon. \tag{5.25}$$

Consequently, there is no error propagation in the null space components and the noise amplification is actually determined by the minimal nonzero eigenvalue.

Matricial Pseudoinverse (Generalized Inverse, Moore–Penrose Inverse). Let $A \in \mathbb{K}^{n\times m}$ be a matrix with n rows and m columns, $y \in \mathbb{K}^m$. Note that A is not required to be square, no rank assumptions are made at this stage. Then we know that the linear system

$$Ax = y, \quad x \in \mathbb{K}^n, \tag{5.26}$$

needs not have a (unique) solution. If (5.26) is unsolvable, a reasonable generalized notion of a solution is a "least square solution", which minimizes the residual $Ax-y$ in the Euclidean norm (once more, note that $\|\ \|$ stands for the Euclidean norm in this subsection): A vector $x \in \mathbb{K}^n$ is called a

(1) *least squares solution of* (5.26) if and only if

$$\|Ax - y\| = \inf\{\|Az - y\| : z \in \mathbb{K}^n\}, \tag{5.27}$$

(2) *best-approximate-solution* (*or minimal norm solution*) *of* (5.26) if and only if x is a least squares solution and

$$\|x\| = \inf\{\|z\| : z \text{ is a least square solution}\}. \tag{5.28}$$

As already mentioned, one could also use other norms in (5.27) and in (5.28), which would lead to different notions of a generalized solution. Also, instead of minimizing $\|z\|$ in (5.28), it is often of interest to minimize $\|Tz\|$ for some prescribed matrix T.

The following results are well known from classical linear algebra (see, e.g., [226]):

(i) A vector x^* is a solution of (5.27) if and only if the "normal equations" $A^H A x^* = A^H y$ are satisfied.

(ii) The problem (5.27) possesses a unique solution if and only if A has full rank.

Our purpose is to show that a best-approximate solution in the sense of (2) always exists and is unique such that the following definition makes sense:

Definition 5.1. If $A^\dagger \in \mathbb{K}^{m\times n}$ is understood as the matrix which assigns to each $y \in \mathbb{K}^n$ the best-approximate solution of (5.26), then it is called the *pseudoinverse* (*Moore–Penrose or generalized inverse*) *of* A.

Our aim is to construct $A^\dagger$ and, hence, best-approximation solutions via the so-called *singular value decomposition* (*SVD*) *of* A. To this end we first recall the definition of the *singular values of a matrix* A.

Definition 5.2. Let $\sigma_1, \ldots, \sigma_r > 0$ be such that $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0$ are the positive eigenvalues of the matrix $A^H A$ (each one written down as often as its multiplicity is). Then $\sigma_1, \ldots, \sigma_r$ are called the "(non-zero) singular values of A".

This definition makes sense, since $A^H A$ is positive semidefinite. Obviously, $r \leq \min\{n, m\}$, where r is the rank of A.

We know that a Hermitian matrix can be diagonalized, where the diagonal elements are its eigenvalues. The following theorem generalizes this result to the non-Hermitian case.

Theorem 5.1. *Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ be the singular values of A. Then there exist unitary matrices $U \in \mathbb{K}^{m\times m}$ and $V \in \mathbb{K}^{n\times n}$ such that*

$$V^H A U = \underbrace{\begin{pmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{pmatrix}}_{m\times n} \tag{5.29}$$

holds true. The columns of U and V are eigenvectors of $A^H A$ and AA^H, respectively. The expression (5.29) *is called "singular value decomposition (SVD) of A".*

The singular value decomposition (SVD) is not unique, since the unitary matrices U and V are not. Obviously, from (5.29), we obtain

$$A = V \begin{pmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{pmatrix} U^H, \tag{5.30}$$

since V and U are invertible, and the rank of A is r, where r is the number of non-zero singular values (counted with multiplicity).

Remark 5.1. In accordance with Theorem 5.1 one has to compute the singular values of A, e.g., as the positive square roots of the eigenvalues of $A^H A$, which can be done, e.g., by the QR-algorithm. However, since $A^H A$ usually has a worse condition than A, one should (in critical cases) use a variant of the QR-method that does not use $A^H A$ explicitly.

Note that with $U = (u_1, \ldots, u_m)$, $V = (v_1, \ldots, v_n)$, we have for $i \in \{1, \ldots, r\}$

$$A u_i = \sigma_i v_i \tag{5.31}$$

and

$$A^H v_i = \sigma_i u_i, \tag{5.32}$$

which follows from the singular value decomposition (5.29) via multiplication by V and U^H, respectively. The system $\{(\sigma_i; u_i, v_i) : i \in \{1, \ldots, r\}\}$ is called a "*singular*

system" for A. Since, as we have seen in the proof of Theorem 5.1, $\{v_1, \ldots, v_r\}$ is an orthonormal basis for the range $\mathcal{R}(A) = \{y : Ax = y\}$, for any $x \in \mathbb{K}^n$,

$$Ax = \sum_{i=1}^{r} \langle Ax, v_i \rangle v_i = \sum_{i=1}^{r} \langle x, A^H v_i \rangle v_i, \tag{5.33}$$

which implies in connection with (5.32) that

$$Ax = \sum_{i=1}^{r} \sigma_i \langle x, u_i \rangle v_i \tag{5.34}$$

holds true. Analogously, for all $y \in \mathbb{K}^m$,

$$A^H y = \sum_{i=1}^{r} \sigma_i \langle y, v_i \rangle u_i. \tag{5.35}$$

Remark 5.2. The notion of a singular system and the expansion (5.34) and (5.35) generalize to compact operators on infinite-dimensional spaces, e.g., integral operators, as we will see later on.

Remark 5.3. Note that if A has real entries, so U and V have.

Now we relate the SVD to the Moore–Penrose inverse. Moreover, we show that SVD can be used to compute the best-approximate solution.

Theorem 5.2 (Pseudoinverse). *Let A have the SVD* (5.29)*. Then*

$$A^\dagger = U \underbrace{\begin{pmatrix} \frac{1}{\sigma_1} & & & & 0 \\ & \ddots & & & \\ & & \frac{1}{\sigma_r} & & \\ & & & \ddots & \\ 0 & & & & 0 \end{pmatrix}}_{n \times m} V^H. \tag{5.36}$$

Theorem 5.2 also implies the existence and uniqueness of a best-approximate solution.

Since $A^\dagger b$ is the least squares solution of minimal norm, we obtain that $A^\dagger b$ is a solution of the normal equations $A^H Ax = A^H b$ with minimal norm, i.e.,

$$A^\dagger y = (A^H A)^\dagger A^H y. \tag{5.37}$$

This means that in order to approximate $A^\dagger b$ we may as well compute an approximation to the minimal-norm solution in the normal equations, a fact we will heavily use in the construction of regularization methods, later on.

For the case of overdetermined linear systems, i.e., (5.26) with $m \geq n$ the null space $\mathcal{N}(A) = \{x : Ax = 0\}$ only consists of $\{0\}$, so that the best-approximation solution of (5.26) can effectively be computed via Householder transformations (see, e.g., [227]) in the following way: One transforms A by $n - 1$ Householder

(or other orthogonal, e.g., Givens) transformations to a $n \times n$-triangular matrix R (and 0 from the $(n+1)$st row on) and applies the same transformations to b. If Q is the unitary matrix representing these transformations, we are led to

$$QA = \begin{pmatrix} R \\ 0 \end{pmatrix}, \quad Qb = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}. \tag{5.38}$$

For any $x \in \mathcal{C}^n$,

$$\|Ax - b\| = \|Q(Ax - b)\| = \left\| \begin{matrix} Rx - h_1 \\ -h_2 \end{matrix} \right\|. \tag{5.39}$$

Consequently, the best-approximate solution can be computed in the form $A^\dagger b = R^{-1}h_1$ (since R is triangular, R^{-1} can be computed easily by substitution).

Continuous Dependence While for the case of an invertible matrix A, $A^\dagger b = A^{-1}b$ depends continuously not only on b, but also on A, this is not the case in general: Consider, for example, the matrix

$$A(\epsilon) = \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix}; \tag{5.40}$$

then

$$A(\epsilon)^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\epsilon} \end{pmatrix} \tag{5.41}$$

for $\epsilon \neq 0$, and we have

$$A(0)^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{5.42}$$

so that, e.g.,

$$A(\epsilon)^\dagger \begin{pmatrix} 0 \\ 1 \end{pmatrix} \nrightarrow A(0)^\dagger \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{5.43}$$

as $\epsilon \to 0$.

Truncated Singular Value Regularization. The identity (5.36) also shows how errors in y affect the result $A^\dagger y$: Errors in components of y corresponding to small singular values are amplified by the large factor of the singular value, so that such data errors are dangerous. This explains the numerical instability of (5.36), if A has small singular values. Although the problem of computing the best-approximate solution is well posed, it is then numerically unstable. The first idea to reduce this instability is to replace (5.36) by

$$x_\alpha = \sum_{\substack{i=1 \\ \sigma_i^2 \geq a}}^{r} \frac{\langle y, v_i \rangle}{\sigma_i} u_i \tag{5.44}$$

with an appropriately chosen value $\alpha > 0$; this truncation is the first example of a "regularization", where the original problem is replaced by a neighboring one, which is more stable. However, the choice of the "regularization parameter" α is quite crucial.

If we use (5.44) with perturbed data y^ε (with $\|y - y^\varepsilon\| \le \varepsilon$), we obtain as the "regularized solution"

$$x_\alpha^\varepsilon = \sum_{\substack{i=1 \\ \sigma_i^2 \ge \alpha}}^{r} \frac{\langle y^\varepsilon, v_i \rangle}{\sigma_i} u_i. \tag{5.45}$$

We estimate the total error between x_α^ε and the sought-for quantity $A^\dagger y$:

$$\|x_\alpha^\varepsilon - A^\dagger y\| \le \left\| \sum_{i=1}^{r} \frac{\langle y, v_i \rangle}{\sigma_i} u_i - \sum_{\substack{i=1 \\ \sigma_i^2 \ge \alpha}}^{r} \frac{\langle y, v_i \rangle}{\sigma_i} u_i \right\| + \left\| \sum_{\substack{i=1 \\ \sigma_i^2 \ge \alpha}}^{r} \frac{\langle y, v_i \rangle - \langle y^\varepsilon, v_i \rangle}{\sigma_i} u_i \right\|. \tag{5.46}$$

For sufficiently small α, the first summand on the right-hand side is empty. For the second summand we have

$$\left\| \sum_{\substack{i=1 \\ \sigma_i^2 \ge \alpha}}^{r} \frac{\langle y, v_i \rangle - \langle y^\varepsilon, v_i \rangle}{\sigma_i} u_i \right\|^2 = \sum_{\substack{i=1 \\ \sigma_i^2 \ge \alpha}}^{r} \frac{|\langle y - y^\varepsilon, v_i \rangle|^2}{\sigma_i^2} \le \frac{1}{\alpha} \sum_{i=1}^{r} |\langle y - y^\varepsilon, v_i \rangle|^2 \le \frac{\varepsilon^2}{\alpha}. \tag{5.47}$$

The sum (5.45) is called *truncated singular value expansion*. It can be interpreted as applying a low-pass filter to the data.

Tikhonov Regularization. Another way of making (5.36) more stable would be to replace it by the sum

$$x_\alpha^\varepsilon = \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 + \alpha} \langle y^\varepsilon, v_i \rangle u_i. \tag{5.48}$$

This is a famous *Tikhonov regularization method* (cf. [211–215]), which we shall consider in more detail in infinite dimensions, too. It is helpful to characterize it in a different way: If x_α^ε is defined by (5.48), then it is not hard to see that

$$\left(A^H A + \alpha I\right) x_\alpha^\varepsilon = A^H y^\varepsilon, \tag{5.49}$$

which is an alternative characterization of the *Tikhonov regularization*. From this "*regularized normal equation*", we can obtain still another characterization of x_α^ε, namely as the unique minimizer of the so-called *Tikhonov functional*

$$x \mapsto \|Ax - y^\varepsilon\|^2 + \alpha \|x\|^2, \tag{5.50}$$

which can be seen by putting the first derivative of the functional in (5.50) to 0, resulting exactly in the linear equation (5.49). The minimization of (5.50) can be seen as a combination of the two minimizations that appear in the definition of a best-approximate solution. It has also interpretations as a penalty method, e.g., via Lagrange multipliers.

The computation explained above can also be performed for $\alpha = 0$ (with (5.30) replaced by (5.36)). In this case it shows that x is the solution of minimal norm of the *normal equations*

$$A^H A x = A^H y, \tag{5.51}$$

which was already attacked by C.F. Gauss [72], (see also the contributions by R.L. Plackett [175], D.W. Robinson [183]). If $A^H A$ is invertible (and hence positive definite), the normal equations (5.51) can be solved by standard Cholesky decomposition, which leads to an alternative method for computing best-approximation solutions, for which no SVD is needed. However, as mentioned above, forming $A^H A$ may seriously decrease the numerical stability. Hence, it should be avoided in ill-conditioned cases.

5.4. Today's functional analytical nomenclature

The following nomenclature is part of any functional analytical textbook (see, e.g., [120, 210, 241]). A mapping $A : X \to Y$ between two normed spaces X and Y is called an *operator*. With

$$\mathcal{N}(A) = \{x \in X \,:\, Ax = 0\} \subset X \tag{5.52}$$

and

$$\mathcal{R}(A) = \{y = Ax \,:\, x \in X\} \subset Y \tag{5.53}$$

we denote the *kernel* (or the *null space*) of A and the *image* of A, respectively.

Definition 5.3. Let X and Y be normed spaces. The operator $A : X \to Y$ is called linear, if

(1) $A(x + y) = Ax + Ay$ for all $x, y \in X$,
(2) $A(\alpha x) = \alpha Ax$ for all $x \in X$ and $\alpha \in \mathbb{K}$.

The operator A is called *bounded,* if there exists a constant $C \geq 0$ such that

$$\|Ax\|_Y \leq C \|x\|_X \quad \text{for all } x \in X.$$

Theorem 5.3. *Let X and Y be normed spaces and $A : X \to Y$ be a linear operator. Then the following statements are equivalent:*

(1) *A is continuous on X.*
(2) *A is continuous in $0 \in X$.*
(3) *A is bounded on X.*

The *space of all continuous linear operators* between X and Y is denoted by $L(X, Y)$. If $Y = X$ we set $L(X) = L(X, X)$. A norm on $L(X, Y)$ is given by

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|_Y}{\|x\|_X} = \sup_{\|x\|_X \leq 1} \|Ax\|_Y = \sup_{\|x\|_X = 1} \|Ax\|_Y \,. \tag{5.54}$$

In order to be more precise we often use the notation $\|A\|_{X \to Y}$ for the norm of an operator $A : X \to Y$.

Theorem 5.4. *Together with the norm* (5.54) *the space $L(X, Y)$ is a normed space. If X is a normed space and Y is a Banach space, then $L(X, Y)$ is a Banach space.*

For combinations of two linear operators we have the following result.

Theorem 5.5. *Let $A \in L(X, Y)$ and $B \in L(Y, Z)$, then $BA \in L(X, Z)$ and we have*

$$\|BA\|_{X \to Z} \leq \|B\|_{Y \to Z} \|A\|_{X \to Y} \,. \tag{5.55}$$

For $A \in L(X)$ we get iteratively

$$\|A^n\| \leq \|A\|^n \tag{5.56}$$

for all $n \in \mathbb{N}$.

For sequences of operators two different terms of convergence are used.

Definition 5.4. Let $\{A_n\}_{n\in\mathbb{N}} \subset L(X,Y)$ be a sequence of operators between X and Y and let $A \in L(X,Y)$.

(1) A sequence of operators $\{A_n\}_{n\in\mathbb{N}}$ is called *pointwise convergent* to an operator A if

$$\lim_{n\to\infty} \|A_n x - Ax\|_Y = 0 \tag{5.57}$$

for all $x \in X$.

(2) A sequence of operators $\{A_n\}_{n\in\mathbb{N}}$ is called *uniformly convergent* to an operator A if

$$\lim_{n\to\infty} \|A_n - A\|_{X\to Y} = 0. \tag{5.58}$$

Uniform convergence implies pointwise convergence but the converse is in general not true.

Since we mainly study inverse problems, the inversion of operators are of particular interest. For the existence of a continuous inverse of an operator we have the following equivalent criterion.

Theorem 5.6. *The linear operator $A : X \to Y$ between the normed spaces X, Y has a continuous inverse $A^{-1} \in L(\mathcal{R}(A), X)$ if and only if there exists a constant $c > 0$ such that*

$$c\,\|x\|_X \leq \|Ax\|_Y \quad \text{for all } x \in X\,. \tag{5.59}$$

In this case,

$$\left\|A^{-1}\right\| \leq c^{-1}. \tag{5.60}$$

As for the existence of a bounded inverse of an operator A we also have an equivalent criterion for the non-existence of a bounded inverse.

Theorem 5.7. *An operator $A \in L(X,Y)$ does not have a continuous inverse if and only if there exists a sequence $\{x_n\} \subset X$ with $\|x_n\|_X = 1$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} Ax_n = 0$.*

Since $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, the space $L(X,\mathbb{K})$ is clearly a Banach space. It is symbolized by X^* and called *dual space* of X. The elements of X^* are called linear functionals.

Well-Posedness in the Sense of Hadamard. A *mathematical model* is a mapping $A : X \to Y$ from the set (of causes) X to the set (of effects) Y. $\mathcal{D}(A) = X$, $\mathcal{R}(A) = \{y = Ax : x \in X\}$ and $\mathcal{N}(A) = \{x \in X : Ax = 0\}$ are the domain (of definition) of A, the range (or image) of A, and the null space (or kernel) of A, respectively. We are interested in investigating the operator equation

$$Ax = y, \quad x \in X,\ y \in Y. \tag{5.61}$$

A *direct problem* amounts to the description and evaluation of y, i.e., Ax. An *inverse problem* is the task to find, for a given element $y \in Y$, an element $x \in X$ such that (5.61) holds true.

Definition 5.5 (Hadamard's Classification). Let X, Y be metric spaces. Suppose that $A : X \to Y$ is a mapping from X to Y. The problem $(A; X, Y)$ is called *well posed* (in the sense of Hadamard) if

(H1) For every $y \in Y$ there exists an $x \in X$ with $Ax = y$ (*Existence of the Inverse*).

(H2) For every $y \in Y$ there exists one and only one $x \in X$ with $Ax = y$ (*Uniqueness of the Inverse*).

(H3) The inverse mapping $A^{-1} : Y \to X$ is continuous, i.e., the solution $x \in X$ of $Ax = y$ depends continuously on $y \in Y$ (*Continuous Dependence*).

If one of the three properties is violated, then the problem is called *ill-posed* (in the sense of Hadamard).

Using a singular system in finite-dimensional context, x can also be written as the sum $\sum_{i=1}^{r} \sigma_i^{-1} \langle y, v_i \rangle u_i$, such that

$$A^{\dagger} y = \sum_{i=1}^{r} \frac{\langle y, v_i \rangle}{\sigma_i} u_i, \tag{5.62}$$

which will also generalize to the infinite-dimensional setting.

The matrix $A^H A$ is invertible if $\mathcal{N}(A^H A) = \mathcal{N}(A) = \{0\}$. Then (cf. [152, 153]) we have

$$A^{\dagger} = (A^H A)^{-1} A^H. \tag{5.63}$$

The SVD of A immediately lead us to the famous "*Moore–Penrose equations*", which also characterize $A^{\dagger}$:

$$A^{\dagger} A A^{\dagger} = A^{\dagger}, \tag{5.64}$$

$$A A^{\dagger} A = A, \tag{5.65}$$

$$(A^{\dagger} A)^H = A^{\dagger} A, \tag{5.66}$$

$$(A A^{\dagger})^H = A A^{\dagger}. \tag{5.67}$$

It can also be seen that $A^{\dagger} A$ and $A A^{\dagger}$ are orthogonal projectors onto $\mathcal{N}(A)^{\perp}$ and $\mathcal{R}(A)$, respectively. These facts, in turn, can be used to characterize $A^{\dagger}$ in yet another way, namely as the unique linear operator satisfying

$$A^{\dagger}|_{\mathcal{R}(A)} = \left[A|_{\mathcal{N}(A)^{\perp}} \right]^{-1} \tag{5.68}$$

and

$$\mathcal{N}(A^{\dagger}) = \mathcal{R}(A)^{\perp}. \tag{5.69}$$

In other words, the Moore–Penrose (generalized) inverse is the unique linear extension of $\left[A|_{\mathcal{N}(A)^\perp}\right]^{-1}$ to

$$A^\dagger = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp \tag{5.70}$$

satisfying (5.69). Moreover, we have

Theorem 5.8 (Pseudosolution). *For each $y \in \mathcal{D}(A^\dagger)$, the equation* (5.26) *has a unique best-approximate-solution* (*pseudosolution*) *given by*

$$x^\dagger = A^\dagger y, \tag{5.71}$$

where the set of all least squares solutions is given by $\{x^\dagger\} + \mathcal{N}(A)$.

5.5. Pseudoinverse for infinite-dimensional operator equations

Next our goal is to extend some material known from matrix analysis in (finite-dimensional) Euclidean spaces to linear operator framework between Hilbert spaces. Central in our considerations is the introduction of the pseudoinverse. We begin with some preparatory remarks clarifying the functional analytic background in order to make the Hilbert space (HS) context of the theory of ill-posed problems (IPP) more transparent:

Let X and Y be Hilbert spaces and let $A : X \to Y$ be a bounded linear operator whose range $\mathcal{R}(A)$ is not necessarily closed. Then we have the orthogonal decompositions

$$X = \mathcal{N}(A) \oplus \mathcal{N}(A)^\perp, \quad A = \overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^\perp \tag{5.72}$$

and

$$\mathcal{N}(A^*) = \mathcal{R}(A)^\perp, \tag{5.73}$$

where $\mathcal{N}(A)$ is the null space of A, $\overline{\mathcal{R}(A)}$ is the closure of the range of A, and A^* is the adjoint operator of A, i.e., $\langle Ax, y\rangle = \langle x, A^*y\rangle$ for all $x \in X$ and $y \in Y$.

Four (mutually exclusive) situations (S1), ..., (S4) arise in considering the operator equation (cf. [165])

$$Ax = y, \quad x \in X, \quad y \in Y, \tag{5.74}$$

involving Hilbert spaces X, Y:

(S1) $\mathcal{R}(A)$ is dense in Y, (so $\mathcal{N}(A^*) = \{0\}$), and $y \in \mathcal{R}(A)$;

(S2) $\mathcal{R}(A)$ is dense in Y, and $y \notin \mathcal{R}(A)$;

(S3) $\overline{\mathcal{R}(A)}$ is a proper subspace of Y, and $y \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$;

(S4) $\overline{\mathcal{R}(A)} \neq Y$, and $y \notin \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$.

In case (S1), the operator equation (5.74) has, of course, a solution in the classical sense; in cases (S2) and (S4), a classical solution does not exist, while in case (S3) a solution need not exist. Later on, in analogy to finite-dimensional settings, we shall discuss that x is a "least squares solution" of the operator equation (5.74) if $\inf\{\|Az - y\| : z \in X\} = \|Ax - y\|$. Since $\|Az - y\|^2 = \|Az - Qy\|^2 + \|y - Qy\|^2$, where Q is the orthogonal projector $P_{\overline{\mathcal{R}(A)}}$ of Y onto $\overline{\mathcal{R}(A)}$, we are led to the fact

that a least squares solution exists if and only if $y \in \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$, which is a dense set in Y. For such y the set of all least squares solutions of $Ax = y$, denoted by $\mathbb{L}(y)$, is a non-empty closed convex set (indeed, $\mathbb{L}(y)$ is the translate of $\mathcal{N}(A)$ by a fixed element of $\mathbb{L}(y)$), hence, has a unique element of minimal norm, denoted by $A^{\dagger}y$.

Hadamard's and Nashed's Classifications. It turns out that the statements (H1), (H2), and (H3) of Hadamard's classification (Definition 5.5) are satisfied by the Moore–Penrose inverse $A^{\dagger}$ if and only if $\mathcal{R}(A) = \overline{\mathcal{R}(A)}$. This observation leads to a new notion of well-posedness that goes back to Nashed (for more details see [166]).

Definition 5.6 (Nashed's Classification). The problem $(A; X, Y)$ is called *well posed in the sense of Nashed*, if $\mathcal{R}(A)$ is closed in Y. Otherwise, i.e., if $\mathcal{R}(A)$ is not closed in Y, the problem $(A; X, Y)$ is called *ill posed in the sense of Nashed*.

In accordance with M.Z. Nashed [164], we are led to say that *the operator equation* (5.74) is *well posed in the least squares* (relative to X and Y) if, for each $y \in Y$, it has a unique least squares solution (of minimal norm), which depends continuously on y; otherwise the problem is ill posed. The advantage of adopting this notion of well-posedness is that it focuses on infinite-dimensional problems (e.g., an inconsistent finite system of linear algebraic equations will not be ill posed in our sense, while it is ill posed in the sense of Hadamard). Furthermore, we are led to show by functional analytic means that the following statements are equivalent:

(a) the operator equation (5.74) is well posed;
(b) $\mathcal{R}(A)$ is closed;
(c) $A^{\dagger}$ is bounded.

As a consequence, it turns out that the *pseudoinverse* $A^{\dagger}$ is the linear operator which assigns to each $y \in \mathcal{D}(A^{\dagger}) = \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$, the unique element in $\mathbb{L}(y) \cap \mathcal{N}(A)^{\perp}$, so that $\mathbb{L}(y) = A^{\dagger}y + \mathcal{N}(A)$. $A^{\dagger}y$ is the minimal-norm solution (equivalently the unique solution in $\mathcal{N}(A)^{\perp}$) of the normal equations

$$A^{*}Ax = A^{*}y \tag{5.75}$$

(the equation obtained with $Q = P_{\overline{\mathcal{R}(A)}}$ by setting the first variation of $\|Ax - y\|^2$ equal to zero). $A^{\dagger} = (A/\mathcal{N}(A)^{\perp})^{-1}Q$ so that $A^{\dagger}$ can be characterized as the linear operator with the function-theoretic properties: $D(A^{\dagger}) = \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$, $\mathcal{N}(A^{\dagger}) = \mathcal{R}(A)^{\perp} = \mathcal{N}(A^{*})$, and $\mathcal{R}(A^{\dagger}) = \mathcal{N}(A)^{\perp}$.

In fact, we are led to the following conclusions: In case (S1) above, $A^{\dagger}$ indeed gives to the minimal-norm solution of $Ax = y$. In case (S3), $Ax = y$ has a least squares solution (which is unique if and only if $\mathcal{N}(A) = \{0\}$). In both cases, the infimum is attained and is equal to zero and $\|y - Qy\|$, respectively. Cases (S2) and (S4) are pathological and are of no deeper interest in pseudoinverse theory and since in both cases $y \notin \mathcal{D}(A^{\dagger})$, and the infimum is not attained.

After the general remarks we are now prepared to characterize least squares solutions of operator equations in more detail: Let X and Y be Hilbert spaces

and $A \in L(X,Y)$, i.e., A is linear and bounded. Our interest is to determine an element $x \in X$ for a given element $y \in Y$ such that the operator equation

$$Ax = y, \quad x \in X,\ y \in Y \tag{5.76}$$

becomes attackable (in least squares sense). An element $x \in X$ is called a

(i) *least squares solution* of (5.76), if

$$\|Ax - y\|_Y = \inf\{\|Az - y\|_Y :\ z \in X\} \tag{5.77}$$

(ii) *best-approximate solution* (*or minimal norm solution*) of (5.76), if $x \in X$ solves (5.77) and

$$\|x\|_X = \inf\{\|z\| : z \text{ is least squares solution of } Ax = y\}.$$

Obviously, the problem in the framework of $L(X,Y)$ involving Hilbert spaces X and Y is to minimize the residual.

Theorem 5.9. *Suppose that $y \in Y$ and $A \in L(X,Y)$ with X, Y Hilbert spaces. Then, the following statements are equivalent:*

(1) *x fulfills*

$$Ax = P_{\overline{\mathcal{R}(A)}} y, \tag{5.78}$$

where $P_{\overline{\mathcal{R}(A)}}$ denotes the orthogonal projection on $\overline{\mathcal{R}(A)}$.

(2) *x minimizes the residual, i.e.,*

$$\|Ax - y\| \le \|Ax' - y\| \quad \text{for all } x' \in X. \tag{5.79}$$

(3) *$x \in X$ solves the so-called normal equations*

$$A^*Ax = A^*y. \tag{5.80}$$

Proof. (1) $\Rightarrow$ (2): Suppose that $x' \in X$ and $y \in Y$. Then we obtain $P_{\overline{\mathcal{R}(A)}} y - y \in \overline{\mathcal{R}(A)}^{\perp}$, and we have

$$\|Ax' - y\|^2 = \left\|Ax' - P_{\overline{\mathcal{R}(A)}} y\right\|^2 + \left\|P_{\overline{\mathcal{R}(A)}} y - y\right\|^2 + 2\left\langle Ax' - P_{\overline{\mathcal{R}(A)}} y, P_{\overline{\mathcal{R}(A)}} y - y\right\rangle \tag{5.81}$$

such that the Pythagorean theorem tells us that

$$\begin{aligned} \|Ax' - y\|^2 &= \left\|Ax' - P_{\overline{\mathcal{R}(A)}} y\right\|^2 + \|Ax - y\|^2 \\ &\ge \|Ax - y\|^2 \qquad \text{for all } x' \in X. \end{aligned} \tag{5.82}$$

(2) $\Rightarrow$ (3): Suppose that $x' \in X$ and $F(\lambda) = \|A(x + \lambda x') - y\|^2$. The condition (2) tells us that F possesses a minimum for $\lambda = 0$. Therefore, we obtain

$$\begin{aligned} 0 = \frac{\partial F}{\partial \lambda}(0) &= \left(\frac{\partial}{\partial \lambda}\left(\langle Ax + \lambda Ax' - y, Ax + \lambda Ax' - y\rangle\right)\right)_{\lambda=0} \\ &= 2\langle Ax', Ax - y\rangle \\ &= 2\langle x', A^*Ax - A^*y\rangle \end{aligned} \tag{5.83}$$

for all $x' \in X$. Hence, we are able to conclude that

$$A^*Ax - A^*y = 0. \tag{5.84}$$

(3) $\Rightarrow$ (1): For $y \in Y$, let $x \in X$ satisfy the equation

$$A^*Ax = A^*y \Leftrightarrow A^*(Ax - y) = 0. \tag{5.85}$$

Then, from functional analysis, it follows that

$$\begin{aligned} & Ax - y \in \mathcal{N}(A^*) = \overline{\mathcal{R}(A)}^{\perp} \\ \Rightarrow \quad & P_{\overline{\mathcal{R}(A)}}(Ax - y) = 0 \\ \Rightarrow \quad & P_{\overline{\mathcal{R}(A)}}Ax = P_{\overline{\mathcal{R}(A)}}y. \end{aligned} \tag{5.86}$$

In other words,

$$Ax = P_{\overline{\mathcal{R}(A)}}y. \tag{5.87}$$

This completes our proof. □

Remark 5.4. The normal equations (5.80) owe their name to the property that $Ax - y \in \mathcal{R}(A)^{\perp}$, i.e., $Ax - y \in \mathcal{N}(A^*)$.

Theorem 5.10. *Assume that y belongs to Y. Then the following statements are true:*

(1) *The set of solutions of the normal equations*

$$\mathbb{L}(y) = \{x \in X : A^*Ax = A^*y\} \tag{5.88}$$

is non-empty if and only if $y \in \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$.

(2) *$\mathbb{L}(y)$ is closed and convex.*

Proof. (1) Assume that $x \in \mathbb{L}(y)$. Then we have $y = Ax + (y - Ax) \in \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$. Moreover, suppose that $y \in \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$. Then there exists elements $x \in X$ and $\widetilde{y} \in \mathcal{R}(A)^{\perp}$ satisfying

$$y = Ax + \widetilde{y} \tag{5.89}$$

and

$$P_{\overline{\mathcal{R}(A)}}y = P_{\overline{\mathcal{R}(A)}}Ax + P_{\overline{\mathcal{R}(A)}}\widetilde{y} = Ax + 0.$$

Thus, x satisfies the normal equations, hence, $x \in \mathbb{L}(y)$.

(2) Let $\{x_n\}_n$ be a sequence in $\mathbb{L}(y)$ which converges to $x \in X$. Both operators A and A^* are continuous. Therefore, we obtain

$$A^*y = A^*Ax_n, \tag{5.90}$$

for all $n \in \mathbb{N}$. Furthermore, for $n \to \infty$,

$$A^*y = A^*Ax, \tag{5.91}$$

which shows us that $x \in \mathbb{L}(y)$, i.e., $\mathbb{L}(y)$ is closed. Suppose now that $x, x' \in \mathbb{L}(y)$ and $0 \leq \lambda \leq 1$. Then the identities

$$\begin{aligned} A^* A(\lambda x + (1-\lambda) x') &= \lambda A^* A x + (1-\lambda) A^* A x' \\ &= \lambda A^* y + (1-\lambda) A^* y \\ &= A^* y, \end{aligned} \tag{5.92}$$

imply that $\mathbb{L}(y)$ is convex. □

Remark 5.5. Generally, we have

$$\mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp} \neq Y. \tag{5.93}$$

However, it should be noted that $\mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp} = Y$ if the dimension of Y is finite.

Lemma 5.1 (Pseudosolution). *Suppose that y belongs to $\mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$. Then $\mathbb{L}(y)$ possesses a unique element $x^{\dagger} \in \mathbb{L}(y)$ satisfying*

$$\left\|x^{\dagger}\right\| < \|x\| \tag{5.94}$$

for all $x \in \mathbb{L}(y) \setminus \left\{x^{\dagger}\right\}$, i.e., there exists an element of minimal norm.

Proof. $\mathbb{L}(y)$ is non-empty, closed and convex. $x^{\dagger}$ is the best-approximate element to $0 \in X$ in $\mathbb{L}(y)$. □

Remark 5.6. The reason for the proof of Lemma 5.1 is the theorem of best-approximate elements. Let $\emptyset \neq U \subset X$ be closed and convex and $x \in X$. Then there exists one and only one $u \in U$ such that

$$\|x - u\|_X < \|x - u'\|_X \tag{5.95}$$

for all $u' \in X \setminus \{x\}$. The proof can be found in any standard textbook on functional analysis (see e.g., [110]).

Pseudoinverse (Moore–Penrose Inverse, Generalized Inverse). After these considerations concerned with least squares approaches we are in position to introduce the *pseudoinverse* (*Moore–Penrose inverse, generalized inverse*) *of operators* $A \in L(x, y)$, where X and Y are Hilbert spaces.

Definition 5.7. The operator $A^{\dagger} : \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp} \to X$ (i.e., $\mathcal{D}\left(A^{\dagger}\right) = \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp} \subset Y$), that maps each element $y \in \mathcal{D}\left(A^{\dagger}\right)$ to the unique element $x^{\dagger} \in \mathbb{L}(y)$ with minimal norm, is called the *pseudoinverse* or *Moore–Penrose inverse, generalized inverse* of A.

Theorem 5.11. *Let y belong to $\mathcal{D}\left(A^{\dagger}\right)$. Then $x^{\dagger} = A^{\dagger} y$ is the best-approximate solution of* (5.76)*. It represents the unique solution of the normal equations in $\mathcal{N}(A)^{\perp}$, i.e., $x^{\dagger} = A^{\dagger} y$ if and only if $A^* A x^{\dagger} = A^* y$ and $x^{\dagger} \in \mathcal{N}(A)^{\perp}$.*

Proof. "⇒" Suppose that $x^\dagger = A^\dagger y$. Then $x^\dagger$ solves the normal equations and has minimal norm by Definition 5.7. It remains to show that $x^\dagger \in \mathcal{N}(A)^\perp$: Assume that $x^\dagger \in \mathcal{N}(A)$ and $x^\dagger \neq 0$. Then it follows that

$$A^*Ax^\dagger = A^*0 = 0 = A^*y, \tag{5.96}$$

$$A^*A\ 0 = 0 = A^*y. \tag{5.97}$$

Consequently, $0 \in \mathbb{L}(y)$. However, $0 = \|0\| < \|x^\dagger\|$ which forms a contradiction to the minimal norm property of $x^\dagger$. Thus, $x^\dagger \in \mathcal{N}(A)^\perp$ since $X = \mathcal{N}(A) \oplus \mathcal{N}(A)^\perp$.

"⇐" Suppose that $x^\dagger \in \mathcal{N}(A)^\perp$ and $A^*Ax^\dagger = A^*y$. Then, $x^\dagger \in \mathbb{L}(y)$. We have to show that $x^\dagger$ has minimal norm. Let $x' \in \mathbb{L}(y)$ be arbitrary. It follows that

$$A(x^\dagger - x') = Ax^\dagger - Ax' = P_{\overline{\mathcal{R}(A)}}y - P_{\overline{\mathcal{R}(A)}}y = 0. \tag{5.98}$$

Thus, $x^\dagger - x' \in \mathcal{N}(A)$. Now we consider the norm of x':

$$\|x'\|^2 = \|x^\dagger + (x' - x^\dagger)\|^2 = \|x^\dagger\|^2 + \|x' - x^\dagger\|^2 + 2\underbrace{\langle x^\dagger, x' - x^\dagger\rangle}_{=0} \geq \|x^\dagger\|. \tag{5.99}$$

Note that the scalar product is 0 since $x^\dagger \in \mathcal{N}(A)^\perp$, whereas $x' - x^\dagger \in \mathcal{N}(A)$. As a consequence, we have $x^\dagger = A^\dagger y$. □

Theorem 5.12 (Properties of the Pseudoinverse). *The generalized inverse $A^\dagger$ possesses the following properties:*

(1) $\mathcal{D}\left(A^\dagger\right) = Y$ *if and only if* $\mathcal{R}(A)$ *is closed.*
(2) $\mathcal{R}\left(A^\dagger\right) = \mathcal{N}(A)^\perp$.
(3) $A^\dagger$ *is linear.*
(4) $A^\dagger$ *is continuous if and only if* $\mathcal{R}(A)$ *is closed.*

Proof. (1) $\mathcal{D}\left(A^\dagger\right) = Y$ is equivalent to $\mathcal{R}(A) \oplus \mathcal{R}(A)^\perp = Y$, i.e., $\mathcal{R}(A)$ is closed.

(2) We have $\mathcal{R}\left(A^\dagger\right) \subset \mathcal{N}(A)^\perp$ due to Theorem 5.11. Choose $x \in \mathcal{N}(A)^\perp$ and set $y = Ax$. Then we obtain

$$P_{\overline{\mathcal{R}(A)}}y = Ax \Rightarrow x \in \mathbb{L}(y). \tag{5.100}$$

Using Theorem 5.11 we are led to $\mathcal{N}(A)^\perp \subset \mathcal{R}\left(A^\dagger\right)$.

(3) Suppose that $y, y' \in \mathcal{D}\left(A^\dagger\right)$. Then we have

$$AA^\dagger y = P_{\overline{\mathcal{R}(A)}}y, \tag{5.101}$$

$$AA^\dagger y' = P_{\overline{\mathcal{R}(A)}}y'. \tag{5.102}$$

Thus it is clear that

$$A\left(A^\dagger y + A^\dagger y'\right) = P_{\overline{\mathcal{R}(A)}}(y + y') = AA^\dagger(y + y') \tag{5.103}$$

and

$$\left(A^{\dagger}y + A^{\dagger}y' - A^{\dagger}\left(y + y'\right)\right) \in \mathcal{N}\left(A\right), \tag{5.104}$$

$$\left(A^{\dagger}y + A^{\dagger}y' - A^{\dagger}\left(y + y'\right)\right) \in \mathcal{R}\left(A^{\dagger}\right) = \mathcal{N}\left(A\right)^{\perp}, \tag{5.105}$$

so that

$$A^{\dagger}y + A^{\dagger}y' - A^{\dagger}\left(y + y'\right) = 0. \tag{5.106}$$

Analogously we are able to show that $A^{\dagger}\left(\alpha y\right) = \alpha A^{\dagger}y$ for all $y \in \mathcal{D}\left(A^{\dagger}\right)$ and $\alpha \in \mathbb{C}$.

(4) Assume that $A^{\dagger}$ is continuous. Then, $\mathcal{D}\left(A^{\dagger}\right) = \mathcal{R}\left(A\right) \oplus \mathcal{R}\left(A\right)^{\perp}$ is dense in Y and $A^{\dagger}$ can be extended continuously to all of Y by $B \in L\left(Y, X\right)$ (in accordance with the continuous extension of operators, see, e.g., [110]). It follows that

$$ABy = P_{\overline{\mathcal{R}(A)}}y \tag{5.107}$$

for all $y \in Y$, i.e.,

$$\overline{\mathcal{R}\left(A\right)} = \mathcal{R}\left(P_{\overline{\mathcal{R}(A)}}\right) \subset \mathcal{R}\left(A\right). \tag{5.108}$$

Therefore,

$$\overline{\mathcal{R}\left(A\right)} = \mathcal{R}\left(A\right). \tag{5.109}$$

On the other hand, let $\mathcal{R}\left(A\right)$ be closed. Consider the operator $\hat{A}$ given by

$$\hat{A} : \mathcal{N}\left(A\right)^{\perp} \to \mathcal{R}\left(A\right), x \mapsto Ax. \tag{5.110}$$

$\hat{A}$ is bijective and bounded. Due to the inverse mapping theorem of functional analysis (see, e.g., [110]) $\hat{A}^{-1}$ is also bounded, and we obtain

$$\begin{aligned} \left\|A^{\dagger}y\right\|_X &= \left\|\hat{A}^{-1}\hat{A}A^{\dagger}y\right\|_X \\ &\leq \left\|\hat{A}^{-1}\right\|_{Y\to X}\left\|\hat{A}A^{\dagger}y\right\|_Y \\ &= \left\|\hat{A}^{-1}\right\|_{Y\to X}\left\|AA^{\dagger}y\right\|_Y \end{aligned} \tag{5.111}$$

for all $y \in \mathcal{D}\left(A^{\dagger}\right) = Y$. Furthermore,

$$\left\|y\right\|_Y \geq \left\|P_{\overline{\mathcal{R}(A)}}y\right\|_Y = \left\|AA^{\dagger}y\right\|_Y \geq \left\|\hat{A}^{-1}\right\|^{-1}_{Y\to X}\left\|A^{\dagger}y\right\|_X \tag{5.112}$$

so that

$$\frac{\left\|A^{\dagger}y\right\|_X}{\left\|y\right\|_Y} \leq \left\|\hat{A}^{-1}\right\| \tag{5.113}$$

for all $y \in Y$. Altogether, we have

$$\left\|A^{\dagger}\right\|_{Y\to X} \leq \left\|\hat{A}^{-1}\right\|_{Y\to X}. \tag{5.114}$$

Thus, $A^{\dagger}$ is bounded. □

An equivalent way to introduce the generalized inverse is the introduction via the four Moore–Penrose conditions listed below.

Theorem 5.13 (Moore–Penrose Conditions). *The generalized inverse is uniquely determined by the four Moore–Penrose conditions:*

(1) $AA^\dagger A = A$,
(2) $A^\dagger A = P_{\overline{\mathcal{R}(A^\dagger)}}$,
(3) $A^\dagger AA^\dagger = A^\dagger$,
(4) $AA^\dagger = P_{\overline{\mathcal{R}(A)}}$,

where the last property is restricted to $\mathcal{D}(A^\dagger)$.

Proof. The proof is split into two parts:

(a) $A^\dagger$ satisfies the conditions (1) to (4),
(b) if $B \in L(Y,X)$ and $C \in L(Y,X)$ satisfy the Moore–Penrose conditions (1) to (4), then $B = C$.

We begin with (a): Clearly, $\mathcal{R}(A^\dagger) = \mathcal{N}(A)^\perp$ and $\mathcal{R}(A^\dagger) = \overline{\mathcal{R}(A^\dagger)}$. Let $y \in \mathcal{D}(A^\dagger)$. Then we are able to conclude

$$AA^\dagger y = Ax \quad \text{with } x = A^\dagger y \quad \Rightarrow A^*Ax = A^*y \quad \Rightarrow Ax = P_{\overline{\mathcal{R}(A)}} y \quad \Rightarrow (4). \tag{5.115}$$

Set $x = \underbrace{x_1}_{\in \mathcal{N}(A)^\perp} + \underbrace{x_2}_{\in \mathcal{N}(A)} = P_{\mathcal{N}(A)(A)^\perp} x + P_{\mathcal{N}(A)} x$. Then, it follows that

$$A^\dagger Ax_2 = A^\dagger 0 = 0 = P_{\mathcal{N}(A)^\perp} x_2 \tag{5.116}$$

and

$$z = A^\dagger Ax_1 \Rightarrow Az = P_{\overline{\mathcal{R}(A)}} Ax_1 = Ax_1 \Rightarrow A(z - x_1) = 0 \Rightarrow z - x_1 \in \mathcal{N}(A)\,. \tag{5.117}$$

But it should be noted that

$$z = A^\dagger Ax_1 \Rightarrow z \in \mathcal{R}(A^\dagger) = \mathcal{N}(A)^\perp \Rightarrow z - x_1 \in \mathcal{N}(A)^\perp\,. \tag{5.118}$$

Thus, from (5.117) and (5.118), it can be deduced that $z = x_1 = P_{\mathcal{N}(A)^\perp} x$. This means that

$$A^\dagger Ax = A^\dagger Ax_1 + \underbrace{A^\dagger Ax_2}_{=0} = P_{\mathcal{N}(A)^\perp} x = P_{\overline{\mathcal{R}(A^\dagger)}} x \quad \Rightarrow (2). \tag{5.119}$$

Now, we easily arrive at the other two Moore–Penrose conditions for $A^\dagger$:

$$A^\dagger AA^\dagger = P_{\overline{\mathcal{R}(A^\dagger)}} A^\dagger = A^\dagger \quad \Rightarrow (3), \tag{5.120}$$

$$AA^\dagger A = P_{\overline{\mathcal{R}(A)}} A = A \quad \Rightarrow (1). \tag{5.121}$$

This concludes the part (a) of the proof.

We continue with part (b): Assume that $B, C \in L(Y, X)$ satisfy (1) to (4). Then we have

$$\begin{aligned} B &= BAB \\ &= BAC \qquad (\text{note that } AB = P_{\overline{\mathcal{R}(A)}} = AC) \\ &= CAC \qquad (\text{note that } BA = P_{\overline{\mathcal{R}(B)}} = P_{\mathcal{N}(A)^\perp} = P_{\overline{\mathcal{R}(C)}} = CA) \\ &= C, \end{aligned} \tag{5.122}$$

where we still need to prove that $\overline{\mathcal{R}(B)} = \mathcal{N}(A)^\perp$ (which also shows that $\overline{\mathcal{R}(C)} = \mathcal{N}(A)^\perp$ since B and C can be exchanged). Let $x \in \overline{\mathcal{R}(B)}$. Assume that $x \in \mathcal{N}(A)$, $x \neq 0$. Then it follows that

$$Ax = 0 \quad \Rightarrow \quad BAx = 0 \quad \Rightarrow \quad P_{\overline{\mathcal{R}(B)}} x = 0 \quad \Rightarrow \quad x \in \overline{\mathcal{R}(B)}^\perp, \tag{5.123}$$

which is a contradiction. Thus, $x \in \mathcal{N}(A)^\perp$.

On the other hand, suppose that $x \in \mathcal{N}(A)^\perp$. Moreover, assume that $x \in \overline{\mathcal{R}(B)}^\perp$. Then

$$0 = P_{\overline{\mathcal{R}(B)}} x = BAx \quad \Rightarrow \quad Ax = ABAx = A0 = 0 \quad \Rightarrow \quad x \in \mathcal{N}(A), \tag{5.124}$$

which again is a contradiction and, therefore, $x \in \overline{\mathcal{R}(B)}$.

Summarizing our results we obtain

$$\overline{\mathcal{R}(B)} = \mathcal{N}(A)^\perp, \tag{5.125}$$

which yields $B = C$, i.e., the Moore–Penrose conditions uniquely determine the pseudoinverse concludes the part (b) of the proof. □

Compact Operators. Next we discuss a certain set of operators, viz. compact operators, that turn out to be prototypes for generating a large class of ill-posed problems.

Definition 5.8. Let X, Y be normed spaces. An operator $A : X \to Y$ is called *compact*, if one of the following equivalent conditions is fulfilled:

(1) Every bounded subset $U \subset X$ possesses an image in Y, which is relatively compact, i.e., $\overline{A(U)}$ is a compact set.
(2) For every bounded sequence $\{x_n\}_n \subset X$ the sequence $\{Ax_n\}_n$ possesses a subsequence that converges in Y.

By convention, we introduce the following settings:

$$K(X, Y) = \{A : X \to Y : A \text{ is linear and compact}\} \tag{5.126}$$

and

$$K(X) = K(X, X). \tag{5.127}$$

Lemma 5.2. *Let X, Y be normed spaces. Then the following statements hold true:*

(1) $K(X, Y) \subset L(X, Y)$.
(2) *If $A \in L(X, Y)$ with $\dim \mathcal{R}(A) < \infty$, then A is compact.*

Proof. (1) Assume that A belongs to $K(X,Y)$. Clearly, the closed unit ball $\overline{B_1(0)} \subset X$ is bounded. Therefore, $A(\overline{B_1(0)})$ is relatively compact and $A(\overline{B_1(0)})$ is bounded, i.e., $\|Ax\| \leq C$ for all $x \in \overline{B_1(0)}$. This implies $\sup_{x\in\overline{B_1(0)}} \|Ax\| \leq C$, so that A is continuous.

(2) $A : X \to Y$, $\dim \mathcal{R}(A) < \infty$. Therefore, each closed and bounded subset of $\mathcal{R}(A)$ is compact. This means that each bounded subset of $\mathcal{R}(A)$ is relatively compact. Now, let $U \subset X$ be bounded. Then $A(U)$ is also bounded and $A(U) \subset \mathcal{R}(A)$ is relatively compact. Thus, A is compact. □

Theorem 5.14. *Let X, Y, Z be normed spaces.*

(1) *If $A : X \to Y$ is compact and $B : Y \to Z$ is continuous or $A : X \to Y$ is continuous and $B : Y \to Z$ is compact, then $AB : X \to Z$ is compact.*
(2) *The identity operator $I : X \to X$ is compact if and only if X is finite-dimensional.*
(3) *Let Y be a Banach space. Then $K(X,Y)$ is closed, i.e., a sequence of compact operators $\{A_n\}_n \subset K(X,Y)$ with $\lim_{n\to\infty} \|A_n - A\|_{X\to Y} = 0$ has a compact limit, i.e., the limit operator A is compact.*
(4) *If A is compact and invertible and X is not finite-dimensional, then A^{-1} is* not *continuous.*

Example. Let $\mathcal{G} \subset \mathbb{R}^q$ be a regular region and let K be of class $C^{(0)}\left(\overline{\mathcal{G}} \times \overline{\mathcal{G}}\right)$. We introduce the integral operator $A : C^{(0)}\left(\overline{\mathcal{G}}\right) \to C^{(0)}\left(\overline{\mathcal{G}}\right)$ by letting

$$(AF)(x) = \int_{\mathcal{G}} K(x,y) F(y)\, dy, \quad F \in C^{(0)}\left(\overline{\mathcal{G}}\right). \tag{5.128}$$

The proof of the compactness of A can be based on a well-known theorem of classical analysis, namely the *Theorem of Arzelà–Ascoli.* This theorem (cf. Theorem 5.15) provides two equivalent properties to the relative compactness of a subset $U \subset C^{(0)}\left(\overline{\mathcal{G}}\right)$.

Theorem 5.15 (Theorem of Arzelà–Ascoli). *Let $\emptyset \neq \mathcal{G} \subset \mathbb{R}^q$ be regular. A subset $U \subset C^{(0)}\left(\overline{\mathcal{G}}\right)$ is relatively compact if and only if the following two statements are valid:*

(1) *U is equicontinuous, i.e., for every $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$, such that for all $F \in U$*
$$|F(x) - F(y)| < \varepsilon \tag{5.129}$$
for all $x, y \in G$ with $\|x - y\| < \delta(\varepsilon)$.
(2) *U is bounded, i.e., there exists an $M > 0$ with $\|F\|_{C^{(0)}\left(\overline{\mathcal{G}}\right)} < M$ for all $F \in U$.*

In accordance with the theorem of Arzelà–Ascoli (Theorem 5.15) we are now interested in applying the conditions (1) and (2) to the context of the integral operator introduced in (5.128):

(1) K is *uniformly continuous* in $\overline{\mathcal{G}} \times \overline{\mathcal{G}}$, i.e., for all $\varepsilon > 0$ exists a $\delta > 0$ such that for all $x, y, z \in \overline{\mathcal{G}}$ with $\|x - y\| < \delta$ holds $|K(x,z) - K(y,z)| < \frac{\varepsilon}{M \cdot \|\mathcal{G}\|}$. Thus

we get, for every bounded subset $U \subset C^{(0)}(\overline{\mathcal{G}})$,

$$\begin{aligned} |AF(x) - AF(y)| &= \left| \int_{\mathcal{G}} K(x,z) F(z)\, dz - \int_{\mathcal{G}} K(y,z) F(z)\, dz \right| \\ &\leq M \int_{\mathcal{G}} |K(x,z) - K(y,z)|\, dz \\ &< M\|\mathcal{G}\| \frac{\varepsilon}{M\|\mathcal{G}\|} \\ &= \varepsilon, \end{aligned} \tag{5.130}$$

provided that $\|x - y\| \leq \delta$. Therefore, AU is equicontinuous.

(2) Suppose that $U \subset C^{(0)}(\overline{\mathcal{G}})$ is *bounded*. Assume that F is of class U. Then we have

$$|AF(x)| = \left| \int_{\mathcal{G}} K(x,y) F(y)\, dy \right| \leq M \max_{x,y \in \overline{\mathcal{G}}} |K(x,y)|\, \|\mathcal{G}\| < \infty. \tag{5.131}$$

In other words, AU is bounded.

As a consequence, by virtue of Theorem 5.15, we are able to conclude that AU is relatively compact.

Theorem 5.16. *The operator A defined by* (5.128) *is compact.*

Example. Let $\mathcal{G} \subset \mathbb{R}^q$ be a regular region. We introduce the operator $A : L^2\left(\overline{\mathcal{G}}\right) \to L^2\left(\overline{\mathcal{G}}\right)$ given by

$$AF(x) = \int_{\mathcal{G}} K(x,y) F(y)\, dy, \quad F \in L^2(\mathcal{G}). \tag{5.132}$$

If K is continuous for $x \neq y$ and *weakly singular*, i.e., there exist a value $\alpha \in (0, q)$ and a constant $C > 0$ such that

$$|K(x,y)| \leq C \frac{1}{|x-y|^{q-\alpha}}, \tag{5.133}$$

then A is compact (for the proof the reader is referred to standard textbooks about integral equations, e.g., [30, 110, 120]). As a consequence, the Newton volume integral occurring in the theory of Earth's gravitation forms a compact operator A.

Singular Value Decomposition. Next we are concerned with the introduction of eigenvalues and eigenfunctions corresponding to an operator $A \in L(X)$.

Definition 5.9. Let X be a normed space. Suppose that A is of class $L(X)$.

(1) $\lambda \in \mathbb{C}$ is called a *regular value* of A if and only if $\lambda I - A$ is continuously invertible. $\rho(A) = \{\lambda \in \mathbb{C} : \lambda I - A$ continuously invertible$\}$ is called the *resolvent set*.

(2) $\sigma(A) = \mathbb{C} \setminus \rho(A)$ is called the *spectrum* of A.

(3) $\lambda \in \sigma(A)$ is called an *eigenvalue* of A if $\mathcal{N}(\lambda I - A) \neq \{0\}$. The elements of $\mathcal{N}(\lambda I - A) \setminus \{0\}$ are called *eigenvectors* of A corresponding to the eigenvalue λ.

We are now in position to verify the following result.

Theorem 5.17. *Let X be a Banach space.*

(1) *If $\lambda \in \sigma(A)$, then $|\lambda| \leq \|A\|$, i.e., the spectrum is bounded.*
(2) *$\sigma(A) \subset \mathbb{C}$ is compact.*

Theorem 5.18. *Let X be a normed space. Suppose that A is a compact operator on X (i.e., $A \in K(X)$).*

(1) *If $\lambda \in \sigma(A) \setminus \{0\}$, then λ is an eigenvalue of A.*
(2) *If λ is an eigenvalue of A, then $\dim \mathcal{N}(\lambda I - A) < \infty$.*
(3) *$\sigma(A)$ is at most countable. Furthermore, $0 \in \sigma(A)$.*
(4) *0 is the only accumulation point of $\sigma(A)$.*

The proofs are standard. They can be found in any textbook of functional analysis (e.g., [110]).

Central in our considerations about compact operator is the following spectral theorem that can be seen in parallel to the finite-dimensional case of matrix operators.

Theorem 5.19 (Spectral Theorem for Compact Self-adjoint Operators). *Let X be a Hilbert space. Assume that A is of class $K(X)$ and that A is self-adjoint, i.e., $A^* = A$. Then there exists an orthonormal system $\{x_i\}_{i\in\mathbb{N}} \subset X$ and a sequence $\{\mu_i\}_{i\in\mathbb{N}} \subset \mathbb{R}$ (finite or countable) with $|\mu_1| \geq |\mu_2| \geq \cdots > 0$, such that*

$$Ax = \sum_{i=1}^{\infty} \mu_i \langle x, x_i \rangle x_i \tag{5.134}$$

holds true for all $x \in X$.

Proof. Once again, the proof is standard. □

Suppose now that X and Y are Hilbert spaces. Furthermore, let A be of class $K(X,Y)$. Then A^*A is also compact and obviously self-adjoint. Due to the spectral theorem (Theorem 5.19) there exist a sequence $\{\lambda_i\}_{i\in\mathbb{k}N} \subset \mathbb{R}$ and an orthonormal system $\{x_i\}_{i\in\mathbb{N}} \subset X$ such that

$$A^*Ax = \sum_{i=1}^{\infty} \lambda_i \langle x, x_i \rangle x_i, \quad x \in X. \tag{5.135}$$

Suppose that $\lambda_i \in \sigma(A^*A) \setminus \{0\}$ and denote, as usual, by x_i its corresponding eigenvector. It follows that

$$\begin{aligned} \lambda_i \|x_i\|^2 &= \lambda_i \langle x_i, x_i \rangle = \langle \lambda_i x_i, x_i \rangle_X = \langle A^*Ax_i, x_i \rangle_X \\ &= \langle Ax_i, Ax_i \rangle_Y = \|Ax_i\|_Y^2 . \end{aligned} \tag{5.136}$$

Therefore we are able to conclude that $\lambda_i > 0$.

In the sequel, we assume that the eigenvalues are listed in the chronological order as follows: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_i \geq \lambda_{i+1} \geq \cdots \geq 0$. Set $\sigma_j = \sqrt{\lambda_j}$. Moreover, let

$y_i = \frac{1}{\sigma_i} A x_i$ i.e., $A x_i = \sigma_i y_i$, $i \in \mathbb{N}$, and $A^* y_i = A^*\left(\frac{1}{\sigma_i} A x_i\right) = \frac{1}{\sigma_i} A^* A x_i = \frac{1}{\sigma_i} \lambda_i x_i = \sigma_i x_i$. Furthermore, we have $\langle y_i, y_k \rangle_Y = \frac{1}{\sigma_i \sigma_k} \langle A x_i, A x_k \rangle_Y = \frac{1}{\sigma_i \sigma_k} \langle A^* A x_i, x_k \rangle_X = \delta_{i,k}$. Thus, $\{y_i\}_{i \in \mathbb{N}} \subset Y$ forms a complete orthonormal system (ONS) in $\overline{\mathcal{R}(A)}$, so that the system $\{x_i\}_{i \in \mathbb{N}}$ is a complete ONS in $\mathcal{N}(A)^\perp$. Now, assume that x is a member of $\mathcal{N}(A)^\perp$. Then it follows that

$$x = \sum_{i=1}^{\infty} \langle x, x_i \rangle_X \, x_i \tag{5.137}$$

and

$$A x = \sum_{i=1}^{\infty} \langle x, x_i \rangle_X \, A x_i = \sum_{i=1}^{\infty} \sigma_i \, \langle x, x_i \rangle_X \, y_i \tag{5.138}$$

for all $x \in \mathcal{N}(A)^\perp$.

Definition 5.10. Let X, Y be Hilbert spaces. The set $\{\sigma_i; x_i, y_i\}_{i \in \mathbb{N}} \subset (0, \infty) \times X \times Y$ is called the *singular system of an operator* $A \in K(X, Y)$. The values σ_i are called the *singular values* of A. The elements x_i, y_i are called the *singular vectors*. Furthermore, the series

$$A x = \sum_{i=1}^{\infty} \sigma_i \, \langle x, x_i \rangle_X \, y_i, \;\; x \in X \tag{5.139}$$

is called the *singular value decomposition* (SVD) of A.

Picard Condition. The following condition plays an essential role in the solvability of inverse problems.

Theorem 5.20 (Picard Condition). *If $A : X \to Y$ is compact with singular value decomposition (SVD) $\{\sigma_i; x_i, y_i\}_{i \in \mathbb{N}}$, then $y \in \overline{\mathcal{R}(A)}$ is an element of $\mathcal{R}(A)$ if and only if*

$$\sum_{i=1}^{\infty} \frac{|\langle y, y_i \rangle|^2}{\sigma_i^2} \tag{5.140}$$

is convergent.

Proof. "$\Rightarrow$" Suppose that y belongs to $\mathcal{R}(A)$. Then there exists a member $x \in X$ with $A x = y$ such that

$$\langle y, y_i \rangle_Y = \langle A x, y_i \rangle_Y = \langle x, A^* y_i \rangle_X \tag{5.141}$$

and

$$\sum_{i=1}^{\infty} \frac{|\langle y, y_i \rangle|^2}{\sigma_i^2} = \sum_{i=1}^{\infty} |\langle x, x_i \rangle|^2 \overset{\text{Bessel}}{\leq} \|x\|_X^2 < \infty. \tag{5.142}$$

"$\Leftarrow$" Suppose that y belongs to $\overline{\mathcal{R}(A)}$. Assume that the series $\sum_{i=1}^{\infty} \frac{|\langle y, y_i \rangle|^2}{\sigma_i^2}$ converges. We let

$$x = \sum_{i=1}^{\infty} \frac{\langle y, y_i \rangle}{\sigma_i} x_i. \tag{5.143}$$

Then it follows that

$$\|x\|_X^2 = \sum_{i=1}^{\infty} \frac{|\langle y, y_i\rangle|^2}{\sigma_i^2} < \infty \tag{5.144}$$

and

$$Ax = \sum_{i=1}^{\infty} \frac{\langle y, y_i\rangle}{\sigma_i} Ax_i = \sum_{i=1}^{\infty} \langle y, y_i\rangle\, y_i = P_{\overline{\mathcal{R}(A)}} y \overset{y\in\overline{\mathcal{R}(A)}}{=} y. \tag{5.145}$$

Therefore, y is an element of $\mathcal{R}(A)$. □

Remark 5.7. Obviously, from the Picard condition, it follows that $\sigma_i^{-2}|\langle y, y_i\rangle|^2 \to 0$ if $i \to \infty$ so that information about the decay of the Fourier coefficients of an element y becomes available.

Theorem 5.21 (Generalized or Pseudoinverse of a Compact Operator). *Let A be a compact operator (i.e., $A \in K(X,Y)$) with SVD $\{\sigma_i; x_i, y_i\}_{i\in\mathbb{N}}$. Then we have*

$$A^\dagger y = \sum_{i=1}^{\infty} \frac{1}{\sigma_i} \langle y, y_i\rangle_Y\, x_i \tag{5.146}$$

for all $y \in \mathcal{D}\left(A^\dagger\right)$.

Proof. Suppose that $y \in \mathcal{D}\left(A^\dagger\right) = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$ such that

$$y = Ax + \tilde{y}. \tag{5.147}$$

By definition, we set

$$\tilde{x} = \sum_{i=1}^{\infty} \frac{1}{\sigma_i} \langle y, x_i\rangle_Y x_i. \tag{5.148}$$

Then it follows that

$$\tilde{x} = \sum_{i=1}^{\infty} \frac{1}{\sigma_i} \langle Ax, y_i\rangle_Y\, x_i = \sum_{i=1}^{\infty} \frac{1}{\sigma_i} \langle x, A^* y_i\rangle_X\, x_i = \sum_{i=1}^{\infty} \langle x, x_i\rangle\, x_i. \tag{5.149}$$

In other words, $\tilde{x} \in X$. Moreover, $\tilde{x} \in \mathcal{N}(A)^\perp$, since $\{x_i\} \subset \mathcal{N}(A)^\perp$. It is not hard to see that

$$\begin{aligned} A^* A\tilde{x} &= \sum_{i=1}^{\infty} \frac{1}{\sigma_i} \langle y, y_i\rangle\, A^* Ax_i = \sum_{i=1}^{\infty} \frac{1}{\sigma_i} \langle y, y_i\rangle\, \sigma_i^2 x_i \\ &= \sum_{i=1}^{\infty} \sigma_i \langle y, y_i\rangle\, x_i = \sum_{i=1}^{\infty} \langle y, y_i\rangle\, A^* y_i = A^* y. \end{aligned} \tag{5.150}$$

Consequently, $\tilde{x}$ satisfies the normal equations, and $\tilde{x} \in \mathcal{N}(A)^\perp$. Thus, $\tilde{x}$ is equal to $A^\dagger y$. □

Corollary 5.1. *If $\mathcal{R}(A)$ is finite dimensional (i.e., there exist only finitely many element y_i), then $\mathcal{R}\left(A^\dagger\right) < \infty$. Therefore, $A^\dagger$ is compact. In particular, $A^\dagger$ is continuous.*

Proof. If there exist only finitely many elements y_i, then $\sum_{i=1}^{\infty} \frac{1}{\sigma_i} \langle y, y_i \rangle_Y x_i$ is a finite sum, hence, $\mathcal{R}\left(A^{\dagger}\right)$ is finite dimensional. □

The representation of the generalized inverse in terms of the singular value decomposition (5.146) opens the perspective to classify ill-posed problems. Indeed, the summands $\sigma_i^{-1} \langle y, y_i \rangle_Y x_i$ occurring in the series (5.146) depend closely on the singular values. If the values σ_i are small, then the contribution by the series (5.146) becomes large. The existence of SVD can be guaranteed for all compact operators. Nevertheless, the concrete knowledge of SVD is critical. Only in rare exceptions, SVD is explicitly known.

Remark 5.8. Singular systems are theoretically nice and easy, but their calculation might be rather tricky.

The specific amount of the growth of the singular values finally leads us to classify ill-posedness.

Definition 5.11 (Classification of Ill-posed Problems for Compact Operators). Let A be a compact operator (i.e., $A \in K(X, Y)$) with SVD $\{\sigma_i; x_i, y_i\}_{i \in \mathbb{N}}$.

(1) If there exists $\alpha > 0$ such that

$$\sigma_i = O(i^{-\alpha}), \tag{5.151}$$

then the operator A is called *ill posed of order* α.

(2) If there exists $\rho > 0$ such that

$$|\ln \sigma_i| \geq c\, i^{\rho}, \tag{5.152}$$

then the operator A is called *exponentially ill posed*.

Remark 5.9. Note that the aforementioned classification is senseful only for linear problems reflecting the representation of $A^{\dagger} y$ by its superposition in terms (5.146).

Truncated Singular Value Regularization. Obviously, the first two criteria (H1) and (H2) determining a well-posed problem in the sense of Hadamard, can always be enforced by considering the generalized inverse $A^{\dagger}$. A violating of the third point, i.e., instability arises if the spectrum of the operator A is not bounded away from zero. Thus, it seems to be natural to construct regularizing approximations via modifying the smallest singular values.

Indeed, in accordance with the singular value decomposition of the generalized inverse, it follows that such a modification of small values and, hence, a construction of regularization operators can be obtained in the form

$$x_\alpha = R_\alpha y = \sum_{i=1}^{\infty} \sigma_i F_\alpha(\sigma_i^2) \langle y, y_i \rangle_Y x_i \quad y \in Y, \tag{5.153}$$

with some function $F_\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$F_\alpha(\lambda) \to \frac{1}{\lambda}, \quad \lambda > 0,\ \alpha \to 0. \tag{5.154}$$

Such an operator R_α as defined by (5.153) and (5.154) may be understood as a *regularization operator* if

$$\lambda |F_\alpha(\lambda)| \leq C_{F_\alpha} < \infty, \quad \lambda > 0. \tag{5.155}$$

If (5.155) is satisfied, then we are able to see that

$$\|R_\alpha y\|_X^2 = \sum_{i=1}^{\infty} \sigma_i^2 (F_\alpha(\sigma_i))^2 |\langle y, y_i \rangle_Y|^2 \leq C_{F_\alpha}^2 \sum_{i=1}^{\infty} |\langle y, y_i \rangle_Y|^2 \leq C_{F_\alpha}^2 \|y\|_Y^2, \tag{5.156}$$

where C_F is a bound for the norm of R_α. Note that the pointwise convergence of F_α immediately implies the pointwise convergence of R_α to $A^\dagger$.

Let A be a compact operator (i.e., $A \in K(X,Y)$) with SVD $\{\sigma_i; x_i, y_i\}_{i \in \mathbb{N}}$. The main idea of truncated singular value decomposition is to ignore all singular values below a certain threshold value, which we can identify with the regularization parameter α, hence, the representation of the regularized solution is given by

$$F_\alpha(\lambda) = \begin{cases} \frac{1}{\lambda}, & \lambda \geq \alpha, \\ 0, & \lambda < \alpha, \end{cases} \tag{5.157}$$

i.e.,

$$x_\alpha = R_\alpha y = \sum_{\sigma_i \geq \alpha} \frac{1}{\sigma_i} \langle y, y_i \rangle_Y \; x_i, \quad y \in Y, \tag{5.158}$$

which explains the name truncated singular value decomposition, since all terms in the sum corresponding to small singular values are truncated. Since 0 is the only accumulation point of the singular values of a compact operator, the sum in (5.158) is always finite for $\alpha > 0$. In particular, only a finite number of singular values and singular vectors has to be computed in order to realize this method. On the other hand it should be mentioned that, for α being small, the number of singular values that need to be computed can increase strongly. Obviously, $C_{F_\alpha} = \alpha^{-1}$.

Tikhonov Regularization. The regularized solution is given by

$$F_\alpha(\lambda) = \frac{1}{\lambda + \alpha}, \quad \lambda > 0, \; \alpha > 0, \tag{5.159}$$

so that

$$x_\alpha = R_\alpha y = \sum_{i=1}^{\infty} \frac{\sigma_i}{\sigma_i^2 + \alpha} \langle y, y_i \rangle_Y \; x_i, \quad y \in Y. \tag{5.160}$$

As in the case of Lavrentiev regularization, we can compute x_α defined by (5.160) without any knowledge of the singular system. In fact, it is easy to see that

$$(A^* A + \alpha I) \; x_\alpha = A^* y \tag{5.161}$$

and, hence, we can solve a well-posed linear system to obtain x_α. From this representation it also follows that Tikhonov regularization is just Lavrentiev regularization applied to the normal equations. It is not hard to see that $\lambda^2 + \alpha \geq 2\lambda\sqrt{\alpha}$, hence, C_{F_α} can be chosen as $2\alpha^{-1/2}$.

We are interested in an estimate of the approximation error, which is independent of the noise level ε:

Theorem 5.22. *Let $F_\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ be a piecewise continuous function satisfying the assumptions*

$$F_\alpha(\lambda) \to \frac{1}{\lambda}, \quad \lambda > 0, \quad \alpha \to 0, \tag{5.162}$$

$$|F_\alpha(\lambda)| \leq M_\alpha < \infty, \quad \lambda > 0, \tag{5.163}$$

and

$$\sup_{\alpha,\lambda}(\lambda F_\alpha(\lambda)) \leq C_F < \infty \tag{5.164}$$

for some constant $C_F > 0$. Moreover, let the regularization operator be defined by (5.153)*. Then for all $y \in \mathcal{D}(A^\dagger)$,*

$$R_\alpha y \to A^\dagger y, \quad \alpha \to 0. \tag{5.165}$$

Proof. From the singular value decomposition (SVD) we obtain

$$\begin{aligned} R_\alpha y - A^\dagger y &= \sum_{i=1}^{\infty} \left(\sigma_i F_\alpha(\sigma_i^2) - \frac{1}{\sigma_i} \right) \langle y, y_i \rangle_Y x_i \\ &= \sum_{i=1}^{\infty} (\sigma_i^2 F_\alpha(\sigma_i^2) - 1) \langle x^\dagger, x_i \rangle_X x_i. \end{aligned} \tag{5.166}$$

Under the assumptions imposed on F_α we easily see that

$$|(\sigma_i^2 F_\alpha(\sigma_i^2) - 1)\langle x^\dagger, x_i \rangle \leq (C_F + 1)\|x^\dagger\|. \tag{5.167}$$

Hence, we are able to deduce that

$$\begin{aligned} \limsup_{\alpha \to 0} \|R_\alpha y - A^\dagger y\|^2 &\leq \limsup_{\alpha \to 0} \sum_{i=1}^{\infty} (\sigma_i^2 F_\alpha(\sigma_i^2) - 1)^2 \langle x^\dagger, x_i \rangle_X^2 \\ &\leq \sum_{i=1}^{\infty} \underbrace{\left(\lim_{\alpha \to 0} (\sigma_i^2 F_\alpha(\sigma_i^2)) - 1 \right)^2}_{=0} \langle x^\dagger, x_i \rangle_X^2. \end{aligned} \tag{5.168}$$

From the pointwise convergence $\lim_{\alpha \to 0}(\lambda F_\alpha(\lambda)) - 1 = 0$ we obtain the desired result. y □

Remark 5.10. The function $t \mapsto F_\alpha(t)$, $t \in \mathbb{R}^+$, converges pointwise to the function

$$F(t) = \begin{cases} 0\,, & t > 0, \\ 1\,, & t = 0. \end{cases} \tag{5.169}$$

Due to the discontinuity at zero, the convergence of $tF_\alpha(t) - 1$ to zero is becoming slower and slower as t decreases to zero. Since it is allowed to specify an arbitrarily small singular value σ_i and the minimal norm solution $x^\dagger = x_i$, the convergence of regularized solutions is arbitrarily slow. On the other hand, we observe from the proof that there is a possibly faster convergence if the components $\langle x^\dagger, x_i \rangle_X$ decay

sufficiently fast compared to the eigenvalues. For example, if we have $|\langle x^\dagger, x_i\rangle_X| \leq c\sigma_i^\mu$ for some constant $c > 0$ and $\mu > 0$, then it follows

$$\begin{aligned}\limsup_{\alpha\to 0} \|R_\alpha y - A^\dagger y\|^2 &\leq \limsup_{\alpha\to 0} c^2 \sum_{n=1}^{\infty} (\sigma_i F_\alpha(\sigma_i) - 1)^2 \sigma_i^{2\mu} \\ &\leq c^2 \sum_{n=1}^{\infty} \lim_{\alpha} (\sigma_i^{1+\mu} F_\alpha(\sigma_i) - \sigma_i^\mu)^2. \end{aligned} \tag{5.170}$$

In other words, one has to consider the limit of the function $t \mapsto |t^{1+\mu} F_\alpha(t) - t^\mu|$ as $t \to \infty$ instead, which is usually much faster. For example, in case of the truncated singular value decomposition, we obtain

$$|t^{1+\mu} F_\alpha(t) - t^\mu| = \begin{cases} 0, & t \geq \alpha, \\ t^\mu, & t < \alpha. \end{cases} \tag{5.171}$$

If the singular values of the operator decay sufficiently fast (which is the typical case for ill-posed problems), e.g., $\sum_{n=1}^{\infty} \sigma_i^\mu < \infty$, we are confronted with the situation

$$\|R_\alpha y - A^\dagger y\|^2 \leq c^2 \sum_{\sigma_i < \alpha} \sigma_i^{2\mu} \leq c^2 \alpha^\mu \sum_{i=1}^{\infty} \sigma_i^\mu \tag{5.172}$$

so that $\|R_\alpha y - A_y^\dagger\|$ is of order $\alpha^{\mu/2}$. Consequently, we somehow need smoothness of the solution (in terms of the smoothing properties of the operator) in order to obtain a convergence rate in terms of α. We shall pursue this idea by introducing spaces of smoothness involving the absolute value of a compact operator.

Next we are concerned with the propagation of the data error through the regularization.

Theorem 5.23. *Let F_α and C_F be as in Theorem* 5.22*, and let $x_\alpha = R_\alpha y$, $x_\alpha^\varepsilon = R_\alpha y^\varepsilon$. Then the estimates*

$$\|Ax_\alpha - Ax_\alpha^\varepsilon\| \leq C_F \varepsilon, \tag{5.173}$$

and

$$\|x_\alpha - x_\alpha^\alpha\| \leq M_\alpha \varepsilon \tag{5.174}$$

are valid.

Proof. From the singular value decomposition it follows directly

$$\begin{aligned}\|Ax_\alpha - Ax_\alpha^\varepsilon\|_Y^2 &\leq \sum_{i=1}^{\infty} \left(\sigma_i^2 F_\alpha(\sigma_i^2)\right)^2 |\langle y - y^\delta, y_i\rangle_Y|^2 \\ &\leq C_F^2 \sum_{n=1}^{\infty} |\langle y - y^\varepsilon, y_i\rangle_Y|^2 = C_F^2 \|y - y^\varepsilon\|^2 \leq (C_F\varepsilon)^2, \end{aligned} \tag{5.175}$$

so that (5.173) is valid. In the same way we obtain

$$\|x_\alpha - x_\alpha^\varepsilon\|^2 \le \sum_{i=1}^{\infty} (F_\alpha(\sigma_i))^2 |\langle y - y^\varepsilon, y_i\rangle_Y|^2$$
$$\le M_\alpha^2 \sum_{i=1}^{\infty} |\langle y - y^\varepsilon, y_i\rangle_Y|^2 = M_\alpha^2 \|y - y^\varepsilon\|^2 \le (M_\alpha \varepsilon)^2, \qquad (5.176)$$

so that (5.174) is implied (note that (5.174) estimates the norm of R_α by C_F). □

As a consequence, the error can be split in the following form

$$\|x_\alpha^\varepsilon - x\|_X = \|R_\alpha y^\varepsilon - x\|_X$$
$$\le \|R_\alpha y^\varepsilon - R_\alpha y\|_X + \|R_\alpha y - x\|_X$$
$$\le \|R_\alpha\|_{Y\to X} \, \|y^\varepsilon - y\|_Y + \|x_\alpha - x\|_X, \qquad (5.177)$$

such that

$$\|x^{\alpha,\varepsilon} - x\|_X \le \|R_\alpha\|_{Y\to X} \, \varepsilon + \|x_\alpha - x\|_X. \qquad (5.178)$$

We see that the error between the exact and the approximate solution consists of two parts: The first term is the product of the bound for the error in the data and the norm of the regularization parameter R_α. This term will usually tend to infinity for $\alpha \to 0$ if the inverse A^{-1} is unbounded and A is compact. The second term denotes the approximation error $\|(R_\alpha - A^{-1})y\|_X$ for the exact right-hand side $y = Ax$. This error tends to zero as $\alpha \to 0$ by the definition of a regularization strategy. Thus, both parts of the error show a diametrically reflected behavior. A typical picture of the errors in dependence on the regularization parameter α is sketched in Figure 5.5. Thus, a strategy is needed to choose α dependent an ε in

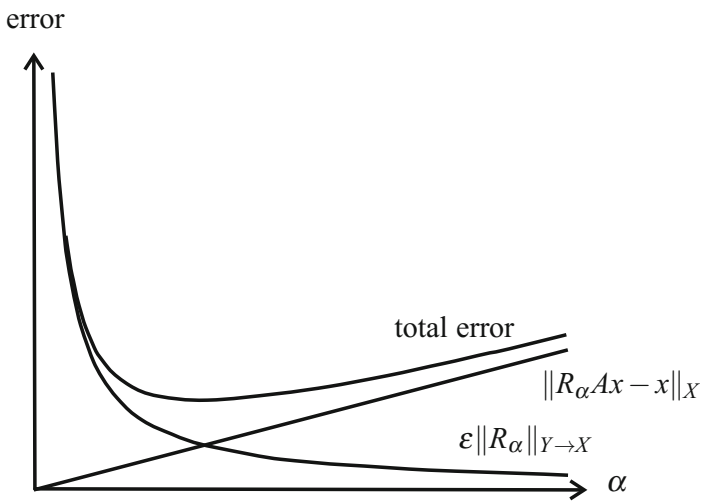


FIGURE 5.5. Typical behavior of the total error in a regularization process.

order to keep the error as small as possible, i.e., we would like to minimize

$$\|R_\alpha\|_{Y\to X}\ \varepsilon + \|R_\alpha Ax - x\|_X. \tag{5.179}$$

In principle, we distinguish two classes of parameter choice rules: If $\alpha = \alpha(\varepsilon)$ does not depend on ε, we call $\alpha = \alpha(\varepsilon)$ an *a priori* parameter choice rule. Otherwise α depends also on y^ε and we call $\alpha = \alpha(\varepsilon, y^\varepsilon)$ an *a posteriori* parameter choice rule. It is conventional to say that a parameter choice rule is convergent, if for $\varepsilon \to 0$ the rule is such that

$$\lim_{\varepsilon\to 0} \sup\{\|R_{\alpha(\varepsilon,y^\varepsilon)}y^\varepsilon - A^\dagger y\|_X\ : y^\varepsilon \in Y,\ \|y^\varepsilon - y\|_Y \le \varepsilon\} = 0 \tag{5.180}$$

and

$$\lim_{\varepsilon\to 0} \sup\{\alpha(\varepsilon, y^\varepsilon)\ : y^\varepsilon \in Y,\ \|y - y^\varepsilon\|_Y \le \varepsilon\} = 0. \tag{5.181}$$

All in all, numerous methods have been proposed for treating and regularizing various types of ill-posed problems. The rationale in most methods for resolution (approximate solvability) of ill-posed problems is to construct a "solution" that is acceptable physically as a meaning field approximation and is sufficiently stable from computational standpoint. The main dilemma of modeling ill-posed problems is that the closer the mathematical model describes the ill-posed problem the worse is the "condition number" of the associated computation problem (i.e., the more sensitive to errors, see [165]). A way out can only be found by additional "exterior" information about the problem to be solved.

5.6. Multi-scale solutions of inverse pseudodifferential equations

All gravitational information under discussion in physical geodesy leads to operator equations relating the disturbing potential to geodetically relevant observables. The most important operators are listed in the so-called "Pocket Guide of Physical Geodesy" (see, e.g., [170, 186]).

Pocket Guide of Physical Geodesy. In physical geodesy, one can think of observables as operating on an "input signal" F (e.g., the disturbing potential) to produce an "output signal" of the form

$$\Lambda F\ =\ G \tag{5.182}$$

(for example, geoidal undulation, gravity anomaly, radial derivative), where Λ is a certain operator (note that we use capital letters $F, G, \ldots$ in this subsection to characterize geodetic quantities). Fortunately, it is the case in geodetic applications involving the disturbing potential that large portions of interest can be well approximated by operators that represent linear, rotation-invariant pseudodifferential operators.

The standard pseudodifferential operators Λ occurring in physical geodesy (cf. [208]) have to reflect the aforementioned Pizzetti concept. As an immediate consequence, for the operator equation relating a geodetic observable $G = \Lambda F$ to the disturbing potential F (see, e.g., [107, 150]), we are led to an operator equation $\Lambda F = G$ which links F and the input function G under the following constraint:

$$\Lambda Y_{n,k} = \Lambda^\wedge(n)\ Y_{n,k}, \quad n = 0, 1, 2, \ldots, \quad k = 1, \ldots, 2n+1, \tag{5.183}$$

such that

$$\Lambda^\wedge(0) = \Lambda^\wedge(1) = 0 \tag{5.184}$$

and

$$\Lambda^\wedge(n) \neq 0, \quad n \geq 2, \tag{5.185}$$

where $\{\Lambda^\wedge(n)\}_{n=2,3,\ldots}$ is a sequence of real values (note that $\{Y_{n,k}\}$ is assumed to constitute a (real) complete system of spherical harmonics in $L^2(\Omega)$). Consequently, we have to discuss the invertibility of the operator Λ on the space $L^2_{2,\ldots}(\Omega)$ defined by

$$L^2_{2,\ldots}(\Omega) \;=\; L^2(\Omega) \setminus \bigoplus_{n=0}^{1} Harm_n = \; L^2(\Omega) \setminus \; Harm_{0,1}\,. \tag{5.186}$$

where $Harm_n$ is the linear space of all spherical harmonics of degree n.

In other words, we have the requirement that the spherical harmonics of degrees $n \geq 2$ are the eigenfunctions of the operator Λ, and the invertibility has to be controlled by the invertibility of the values $\Lambda^\wedge(n), n \geq 2$.

Definition 5.12 (Definition of Pseudodifferential Operators). Let $\{\Lambda^\wedge(n)\}_{n=2,3,\ldots}$ be a sequence of real numbers $\Lambda^\wedge(n)$ satisfying

$$\lim_{n\to\infty} \frac{|\Lambda^\wedge(n)|}{(n+\frac{1}{2})^t} = \text{const}\,. \;\neq 0 \tag{5.187}$$

for some $t \in \mathbb{R}$. Then the operator Λ defined by

$$\Lambda(F) = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \Lambda^\wedge(n) \underbrace{\int_\Omega F(\eta)) Y_{n,j}(\eta)\; dS(\eta)}_{=F^\wedge(n,j)} \; Y_{n,j} \tag{5.188}$$

is called (invariant) *pseudodifferential operator of order* t. $\{\Lambda^\wedge(n)\}_{n=2,3,\ldots}$ is called *spherical symbol* of Λ. Moreover, if

$$\lim_{n\to\infty} \frac{|\Lambda^\wedge(n)|}{(n+\frac{1}{2})^t} = 0 \tag{5.189}$$

for all $t \in \mathbb{R}$, then the operator Λ is called *pseudodifferential operator of order* $-\infty$.

The spherical symbol has many appealing properties: It is easily seen that

$$(\Lambda' + \Lambda'')^\wedge(n) = (\Lambda')^\wedge(n) + (\Lambda'')^\wedge(n), \tag{5.190}$$

$$(\Lambda'\Lambda'')^\wedge(n) = (\Lambda')^\wedge(n)(\Lambda'')^\wedge(n) \tag{5.191}$$

for all $n = 2, 3, \ldots$.

As any "output function" (output signal) can be expanded into an orthogonal series of surface spherical harmonics

$$G = \Lambda F = \sum_{n=2}^{\infty} \sum_{k=1}^{2n+1} \Lambda^\wedge(n) F^\wedge(n,k) Y_{n,k} = \sum_{n=2}^{\infty} \sum_{k=1}^{2n+1} G^\wedge(n,k) Y_{n,k} \tag{5.192}$$

in the sense of $\|\cdot\|_{L^2(\Omega)}$, we are confronted with a spectral representation of the form

$$G^\wedge(n,k) = (\Lambda F)^\wedge(n,k) = \Lambda^\wedge(n)\ F^\wedge(n,k), \quad n = 2,3,\ldots,\ k = 1,\ldots,2n+1. \tag{5.193}$$

This means that the "amplitude spectrum" $\{G^\wedge(n,k)\}$ of the response of Λ is described in terms of the amplitude spectrum of functions (signals) F by a simple multiplication by the "transfer" $\Lambda^\wedge(n)$. If a comparison of the "output function" with the actual value is done, discrepancies would be observed. A mathematical description of these discrepancies has to follow the laws of probability theory in a stochastic model. According to this approach we again assume that we have

$$G^\varepsilon = G + \varepsilon = \Lambda F + \varepsilon, \tag{5.194}$$

where ε is the *observation noise*.

Table 2 shows the so-called *Integrated Spherical Harmonic Model of Physical Geodesy* or *"Meissl Scheme"*, see [150, 170, 185] (earlier already called *"Pocket Guide of Physical Geodesy"*). Herein, R designates the Earth's mean radius, H is the satellite height.

operator/quantity Λ	term	$\Lambda^\wedge(n)$	order
gravity anomaly	Λ_A	$\frac{n-1}{R}$	1
geoid undulations	Λ_U	R^2	0
Stokes operator	Λ_{St}	$\frac{R}{n-1}$	-1
single layer	Λ_S	$\frac{R}{n+\frac{1}{2}}$	-1
double layer	Λ_D	$-\frac{R}{2n+1}$	-1
first radial derivative	Λ_{FND}	$-\frac{n+1}{R}$	1
second radial derivative	Λ_{SND}	$\frac{(n+1)(n+2)}{R^2}$	2
upward continuation	Λ_{UPC}	$\left(\frac{R}{R+H}\right)^{n+1}$	$-\infty$
satellite gravity gradiometry	Λ_{SGG}	$\left(\frac{R}{R+H}\right)^{n+1}\frac{(n+1)(n+2)}{(R+H)^2}$	$-\infty$

TABLE 2. Geodetic Operators in Spherical Nomenclature.

In spherically reflected satellite problems, the orbits are quite attractive for mathematical modeling: A circular orbit implies that the data are lying on a sphere; the measurements offer a global data coverage and an extremely dense and uniform distribution; the measurements (achieved by employing the significant principles of, e.g., satellite gravity gradiometry (SGG)) provide global information about the second radial derivatives of the gravitational potential at a moderate altitude.

Table 2 shows that the radial derivatives on spherical orbits are representable by rotation-invariant pseudodifferential equations (for more details on pseudodifferential operators on the sphere see, e.g., [208]; modeling concepts in modern satellite problems are described in [45]).

Multi-scale Regularizations of Inverse Geodetic Pseudodifferential Equations. Next we are interested in discussing a wavelet sampling solution of pseudodifferential equations. Since well-posed problems can be solved in a more or less straightforward way, we restrict ourselves to the solution of ill-posed pseudodifferential equations. We shall see that a sufficiently fast decay of the Legendre transform of the scaling function leads to regularization strategies for ill-posed problems involving pseudodifferential operators of finite order. For the exponentially ill-posed problems it is particularly adequate to use bandlimited scaling functions.

Given $G \in L^2_{2,\dots}(\Omega)$, find $F \in L^2_{2,\dots}(\Omega)$ so that

$$\Lambda F \;=\; G, \tag{5.195}$$

where $\Lambda : L^2_{2,\dots}(\Omega) \to L^2_{2,\dots}(\Omega)$ is an isotropic pseudodifferential operator of order $s < 0$ or $s = -\infty$.

We assume in the following that $\Lambda^\wedge(n) \neq 0$ for all $n = 2, 3, \dots$, which makes Λ injective. So, Λ is a linear bounded injective compact operator so that problem (5.195) is ill-posed. In the nomenclature of the theory of ill-posed problems we are able to say that Λ possesses the singular system $(\sigma_{n,k}; Y_{n,k}, Y_{n,k})_{n=2,3,\dots,k=1,\dots,2n+1}$ with $\sigma_{n,k} = \Lambda^\wedge(n)$.

We start our considerations with the regularization of problem (5.195).

Roughly speaking we call a regularization a family of bounded linear operators $R_\alpha : L^2_{2,\dots}(\Omega) \to L^2_{2,\dots}(\Omega)$ which approximates the inverse Λ^{-1}.

In more detail,

Definition 5.13 (Regularization Strategy). A regularization strategy for the problem (5.195) is a family of linear bounded pseudodifferential operators

$$R_\alpha : L^2_{2,\dots}(\Omega) \to L^2_{2,\dots}(\Omega), \quad \alpha > 0, \tag{5.196}$$

with symbol $\{(R_\alpha)^\wedge(n)\}_{n=2,3,\dots}$, so that

$$\lim_{n\to\infty} R_\alpha \Lambda F = F \tag{5.197}$$

for all $F \in L^2_{2,\dots}(\Omega)$, i.e., the operators $R_\alpha\Lambda$ converge in pointwise sense to the identity operator in $L^2_{2,\dots}(\Omega)$.

The following result immediately follows from the theory of inverse problems.

Theorem 5.24 (Tikhonov Regularization Strategy). *Suppose that the pseudodifferential operator Λ of type (5.195) is of order $s < 0$ or $-\infty$. Assume that the (non-bandlimited) Tikhonov kernel Φ_j is given by*

$$\Phi_j(\xi \cdot \eta) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} (\Phi_j)^\wedge(n) P_n(\xi \cdot \eta), \quad \xi, \eta \in \Omega \tag{5.198}$$

with

$$(\Phi_j)^\wedge(n) = \frac{(\Lambda^\wedge(n))^2}{(\Lambda^\wedge(n))^2 + \gamma_j^2}, \quad n \in \mathbb{N},\ j \in \mathbb{N}_0, \tag{5.199}$$

where $\{\gamma_j\}$, $j \in \mathbb{N}_0$ *is a sequence of real numbers satisfying* $\lim_{j\to\infty} \gamma_j = 0$. *Then the operators*

$$R_j = \Phi_j * \Lambda^{-1} \tag{5.200}$$

constitute a regularization strategy in $L^2_{2,\dots}(\Omega)$.

Since the boundedness of regularization operators is an important property, we are led to the following characterization within the framework of $L^2_{2,\dots}(\Omega)$.

Lemma 5.3. *The pseudodifferential operator* $\Lambda : L^2_{2,\dots}(\Omega) \to L^2_{2,\dots}(\Omega)$ *with symbol* $\{(\Lambda)^\wedge(n)\}_{n=2,3,\dots}$ *is bounded, if*

$$\sum_{n=2}^{\infty} \frac{2n+1}{4\pi} |\Lambda^\wedge(n)|^2 \ < \ \infty. \tag{5.201}$$

Proof. Let F be of class $L^2_{2,\dots}(\Omega)$, i.e.,

$$\|F\|^2_{L^2_{2,\dots}(\Omega)} = \sum_{n=2}^{\infty} \sum_{m=1}^{2n+1} |F^\wedge(n,m)|^2 < \infty. \tag{5.202}$$

Now,

$$(\Lambda F)^\wedge(n,m) = \Lambda^\wedge(n) F^\wedge(n,m). \tag{5.203}$$

Thus, for $N \geq 2$, we are able to deduce from the Cauchy–Schwarz inequality that

$$\begin{aligned} \sum_{n=2}^{N} \sum_{m=1}^{2n+1} |\Lambda^\wedge(n) F^\wedge(n,m)|^2 &\leq \left(\sum_{n=2}^{N} \sum_{m=1}^{2n+1} |\Lambda^\wedge(n)|^2 \right) \left(\sum_{n=2}^{N} \sum_{m=1}^{2n+1} |F^\wedge(n,m)|^2 \right) \\ &\leq \left(\sum_{n=2}^{N} (2n+1) |\Lambda^\wedge(n)|^2 \right) \|F\|^2_{L^2_{2,\dots}(\Omega)}. \end{aligned} \tag{5.204}$$

Taking the limit $N \to \infty$ we get the desired result. □

We are interested in regularizations by use of isotropic filtering, i.e., the filter Q_α is chosen as a pseudodifferential operator with symbol $\{(Q_\alpha)^\wedge(n)\}_{n=2,3,\dots}$ (see Figure 5.6 for an example of SGG). The regularization operator $R_\alpha = Q_\alpha \Lambda^{-1}$ has to be bounded, i.e., in view of Lemma 5.3 we arrive at the condition

$$\sum_{n=2}^{\infty} \frac{2n+1}{4\pi} \left| \frac{(Q_\alpha)^\wedge(n)}{\Lambda^\wedge(n)} \right|^2 < \infty. \tag{5.205}$$

Theorem 5.25. *Let* $\Lambda : L^2_{2,\dots}(\Omega) \to L^2_{2,\dots}(\Omega)$ *be a pseudodifferential operator of order* $s < 0$ *or of order* $-\infty$. *Assume that the family of pseudodifferential operators* Q_α *with symbol* $\{(Q_\alpha)^\wedge(n)\}_{n=2,3,\dots}$, $\alpha > 0$ *has the following properties:*

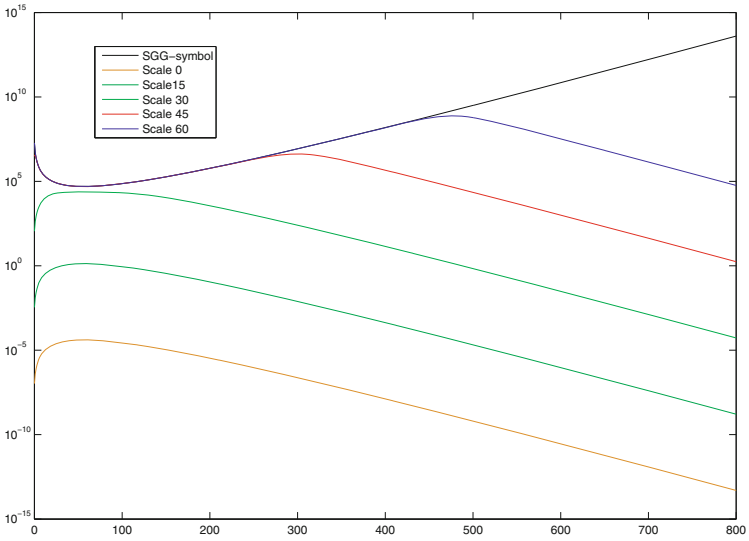


FIGURE 5.6. Symbol of the SGG-operator Λ_{SGG} with $R = 6378.127$ [km], $H = R+200$ [km] (black) and the operators R_j in case of Tikhonov regularization for different dyadic scales $\gamma_j = 2^{-j}$.

(i) $|(Q_\alpha)^\wedge(n)| \leq 1$ *for all* $\alpha > 0$ *and* $n = 2, 3, \ldots$
(ii) *For every* $\alpha > 0$,

$$\sum_{n=2}^{\infty} \frac{2n+1}{4\pi} \left| \frac{(Q_\alpha)^\wedge(n)}{\Lambda^\wedge(n)} \right|^2 < \infty. \tag{5.206}$$

(iii) *For every* $n = 2, 3, \ldots$,

$$\lim_{\alpha \to 0} (Q_\alpha)^\wedge(n) = 1. \tag{5.207}$$

Then the operator $R_\alpha = Q_\alpha \Lambda^{-1}$ *is a regularization strategy for* Λ.

Proof. From the estimate (5.206) and Lemma 5.3 we are able to conclude that the operators $R_\alpha : L^2_{2,\ldots}(\Omega) \to L^2_{2,\ldots}(\Omega)$ are bounded. For $F \in L^2_{2,\ldots}(\Omega)$, we have

$$R_\alpha \Lambda F = Q_\alpha \Lambda^{-1} \Lambda F = Q_\alpha F. \tag{5.208}$$

By virtue of the theory of singular integrals it thus follows that

$$\lim_{\alpha \to 0} R_\alpha \Lambda F = F. \tag{5.209}$$

□

The proof of Theorem 5.25 demonstrates the close relationship between the concepts of regularization and singular integrals.

In order to deal with *pseudodifferential operators* Λ *of order* $-\infty$ we are concerned with the fully discrete wavelet transform as presented in [60], where the following

properties imposed on functions $\varphi_0 : [0,\infty) \to \mathbb{R}$, $\psi_0 : [0,\infty) \to \mathbb{R}$, respectively, are given to establish spherical scaling – and wavelet functions:

(i) $\varphi_0(0) = 1$,
(ii) φ_0 is monotonously decreasing,
(iii) φ_0 is continuous at 0.
(iv) $\varphi_0 : [0,\infty) \to \mathbb{R}$ has a local support, i.e., supp $\varphi_0 \subset [0,1]$.
(v) The generator $\psi_0 : [0,\infty) \to \mathbb{R}$ of the mother wavelet given by the so-called refinement equation

$$(\psi_0(t))^2 = \left(\varphi_0\left(\frac{t}{2}\right)\right)^2 - (\varphi_0(t))^2\,,\ t \in [0,\infty), \tag{5.210}$$

also possesses a local support, i.e.,

$$\text{supp } \psi_0 \subset [0,1]. \tag{5.211}$$

So we are canonically led to the scale-discrete (zonal) scaling and wavelet functions, respectively,

$$\Phi_j(\xi\cdot\eta) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \underbrace{(\Phi_j)^\wedge(n)}_{=\varphi_0(2^{-j}n)} P_n(\xi\cdot\eta), \quad \xi,\eta\in\Omega, \tag{5.212}$$

and

$$\Psi_j(\xi\cdot\eta) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \underbrace{(\Psi_j)^\wedge(n)}_{=\psi_0(2^{-j}n)} P_n(\xi\cdot\eta), \quad \xi,\eta\in\Omega, \tag{5.213}$$

where

$$0 \le (\Phi_j)^\wedge(n) = \varphi_0(2^{-j}n) \le 1, \qquad n\in\mathbb{N}_0, \tag{5.214}$$

and

$$0 \le (\Psi_j)^\wedge(n) = \psi_0(2^{-j}n) \le 1, \qquad n\in\mathbb{N}_0 \tag{5.215}$$

with

$$\lim_{j\to\infty} (\Phi_j)^\wedge(n) = \lim_{j\to\infty} \varphi_0(2^{-j}n) = 1, \qquad n\in\mathbb{N}_0, \tag{5.216}$$

and

$$\lim_{j\to\infty} (\Psi_j)^\wedge(n) = \lim_{j\to\infty} \psi_0(2^{-j}n) = 0, \qquad n\in\mathbb{N}_0, \tag{5.217}$$

such that the "approximate identity"

$$\Phi_j * F = \int_\Omega \Phi_j(\,\cdot\eta) F(\eta)\, dS(\eta) \quad \to \quad F, \qquad j\to\infty, \tag{5.218}$$

holds for F being of class $L^2_{2,\dots}(\Omega)$ (in the topologies $\|\cdot\|_{L^2(\Omega)}$ and $\|\cdot\|_{C^{(0)}(\Omega)}$).

Note that the compact support of φ_0 implies that only finitely many $(\Phi_j)^\wedge(n)$ are different from 0. This analogously holds true for ψ_0.

Moreover, for the functions φ_j and ψ_j, defined by $\varphi_j = \varphi_0(2^{-j}\,\cdot)$ and $\psi_j = \psi_0(2^{-j}\,\cdot)$, respectively, we have

$$\text{supp}\ \varphi_j \subset [0, 2^j], \tag{5.219}$$

$$\text{supp}\ \psi_j \subset [0, 2^j]. \tag{5.220}$$

Since there are only a few conditions for a function $\varphi_0 : [0, \infty)$, there are various possibilities for its bandlimited as well as non-bandlimited realizations (cf. [60]). In our geodetically oriented framework we restrict ourselves to bandlimited cases.

Example. The *generator of the Shannon scaling function* is given by

$$\varphi_0(t) = \begin{cases} 1, & \text{for} \quad t \in [0, 1), \\ 0, & \text{for} \quad t \in [1, \infty), \end{cases} \tag{5.221}$$

so that

$$\varphi_j(t) = \begin{cases} 1, & \text{for} \quad t \in [0, 2^j), \\ 0, & \text{for} \quad t \in [2^j, \infty). \end{cases} \tag{5.222}$$

It is easy to see that all conditions for φ_0 to be a generator of a scaling function are fulfilled. We have

$$\Phi_j(t) = \sum_{n=0}^{2^j-1} \frac{2n+1}{4\pi} P_n(t)\ ,\ t \in [-1, +1]. \tag{5.223}$$

A remarkable property is that Φ_j coincides with its iterations, i.e.,

$$\Phi_j^{(k)}(\xi \cdot \eta) = \Phi_j * \Phi_j^{(k-1)}(\xi \cdot \eta) = \int_\Omega \Phi_j(\xi \cdot \zeta) \Phi_j^{(k-1)}(\zeta \cdot \eta)\, dS(\zeta),\ \xi, \eta \in \Omega\ , \tag{5.224}$$

$k = 2, 3, \ldots$, in particular, $\Phi_j^{(2)}(\xi \cdot \eta) = \Phi_j(\xi \cdot \eta),\ \xi, \eta \in \Omega$. The construction of wavelets is straightforward (cf. Figure 5.7).

$$\psi_j(t) = \begin{cases} 1, & \text{for } x \in [2^j, 2^{j+1}), \\ 0, & \text{elsewhere.} \end{cases} \tag{5.225}$$

Hence,

$$\Psi_j(t) = \sum_{n=2^j}^{2^{j+1}-1} \frac{2n+1}{4\pi} P_n(t)\ ,\ t \in [-1, +1]. \tag{5.226}$$

Example. We consider a somehow "smoothed" version of the generator of the Shannon wavelets, called *de la Vallée Poussin generator*, (dependent on a parameter

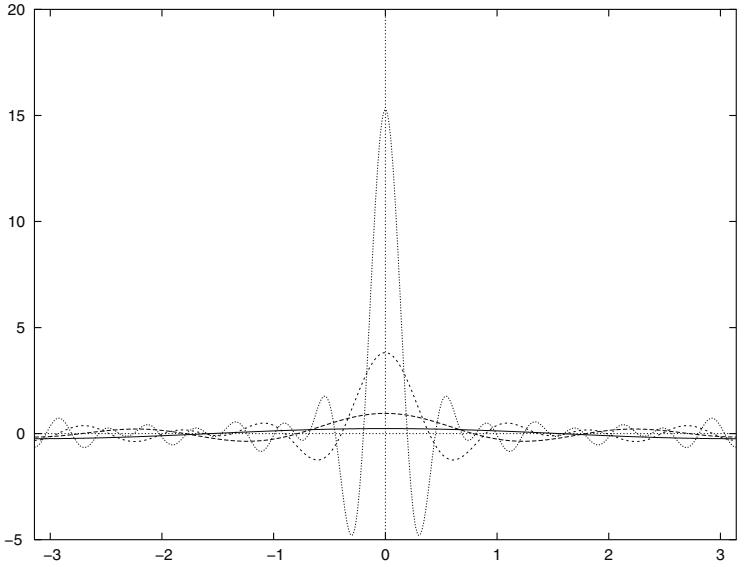


FIGURE 5.7. Shannon wavelets $\Psi_j(\cos\vartheta)$, $\vartheta \in [-\pi,\pi]$, $j = 0,\dots,3$ (sectional illustration).

$h \in (0,1)$)

$$\varphi_0(t) = \begin{cases} 1 \ , & \text{for} \quad t \in [0,h), \\ \frac{1-t}{1-h} \ , & \text{for} \quad t \in [h,1), \\ 0 \ , & \text{for} \quad t \in [1,\infty). \end{cases} \tag{5.227}$$

With the definition (5.227) the "dilates" have the form

$$\varphi_j(t) = \begin{cases} 1 \ , & \text{for} \quad t \in [0,2^j h), \\ \frac{1-2^{-j}t}{1-h} \ , & \text{for} \quad t \in [2^j h,2^j), \\ 0 \ , & \text{for} \quad t \in [2^j,\infty), \end{cases} \tag{5.228}$$

$j \in \mathbb{N}_0$. For the formulation of the wavelets corresponding to the "de la Vallée Poussin generator" we distinguish three cases:

- $0 < h < \frac{1}{2}$

$$\Psi_j^{\wedge}(n) = \begin{cases} 0, & \text{for} \quad n < 2^j h, \\ \left(1 - \left(\frac{1-2^{-j}n}{1-h}\right)^2\right)^{1/2}, & \text{for} \quad 2^j h \le n < 2^{j+1}h, \\ \left(\left(\frac{1-2^{-j-1}n}{1-h}\right)^2 - \left(\frac{1-2^{-j}n}{1-h}\right)^2\right)^{1/2}, & \text{for} \quad 2^{j+1}h \le n < 2^j, \\ \frac{1-2^{-j-1}n}{1-h}, & \text{for} \quad 2^j \le n < 2^{j+1}, \\ 0, & \text{for} \quad 2^{j+1} \le n < \infty. \end{cases} \tag{5.229}$$

- $1 > h > \frac{1}{2}$

$$\Psi_j^{\wedge}(n) = \begin{cases} 0, & \text{for} \quad 0 \le n < 2^j h, \\ \left(1 - \left(\frac{1-2^{-j}n}{1-h}\right)^2\right)^{1/2}, & \text{for} \quad 2^j h \le n < 2^j, \\ 1, & \text{for} \quad 2^j \le n < 2^{j+1}h, \\ \frac{1-2^{-j-1}n}{1-h}, & \text{for} \quad 2^{j+1}h \le n < 2^{j+1}, \\ 0, & \text{for} \quad 2^{j+1} \le n < \infty. \end{cases} \tag{5.230}$$

- $h = \frac{1}{2}$

$$\Psi_j^{\wedge}(n) = \begin{cases} 0, & \text{for} \quad 0 \le n < 2^{j-1}, \\ \left(1 - \left(2 - 2^{j+1}n\right)^2\right)^{1/2}, & \text{for} \quad 2^{j-1} \le n < 2^j, \\ 2 - 2^{-j}n, & \text{for} \quad 2^j \le n < 2^{j+1}, \\ 0, & \text{for} \quad 2^{j+1} \le n < \infty. \end{cases} \tag{5.231}$$

Compared with the Shannon wavelets there generally are more non-vanishing Legendre coefficients of Ψ_j. This explains the suppressing frequency effect. An illustration is given by Figure 5.8.

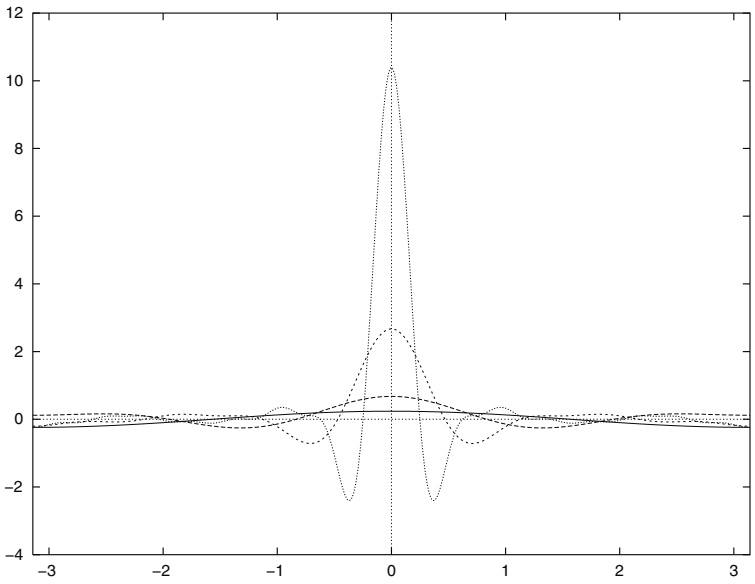


FIGURE 5.8. De la Vallée Poussin wavelets $\Psi_j(\cos\vartheta)$, $\vartheta \in [-\pi, \pi]$, $j = 0, \ldots, 3$, $h = 0.5$ (sectional illustration).

Theorem 5.26. *Let $\Lambda : L^2_{2,\dots}(\Omega) \to L^2_{2,\dots}(\Omega)$ be a pseudodifferential operator of order $s < 0$ or of order $-\infty$. Assume that the bandlimited kernels Φ_j are defined via a generator φ_0 satisfying the properties* (i)–(iv) *as stated above. Then the operators*

$$R_j = \Phi_j * \Lambda^{-1} \tag{5.232}$$

constitute a regularization strategy in the following sense: $R_j : L^2_{2,\dots}(\Omega) \to L^2_{2,\dots}(\Omega)$ is bounded, and the limit relation

$$\lim_{j\to\infty} R_j \Lambda F = F \tag{5.233}$$

holds true for all $F \in L^2_{2,\dots}(\Omega)$.

Remark 5.11. Analogously, $\{\Phi_j^{(2)} * \Lambda^{-1}\}_{j\in\mathbb{N}_0}$ constitutes a regularization strategy, provided that $\{B_j\}_{j\in\mathbb{N}_0}$ defines a (scale-discrete) scaling function.

5.7. Multi-scale signal-to-noise ratio and tree sampling

Usually, observations in geosciences are looked upon as a function G^ϵ on the sphere Ω so that

$$G^\epsilon = G + \varepsilon, \tag{5.234}$$

where ε is the *observation noise*. We suppose the *covariance* to be known

$$E\left[\tilde{\varepsilon}(\xi), \tilde{\varepsilon}(\eta)\right] = K(\xi \cdot \eta), \qquad (\xi, \eta) \in \Omega \times \Omega, \tag{5.235}$$

where the following conditions (cf. [54]) are imposed on the symbol $\{K^\wedge(n)\}_{n=0,1,}$ of the kernel $K : \Omega \times \Omega \to \mathbb{R}$:

(C1) $K^\wedge(n,k) \geq 0$ for all $n = 0, 1, \dots, k = 1, \dots, 2n+1$,
(C2) $\sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \left(K^\wedge(n)\right)^2 < \infty$.

Condition (C2), indeed, implies the $\ell^{(2)}$-summability of the symbol, i.e.,

$$\sum_{n=0}^{\infty} \frac{2n+1}{4\pi} (K^\wedge(n))^2 < \infty. \tag{5.236}$$

The error-affected J-scale approximation provided by a bandlimited/non-bandlimited regularization strategy of aforementioned type is given by

$$R_J(G^\varepsilon)(\xi) = \int_\Omega R_{J_0}(\xi \cdot \zeta) G^\varepsilon(\zeta)\, dS(\zeta) + \sum_{j=J_0}^{J-1} \int_\Omega S_j(\xi \cdot \zeta) G^\varepsilon(\zeta)\, dS(\zeta), \quad J > J_0, \tag{5.237}$$

where S_j designates the difference

$$S_j \;=\; R_{j+1} \;-\; R_j. \tag{5.238}$$

Evidently, the computation of the occurring integrals will require methods of numerical cubature. We base the integration on approximate formulas associated to

known weights $w_i^{N_j} \in \mathbb{R}$ and knots $\eta_i^{N_j} \in \Omega$

$$\int_\Omega R_{J_0}(\xi \cdot \zeta)\ G^\varepsilon(\zeta)\ dS(\zeta) \backsimeq \sum_{i=1}^{N_{J_0}} w_i^{N_{J_0}} R_{J_0}(\xi \cdot \eta_i^{N_{J_0}})\ G^\varepsilon(\eta_i^{N_{J_0}}), \tag{5.239}$$

$$\int_\Omega S_j(\xi \cdot \zeta)\ G^\varepsilon(\zeta)\ dS(\zeta) \backsimeq \sum_{i=1}^{N_j} w_i^{N_j} S_j(\xi \cdot \eta_i^{N_j})\ G^\varepsilon(\eta_i^{N_j}), \quad j = J_0, \ldots, J-1 \tag{5.240}$$

(the symbol "$\backsimeq$" always means that the error is assumed to be negligible, even better, in case of a bandlimited regularization strategy the integration error can be guaranteed to be 0). Since the "true" coefficients of (5.239) and (5.240) are the ones that should be included in a selective reconstruction of G from G^ε, in estimating the unknown function F it is natural to include only coefficients larger than some specified threshold value. The threshold value is understood to be the *scale and space error covariance at* $\eta_i^{N_j}$ with respect to the (scale discrete) wavelet function $\{S_j\}_{j\in\mathbb{N}_0}$

$$\begin{aligned}\mathrm{Cov}^{S_j}_{j,\eta_i^{N_j}}(K) &= \int_\Omega \int_\Omega K(\xi \cdot \zeta)\ S_j(\xi \cdot \eta_i^{N_j})\ S_j(\zeta \cdot \eta_i^{N_j})\ \ dS(\xi)\ \ dS(\zeta) \\ &\backsimeq \sum_{p=1}^{N_j} w_p^{N_j} \sum_{r=1}^{N_j} w_r^{N_j} K(\eta_p^{N_j} \cdot \eta_r^{N_j})\ S_j(\eta_p^{N_j} \cdot \eta_i^{N_j})\ S_j(\eta_r^{N_j} \cdot \eta_i^{N_j}).\end{aligned} \tag{5.241}$$

We compare the scale and space error covariance at $\eta_i^{N_j}$ with the *scale and space error variance of* G^ε *at* $\eta_i^{N_j}$ with respect to the (scale discrete) scaling function $\{\Phi_{h_j}\}_{j\in\mathbb{N}_0}$

$$\begin{aligned}\mathrm{Var}^{S_j}_{j,\eta_i^{N_j}}(G^\varepsilon) &= \int_\Omega \int_\Omega G^\varepsilon(\xi)\ G^\varepsilon(\zeta)\ S_j(\xi \cdot \eta_i^{N_j})\ S_j(\zeta \cdot \eta_i^{N_j})\ \ dS(\xi)\ \ dS(\zeta) \\ &\backsimeq \sum_{p=1}^{N_j} w_p^{N_j} \sum_{r=1}^{N_j} w_r^{N_j} G^\varepsilon(\eta_p^{N_j}) G^\varepsilon(\eta_r^{N_j})\ S_j(\eta_p^{N_j} \cdot \eta_i^{N_j})\ S_j(\eta_r^{N_j} \cdot \eta_i^{N_j}).\end{aligned} \tag{5.242}$$

Signal and noise scale "intersect" at the so-called *scale and space resolution set* $\mathcal{Z}^{S_j}_{\mathrm{res}}, j = J_0, \ldots, J-1$. We distinguish the following cases for signal-to-noise ratio:

(i) *Signal dominates noise*

$$\mathrm{Var}^{S_j}_{j,\eta_i^{N_j}}(G^\varepsilon) \geq \mathrm{Cov}^{S_j}_{j,\eta_i^{N_j}}(K), \quad (j, \eta_i^{N_j}) \in \mathcal{Z}^{S_j}_{\mathrm{res}}\ , i \in \{1, \ldots, N_j\}.$$

(ii) *Noise dominates signal*

$$\mathrm{Var}^{S_j}_{j,\eta_i^{N_j}}(G^\varepsilon) < \mathrm{Cov}^{S_j}_{j,\eta_i^{N_j}}(K), \quad (j, \eta_i^{N_j}) \notin \mathcal{Z}^{S_j}_{\mathrm{res}}\ ,\ i \in \{1, \ldots, N_j\}.$$

An *estimator of the "true" solution* $F = \Lambda^{-1}G$ can be determined via the indicator function I in the form

$$\begin{aligned} R_J^{est}(G^\varepsilon)(\xi) \qquad\qquad (5.243)\\ = \sum_{i=1}^{N_{J_0}} I\left\{ \mathrm{Var}^{R_{J_0}}_{J_0,\eta_i^{N_{J_0}}}(G^\varepsilon) \geq \mathrm{Cov}^{R_{J_0}}_{J_0,\eta_i^{N_{J_0}}}(K) \right\} w_i^{N_{J_0}} R_{J_0}(\xi\cdot\eta_i^{N_{J_0}})\, G^\varepsilon(\eta_i^{N_{J_0}}) \\ + \sum_{j=J_0}^{J-1}\sum_{i=1}^{N_j} I\left\{ \mathrm{Var}^{S_j}_{j,\eta_i^{N_j}}(G^\varepsilon) \geq \mathrm{Cov}^{S_j}_{j,\eta_i^{N_j}}(K) \right\} w_i^{N_j}\, S_j(\xi\cdot\eta_i^{N_j})\, G^\varepsilon(\eta_i^{N_j})\ . \end{aligned}$$

In other words, the large coefficients are kept intact and the small coefficients are set to zero. The *thresholding estimator* of "true" coefficients are thus characterized by

$$\begin{aligned} &\int_\Omega R_{J_0}(\xi\cdot\zeta)\, G^\varepsilon(\zeta)\, dS(\zeta) \\ &\quad \backsimeq \sum_{i=1}^{N_{J_0}} \delta^{\mathrm{hard}}_{\mathrm{Cov}^{R_{J_0}}_{J_0,\eta_i^{N_{J_0}}}} (\mathrm{Var}^{R_{J_0}}_{J_0,\eta_i^{N_{J_0}}}(G^\varepsilon))\, w_i^{N_{J_0}} R_{J_0}(\xi\cdot\eta_i^{N_{J_0}})\, G^\varepsilon(\eta_i^{N_{J_0}}), \qquad (5.244) \\ &\int_\Omega S_j(\xi\cdot\zeta)\, G^\varepsilon(\zeta)\, dS(\zeta) \\ &\quad \backsimeq \sum_{i=1}^{N_j} \delta^{\mathrm{hard}}_{\mathrm{Cov}^{S_j}_{j,\eta_i^{N_j}}} (\mathrm{Var}^{S_j}_{j,\eta_i^{N_j}}(G^\varepsilon))\, w_i^{N_j} S_j(\xi\cdot\eta_i^{N_j})\, G^\varepsilon(\eta_i^{N_j}), \qquad (5.245) \end{aligned}$$

$j = J_0, \ldots, J-1$, where the function $\delta_\lambda^{\mathrm{hard}}$ is the *hard thresholding function*

$$\delta_\lambda^{\mathrm{hard}}(x) = \begin{cases} 1, & \text{if } |x| \geq \lambda, \\ 0, & \text{otherwise}\ . \end{cases} \qquad (5.246)$$

The "keep or kill" hard thresholding operation is not the only reasonable way of estimating the coefficients. Recognizing that each coefficient consists of both a signal portion and a noise portion, it might be desirable to attempt to isolate the signal contribution by removing the noisy part. This idea leads to the *soft thresholding function*

$$\delta_\lambda^{\mathrm{soft}}(x) = \begin{cases} \max\{0, 1-\frac{\lambda}{|x|}\}, & \text{if} \quad x \neq 0, \\ 0, & \text{if} \quad x = 0, \end{cases} \qquad (5.247)$$

which can also be used in the coefficients of (5.244) and (5.245). When soft thresholding is applied to a set of empirical coefficients, only coefficients greater than the threshold (in absolute value) are included, but their values are 'shrunk' toward zero by an amount equal to the threshold λ. In other words, an estimator $R_J^{est}(G^\varepsilon)(\xi)$ of the "true" solution F is first approximated by a thresholded version of (5.244), which represents the trend (smooth) components of the data. Then the

coefficients at higher resolutions are thresholded, so that the noise is suppressed but the fine-scale details are included in the calculation.

Tree Sampling. Let us again consider an ill-posed pseudodifferential equation of the form $\Lambda : L^2_{2,\dots}(\Omega) \to L^2_{2,\dots}(\Omega), \Lambda F = G$, with a given error-affected right-hand side G^ε instead of $G \in L^2_{2,\dots}(\Omega)$. We assume that the operator Λ is of finite order $s < 0$ or of order $-\infty$. We have seen in Subsection 5.6, that a bandlimited regularization strategy based on a scaling function (5.212) can be given by $R_j = \Phi_j^{(2)} * \Lambda^{-1},\ j = 1, 2, \dots$. For more the decorrelation of the structural content in the solution F it is important to become more detailed insight into the regularized solutions $F_j = R_j G$ at many levels j. To this end, we present a tree algorithm which allows an efficient estimation of F_j at different scales from G^ε.

Once more, the assumptions on the generator φ_0 of the bandlimited scale discrete scaling function allow a *refinement equation* of the form

$$\Phi_j^{(2)} \ = \ \Xi_j * \Phi_{j+1}^{(2)}, \tag{5.248}$$

where

$$\Xi_j \ = \ \sum_{\substack{n\in\mathbb{N}_0 \\ \varphi_0^2(2^j n)\neq 0}} \frac{2n+1}{4\pi} \frac{\varphi_0^2(2^j n)}{\varphi_0^2(2^{j+1} n)}\ P_n \tag{5.249}$$

(note that the monotonicity and the compactness of φ_0 imply $\varphi_0^2(2^j n) = 0$ for $n \in \mathbb{N}_0$ provided that $\varphi_0^2(2^{j+1} n) = 0$ for that n, hence, the kernel (5.249) is well defined as finite sum). This observation enables us to realize a *tree algorithm for the decomposition of a signal* $G^\varepsilon \in L^2_{2,\dots}(\Omega)$: Starting from a (sufficiently large) $J \in \mathbb{N}$, such that the sampling formula

$$\begin{aligned} F^\varepsilon(\xi) \ \simeq \ R_J G^\varepsilon &= \left(\Phi_J^{(2)} * \Lambda^{-1}\right) * G^\varepsilon \ (\xi) \\ &= \sum_{i=1}^{(2m_J+1)^2} w_i^J \ \left(\Phi_J^{(2)} * \Lambda^{-1}\right)(\eta_i^J \cdot \xi), \quad \xi \in \Omega, \end{aligned} \tag{5.250}$$

with

$$w_i^J = a_i^J \ G^\varepsilon(\eta_i^J), \quad i = 1, \dots, (2m_J + 1)^2, \tag{5.251}$$

is valid, we are able to conclude that the coefficient vectors

$$w^j = (w_1, \dots, w_{(2m_j-1)^2})^T, \quad j = 0, \dots, J-1 \tag{5.252}$$

(being, of course, dependent on the bandlimited "replacement" of G^ε) can be calculated in the following way:

(i) The vectors w^j, $j = 0, \dots, J-1$, with

$$w_i^j = a_i^j \ \left(\Phi_J^{(2)} * \Lambda^{-1}\right) * G^\varepsilon \ (\eta_i^j), \ \ i = 1, \dots, (2m_j + 1)^2, \tag{5.253}$$

are subsequently obtainable by recursion from the values w_i^J in (5.251).

(ii) For $j = 0, \ldots, J-1$, we have

$$\left(\Phi_j^{(2)} * \Lambda^{-1}\right) * G^\varepsilon\ (\xi) = \sum_{i=1}^{(2m_j+1)^2} w_i^j\ \left(\Phi_j^{(2)} * \Lambda^{-1}\right)\ (\eta_i^j \cdot \xi),\quad \xi \in \Omega. \tag{5.254}$$

Our approach is divided into two parts, viz. an initial sampling step concerning the (sufficiently large) scale level J and the recursion step:

The initial sampling step: We just read in the given data $w_i^J = a_i^J\ \tilde{G}(\eta_i^J),\ i = 1, \ldots, (2m_J+1)^2$, to get the representation (5.250).

The recursion step: For $j = 0, \ldots, J-1$ it follows that

$$\begin{aligned} w_i^j &= a_i^j \quad \left(\Phi_j^{(2)} * \Lambda^{-1}\right) * G^\varepsilon\ (\eta_i^j) \\ &= a_i^j \quad \Xi_j * \left(\Phi_{j+1}^{(2)} * \Lambda^{-1}\right) * G^\varepsilon\ (\eta_i^j) \\ &= a_i^j \quad \sum_{i=1}^{(2m_{j+1}+1)^2} w_i^{j+1}\ \Xi_j(\eta_i^j \cdot \eta_i^{j+1})((\Phi_{j+1}^{(2)} * \Lambda^{-1}) * G^\varepsilon)\ (\eta_i^{j+1}) \\ &= a_i^j \quad \sum_{i=1}^{(2m_{j+1}+1)^2} w_i^{j+1}\ \Xi_j(\eta_i^j \cdot \eta_i^{j+1})\ w_i^{j+1}. \end{aligned} \tag{5.255}$$

In other words, the coefficients w_i^{J-1} can be calculated recursively starting from the data w_i^J for the initial level J, w_i^{J-2} can be deduced recursively from w_i^{J-1}, etc. Moreover, the coefficients are independent of the special choice of the kernel (observe that (5.250) is equivalent to

$$G^{\varepsilon\wedge}(n,k) = \sum_{i=1}^{(2m_j+1)^2} w_i^j\ Y_{n,j}(\eta_i^{N_j}) \tag{5.256}$$

for $n = 0, 1, \ldots,\ \ k = 1, \ldots, 2n+1$). This fact finally leads us to the formulas

$$(\Phi_J^{(2)} * \Lambda^{-1}) * G^\varepsilon\ (\xi) = \sum_{i=1}^{(2m_J+1)^2} w_i^J\ (\Phi_j^{(2)} * \Lambda^{-1})(\eta_i^J \cdot \xi),\quad \xi \in \Omega, \tag{5.257}$$

$$(\Phi_J * \Lambda^{-1}) * G^\varepsilon\ (\xi) = \sum_{i=1}^{(2m_J+1)^2} w_i^J\ (\Phi_j * \Lambda^{-1})(\eta_i^J \cdot \xi),\quad \xi \in \Omega, \tag{5.258}$$

and

$$(\Psi_J * \Lambda^{-1}) * G^\varepsilon\ (\xi) = \sum_{i=1}^{(2m_J+1)^2} w_i^J\ (\Psi_j * \Lambda^{-1})(\eta_i^J \cdot \xi),\quad \xi \in \Omega, \tag{5.259}$$

$$((\Psi_J * \tilde{\Psi}_j) * \Lambda^{-1}) * G^\varepsilon\ (\xi) = \sum_{i=1}^{(2m_J+1)^2} w_i^J\ ((\Psi_J * \tilde{\Psi}_j) * \Lambda^{-1})(\eta_i^J \cdot \xi),\quad \xi \in \Omega, \tag{5.260}$$

for $j = 0, \ldots, J$ with coefficients w_i^j given by (5.251) and (5.255).

The recursion step leads to the following *decomposition scheme*:

$$\begin{array}{ccccccc} G^\varepsilon & \to & w^J & \to & w^{J-1} & \to \cdots \to & w^0 \\ & & \downarrow & & \downarrow & & \downarrow \\ & & (\Phi_J * \Lambda^{-1}) * G^\varepsilon & & (\Phi_{J-1} * \Lambda^{-1}) * G^\varepsilon & & (\Phi_0 * \Lambda^{-1}) * G^\varepsilon \ . \end{array}$$

The coefficient vectors $w^0 = (w_1^0, \ldots, w_{(2m_0+1)^2})^T$, $w^1 = (w_1^1, \ldots, w_{(2m_1+1)^2})^T, \ldots$ allow the following *reconstruction scheme* of F:

$$\begin{array}{ccccccc} & w_0 & & w_1 & & w_2 & \\ & \downarrow & & \downarrow & & \downarrow & \\ (\Psi_0 * \Lambda^{-1}) * G^\varepsilon & & (\Psi_1 * \Lambda^{-1}) * G^\varepsilon & & (\Psi_2 * \Lambda^{-1}) * \tilde{G} & & \\ & \searrow & & \searrow & & \searrow & \\ (\Phi_0 * \Lambda^{-1}) * G^\varepsilon & \to + \to & (\Phi_1 * \Lambda^{-1}) * G^\varepsilon & \to + \to & (\Phi_2 * \Lambda^{-1}) * G^\varepsilon & \to + \to \cdots . \end{array}$$

Scale Thresholding. As we have seen, the coefficient vectors $w^j = \left(w_1^j, \ldots, w_j^{N_j}\right)^T \in \mathbb{R}^{N_j}$ $j = J_0, \ldots, J-1$, consists of the components

$$w_i^j = a_i^j \quad \left(\Phi_j^{(2)} * \Lambda^{-1}\right) * G^\varepsilon(\eta_i^j), \quad j = J_0, \ldots, J-1 \tag{5.261}$$

with w_i^j, η_i^j being known weights and knots of the applied approximate integration formula.

Since the large "true" coefficients are the ones that should be included in a selective approximation, in estimating an unknown function it is natural to include only coefficients larger than some specified threshold value.

In our context a "larger" coefficient is taken to mean one that satisfies for $j = J_0, \ldots, J$ and $i = 1, \ldots, N_j$

$$\begin{aligned} \left(w_i^{N_j}\right)^2 &= \left(a_i^j \quad \left(\Phi_j^{(2)} * \Lambda^{-1}\right) * G^\varepsilon(\eta_i^j)\right)^2 \\ &= (a_i^{N_j})^2 \int_\Omega \int_\Omega G^\varepsilon(\xi)\ G^\varepsilon(\zeta) \quad \left(\Phi_j^{(2)} * \Lambda^{-1}\right)(\xi, \eta_i^{N_j}) \\ &\qquad\qquad \left(\Phi_j^{(2)} * \Lambda^{-1}\right)(\zeta, \eta_i^{N_j})\ dS(\xi)\ dS(\zeta) \\ &\geq (a_i^{N_j})^2 \int_\Omega \int_\Omega K(\xi \cdot \zeta) \quad \left(\Phi_j^{(2)} * \Lambda^{-1}\right)(\xi, \eta_i^{N_j}) \\ &\qquad\qquad \left(\Phi_j^{(2)} * \Lambda^{-1}\right)(\zeta, \eta_i^{N_j})\ dS(\xi)\ dS(\zeta) \\ &= (k_i^j)^2. \end{aligned} \tag{5.262}$$

For the given threshold values k_i^j such an estimator can be written in explicit form:

$$\begin{aligned} \hat{F}_J = &\sum_{i=1}^{N_{J_0}} I_{\{(w_i^{J_0})^2 \geq (k_i^{J_0})^2\}} \left(\Phi_{J_0}^{(2)} * \Lambda^{-1}\right)(\cdot, \eta_i^{J_0}) \quad w_i^{J_0} \\ &+ \sum_{j=J_0}^{J-1} \sum_{i=1}^{N_j} I_{\{(w_i^j)^2 \geq (k_i^j)^2\}} \left(\Psi_{J_0}^{(2)} * \Lambda^{-1}\right)\left(\cdot, \eta_i^{N_j}\right) \quad w_i^j \ . \end{aligned} \tag{5.263}$$

We conclude our considerations on multi-scale approximation with the decorrelation of the geoidal undulations into wavelet signatures for the Earth's Gravitational Model EGM96) via a tree algorithm using Shannon scaling functions of scales $8, 7, 6, 5, 4$ (left) and the Shannon wavelet functions of scales $7, 6, 5, 4, 3$ (right) in [m] (see Figure 5.9). The illustrations (left) provide low-pass filtered geoidal heights from a fine ($j = 8$) down to a rough ($j = 4$) resolution, while the illustrations (right) show the corresponding band-pass filtered geoidal heights. Each band-pass filtered illustration yields detail information contained in the low-pass filtered version of scale $j + 1$, but not in the low-pass filtered version of scale j for $7, 6, 5, 4, 3$.

Figure 5.9 shows that geoidal undulations (and, by virtue of Bruns's formula, the Earth's disturbing potential) are "smooth" functions for large parts, so that they can be approximated efficiently and economically by a multi-scale procedure in form of a "read in" (tree) algorithm (see the low-passed filtered version of scale $j = 8$, i.e., the topmost illustration (left)). However, it also becomes obvious from Figure 5.9 that parts of particular geodetic interest, e.g., subduction zones, orogenetic areas, etc. are not sufficiently reflected by the global (spherical harmonic) EGM96-model (see the band-passed filtered version of scale $j = 7$, i.e., the topmost illustration (right) characterizing the non-green areas).

Geoid undulations (cf. Figure 5.10) may be understood as a measure for the perturbations of the Earth from a hydrostatic equilibrium (see, e.g., [185] for a more detailed geodetic interpretation). They form the deviations of the equipotential surfaces at mean sea level from the reference ellipsoid. Geoid undulations show no essential correlation to the distributions of the continents. They seem to be generated by density contrasts deeper inside the Earth.

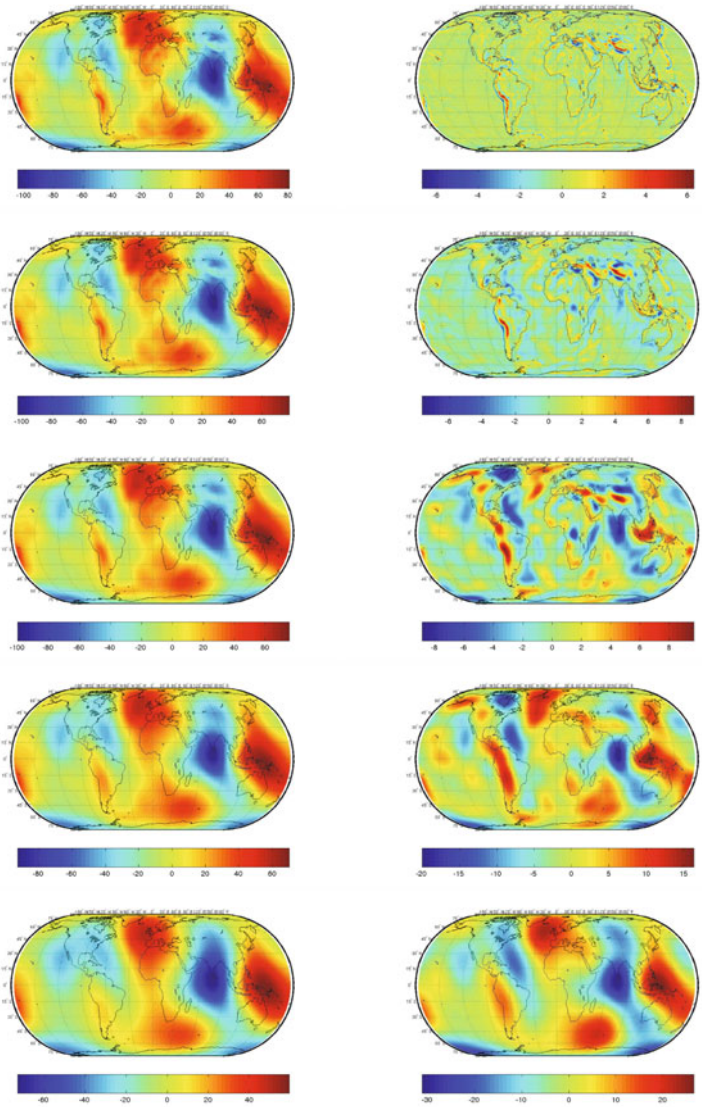


FIGURE 5.9. Multi-scale decomposition of the geoidal undulations (from Earth's Gravitational Model EGM96) via a tree algorithm using Shannon scaling functions of scales 8, 7, 6, 5, 4 (left) and the Shannon wavelet functions of scales 7, 6, 5, 4, 3 (right) in [m] (Illustration taken from W. Freeden, M.Z. Nashed, M. Schreiner (2018): Spherical Sampling, Geosystems Mathematics, Birkhäuser, Basel).

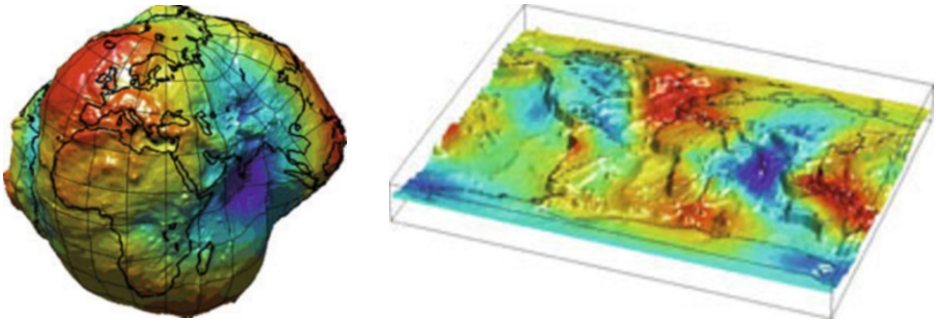


FIGURE 5.10. 3D (left) and 2D (right) graphical illustrations of the EGM96 – geoidal surface (taken from [48]).

6. Conclusions: Gaussian geometry and geodetic surveying

We have singled out only four fields of the wide spectrum of Gauss's works which are still relevant in mathematics as well as in geodesy up to our times. These four fields serve us as examples for building bridges between Gauss's world of ideas and modern scientific developments. A further important field, the foundation of surveying, will only be touched upon here. These important foundations for geodetic practice even today rest essentially on the mathematical foundations created by Gauss who was led fairly early to work on the theory of surfaces. Gauss had planned a summarizing publication on "Higher Geodesy" but he did eventually not finish it, as also a work on his trigonometric survey in the Kingdom of Hanover never saw the light of day (Gauss Werke, Vol. VIII, p. 400 and Vol. IX, p. 401).

6.1. The geodesic

Gauss's investigations concerning the theory of surfaces, the *Disquisitiones generales circa superficies curvas* (Gauss Werke, Vol. IV) were published in 1828 and are dealing extensively among other things with the theory of the geodesic, which is directly of great importance for geodetic surveys. Indeed, mathematicians like Bernoulli, Euler, and Clairaut had dealt with this topic before Gauss, but nobody so completely and profoundly. The findings in this publication are the basis for the *Untersuchungen über Gegenstände der Höheren Geodäsie*, published in 1843 and 1846, which give plenty of applications for geodetic tasks and computations on the ellipsoid of rotation. In the second publication Gauss explains above all the transfer of geographic coordinates on the ellipsoid from one given point to another one (the direct problem) using the azimuth and both of the points connecting a geodesic. In 1806 Legendre was the first to use the geodesic for the transfer of geographic coordinates. For the solution of this task he developed the so called Legendre series. Gauss's solution employs arguments of the means for the series so that the Taylor-type series converge faster and therefore less elements have to be computed. For the inversion of the direct problem the determination of the arc length of the geodesic and its azimuths from the given geographic coordinates of

both points Gauss's idea proved to be very fertile. His formulas are distinguished from other solutions due to their superior convergence, but also because Gauss had proven their accuracy and the size of the neglected terms. Gauss derived these formulas in two different ways: through a conformal mapping of the ellipsoid to the sphere and through power series expansions stemming from the differential equations of the geodesic on the ellipsoid of rotation. Gauss, however, did not present the formulas. This was done by F.R. Helmert [108] based on Gauss's formulas of the mean width (cf. [141]). A disadvantage of the series ansatz using arguments of means lies in the fact that the coefficients of the series have to be newly computed for each new pair of points, while the coefficients of the Legendre series have to be computed only once if the coordinates of multiple points have to be determined with respect to the same pole.

Approximately since 1965 the possibilities of numerical methods to solve the main tasks of geodesy, provided by standard methods of numerical mathematics, can be exploited due to the beginning development of electronic data processing computers. The variety of potential solutions can be divided roughly in four categories (according to B. Heck [105]):

(i) The first kind of approach rests on classical solutions of O. Schreiber [196] and H. Boltz [12]. Since poles are singular points of the geographical coordinate system Legendre's series cannot be used in the vicinity of poles. Moreover, these series show slow convergence in case of large distances so that this approach makes sense only in case of a short geodesic up to 150 km and up to 400 km if Gauss's formulas are used.

(ii) The second category rests on Bessel's approach (cf. [6]) of exploiting elliptic integrals which are today directly computed by numerical quadrature rules (see, e.g., [193]). This approach can also be used in the vicinity of poles and is accurate and advisable even in case of large distances of > 500 km.

(iii) The solution approaches in the third category are based on Gauss's conformal mapping of the ellipsoid to the sphere or the plane (cf. [84]) which are considered as auxiliary surfaces on which the main tasks of geodesy can be accomplished by means of closed formulas from spherical and plane trigonometry, respectively. Subsequently, the inverse mapping may be used to transfer back to the ellipsoid.

(vi) In the fourth category we find methods which were already used by A.M. Legendre [139] in a special case, namely that a given geodetic line is replaced by another line connecting start and end point of the geodesic. This other line may be a spatial chord, for example. Due to geodetic use of Earth satellites this three-dimensional ansatz has grown in importance.

6.2. Gaussian conformal mapping of the Earth's surface

For quite some time Gauss had dealt with the question of the method best suited to coordinate triangulation points because the knowledge of geographical coordinates is not sufficient. It seemed most convenient to him to exploit the coordinates with which every point could be represented in a plane (Gauss Werke, Vol. V, p. 367).

About 1815 he recognized the conformal mapping of the triangular points onto the plane as being the most appropriate solution to the problem. The term "konform" was used by him since 1843. The task formulated by Gauss, to project a given surface onto another one in a fairly general manner so that the image becomes similar to the preimage in its smallest parts was formulated as a contest question for the year 1821 by the Scientific Society in Copenhagen and reformulated the next year. Gauss submitted his solution (cf. [78]) on the 11, December 1822 and won the prize. His solution was distinguished from other known work in that it was valid in case of arbitrary surfaces and that he had fully laid open the conditions of conformity, the constancy of the augmenting relation in a certain point, and the equal angle condition (cf. [141]). Gauss applied his general solution to the following particular cases:

- *Conformal mapping between two planes.* As Gauss himself remarked (cf. Art. 8 in [78]), this is a useful method for the geodetic praxis, in order "to transform a map based on mediocre measurements which may be good in small details but is generally somehow distorted into a better one, if one knows the correct loci of a number of points."
- Conformal mapping of the sphere by transversal Mercator's projection or other mappings, e.g., the *conforming Lambert mapping.*
- *Conformal mappings of the ellipsoid of rotation onto the sphere.* In this mapping one can choose a suitable radius of the sphere to keep the differences between the ellipsoidal and spherical geometries relatively small. Hence, for practical applications spherical relations can be used which allow closed form solution. In case of local computations reasonable accuracy can be gained by letting the radius of the sphere depend on the geographic latitude. To this end two "replacement spheres" have proved their worth, namely the image sphere of Soldner, formerly used in Bavarian land surveying, and the Gaussian osculating sphere. The latter was used by Gauss (see [84]) with the radius M_0N_0 with regard to a point P_0, M_0 being the radius of curvature of the meridian and N_0 being the oblique radius of curvature.

  Since the ellipsoid shows the same measure of curvature in the central point P_0 in this mapping the differences in scale between ellipsoid and sphere stay small for regions not too large. As Gauss has shown the metric on a surface is determined by the Gaussian curvature. Hence, line segments and angles agree on two surfaces with the same Gaussian curvature. The surface of an ellipsoid, however, shows a variable Gaussian curvature so that the relationship of curvatures are only equivalent in the small (cf. Figures 6.2 and 6.3).

Under all mappings associated with Gauss's name the so-called *Gauss–Krüger projection* (and similar ones) have enjoyed wide distribution. In this projection the main meridian is mapped length-preserving to a straight line and is the abscissa of the plane system. Gauss had chosen this projection for the land surveying of Hanover (1828–1844) (cf. [141]). The size of the distortion depends only on the

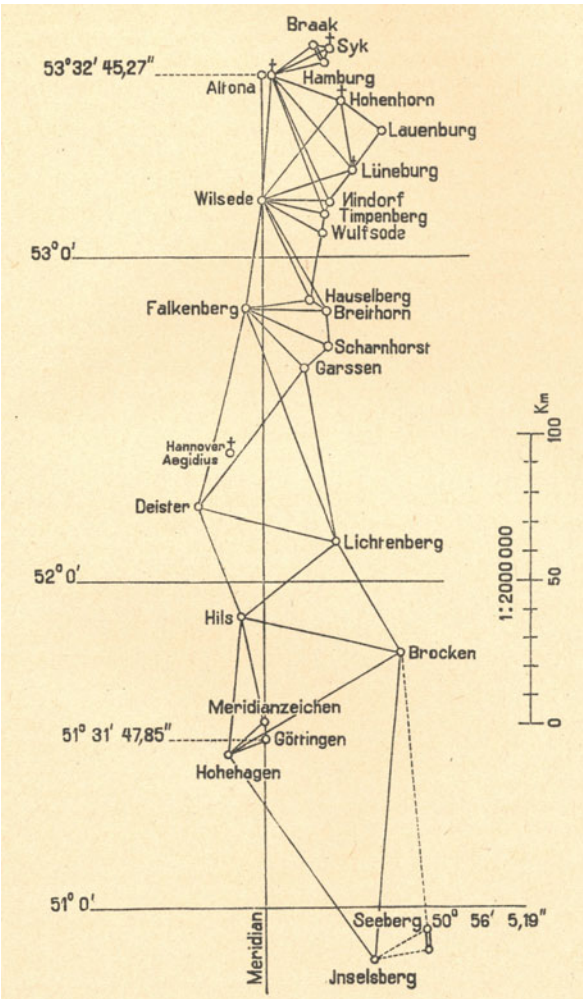


FIGURE 6.1. Triangulation of Hanoverian arc measurement between Göttingen and Altona (now a suburb of Hamburg) carried out under Gauss's supervision 1821–1825, Collected Works, Vol. 9, p. 347 (Illustration taken from [118]).

distance between the points and the main meridian and is independent from the north-south dilatation. Today one follows the idea that meridian strips of 6° width with a difference in longitude of 3° from the main meridian to the boundary meridians suffice for tolerable distortions (this is today's UTM-coordinate system).

Gauss's work on surface theory was important for the later nineteenth century in the sense of Sophus Lie, while until his time in geometry only finite groups of transformations had been considered. Gauss paved the way for the general theory

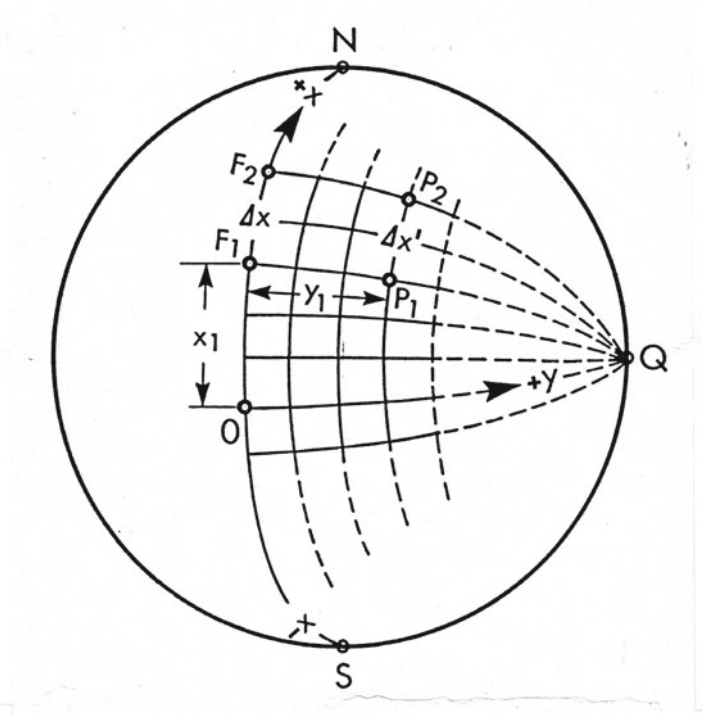


FIGURE 6.2. Ellipsoidal orthogonal coordinates (Illustration taken from [233]).

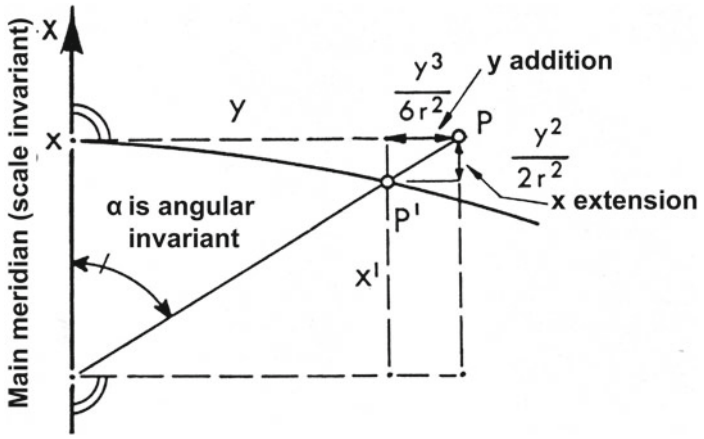


FIGURE 6.3. Differential projection distortions (Illustration taken from [233]).

of the multiply extended manifolds, or n-dimensional space (see [27] for more details).

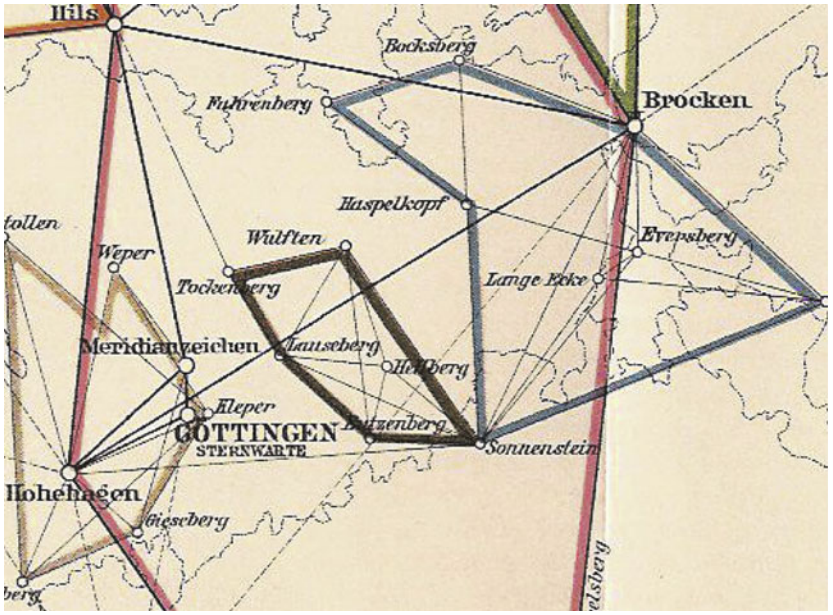


FIGURE 6.4. A detail of the triangulation of Hanover carried out under Gauss's supervision 1821–1825 (Collected Works, Vol. 9, p. 347). Note that the three lines joining the locations Hohenhagen, Brocken, and Inselsberg (this point does not appear here, only a part of the connecting lines is displayed) form a great triangle, the angle at Hohenhagen is close to a right angle, so that the area of the triangle is close to half the product of the two short sides.

The Hanover surveying work also stimulated Gauss's interest in the study of curves and surfaces in three-dimensional differential geometry in Euclidean space (today usually called *Gaussian geometry*). Gauss also was led to the Gaussian curvature (an intrinsic measure of curvature, dependent only on how distances are measured on the surface, not on the way it is embedded in space).

6.3. Claims and perspectives

All in all, while engaged on a surveying task for the Royal House of Hanover in the years after 1818, Gauss was also concerned with the shape of the Earth. He started to formulate revolutionary ideas, like the concept of the geoid. He questioned one of the central tenets of the whole of mathematics, Euclidean geometry, which was clearly premised on a flat, and not a curved, universe. He later claimed to have considered a non-Euclidean geometry, which was internally consistent and free of contradiction. Unwilling to court controversy, however, Gauss decided not to pursue or publish any of his far-reaching concepts in non-Euclidean geometry. János Bolyai independently discovered non-Euclidean geometry in 1829; his work

was published in 1832. After seeing it, Gauss wrote to Farkas Bolyai, the father of János Bolyai:

> *"To praise it would amount to praising myself. For, the entire content of the work ... coincides almost exactly with my own meditations which have occupied my mind for the past thirty or thirty-five years."*

The long history of the discussion of the parallel postulate started probably sooner than Euclid published his *Elements* about 300 BC as his fifth postulate [32], p. 155:

> *"That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles."*

Many modern authors have commented on the meaning of this postulate; we mention only R. Bonola [13], J. Gray [90], B.A. Rosenfeld [184]. In fact, even in our time the fifth Euclidean postulate seems odd and we may understand why so many authors after Euclid tried to prove the fifth postulate (and thus making it a theorem) from the other axioms and postulates. Already Proclus (412–485), who wrote a Commentary on the First Book of Euclid's Elements, tried to prove the fifth postulate exploiting an argument given by Aristoteles to show the finiteness of the universe [176], p. 291, and failed, cf. [90], p. 39. Proclus informs us, that the first of the Ptolemies under whose reign Euclid lived, wrote himself a book on the fifth postulate and proved it, but Proclus points out the fallacy of this attempt. John Playfair (1748–1819) gave an alternative formulation of the fifth postulate in 1795 which is now known as *Playfair's Axiom* but which he himself attributed to William Ludlam (1717–1788). This formulation is the one which most people refer to nowadays when speaking of the "axiom of parallels" [239, p. 16]:

> *"Given a line a and a point A not lying on a, then there exists, in the plane determined by a and A, one and only one line which contains A but not any point of a."*

Or, even shorter,

> *"Through a given point can be drawn only one parallel to a given line."*

Many mathematicians tried to get hold of a proof of the fifth postulate; famous names over the centuries being Nasir al-Din al-Tusi (1201–1274) in Persia, John Wallis (1616–1703) in England, Giovanni Girolamo Saccheri (1667–1733) in Italy, the Swiss mathematician Johann Heinrich Lambert (1728–1777) in Germany, and Adrien-Marie Legendre (1752–1833) in France. At the beginning of the 19th century the invention of a Non-Euclidean Geometry seemed to have been in the air. If so many and different attempts to prove the fifth postulate a mere conclusion of other axions had failed the question whether there exists a geometry in which the fifth postulate was wrong became sensible. This may explain why it was not only Gauss who found the key to this new geometry but also Nikolai Lobachevsky (1792–1856) and János Bolyai (1802–1860) independently. However, Gauss was

the first, although he did not publish his results. It is now known that it was he who coined the denotation "Non-Euclidean" for the new geometries which emerge out of a neglect of the fifth postulate. Already in 1804 Gauss received a letter of his friend Farkas Bolyai (1775–1856), father of János Bolyai, in which Bolyai presented a proof that the fifth postulate could be deduced from other axioms of geometry. Gauss praised the work of his friend, but found a flaw in the argument. He wrote (cf. [16, p. 100]):

> *"You want to know my sincere and frank opinion. And this is that your method does not satisfy me. I will try to make the critical point (which belongs to the same kind of obstacles which made my own efforts so futile) as clear as I can. I still hope that these cliffs will be navigated eventually, and this, before I die. For now, I am, however, extremely busy with other things..."*

Although Gauss remarked in 1846 that he knew about the existence of Non-Euclidean geometries for the last fifty years it is not before 1816 that we see written evidence. In that year Gauss reflected on different attempts to prove the fifth postulate in a book review. Although he was too careful to express his own opinion we can be sure that the reason why he reviewed these attempts can be seen in the fact that Gauss was already convinced in 1816 that Non-Euclidean geometries existed.

Perhaps the Non-Euclidean geometry which can be most easily understood is the spherical geometry. The navigators, map makers, and naval mathematicians of the 16th century were very well aware of this type of geometry in which every triangle (build from parts of three great circles on the sphere) shows a sum of inner angles of larger than 180° and is thus today seen as a simple model of elliptic geometry. The Non-Euclidean geometries by Gauss, Lobachevsky, and Bolyai are nowadays classified as being hyperbolic geometries, in which a triangle shows an angular sum of less than 180°.

Spherical geometry was Gauss's bread-and-butter occupation as early as 1803 when he started to survey the Duchy of Brunswieck by means of a triangulation and met his wife to be, Johanna Osthoff. After Heinrich Christian Schumacher (1780–1850), astronomer and geodesist, informed Gauss of the Danish arc measurement and stimulated a surveying of the Hanoverian lands to fit the Danish measurements in the north, Gauss became enthusiastic. Eventually, Schumacher succeeded in persuading the English king Georg IV, who was head of the house of Hanover, to authorize Gauss with the surveying work, and Gauss started field work in 1821 (see Figures 6.1 and 6.4) concerned with measurements from three mountains in Germany, Hohenhagen, near Göttingen, Brocken in the Harz Mountains and Inselberg in the Thüringer Wald to the south. The three lines joining these locations form a great triangle, the angle at Hohenhagen is close to a right angle, so the area of the triangle is close to half the product of the two short sides). Gauss made the important invention of the heliotrope (cf. Figure 6.5) to ensure measurements of hitherto unknown accuracy. In fact, heliotropes were used

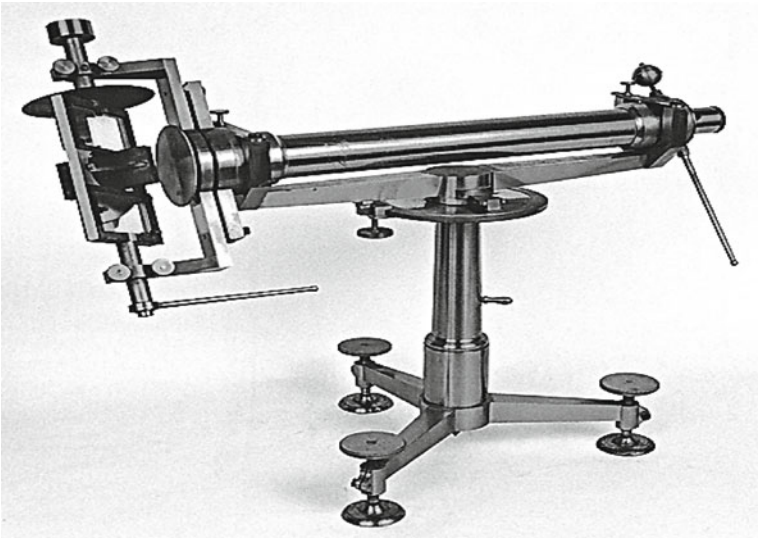


FIGURE 6.5. Heliotrope (Geophysical Institute of Göttingen University).

in surveys from Gauss's survey in Germany in 1821 through the late 1980s, when GPS measurements replaced the use of the heliotrope in long distance surveys. Surveying field work lasted until 1825 when Gauss withdraw from working in the countryside. However, he oversaw the further surveying work going on in 1828 and lasting until 1844 and did all the number crunching necessary. Friends, including Friedrich Wilhelm Bessel (1784–1846), criticized him for wasting his time in these computations instead of creating new theorems, but Gauss saw deeper. In a letter to Bessel dated March 14, 1824, he wrote (cf. [8]):

> *"... you accused me of loosing my time and wished me luck that the loss of time might be over soon. Great God, how wrong you judge me. ... Certainly, I also think like you in that matter. All measurements in the world do not outweigh a single theorem, with which the science of eternal truths will be truly advanced. But you should not judge over the absolute, but over the relative worth* [of measurements]. *... And however small you estimate this worth, in my eyes it is higher than those concerns which are interrupted by them."*

Surveying problems also motivated Gauss to develop his thoughts on least squares and more general problems of what is now called mathematical statistics. The result was the definitive exposition of his mature concepts in the note "Theoria combinationis observationum erroribus minimis obnoxiae" (1823, with supplement in 1826). In "Bestimmung des Breitenunterschiedes zwischen den Sternwarten von Göttingen and Altona durch Beobachtungen am Ramsdenschen Zenithsector" of the year 1828 he summed up his ideas on the figure of the Earth, instrumental

errors, and the calculus of observations. Furthermore, his publication "Disquisitiones generales circa superficies curvas" (1828), which grew out of his meditations in surveying and geodesy of three decades, represented the seed of more than a century of work on differential geometry.

Finally it should be mentioned that resulting research led to, among other things, Einstein's theory of general relativity, which describes the universe as non-Euclidean.

References

[1] Adams, D.R., Hedberg, L.I.: *Function Spaces and Potential Theory*. Grundlehren der Mathematischen Wissenschaften, 314, Springer, Berlin, (1996).

[2] Adrain, R.: *Research Concerning the Probabilities of the Errors which Happen in Making the Observations*. The Analyst of Mathematical Museum, Vol. I, W.P. Farrand and Co., Philadelphia, (1808).

[3] Augustin, M., Freeden, W.: A Survey on Classical Boundary Value Problems in Physical Geodesy. Encyclopedia of Geodesy, E.W. Grafarend, ed., Springer Int. Publ. Switzerland, DOI 10.1007/978-3-319-02370-0-1117-1, (2015).

[4] Baer, R.: *Linear Algebra and Projective Geometry*. Academic Press, New York, (1952).

[5] Ben-Israel, A., Greenville, T.N.E., *Generalized Inverses: Theory and Applications*. New York: Wiley-Interscience, (1974).

[6] Bessel, W.: Über die Berechnung der geographischen Längen und Breiten aus geodätischen Vermessungen. Astronomische Nachrichten, Bd. 4, Nr. 86, Altona, (1826).

[7] Bialas, V.: *Erdgestalt, Kosmologie und Weltanschauung*. Konrad Wittwer Stuttgart, (1982).

[8] Biermann, K.-R: *Carl Friedrich Gauss – Der "Fürst der Mathematiker" in Briefen und Gesprächen*. Verlag C.H. Beck, München, (1990).

[9] Bjerhammar, A.: Rectangular Reciprocal Matrices with Special Reference to Geodetic Calculations. Bulletin Géodesique, 188–220, (1951).

[10] Bjerhammar, A.: *Gravity Reduction to an Internal Sphere*. Division of Geodesy, Stockholm, Sweden, (1962).

[11] Börsch, A.: Simon, P.: Abhandlungen zur Methode der kleinsten Quadrate von Carl Friedrich Gauss (in deutscher Sprache herausgegeben). Berlin, (1887).

[12] Boltz, H.: *Formeln und Tafeln zur numerischen* (*nicht logarithmischen*) *Berechnung geographischer Koordinaten aus den Richtungen und Längen der Dreiecksseiten erster Ordnung*. Veröff. Geod. Inst. Potsdam, Neue Folge, Nr. 110, (1942).

[13] Bonola, R: *Non-Euclidean Geometry – With a Supplement Containing 'The Theory of Parallels' by Nicholas Lobachevski and 'The Science of Absolute Space' by John Bolyai*. Engl. transl. of the first edition 1912. Dover Publication, New York, (1955).

[14] Boullion, T.L., Odell, P.L.: *Generalized Inverse Matrices*. Wiley-Interscience, New York, (1971).

[15] Bruns, H.: Die Figur der Erde "Ein Beitrag zur europäischen Gradmessung". P. Stankiewicz, Berlin, (1878).

[16] Bühler, W.K.: *Gauss. A Biographical Study.* Springer-Verlag, Berlin, Heidelberg, New York, (1981).

[17] Bühler, W.K.: *Gauss – Eine Biographische Studie.* Springer, Berlin. Heidelberg, (1987).

[18] Chen, J.-R.: The Lattice Points in a Circle. Sci. Sinica, **12**:633–649, (1963).

[19] Christoffel, E.B.: Über die Gaussische Quadratur und eine Verallgemeinerung derselben. J. Reine Angew. Math., **55**:61–82. (Ges. Math. Abhandlungen I, 65–87), (1858).

[20] Cui J., Freeden W.: Equidistribution on the Sphere. SIAM J. Sci. Stat. Comp., **18**:595–609, (1997).

[21] Davis, C.H.: Theory of the Motion of Heavenly Bodies Moving about the Sun in Conic Sections, translation of "Theoria motus corporum coelestium in sectionibus conicis solem ambientium" by C.F. Gauss, reprinted Dover, New York, (1963).

[22] Davis, C.H., Robinson, D.W.: Generalized Inverses of Morphisms, Linear Algebra and Its Applications, **5**:319–328, (1972).

[23] Davis, P.J.: *Interpolation and Approximation.* Blaisdell, New York, (1963).

[24] Davis, P.J., Rabinowitz, P.: *Numerical Integration.* Blaisdell, Toronto, London, (1967).

[25] Dedekind, R.: Gauß in seiner Vorlesung über die Methode der kleinsten Quadrate. Festschrift zur Feier des hundertfünfzigjährigen Bestehens der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Weidmannsche Buchhandlung, (1901).

[26] Driscoll, J.R., Healy, R.M.: Computing Fourier Transforms and Convolutions on the 2-Sphere. Advances in Applied Mathematics, **15**:202–250, (1994).

[27] Dunnington, G.W.: Carl Friedrich Gauss: Titan of Science A Study of his Life and Work. Exposition Press, New York, (1955).

[28] Edgeworth, F.Y.: On Observations Relating to Several Quantities. Hermathena, **6**:279–285, (1887).

[29] Engl, H.: Discrepancy Principles for Tikhonov Regularization of Ill-posed Problems, Leading to Optimal Convergence Rates. J. Optim. Theory Appl., **52**:209–215, (1987).

[30] Engl, H.: *Integralgleichungen.* Springer, Berlin, (1997).

[31] Eisenhart, C., Carl Friedrich Gauss. International Encyclopedia of Social Sciences, New York, Vol. VI, 74–81, (1986)

[32] Euclid: *The Thirteen Books of The Elements.* Vol. 1. 2nd Edition, translated with introduction and commentary by Sir Thomas L. Heath. Dover Publication, New York, (1956).

[33] Euler, L.: Methodus universalis serierum convergentium summas quam proxime inveniendi, Commentarii Academiae Scientiarum Petropolitanae. **8**:3–9, Opera Omnia (XIV), 101–107, (1736).

[34] Fredholm I.: Sur une classe d'équations functionelles. Acta Mathematica **27**:365–390, (1903).

[35] Freeden, W.: Über eine Verallgemeinerung der Hardy–Landauschen Identität. Manuscr. Math., **24**:205–216, (1978).

[36] Freeden, W.: Über die Gaußsche Methode zur angenäherten Berechnung von Integralen. Math. Method. Appl. Sci., **2**:397–409, (1980).

[37] Freeden, W.: On the Approximation of External Gravitational Potential with Closed Systems of (Trial) Functions. Bull. Géod., **54**, 1–20, (1980a).

[38] Freeden, W.: On Integral Formulas of the (Unit) Sphere and Their Application to Numerical Computation of Integrals. *Computing*, **25**:131–146, 1980b.

[39] Freeden, W.: On Spherical Spline Interpolation and Approximation. Math. Meth. in the Appl. Sci., **3**:551–575, (1981a).

[40] Freeden, W.: On Approximation by Harmonic Splines. Manuscr. Geod., **6**:193–244, (1981b).

[41] Freeden, W.: Least Squares Approximation by Linear Combinations of (Multi)-Poles. Report **344**, Department of Geodetic Science and Surveying, The Ohio State University, Columbus, (1983).

[42] Freeden, W.: Spherical Spline Interpolation: Basic Theory and Computational Aspects. J. Comput. Appl. Math., **11**:367–375, (1984).

[43] Freeden, W.: A Spline Interpolation Method for Solving Boundary Value Problems of Potential Theory From Discretely Given Data. Math. Part. Diff. Equations, **3**:375–398, (1987).

[44] Freeden, W.: Interpolation by Multidimensional Periodic Splines. J. Approx. Theory **55**:104–117, (1988).

[45] Freeden, W.: *Multiscale Modelling of Spaceborne Geodata*. B.G. Teubner, Stuttgart, Leipzig, (1999).

[46] Freeden, W., Schreiner, M.: Spherical Functions of Mathematical Geosciences – A Scalar, Vectorial, and Tensorial Setup. Springer, Heidelberg, (2009).

[47] Freeden, W.: *Metaharmonic Lattice Point Theory*. CRC Press, Taylor & Francis Group, Boca Raton, (2011).

[48] Freeden, W.: Geomathematics: Its Role, Its Aim, and Its Potential. in: Handbook of Geomathematics, Vol. 1, 2nd ed., W. Freeden, M.Z. Nashed, T. Sonar, Eds., 3–79, Springer, New York, Berlin, Heidelberg, (2015).

[49] Freeden, W., Gerhards, C.: *Geomathematically Oriented Potential Theory*. Chapman and Hall, CRC Press, Boca Raton, London, New York, (2013).

[50] Freeden, W., Gutting, M.: *Special Functions of Mathematical (Geo)Physics*. Birkhäuser, Basel, (2013).

[51] Freeden, W., Kersten, H.: *The Geodetic Boundary-Value Problem Using the Known Surface of the Earth*. Veröff Geod Inst RWTH Aachen, **29**, (1980).

[52] Freeden, W., Kersten, H.: A Constructive Approximation Theorem for the Oblique Derivative Problem in Potential Theory. Math. Meth. Appl. Sci., **4**:104–114, (1981).

[53] Freeden, W., Kersten, H.: An Extended Version of Runge's Theorem. *Manuscr. Geod.*, **7**:267–278, (1982).

[54] Freeden, W., Maier, T.: On Multiscale Denoising of Spherical Functions: Basic Theory and Numerical Aspects. Electronic Transactions on Numerical Analysis (ETNA), **14**:40–62, (2002).

[55] Freeden, W., Mayer, C.: Multiscale Solution for the Molodensky Problem on Regular Telluroidal Surfaces, Manuscr. Math., Acta Geod. Geophys. Hung., **41**:55–86, (2006).

[56] Freeden, W., Michel, V.: *Multiscale Potential Theory* (*With Applications to Geoscience*). Birkhäuser, Boston, (2004).

[57] Freeden, W., Nashed, M.Z.: Multi-Variate Hardy-Type Lattice Point Summation and Shannon-Type Sampling. Int. J. Geomath., **6**:163–249, (2015).

[58] Freeden, W., Schneider, F.: Regularization Wavelets and Multiresolution. Inverse Problems, **14**:225–243, (1998).

[59] Freeden, W., Schneider, F.: Runge–Walsh Wavelet Approximation for the Helmholtz Equation. J. Math. Anal. Appl., **235**, 533–566, (1999).

[60] Freeden, W., Schreiner, M.: Orthogonal and Non-Orthogonal Multiresolution Analysis, Scale Discrete and Exact Fully Discrete Wavelet Transform on the Sphere. Const. Approx., **14**:493–515, (1998).

[61] Freeden, W., Schreiner, M.: Local Multiscale Modelling of Geoid Undulations from Deflections of the Vertical. J. Geod., **79**:641–651, (2006).

[62] Freeden, W. and Witte, B.: A Combined (Spline-)Interpolation and Smoothing Method for the Determination of the Gravitational Potential from Heterogeneous Data. Bull. Geod., **56**:53–62, (1982).

[63] Freeden, W., Wolf, K.: Klassische Erdschwerefeldbestimmung aus der Sicht moderner Geomathematik. *Math. Semesterb.*, **56**: 53–77, (2008).

[64] Freeden, W., Gervens, T., Schreiner, M.: *Constructive Approximation on the Sphere* (*With Applications to Geomathematics*). Oxford Science Publications, Clarendon, Oxford, (1998).

[65] Freeden, W., Fehlinger, T., Klug, M., Mathar, D., Wolf, K.: Classical Globally Reflected Gravity Field Determination in Modern Locally Oriented Multiscale Framework. *J. Geodesy*, **83**:1171–1191, (2009).

[66] Fricker, F.: *Einführung in die Gitterpunktlehre*, Birkhäuser, Basel, Boston, Stuttgart, (1981).

[67] Friedrich, K.: Allgemeine für die Rechenpraxis geeignete Lösung für die Aufgaben der kleinsten Absolutsumme und der günstigsten Gewichtsverteilung, Zeitschrift für Vermessungswesen, 337–358, (1937).

[68] Friedrichs, K.O.: *Functional Analysis and Applications.* pp. 144–147, 151–153, New York University, mimeographed lecture notes.

[69] Galle, A.: Über die geodätischen Arbeiten von Gauß, I. Abschnitt: Die Methode der kleinsten Quadrate. Berlin, (1924), (Gauß Werke Bd. XI 2. Abteilung).

[70] Garland, G.D.: The Contributions of Carl Friedrich Gauss. Historia Mathematica, **6**:5–29, (1979).

[71] Gauss, C.F.: Untersuchungen über Höhere Arithmetik (Disquisitiones Arithmeticae (Latin)), Gerhard Fleischer jun., Leipzig, auch: Gauss Werke, Band 1, (1801).

[72] Gauss, C.F.: Theoria motus corporum coelestium in sectionibus conicis solem ambientium, Hamburg, (1809), Werke 7. Translated into English by C.H. Davis, (1963).

[73] Gauss, C.F.: Disquisitio de elementis ellipticis Paladis ex opositionibus annorum 1803, 1804, 1805, 1807, 1808, 1809, Göttingen, (1810).

[74] Gauss, C.F.: Disquisitiones generales circa seriem infinitam $1+((\alpha\cdot\beta)/\ (1-\gamma))x+\cdots$ etc., "Commentationes Societatis Regiae Scientiarium Gottingensis Recentiores", 2. (Werke Ill, 123–162], (1812).

[75] Gauss, C.F.: Theoria attractionis corporum sphaeroidicorum ellipticorum homogeneorum methodo nova tractata, Commentationes societatis regiae scientiarium Gottingensis recentiores, 2, 355–378, (1813).

[76] Gauss, C.F.: Methodus nova integralium valores per approximationem inveniendi. Commentationes Societatis Regiae Scientiarium Gottingensis Recentiores 3, (Werke III, 163–196), (1814).

[77] Gauss, C.F.: Theoria combinationis observationum erroribus minimis obnoxiae, 1. Teil, Göttingen, (1821).

[78] Gauss, C.F.: Allgemeine Auflösung der Aufgabe: Die Teile einer gegebenen Fläche auf einer andern gegebenen Fläche so abzubilden, dass die Abbildung dem Abgebildeten in den kleinsten Theilen ähnlich wird. In H.C. Schumacher (Hrsg.): Astronomische Abhandlungen, Heft 3, Altona 1825, S. 5–30, (1822).

[79] Gauss, C.F.: Theoria combinationis observationum erroribus minimis obnoxiae. 2. Teil, Göttingen, (1823).

[80] Gauss, C.F.: Supplementum theoriae combinationis observationum erroribus minimis obnoxiae. Göttingen, (1826).

[81] Gauss, C.F.: Allgemeine Theorie des Erdmagnetismus. Resultate aus den Beobachtungen des magnetischen Vereins, Göttingen, (1838).

[82] Gauss, C.F.: Werke, Vierter Band. Hrsg. Königl. Gesellschaft der Wissenschaft zu Göttingen, 109–117, B.G. Teubner, Leipzig, (1880).

[83] Gauss, C.F.: Werke, Neunter Band, Hrsg. Königl. Gesellschaft der Wissenschaften zu Göttingen, B.G. Teubner, Leipzig, (1903).

[84] Gauss, C.F.: Untersuchungen über Gegenstände der Höheren Geodäsie. Hrsg. J. Frischauf, Ostwald's Klassiker der exakten Wissenschaften, Nr. 177, Leipzig, (1910).

[85] Gautschi, W.: A Survey of Gauss–Christoffel Quadrature Formulas. In: E.B. Christoffel: The Influence of His Work on Mathematics and the Physical Sciences. P.L. Butzer and F. Feher, Eds., Birkhäuser, Basel, (1981).

[86] Gerling, C.L.: Ausgleichungsrechnungen der praktischen Geometrie oder die Methode der kleinsten Quadrate. Hamburg und Gotha, (1843).

[87] Gnedenko, B.W.: Über die Arbeiten von C.F. Gauß zur Wahrscheinlichkeitsrechnung, 193–204, in: C.F. Gauß Leben und Werk, Hrsg. H. Reichardt, Haude & Spenersche Verlagsbuchhandlung, Berlin, (1960).

[88] Goldstine, H.H.: A History of Numerical Analysis from the 16th Through the 19th Century. Springer-Verlag, New York, (1977).

[89] Grafarend, C.F., Klapp, M., Martinec, M.: Spacetime Modelling of the Earths Gravity Field by Ellipsoidal Harmonics. Handbook of Geomathematics, W. Freeden, Z. Nashed, T. Sonar, Eds., Springer, Heidelberg, 2nd Edition, 381–496, (2015).

[90] Gray, J: *Ideas of Space. Euclidean, Non-Euclidean, and Relativistic.* Clarendon Press, Oxford, (1979).

[91] Green, G.: *An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism.* T. Wheelhouse, Nottingham, England, (1838).

[92] Groetsch, C.W.: *Generalized Inverses of Linear Operators*. Marcel Dekker. Inc., New York, (1977).

[93] Groetsch, C.W.: *Inverse Problems in the Mathematical Science*. Vieweg, Braunschweig, (1993).

[94] Groten, E.: *Geodesy and the Earth's Gravity Field* I + II. Dümmler, Bonn, (1979).

[95] Gutting, M.: *Fast Multipole Methods for Oblique Derivative Problems*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern, (2007).

[96] Gutting, M.: *Fast Multipole Accelerated Solution of the Oblique Derivative Boundary Value Problem*. GEM Int. J. Geomath., **3**:223–252, (2012).

[97] Hammer, E.: Beitrag zur Geschichte der Ausgleichungsrechnung. *Zeitschrift für Vermessungswesen*, 613–628, (1900).

[98] Hampel, F.R., Ronschetti, E.M, Rousseeuw, P.R., Stahel, W.A.: *Robust Statistics*. Wiley, New York, (1986).

[99] Hansen, P.A.: Von der Methode der kleinsten Quadrate im Allgemeinen und in ihrer Anwendung auf die Geodäsie. *Abhandlungen der Mathematisch-Physikalischen Klasse der Königlich-Sächsischen Gesellschaft der Wissenschaften, Band* 8, 571–806, S. Hirzel, Leipzig, (1867).

[100] Hansen, P.A: Kurz gefaßte rationelle Ableitung des Ausgleichsverfahrens eines Dreiecksnetzes nach der Methode der kleinsten Quadrate. Berichte der Mathematisch-Physikalischen Klasse der Königlich-Sächsischen Gesellschaft der Wissenschaften, Band 20, 129–150, S. Hirzel, Leipzig, (1868).

[101] Hansen, P.A: Fortgesetzte geodätische Untersuchungen bestehend aus zehn Supplementen zur Abhandlung von der Methode der kleinsten Quadrate. *Abhandlungen der Mathematisch-Physikalischen Klasse der Königlich-Sächsischen Gesellschaft der Wissenschaften, Band* 9, 1–185, S. Hirzel, Leipzig, (1871).

[102] Hansen, P.A.: Entwicklung eines neuen veränderten Verfahrens zur Ausgleichung eine Dreiecksnetzes mit besonderer Beachtung des Falles, in welchem gewisse Winkel im voraus bestimmte Werte erhalten sollen. *Abhandlungen der Mathematisch-Physikalischen Klasse der Königlich-Sächsischen Gesellschaft der Wissenschaften, Band* 9, 185–289, S. Hirzel, Leipzig, (1871).

[103] Hardy, G.H.: On the Expression of a Number as the Sum of Two Squares. Quart. J. Math. (Oxford), **46**:263–283, (1915).

[104] Harter, H.L.: Nonuniqueness of Least Absolute Values Regression. Commun. Statist. Theor. Meth., **A6** (9):829–838, (1977).

[105] Heck, B.: *Rechenverfahren und Auswertemodelle der Landesvermessung Klassische und moderne Methoden*. 3. Auflage, Wichmann Verlag, Heidelberg, (2003).

[106] Heine, E.: *Anwendungen der Kugelfunctionen und der verwandten Functionen*. 2nd Edition, Reimer, Berlin. (I. Theil: Mechanische Quadratur, 1–31), (1881).

[107] Heiskanen, W.A., Moritz, H.: *Physical Geodesy*. Freeman, San Francisco, (1967).

[108] Helmert, F.R.: Lotabweichungen. Heft 1, Veröff. d. Kgl. Preuss. Geod. Inst., Berlin, (1886).

[109] Heun, K.: Neue Methode zur approximativen Integration der Differentialgleichungen einer unabhängigen Veränderlichen. Z. Math. Phys., **45**:23–38, (1900).

[110] Heuser, H.: *Funktionalanalysis*. 4. Auflage, Teubner, (1975).

[111] Hilbert, D.: *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*. Teubner, Leipzig, (1912).

[112] Hofmann-Wellenhof, B., Moritz, H.: *Physical Geodesy*. Springer, Wien, New York, 2nd Edition, (2006).

[113] Hörmander, L.: The Boundary Problems of Physical Geodesy. The Royal Institute of Technology, Division of Geodesy, Report 9, Stockholm, (1975).

[114] Hotine, M.: *Mathematical Geodesy*. ESSA Monography 2, U.S. Department of Commerce, Washington (Reprint 1992 by Springer), (1969).

[115] Huber, P.J.: *Robust Statistics*. Wiley, New York, (1981).

[116] Hurwitz, W.A.: On the Pseudo-Resolvent to the Kernel of an Integral Equation Transactions of the American Mathematical Society **13**: 405–418, (1912).

[117] Huxley, M.N.: Exponential Sums and Lattice Points III. Proc. London Math., **87**:591–609, (2003).

[118] Jordan, W., Eggert, O.: *Handbuch der Vermessungskunde*. Erster Band, Ausgleichungsrechnung nach der Methode der kleinsten Quadrate, 9. Aufl., J.B. Metzler, Stuttgart, (1948).

[119] Kanani, E.: *Robust Estimators for Geodetic Transformations and GIS*. Ph.D.-thesis, ETH No. 13521, Institut für Geodäsie und Photogrammetrie, Eidg. Technische Hochschule Zürich, Mitteilung Nr. 70, Zürich, (2000).

[120] Kantorowitsch, L.W., Akilow, G.P.: *Funktionalanalysis in normierten Räumen*. Akademie-Verlag, Berlin, (1964).

[121] Kellogg, O.D.: *Foundations of Potential Theory*. Frederick Ungar Publishing Company, New York, (1929).

[122] Kertscher, D.: Carl Friedrich Gauss – "Genie, Gigant, Titan". Z. f. Vermessungswesen, 130. Jg., **6**:372–378, (2005).

[123] Koch, K.R.: *Parameterschätzung und Hypothesentests*. Dümmler Verlag, Bonn (1997).

[124] Koch, K.R.: *Parameter Estimation and Hyothesis Testing in Linear Models*. 2nd ed., Springer, Berlin, (1999).

[125] Koch, K.R., Pope, A.J.: Uniqueness and Existence for the Geodetic Boundary Value Problem Using the Known Surface of the Earth. Bulletin Géodésique, **106**:467–476, (1972).

[126] Koch, K.R., Witte, B.: The Earth's Gravity Field Represented by a Simple Layer Potential From Doppler Tracking of Satellites. NOAA Technical Memorandum NOS 9, National Oceanic and Atmospheric Administration, National Ocean Service, Rockville, Md., (1971).

[127] Koch, K.R., Witte, B.: The Earth's Gravity Field Represented by a Simple Layer Potential From Doppler Tracking of Satellites. Journal of Geophysical Research, 76(35): 8471–8479, The American Geophysical Union, (1971).

[128] Kolesnik, G.: On the Method of Exponent Pairs. Acta Arith., **45**: 115–143, (1985).

[129] Kowalewski, A.: *Newton, Cotes, Gauss, Jacobi: Vier grundlegende Abhandlungen über Interpolation und genäherte Quadratur*. von Veit, Leipzig, (1917).

[130] Krarup, T.: *A Contribution to the Mathematical Foundation of Physical Geodesy*. Danish Geodetic Institute, Report No. 44, Copenhagen, (1969).

[131] Krarup, T.: *Letters on Molodensky's Problem* I–IV. Communication to the members of the IAG Special Study Group 4.31, (1973).

[132] Krüger, L.: Über die Ausgleichung von bedingten Beobachtungen in zwei Gruppen. Veröff. d. Kgl. Preu. Inst., N. F., Nr. 18, Potsdam, (1905).

[133] Kutta, W.: Beitrag zur näherungsweisen Integration totaler Differentialgleichungen. Z. Math. Phys., **46**: 435–453, (1901).

[134] Landau, E., Über die Gitterpunkte in einem Kreis (Erste Mitteilung). Gött. Nachr., 148–160, (1915).

[135] Landau, E.: Über die Gitterpunkte in einem Kreis IV. Gött. Nachr., 58–65, (1924).

[136] Laplace, P.S.: Traité de mécanique céleste. tome 2, Paris, (1799).

[137] Laplace, P.S.: Théorie analytique des probabilités. Livre II, Chap. IV. Paris, (1812).

[138] Laplace, P.S.: Théorie analytique des probabilités. Œuvres, tome VII, p. 353, (1812).

[139] Legendre A.M.: Nouvelles méthodes pour la détermination des orbites des comètes. Paris, (1806).

[140] Legendre, A.M.: Analyse des triangles tracés sur la surface d'un sphéroide. Tome VII de la I Série des mémoires de l'Académie des Sciences, Paris, 131, (1806).

[141] Lehmann, G.: Gauss theoretische Arbeiten. In C.F. Gauss und die Landesvermessung in Niedersachsen, herausgegeben von der Niedersächsischen Vermessungs- und Katasterverwaltung, (1955).

[142] Lehmann, R., Lösler, M.: Multiple Outlier Detection: Hypothesis Tests versus Model Selection by Information Criteria. Journal of Surveying Engineering, **142**: DOI: http://dx.doi.org/10.1061/(ASCE)SU.1943-5428.0000189, (2016).

[143] Listing J.B.: Über unsere jetzige Kenntnis der Gestalt und Größe der Erde. Dietrichsche Verlagsbuchhandlung, Göttingen, (1873).

[144] Magnus, W., Oberhettinger, F., Soni, R.P.: *Formulas and Theorems for the Special Functions of Mathematical Physics*. In: Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Band 52, Springer, Berlin, 3. Auflage, (1966).

[145] Maire, Ch., Boscović, R.J.: De Litteraria Expeditione per Pontificiam Ditionem ad Dimentiendos duos Meridiani Gradus, Rom, 409–503, (1755).

[146] Markina, I., Vodopyanov, S.K.: Fundamentals of the Nonlinear Potential Theory for Subelliptic Equations. I, II. Siberian Adv. Math., **7**:32–62, (1997).

[147] Marx, C.: Outlier Detection by Means of Monte Carlo Estimation Including Resistant Scale Estimation. J. of Applied Geodesy, **9**:123–141, (2015).

[148] McEwen, J.D., Puy, G., Thiran, J.-P., Vandergheinst, P.: Sampling Theorems and Compressive Sensing on the Sphere. Wavelets and Sparsity XIV, Papadakis, M., van de Ville, D., Goyal, V.K., Eds., Proc. of SPIE, Vol. 8138, F1–F9, (2011).

[149] Mehler, F.G.: Bemerkungen zur Theorie der mechanischen Quadraturen. J. Reine Angew. Math., **63**:152–157, (1864).

[150] Meissl, P.: *On the Linearization of the Geodetic Boundary Value Problem*. Department of Geodetic Science, No. 152, The Ohio State University, Columbus, OH, (1971).

[151] Molodensky, M.S., Eremeev, V.F., Yurkina, M.I.: *Methods for Study of the External Gravitational Field and Figure of the Earth*. Trudy TSNIIGAiK, Geodezizdat,

Moscow, 131, (1960), English translat.: Israel Program for Scientific Translation, Jerusalem, (1962).

[152] Moore, E.H.: On the Reciprocal of the General Algebraic Matrix. Bulletin of the American Mathematical Society **26**:394–395, (1920).

[153] Moore, E.H.: General Analysis. Memoirs of the American Mathematical Society, **1**:197–209, (1935).

[154] Moritz, H.: Statistische Methoden in der gravimetrischen Geodäsie. Zeitschrift für Vermessungswesen, 409–416, (1963).

[155] Moritz, H.: Der Begriff der mathematischen Erdgestalt seit Gauss. Allgemeine Vermessungs-Nachrichten, 133–138, (1977).

[156] Moritz, H.:*Advanced Physical Geodesy*. Wichmann Verlag, Karlsruhe, (1980).

[157] Moritz, H.: Geodesy and Mathematics. Zeszyty Naukowe Akademii Górniczo-Hutniczej IM. Stanislawa Staszica, No. 780, Geodezja, 63, Kraków, 38–43, (1981).

[158] Moritz, H.: *Advanced Physical Geodesy*, Wichmann, Karlsruhe, 2nd Edition, (1989).

[159] Moritz, H.: Classical Physical Geodesy. Handbook of Geomathematics, W. Freeden, Z. Nashed, T. Sonar, Eds., Springer, Heidelberg, 2nd Edition, 253–290, (2015).

[160] Müller, C.: Eine Verallgemeinerung der Eulerschen Summenformel und ihre Anwendung auf Fragen der analytischen Zahlentheorie. Abh. Math. Sem. Univ. Hamburg, **19**:41–61, (1954).

[161] Müller, C., *Analysis of Spherical Symmetries in Euclidean Spaces*. Springer, New York, Berlin, Heidelberg, (1998).

[162] Nashed, M.Z.: Generalized Inverses, Normal Solvability and Iteration for Singular Operator Equations. In: *Nonlinear Functional Analysis and Applications*, L.B. Rall, ed., Academic, New York, 311–359, (1971).

[163] Nashed, M.Z.: Aspects of Generalized Inverses in Analysis and Regularization. In: Generalized Inverses and Applications. Academic Press, New York, 193–244, (1976).

[164] Nashed, M.Z.: *Generalized Inverses and Applications*. Academic Press, New York, (1976).

[165] Nashed, M.Z.: Operator-Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory. IEEE Transactions on Antennas and Propagation, **29**:220–231, (1981).

[166] Nashed, M.Z.: Inner, Outer, and Generalized Inverses in Banach and Hilbert Spaces. Numer. Funct. Anal. Optim., **9**:261–326, (1987).

[167] Nashed, M.Z., Walter, G.G.: General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces. Math. Control Signals Systems, **4**:363–390, (1991).

[168] Neumann, F., *Vorlesungen über die Theorie des Potentials und der Kugelfunktionen*, Teubner, Leipzig, 135–154, (1887).

[169] Neumann, J.V.: On Regular Rings. Proceedings of the National Academy of Science of the United States of America, **22**:707–713, (1936).

[170] Nutz, H.: *A Unified Setup of Gravitational Field Observables*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen, (2002).

[171] Palczewski, A., Schneider, J., Bobylev, A.: A Consistency Result for a Discrete Velocity Model of the Boltzmann Equation. SIAM J. Numer. Anal., **34**:1865–1883, (1997).

[172] Penrose, R.: A Generalized Inverse for Matrices. Proceedings of the Cambridge Philosophical Society 51, 406–413, (1955).

[173] Petrini, H.: Sur l'existence des dérivées secondes du potentiel. C.R. Acad. Sci. Paris, **130**:233–235, (1900).

[174] Pizzetti, P.: Sopra il Calcoba Tesrico delle Deviazioni del Geoide dall'Ellissoide. *Att. R Accad. Sci. Torino*, **46**:331–350, (1910).

[175] Plackett, R.L.: A Historical Note on the Method of Least Squares. Biometrika, **36**:458–460, (1949).

[176] Proclus: *A Commentary on the First Book of Euclid's Elements*. Translated with Introduction and Notes, by Glenn R. Morrow. Princeton University Press, (1970).

[177] Radau, R.: Étude sur les formules d'approximation qui servent à calculer la valeur numérique d'une intégrale définie. J. Math. Pures Appl., **3**(6):283–336, (1880).

[178] Rado, R.: Note on Generalized Inverses of Matrices. Proceedings of the Cambridge Philosophical Society **52**:600–601, (1956).

[179] Reich, K., Roussanova, E.: Visualising Geomagnetic Data by Means of Corresponding Observations. GEM International J. Geomath., **4**:1–25, (2013).

[180] Reid, W.T.: Generalized Inverses of Differential and Integral Operators. Theory and Applications of Generalized Inverses of Matrices (T.L. Boullion and P.L. Odell 1–25). Symposium Proceedings, Texas Tech University Mathematics Series, Vol. 4., Lubbock; Texas Tech Press, (1965).

[181] Reinsch, C.: Smoothing by Spline Functions. Numer. Math., **10**:177–183, (1967).

[182] Reuter, R.: *Über Integralformeln der Einheitssphäre und harmonische Splinefunktionen*. Ph.D.-thesis, Veröff. Geod. Inst. RWTH Aachen, Report No. 33, (1982).

[183] Robinson, D.W.: Gauss and Generalized Inverses. Historia Mathematics, **7**:118–125, (1980).

[184] Rosenfeld, B.A.: *A History of Non-Euclidean Geometry – Evolution of the Concept of a Geometric Space*. Springer Science+Business Media, (1988).

[185] Rummel, R.: Geodesy. Encyclopedia of Earth System Science, W.A. Nierenberg, ed., Academic Press, **2**:253–262, (1992).

[186] Rummel, R.: Spherical Spectral Properties of the Earth's Gravitational Potential and Its First and Second Derivatives. In: Lecture Notes in Earth Science, Springer, Berlin, Vol. **65**:359–404, (1997).

[187] Runge, C.: Zur Theorie der eindeutigen analytischen Funktionen, *Acta Math.* 6, 229–234, (1885).

[188] Runge, C., Willers, F.A.: Numerische und graphische Integration. Encyklopädie der Math. Wiss., **2**(3):47–176, Leipzig, (1915).

[189] Sanso, F.: A Note on Density Problems and the Runge Krarup's Theorem. Bolletino di Geodesia e Science Affini **41**, 422–477, (1982).

[190] Sard, A.: Integral Representations of Remainders. Duke Math. J. **15**:333–345, (1948).

[191] Schauder, J.: Potentialtheoretische Untersuchungen, Math. Z., **35**:536–538, (1931).

[192] Schlossmacher, E.J.: An Iterative Technique for Absolute Deviations Curve Fitting. J. Americ. Statistical Association, **68**:857–859, (1973).

[193] Schmidt; H.: Lösung der geodätischen Hauptaufgaben auf dem Rotationsellipsoid mittels numerischer Integration. Zeitschrift für Vermessungswesen **123**:121–128, (1999).

[194] Schreiber, O.: Anordnung der Winkelbeobachtungen im Göttinger Basisnetz. Zeitschrift für Vermessungswesen, 129–161, (1882).

[195] Schreiber, O.: *Die konforme Doppelprojektion der trigonometrischen Abteilung der Königlich Preussischen Landesaufnahme*. Formeln und Tafeln. Selbstverlag, Berlin, (1897).

[196] Schreiber, O.: Formeln und Tafeln zur Berechnung der geographischen Coordinaten aus Richtungen und Längen der Dreiecksseiten, Königl. Preuss. Landesaufn., Berlin, (1878).

[197] Schreiner, M.: *Wavelet Approximation by Spherical Up Functions*. Habilitation Thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen, (2004).

[198] Shannon, C.E.: Communication in the Presence of Noise. Proc. Institute of Radio Engineers, **37**:10–21, (1949).

[199] Shure, L., Parker, R.L., Backus, G.E.: Harmonic Splines for Geomagnetic Modelling. Phys. Earth Planet. Inter. **28**:215–229, (1982).

[200] Siegel, C.L.: Über die analytische Theorie der quadratischen Formen, III. Annals of Mathematics **38**:212–291, (1937).

[201] Sierpinski, W.: O pewnem zagadnieniu z rachunku funckcyj asmptotycznych (Über ein Problem des Kalküls der asymptotischen Funktion (polnisch)), Prace Math.-Fiz., **17**:77–118, (1906).

[202] Simons, F.J., Plattner, A.: Scalar and Vector Slepian Functions, Spherical Signal Estimation and Spectral Analysis. Handbook of Geomathematics (W. Freeden, M.Z. Nashed, T. Sonar, Eds.), 2nd Edition, Vol. 3, 2563–2608, (2015).

[203] Slepian, D.: Some Comments on Fourier Analysis, Uncertainty and Modeling. SIAM Rev., **25**(3):379–393, (1983).

[204] Sonar, T., 3000 *Jahre Analysis*. Springer, Berlin, Heidelberg, New York, (2011).

[205] Stewart G.W.: Theoria Combinationis Observationum Erroribus Minimis Obnoxiae. Pars Prior et Pars Posterior and Supplementum, Theory of the Combination of Observations Least subject to Errors, Part One, Part Two, Supplement by C.F. Gauß, Society for Industrial and Applied Mathematics, (1995).

[206] Stoer, J., *Numerische Mathematik* 1. 5. Auflage, Springer-Verlag, Berlin, Heidelberg, New York, (1989).

[207] Stokes, G.G.: On the Variation of Gravity on the Surface of the Earth. Trans. Camb. Phil. Soc., **8**:672–695, (1849).

[208] Svensson, S.L.: Pseudodifferential Operators. A New Approach to the Boundary Value Problems of Physical Geodesy. Manuscr. Geod., **8**:1–40, (1983).

[209] Szegö G.: *Orthogonal Polynomials*. American Mathematical Society Colloquium Publications, 4th edition, Volume 23, American Mathematical Society, Providence, Rhode Island, (1975).

[210] Taylor, A.E., Lay, D.: *Functional Analysis*. Wiley, New York, (1979).

[211] Tikhonov, A.N.: On the Stability of Inverse Problems. *Dokl. Akad. Nauk. SSSR*, **39**:195–198, (1943).

[212] Tikhonov, A.N.: On the Solution of Incorrectly Formulated Problems and the Regularization Method, *Dokl. Akad Nauk SSSR*, **151**:501–504, (1963).

[213] Tikhonov, A.N.: On Methods of Solving Incorrect Problems, *Amer. Math. Soc. Transl.*, **2**, 70: 222–224, (1968).

[214] Tikhonov, A.N., Arsenin V.Ya.: *Solutions of Ill-Posed Problems*. Wiley, New York, (1977).

[215] Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V., Yagola, A.G.: *Numerical Methods for the Solution of Ill-Posed Problems*. Kluwer, Dordrecht, (1995).

[216] Todhunter, I., A History of the Mathematical Theories of Attraction and of the Figure on the Earth from the Time of Newton to that of Laplace, Macmillan, London (reprint 1962 by Dover Publications, New York), (1878).

[217] Vening Meinesz, F.A.: A Formula Expressing the Deflection of the Plumb-Line in the Gravity Anomalies and some Formulae for the Gravity Field and the Gravity Potential Outside the Geoid. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, **31**(3):315–331, (1928).

[218] Vekua, I.N.: Über die Vollständigkeit des Systems harmonischer Polynome im Raum. *Dokl. Akad. Nauk* 90, 495–498, (1953).

[219] Vogler, C.A.: Lehrbuch der praktischen Geometrie. Bd. II 1. Halbband, Braunschweig, (1894).

[220] Voronoi, G.: Sur un problème du calcul des fonctions asymptotiques. JRAM, **126**:241–282, (1903).

[221] Wahba, G.: *Spline Models for Observational Data*. SIAM Philadelphia, (1990).

[222] Walsh, J.L.: The Approximation of Harmonic Functions by Harmonic Polynomials and by Harmonic Rational Functions, *Bull. Amer. Math. Soc.* 35, 499–544, (1929).

[223] Wangerin, A.: *Theorie des Potentials und der Kugelfunktionen*, Walter de Gruyter & Co, Berlin, Leipzig, (1921).

[224] Weizel, R.: C.F. Gauß, Angewandte Mathematik 200 Jahre nach seiner Geburt. Zeitschrift für Vermessungswesen, 520–525, (1977).

[225] Wermer, J.: *Potential Theory*. Lecture Notes in Mathematics, 408, Springer, Berlin, (1974).

[226] Werner, J., Numerische Mathematik 1. Vieweg Studium, Braunschweig, (1991).

[227] Werner, D.: *Funktionalanalysis*. 4. Auflage; Springer, Berlin, (2002).

[228] Weyl, H.: Über die Gleichverteilung von Zahlen mod. Eins. Math. Ann., **77**:313–352, (1916).

[229] Witte, B.: Die Berechnung von Schwerestörungen im Außenraum mit Hilfe des Greenschen Satzes. Zeitschrift für Vermessungswesen (ZfV), 135–141, Stuttgart, (1969).

[230] Witte, B.: Vergleich verschiedener Methoden für die Berechnung von Schwerestörungen im Außenraum. Schweizerische Zeitschrift für Vermessung, Photogrammetrie und Kulturtechnik, 161–168, (1969).

[231] Witte, B.: Die Bestimmung von Horizontalableitungen der Schwere im Außenraum aus einer Weiterentwicklung der Stokesschen Funktion. Gerlands Beiträge zur Geophysik, 87–94, Leipzig, (1970).

[232] Witte, B.: Computational Procedures for the Determination of a Simple Layer Model of the Geopotential From Doppler Observations. NOAA Technical Report NOS-42, National Oceanic and Atmospheric Administration, National Ocean Service, Rockville, Md., (1971).

[233] Witte, B., Sparla, P.: *Vermessungskunde und Grundlagen der Statistik für das Bauwesen*. 8. Auflage, Wichmann – VDE-Verlag, Berlin, (2015).

[234] Wolf, H.: *Ausgleichungsrechnung nach der Methode der kleinsten Quadrate*. Dümmler Verlag, Bonn, (1968).

[235] Wolf, H.: *Ausgleichungsrechnung. Formeln zur praktischen Anwendung*. Dümmler Verlag, Bonn, (1975).

[236] Wolf, H.: C.F. Gauß und die Methode der kleinsten Quadrate. Allgemeine Vermessungs-Nachrichten, 125–133, (1977).

[237] Wolf, H.: Die deutsche Landesvermessung in ihrer historischen Entwicklung. Festschrift 100 Jahre Geodätische Lehre und Forschung in Hannover. Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover, Nr. 100, 47–53. Hannover, (1981).

[238] Wolf, H.: Neues Altes in der Ausgleichungsrechnung. Vermessung, Photogrammetrie, Kulturtechnik, 233–240, (1983).

[239] Wolfe, H.E.: *Introduction to Non-Euclidean Geometry*. Reprint of the first edition 1945. Dover Publications, Mineola, New York, (2012).

[240] Wußing, H.: *Carl Friedrich Gauss*. B.G. Teubner Verlag, Leipzig, (1989).

[241] Yoshida, K.: *Functional Analysis*. 5th Edition, Springer, Berlin, (1980).

Willi Freeden
Geomathematics Group
University of Kaiserslautern
MPI-Gebäude, Paul-Ehrlich-Str. 26
D-67663 Kaiserslautern, Germany
e-mail: freeden@rhrk.uni-kl.de

Thomas Sonar
Institute for Computational Mathematics
University of Brunswick
Pockelsstraße 14
D-38106 Braunschweig, Germany
e-mail: t.sonar@tu-bs.de

Bertold Witte
Institute for Geodesy and Geoinformation
University of Bonn
Nussallee 17
D-53115 Bonn, Germany
e-mail: bertold.witte@uni-bonn.de

Handbook of Mathematical Geodesy
Geosystems Mathematics, 165–199

An Overview on Tools from Functional Analysis

Matthias Augustin, Sarah Eberle, and Martin Grothaus

Abstract. Many modern mathematical methods treat geodetic problems in terms of functions from certain spaces, proving convergence properties of such functions and regard the evaluation of such functions or their derivatives at given points as operators. In doing so, knowingly or unknowingly, they use the language of functional analysis.

This contribution aims at summarizing some fundamental concepts from functional analysis which are used throughout this book. In this way, it tries to add a layer of self-sufficiency and to act as supplement to other contributions for those readers who are not familiar with functional analytic tools. For this purpose, we introduce, among others, the general ideas of vector spaces, norms, metrics, inner products, orthogonality, completeness, Banach spaces, Hilbert spaces, functionals, linear operators, different notions of convergence. Then we show how functions can be interpreted as vectors in different kind of function spaces, e.g., spaces of continuous functions, Lebesgue spaces, or Sobolev spaces and how the more general concepts come into play here. Moreover, we have a brief glimpse at differential equations and how functional analytic tools provide the necessary background to discuss them, and at the idea of reproducing kernels and the corresponding reproducing kernel Hilbert spaces.

Keywords. Functional analysis, metric spaces, normed spaces, function spaces, Sobolev spaces, reproducing kernel Hilbert spaces, basis systems, operators, convergence, weak derivatives, distributions, partial differential equations.

1. Introduction

Readers of this book who do not have a strong mathematical background might wonder why it includes this chapter. Let us give a short motivation.

Most readers are probably familiar with the Fourier expansion and its spherical relative, the expansion of a function in terms of spherical harmonics as it is used, for example, to determine the gravitational potential of the Earth. This

leads to the question which functions can be expanded in a Fourier series. Functional analysis provides the context to discuss this question. As the determination of Fourier coefficients requires the computation of integrals, the resulting answer introduces the space of rapidly decreasing functions. However, this is probably not the space a practitioner prefers. Dealing with signals – and in this context, the gravitational field of the Earth might be considered as a signal – it is natural to demand that a solution has finite energy, as it is otherwise physically impossible. The energy density of a signal is proportional to the square of its amplitude. Thus, the total energy is given by the integral of the square of the amplitude and has to be finite. This yields the concept of square-integrable functions and the more general Lebesgue spaces. It would be desirable to extend the concept of Fourier expansions from the rapidly decreasing to square integrable functions.

The tools to discuss all of the above are given by functional analysis.

Let us consider another, more ambitious task. One of the main topics of geodesy is the determination of the gravity field in the exterior of the Earth. As the density distribution inside the Earth is not accessible, we cannot use Newton's volume potential for this purpose. Instead, early geodesists could only use data from measurements at the Earth's surface, yielding such results as Stokes' famous integral formula to compute the disturbing potential from gravity anomalies ([20]). However, with the advancement of satellite technology, the situation changed dramatically in several ways. One of them is that, with modern GPS, gravity disturbances become available in more and more areas, leading to the conclusion that instead of Stokes' formula, a similar integral formula due to von Neumann becomes applicable. Nowadays, geodesists are confronted with a plethora of different quantities, as not only gravity anomalies or gravity disturbances, but also, among others, gravitational gradients, gravitational tensors, deflections of the vertical, or height anomalies can be measured. Additionally, some of these quantities are available from terrestrial measurements while others are collected by satellites. This yields two questions:

1. Is it possible to find a framework which allows a common interpretation for all of these different measurements?
2. Can we combine different types of data to get more accurate results?

If, for a moment, we consider only gravity anomalies and gravity disturbances, the first of these two questions might be reformulated: what are the common properties of the integral formulas by Stokes and by Neumann? Both formulas take a certain function, assumed to be given on a sphere, as input and give another function, the disturbing potential in the exterior of the Earth, as output. As we try to deduce which kinds of inputs are allowed and how properties of the input and output are linked to each other, we arrive again at the concepts of function spaces, (integral) operators, limits and convergence. Once more, we have reached the mathematical domain of functional analysis.

It is only a small step from opening the rich toolbox of functional analysis to interpreting the above mentioned measurements as the application of operators.

Some of these operators are linear, some non-linear and some derived by linearizing non-linear operators. But it would be shortsighted to assume that functional analysis is only useful to describe the setting of the problem. It also provides tools to solve those problems efficiently. However, the aim of this chapter is not to present solutions to geodetic problems, as this is done in other chapters, but to provide the reader with the necessary background to understand modern approaches in mathematical geodesy.

For this purpose, we start by recalling definitions and properties of metric spaces, normed spaces, Banach spaces, and Hilbert spaces as well as linear operators and sesquilinear or bilinear forms. This leads us to the consideration of different kinds of convergence, e.g., with respect to a given norm, weak convergence and weak* convergence. Next, we specifically consider function spaces, starting with classical spaces of continuous or continuously differentiable functions, but also taking into account spaces of distributions, Lebesgue spaces, and Sobolev spaces. The latter ones are particularly useful when dealing with differential equations in weak form, yielding results on existence, uniqueness and regularity of solutions. The chapter is completed by a short discussion of reproducing kernel Hilbert spaces.

2. Basic concepts

This section summarizes some basic concepts from functional analysis taken from [2, 4, 10].

2.1. Metric spaces, normed spaces, Banach spaces, and linear operators

Let $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{R}^+$, $\mathbb{R}_0^+$, and $\mathbb{C}$ denote the set of positive integers, non-negative integers, integers, real numbers, positive real numbers, non-negative real numbers, and complex numbers, respectively.

In the following, V, W, and Z are $\mathbb{K}$-vector spaces over the field $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. As a particular space, we have for any $n \in \mathbb{N}$, the n-dimensional vector spaces $\underbrace{\mathbb{K} \times \cdots \times \mathbb{K}}_{n\text{-times}} = \mathbb{K}^n$.

For the elements $x = (x_1, x_2, \ldots, x_n)^T, y = (y_1, y_2, \ldots, y_n)^T \in \mathbb{K}^n$ we define the (Euclidean) inner product $x \cdot y$ and its induced (Euclidean) norm $\|x\|$ by

$$x \cdot y = \sum_{i=1}^{n} x_i \overline{y}_i, \tag{1}$$

$$\|x\| = \sqrt{x \cdot x} = \sqrt{\sum_{i=1}^{n} |x_i|^2}, \tag{2}$$

where $\overline{z}$ denotes the complex conjugate of z for any $z \in \mathbb{C}$ and $|\cdot|$ is the absolute value in $\mathbb{K}$, i.e., for $x \in \mathbb{R}$,

$$|x| = \begin{cases} x, & x > 0, \\ 0, & x = 0, \\ -x, & x < 0, \end{cases} \tag{3}$$

and for $z = x + iy \in \mathbb{C}$, $x, y \in \mathbb{R}$, i the imaginary unit,

$$|z| = \sqrt{x^2 + y^2}. \tag{4}$$

The canonical orthonormal basis vectors are denoted by e^1, e^2,..., e^n. In $\mathbb{R}^3$ we have

$$e^1 = (1,0,0)^T,\ e^2 = (0,1,0)^T,\ e^3 = (0,0,1)^T. \tag{5}$$

Furthermore, let $\delta_{n,m}$ given by

$$\delta_{n,m} = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases} \tag{6}$$

for $n, m \in \mathbb{N}_0$, denote the Kronecker delta.

For any subset $\mathcal{G} \subset \mathbb{R}^n$, $\partial\mathcal{G}$ denotes the boundary and its closure is given by $\overline{\mathcal{G}} = \mathcal{G} \cup \partial\mathcal{G}$.

In addition, an open ball of radius $r \in \mathbb{R}^+$ centered at $x \in \mathbb{K}^n$ is defined by

$$B_r(x) = \{y \in \mathbb{K}^n : \|x - y\| < r\}. \tag{7}$$

The corresponding sphere is given by

$$\partial B_r(x) = \{y \in \mathbb{K}^n : \|x - y\| = r\}. \tag{8}$$

Definition 2.1 (Metric Space). A function $\rho : V \times V \to \mathbb{R}_0^+,\ u, v \mapsto \rho(u, v)$ is called a metric if it has the following properties:

(a) for all $u, v \in V$, $\rho(u, v) \geq 0$ and $\rho(u, v) = 0 \iff u = v$,
(b) for all $u, v \in V$, $\rho(u, v) = \rho(v, u)$,
(c) for all $u, v, w \in V$ we have the triangle inequality $\rho(u, v) \leq \rho(u, w) + \rho(v, w)$.

Definition 2.2 (Convergence in Metric Spaces). A sequence $(v_n)_{n\in\mathbb{N}}$ in V is called convergent with limit $v_0 \in V$ if $\lim_{n\to\infty} \rho(v_n, v_0) = 0$.

Definition 2.3 (Cauchy Sequence). A sequence $(v_n)_{n\in\mathbb{N}}$ in V is called Cauchy sequence if for any $\varepsilon \in \mathbb{R}^+$ exists a $n_0 \in \mathbb{N}$ such that for all $n, m \in \mathbb{N}$, $n \geq m \geq n_0$, we have $\rho(v_n, v_m) \leq \varepsilon$.

Definition 2.4 (Completeness). A metric space V is called complete if every Cauchy sequence is convergent.

Definition 2.5 (Density). Let W, Z be two sets in a metric space V such that $W \subset Z \subset V$. W is said to be dense in Z if for every z in Z and every $\varepsilon \in \mathbb{R}^+$ exists a $w \in W$ such that $\rho(w, z) \leq \varepsilon$.

Definition 2.6 (Separability). A metric space V is called separable if it contains a countable, dense subset.

Definition 2.7 (Normed Space). A function $\|\cdot\| : V \to \mathbb{R}_0^+$, $v \mapsto \|v\|$ is called a norm if it has the following properties for all $u, v \in V$, $\alpha \in \mathbb{K}$:

(a) $\|v\| \geq 0$ and $\|v\| = 0 \iff v = 0$,
(b) $\|\alpha v\| = |\alpha| \|v\|$,
(c) $\|u + v\| \leq \|u\| + \|v\|$.

From every norm $\|\cdot\|$ on V, a metric can be constructed by

$$\rho(u, v) = \|u - v\| = \|v - u\|, \quad u, v \in V. \tag{9}$$

Additionally to the defining properties of a metric, a metric defined from a norm also satisfies for all $u, v \in V$, $\alpha \in \mathbb{K}$:

$$\rho(u + w, v + w) = \rho(u, v), \tag{10}$$

$$\rho(\alpha u, \alpha v) = |\alpha| \rho(u, v). \tag{11}$$

Vice versa, a metric that also has properties (10) and (11) can be used to define a norm via

$$\|v\| = \rho(v, 0), \quad v \in V. \tag{12}$$

Definition 2.8 (Complete System of Elements). Let V be a normed space and I a suitable set to index a system of elements $\{v_i : i \in I\} \subset V$. The system $\{v_i : i \in I\}$ is called complete if span $\{v_i : i \in I\}$, i.e., the set of all finite linear combinations of the elements $\{v_i : i \in I\}$, is dense in V.

Theorem 2.9. *Any finite-dimensional normed space is complete.*

Theorem 2.10. *Any finite-dimensional subspace of a normed space V is closed.*

Definition 2.11 (Banach Space). A normed space $(V, \|\cdot\|_V)$ is called Banach space if V is complete with respect to $\|\cdot\|_V$.

If there is no risk of confusion, we usually do not state the norm explicitly and also omit to index the norm by the corresponding space.

Definition 2.12 ((Schauder) Bases in Banach Spaces). Let V be a Banach space. A sequence $(v_n)_{n \in \mathbb{N}}$ in V is called a basis, if every element $v \in V$ can be written in the form $v = \sum_{n=1}^{\infty} a_n v_n = \lim_{N \to \infty} \sum_{n=1}^{N} a_n v_n$ with coefficients $a_n \in \mathbb{K}$ and the elements of every finite subset of $(v_n)_{n \in \mathbb{N}}$ are linearly independent.

Whereas every Banach space with a countable basis is separable, not every separable Banach space has a countable basis (see [5]).

Definition 2.13 (Operators and Functionals). Let V, W be two spaces and U an arbitrary subset of V. A function $L : U \to W$, $U \ni v \mapsto Lv \in W$ that maps every

element $v \in U$ on a unique element $Lv \in W$ is called an operator. The set U is called the domain of the operator L and denoted by $D(L)$. The set

$$R(L) = \{w \in W : w = Lv, v \in D(L)\} \tag{13}$$

is called the range of L. The set

$$N(L) = \{v \in D(L) : Lv = 0\} \tag{14}$$

is called the kernel or null space of L.

If $R(L) \subset \mathbb{K}$, the operator is called a functional.

Definition 2.14 (Continuous Operator). Let (V, ρ_V) and (W, ρ_W) be two metric spaces. An operator $L : V \supset D(L) \to R(L) \subset W$ is called continuous in $v_0 \in D(L)$ if for every sequence $(v_n)_{n\in\mathbb{N}} \subset D(L)$ with $\lim_{n\to\infty} \rho_V(v_n, v_0) = 0$ we have $\lim_{n\to\infty} \rho_W(Lv_0, Lv_n) = 0$.

L is said to be continuous on $D(L)$ if L is continuous for all $v_0 \in D(L)$.

Definition 2.15 (Linear Operator). Let $D(L) \subset V$ be $\mathbb{K}$-vector spaces and $L : D(L) \mapsto R(L) \subset W$. L is called a linear operator if for arbitrary $\alpha \in \mathbb{K}$, and $v_1, v_2 \in D(L)$ we have

$$L(\alpha v_1 + v_2) = \alpha L v_1 + L v_2. \tag{15}$$

It follows from Definition 2.15 that $N(L)$ is a linear subspace of $D(L)$ and $R(L)$ is a linear subspace of W.

Definition 2.16 (Isometric Linear Operator). Let $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ be two normed spaces and $L : V \supset D(L) \mapsto R(L) \subset W$ a linear operator. L is called isometric if for every $v \in D(L)$ we have $\|Lv\|_W = \|v\|_V$.

In the following, we assume every operator to be linear.

Definition 2.17 (Inverse Operator). Let $L : V \supset D(L) \mapsto R(L) \subset W$ such that for every $w \in R(L)$ exists a unique $v \in D(L)$ with $Lv = w$. The operator that assigns to every $w \in R(L)$ its inverse image $v \in D(L)$ is called the inverse operator to L, denoted by L^{-1}, and is a linear operator from $R(L)$ to $D(L)$. We have

$$L^{-1}(Lv) = v \quad \text{for all } v \in D(L), \tag{16}$$

$$L(L^{-1}w) = w \quad \text{for all } w \in R(L). \tag{17}$$

As the existence of the inverse operator requires the uniqueness of the inverse image, a linear operator L has an inverse operator L^{-1} if and only if $N(L) = \{0\}$.

Definition 2.18 (Bounded Operator and Operator Norm). Let $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ be two normed spaces and $L : V \supset D(L) \mapsto R(L) \subset W$ a linear operator. L is bounded if it exists a $C \in \mathbb{R}_0^+$ such that

$$\|Lv\|_W \leq C \|v\|_V \quad \text{for all } v \in D(L). \tag{18}$$

The space of all linear bounded operators with domain $D(L)$ and range $R(L)$ is denoted by $\mathscr{L}(D(L), R(L))$. This space can be equipped with the operator norm, given as

$$\|L\|_{\mathscr{L}(D(L),R(L))} = \sup_{\substack{v\in D(L)\\ v\neq 0}} \frac{\|Lv\|_W}{\|v\|_V} = \sup_{\substack{v\in D(L)\\ \|v\|_V=1}} \|Lv\|_W = \sup_{\substack{v\in D(L)\\ \|v\|_V\leq 1}} \|Lv\|_W . \tag{19}$$

We have $\|Lv\|_W \leq \|L\|_{\mathscr{L}(D(L),R(L))} \|v\|_V$.

The space $V' = \mathscr{L}(V, \mathbb{K})$ is called the (topological) dual space to V and consists of all continuous linear functionals on V. Elements of V' are denoted by v'. A norm on V' is defined in the same way as the general operator norm.

The dual space of V' is denoted by V'' and called the bidual space of V.

A space is called reflexive, if V'' and V are isomorph.

Theorem 2.19 (Continuity of Linear Operators). *Let $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ be two normed spaces and $L: V \supset D(L) \mapsto R(L) \subset W$ a linear operator. L is continuous if and only if L is bounded.*

Definition 2.20 (Dual Space Adjoint Operator). Let $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ be two normed spaces and $L \in \mathscr{L}(V, W)$. The dual space adjoint operator $L': W' \to V'$ is defined by

$$(L'w')(v) = w'(Lv), \quad v \in V, w \in W'. \tag{20}$$

As

$$|(L'w')v| \leq \|w'\|_{W'} \|L\|_{\mathscr{L}(V,W)} \|v\|_V \tag{21}$$

we have $L' \in \mathscr{L}(W', V')$ and $\|L'\|_{\mathscr{L}(W',V')} \leq \|L\|_{\mathscr{L}(V,W)}$.

2.2. Sesquilinear forms, inner products, and Hilbert spaces

Definition 2.21 (Sesquilinear Form, Bilinear Form and Inner Product Space). A function $a(\cdot,\cdot): V \times V \to \mathbb{K}$ is called a sesquilinear form if for every u, u_1, u_2, v, v_1, $v_2 \in V$ and ever $\alpha \in \mathbb{K}$ we have

$$a(\alpha\, u, v) = \alpha\, a(u, v), \tag{22}$$

$$a(u, \alpha\, v) = \overline{\alpha}\, a(u, v), \tag{23}$$

$$a(u_1 + u_2, v) = a(u_1, v) + a(u_2, v), \tag{24}$$

$$a(u, v_1 + v_2) = a(u, v_1) + a(u, v_2). \tag{25}$$

If instead of (23), we have

$$a(u, \alpha\, v) = \alpha\, a(u, v), \tag{26}$$

then $a(\cdot,\cdot)$ is called a bilinear form.

A sesquilinear form is called Hermitian if

$$a(u, v) = \overline{a(v, u)}. \tag{27}$$

A bilinear form is called symmetric if

$$a(u, v) = a(v, u). \tag{28}$$

A Hermitian sesquilinear form is called positive definite if

$$a(u,u) \geq 0 \text{ and } a(u,u) = 0 \text{ if and only if } u = 0. \tag{29}$$

A function $(\cdot,\cdot)_V : V \times V \to \mathbb{K}$ is called an inner product if it is a positive definite, Hermitian sesquilinear form. The space V equipped with $(\cdot,\cdot)_V$ is called an inner product space.

Lemma 2.22 (Cauchy–Schwarz Inequality). *Let $(V,(\cdot,\cdot)_V)$ be an inner product space. For ever $u,v \in V$, we have*

$$|(u,v)_V|^2 \leq (u,u)_V\,(v,v)_V. \tag{30}$$

Definition 2.23 (Pre-Hilbert Space and Hilbert Space). An inner product space $(V,(\cdot,\cdot)_V)$ is called a pre-Hilbert space if it is equipped with the norm defined by

$$\|u\|_V = \sqrt{(u,u)_V}. \tag{31}$$

If V is complete with respect to the norm $\|\cdot\|_V$ defined by its inner product, it is called a Hilbert space.

Definition 2.24 (Properties of Sesquilinear Forms). Let $(V,(\cdot,\cdot)_V)$ be a pre-Hilbert space. A sesquilinear form $a(\cdot,\cdot) : V \times V \to \mathbb{K}$ is called

(a) bounded or continuous, if there is a $C \in \mathbb{R}^+$ such that $|a(u,v)| \leq C\,\|u\|_V\,\|v\|_V$ for all $u,v \in V$,
(b) coercive, if there is a $c \in \mathbb{R}^+$ such that $a(u,u) \geq c\,\|u\|_V^2$ for all $u \in V$.

The norm of a bounded sesquilinear form is defined as

$$\|a\|_{\mathscr{L}(V\times V,\mathbb{K})} = \sup_{\|u\|_V\leq 1,\ \|v\|_V\leq 1} |a(u,v)|\,. \tag{32}$$

Lemma 2.25. *Let $(V,(\cdot,\cdot)_V)$ be a pre-Hilbert space. Then the inner product is continuous in both components and the norm satisfies the parallelogram identity*

$$\|u+v\|_V^2 + \|u-v\|_V^2 = 2\left(\|u\|_V^2 + \|v\|_V^2\right) \tag{33}$$

for all $u,v \in V$.

Definition 2.26 (Orthogonality, Orthonormal System). Let $(V,(\cdot,\cdot)_V)$ be a pre-Hilbert space. Two elements u and v of V are orthogonal, if $(u,v)_V = 0$.

A finite or countable set of elements, $\{u_n : n \in N\} \subset V$, $N \subset \mathbb{N}$, is called an orthonormal system if $(u_i,u_k)_V = \delta_{i,k}$ for all $i,k \in \mathbb{N}$.

Theorem 2.27 (Approximation in Pre-Hilbert Spaces, Bessel's Inequality). *Let $(V,(\cdot,\cdot)_V)$ be a pre-Hilbert space and $\{e_k : k = 1,\dots,n\} \subset V$, $n \in \mathbb{N}$ an orthonormal system. For every $u \in V$, the element $a = \sum_{k=1}^n (u,e_k)_V\,e_k$ satisfies $\|u-a\|_V \leq \|u-v\|_V$ for all $v \in \operatorname{span}\{e_k : k = 1,\dots,n\}$.*

The coefficients $(u,e_k)_V$ are called Fourier coefficients of u with respect to $\{u_k : k = 1,\dots,n\}$.

Moreover, from $\|u-a\|_V \geq 0$ *follows Bessel's inequality*

$$\sum_{k=1}^{n} |(u,e_k)_V|^2 \leq \|u\|_V^2 . \tag{34}$$

Theorem 2.28 (Orthonormal Hilbert Basis). *Let* $(V,(\cdot,\cdot)_V)$ *be a pre-Hilbert space and* $\{e_k : k \in \mathbb{N}\} \subset V$ *an orthonormal system. This orthonormal system is a Hilbert basis if it satisfies one of the following equivalent properties:*

(a) $\operatorname{span}(\{e_k\}_{k\in\mathbb{N}})$ *is dense in* V.
(b) *The system* $\{e_k : k \in \mathbb{N}\}$ *is closed, i.e., for all* $u \in V$, *there is a unique sequence* $(a_k)_{k\in\mathbb{N}}$ *in* $\mathbb{K}$, *such that* $u = \lim\limits_{n\to\infty} \sum\limits_{k=1}^{n} a_k e_k$.
(c) $\{e_k : k \in \mathbb{N}\}$ *is complete, i.e.,* $u = \sum\limits_{k=1}^{\infty} (u,e_k)_V \, e_k$ *for all* $u \in V$.
(d) *The Parseval identity holds for* $\{e_k : k \in \mathbb{N}\}$, *i.e., for all* $u \in V$ *we have* $\|u\|_V^2 = \sum\limits_{k=1}^{\infty} |(u,e_k)_V|^2$
(e) *The extended Parseval identity holds for* $\{e_k : k \in \mathbb{N}\}$, *i.e., for all* $u, v \in V$ *we have* $(u,v)_V = \sum\limits_{k=1}^{\infty} (u,e_k)_V \, \overline{(v,e_k)_V}$.

If $(V,(\cdot,\cdot)_V)$ *is a Hilbert space, the above properties are equivalent to the property that there exists no* $0 \neq u \in V$ *such that* $(u,e_k)_V = 0$ *for all* $k \in \mathbb{N}$.

Theorem 2.29 (Existence of an Orthonormal Basis). *Every separable Hilbert space contains an orthonormal basis.*

Remark 2.30. If $\{u_k : k \in N\} \subset V$, $N \subset \mathbb{N}$, is an orthonormal system in a separable Hilbert space V which is not complete, we can find a system $\{v_k : k \in \mathbb{N} \setminus N\} \subset V$ such that both systems together form an orthonormal basis.

Theorem 2.31 (Riesz Representation Theorem). *Let* $(V,(\cdot,\cdot)_V)$ *be a Hilbert space. For every bounded linear functional* $f \in V'$ *exists a unique element* $w \in V$ *such that* $f(v) = (v,w)_V$ *for every* $v \in V$.

The function $J : V \to V'$, $w \mapsto J(w) = (\cdot,w)_V$ *is an isometric, conjugate linear isomorphism. Consequently,* $\|f\|_{V'} = \|w\|_V$.

Remark 2.32. With the extended Parseval identity from Theorem 2.28, we get in separable Hilbert spaces V

$$f(v) = (v,w)_V = \sum_{k=1}^{\infty} (v,e_k)_V \, \overline{(w,e_k)_V} = \sum_{k=1}^{\infty} (v,e_k)_V \, (e_k,w)_V = \sum_{k=1}^{\infty} (v,e_k)_V \, f(e_k) \tag{35}$$

and thus

$$w = \sum_{k=1}^{\infty} \overline{f(e_k)} \, e_k . \tag{36}$$

Definition 2.33 (Hilbert Space Adjoint Operator). Let $(V, (\cdot,\cdot)_V)$, $(W, (\cdot,\cdot)_W)$ be Hilbert spaces, J_V and J_W the corresponding Riesz isomorphisms according to Theorem 2.31 and $L \in \mathscr{L}(V, W)$. The Hilbert space adjoint operator L^* of L is defined by

$$L^* = J_V^{-1}\, L'\, J_W, \tag{37}$$

whereas L' is the dual space adjoint operator to L according to Definition 2.20.

L is called self-adjoint, if $L^* = L$.

Lemma 2.34. *Let $(V, (\cdot,\cdot)_V)$, $(W, (\cdot,\cdot)_W)$ be Hilbert spaces and $L \in \mathscr{L}(V, W)$. Then we have for the Hilbert space adjoint operator $L^* \in \mathscr{L}(W, V)$ and*

$$(v, L^* w)_V = (Lv, w)_W. \tag{38}$$

Moreover, we have $L^{**} = L$, $\|L^*\|_{\mathscr{L}(W,V)} = \|L\|_{\mathscr{L}(V,W)}$ and $\|L^*L\|_{\mathscr{L}(V,V)} = \|L\|^2_{\mathscr{L}(V,W)}$.

Theorem 2.35 (Lax–Milgram Theorem). *Let $(V, (\cdot,\cdot)_V)$ be a Hilbert space and $a(\cdot,\cdot) : V \times V \to \mathbb{K}$ a sesquilinear form. If $a(\cdot,\cdot)$ is continuous, i.e., bounded with the constant $C \in \mathbb{R}^+$, then there exists a unique bounded linear operator $A \in \mathscr{L}(V, V)$ such that*

$$a(u, v) = (u, Av)_V \quad \text{for all } u, v \in V. \tag{39}$$

We have $\|A\|_{\mathscr{L}(V,V)} \le C$ or, more precisely, $\|A\|_{\mathscr{L}(V,V)} = \|a(\cdot,\cdot)\|_{\mathscr{L}(V\times V,\mathbb{K})}$.

If $a(\cdot,\cdot)$ is also coercive with coercivity constant $c \in \mathbb{R}^+$, then A is bijective, i.e., there exists a bounded linear operator $A^{-1} \in \mathscr{L}(V, V)$ which is the inverse operator to A and $\left\|A^{-1}\right\|_{\mathscr{L}(V,V)} \le \frac{1}{c}$.

Corollary 2.36. *Let $(V, (\cdot,\cdot)_V)$ be a Hilbert space and $a(\cdot,\cdot) : V \times V \to \mathbb{K}$ a continuous, coercive sesquilinear form, A the corresponding operator according to Theorem 2.35 and $J : V \to V'$ the Riesz isomorphism according to Theorem 2.31. For $f \in V'$ let $u = A^{-1}J^{-1}f$.*

The so-defined u is the unique solution to the problem

$$a(u, v) = f(v) \quad \text{for all } v \in V \tag{40}$$

and satisfies

$$\|u\|_V \le \frac{1}{c}\, \|f\|_{V'} \tag{41}$$

with $c \in \mathbb{R}^+$ being the coercivity constant to the sesquilinear form.

If, additionally, $a(\cdot,\cdot)$ is Hermitian, i.e., $a(\cdot,\cdot)$ is an inner product, then u is also the unique minimizer of the functional

$$V \ni v \mapsto F(v) = \frac{1}{2} a(v, v) - \operatorname{Re}(f(v)) \ \in \mathbb{R}. \tag{42}$$

2.3. Weak and weak* convergence

Definition 2.37 (Duality). Let V, W be normed spaces. A bilinear form $\langle \cdot, \cdot \rangle : V \times W \to \mathbb{K}$ is called duality, if

(a) for all $0 \neq v \in V$ exists a $w \in W$ such that $\langle v, w \rangle \neq 0$ and
(b) for all $0 \neq w \in W$ exists a $v \in V$ such that $\langle v, w \rangle \neq 0$.

The most common duality is the duality between a linear space V and its dual space V'. For a functional $v' \in V'$ and a $v \in V$ this duality is defined by $\langle v', v \rangle = v'(v)$.

Definition 2.38 (Weak and Weak* Convergence and Compactness). Let V be a Banach space.

(a) A sequence $(v_n)_{n\in\mathbb{N}}$ in V converges weakly to $v \in V$, also written

$$v_n \xrightarrow{n\to\infty} v,$$

if and only if

$$\lim_{n\to\infty} \langle v_n, v' \rangle = \langle v, v' \rangle \quad \text{for all } v' \in V'. \tag{43}$$

(b) A sequence $(v'_n)_{n\in\mathbb{N}}$ in V' converges weakly* to $v' \in V'$, also written

$$v'_n \xrightarrow{n\to\infty}{}^{\star} v',$$

if and only if

$$\lim_{n\to\infty} \langle v, v'_n \rangle = \langle v, v' \rangle \quad \text{for all } v \in V. \tag{44}$$

(c) Weak and weak* Cauchy sequences are defined correspondingly.
(d) A subset $U \subset V$ is called weak sequentially compact if and only if each sequence in U possesses a weak convergent subsequence whose weak limit is also in U.
(e) Weak* compactness for subsets of V' is defined analogously.
(f) Convergence with respect to the norm is subsequently called strong convergence.

Lemma 2.39. *Let V be a Banach space.*

(a) *Weak and weak* limits are unique and weakly or weakly* convergent sequences are bounded in the norms of the corresponding spaces.*
(b) *There exists an isometric mapping $\widetilde{J} \in \mathscr{L}(V, V'')$ that can be defined by $\left\langle v', \widetilde{J}v \right\rangle = \langle v, v' \rangle$ for any $v \in V$ and every $v' \in V'$. V is called reflexive if and only if $\widetilde{J}$ is surjective.*
(c) *A sequence $(v_n)_{n\in\mathbb{N}}$ in V converges weakly to $v \in V$ if and only if $\left(\widetilde{J}v_n\right)_{n\in\mathbb{N}}$ converges weakly* to $\widetilde{J}v \in V''$.*
(d) *From $v_n \xrightarrow{n\to\infty} v$ follows $\|v\|_V \leq \liminf_{n\to\infty} \|v_n\|_V$ and from $v'_n \xrightarrow{n\to\infty}{}^{\star} v'$ follows $\|v'\|_{V'} \leq \liminf_{n\to\infty} \|v'_n\|_{V'}$.*

Remark 2.40.

(i) If V is a Hilbert space, the Riesz representation Theorem 2.31 implies that $v_n \xrightarrow{n\to\infty} v$ means $\lim_{n\to\infty} (v_n, u) = (v, u)$ for all $u \in V$.
(ii) Strong convergence implies weak or weak* convergence, respectively.
(iii) In the dual space V' of a Banach space V, we now have three different concepts of convergence:
   (a) strong convergence: $v'_n \xrightarrow{n\to\infty} v'$, i.e., $\lim_{n\to\infty} \|v'_n - v'\|_{V'} = 0$,
   (b) weak* convergence: $v'_n \xrightarrow{n\to\infty}{}^\star\, v'$, i.e., $\lim_{n\to\infty}\langle v, v'_n\rangle = \langle v, v'\rangle$ for all $v \in V$, and
   (c) weak convergence: $v'_n \xrightarrow{n\to\infty} v'$, i.e., $\lim_{n\to\infty}\langle v'_n, v''\rangle = \langle v', v''\rangle$ for all $v'' \in V''$.

   However, as V is always isomorph to at least a subset of V'' due to the mapping constructed in Lemma 2.39, weak convergence in V' always implies weak* convergence in V'. If V is reflexive, both concepts are identical on V'.
(iv) In a finite-dimensional normed space, strong and weak convergence coincide.

Lemma 2.41. *Let V be a Banach space, $W \subset V$ dense, $Z' \subset V'$ dense, $v \in V$, $v' \in V'$, $(v_n)_{n\in\mathbb{N}}$ in V and $(v'_n)_{n\in\mathbb{N}}$ in V'.*

(a) *$v_n \xrightarrow{n\to\infty} v$ if and only if there is a $C \in \mathbb{R}^+$ such that $\|v_n\|_V \le C$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty}\langle v_n, v'\rangle = \langle v, v'\rangle$ for all $v' \in Z'$.*

(b) *$v'_n \xrightarrow{n\to\infty}{}^\star\, v'$ if and only if there is a $C' \in \mathbb{R}^+$ such that $\|v'_n\|_{V'} \le C'$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty}\langle v, v'_n\rangle = \langle v, v'\rangle$ for all $v \in W$.*

Lemma 2.42 (Mazur). *Let V be a normed space and $(v_n)_{n\in\mathbb{N}}$ a sequence in V with $v_n \xrightarrow{n\to\infty} v$. For every $\varepsilon \in \mathbb{R}^+$ exists a linear combination $u = \sum_{k=1}^{N} a_k v_k$, $N \in \mathbb{N}$, $\{a_k\}_{k=1}^N \subset \mathbb{R}$, $a_k \ge 0$ for all k, $\sum_{k=1}^{N} a_k = 1$ such that $\|u - v\|_V \le \varepsilon$.*

3. Function spaces

In order to deal with differential equations, we have to introduce some notation for differentiation and integration. As it turns out, the classical strong concept of differentiability is too restrictive. This leads to the definition of weak differentiability. Different kinds of requirements on the differentiability of functions yield different sets of functions which can be shown to be normed vector spaces. There are some more remarkable properties of these spaces and the functions they contain as well as interesting and useful relations between them.

Definition 3.1 ((Strong) Derivative). Let $\mathcal{G}$ be a bounded open subset of $\mathbb{R}^n$, $n \in \mathbb{N}$, $u : \mathcal{G} \to \mathbb{R}$, $\gamma \in \mathbb{N}_0^n$, and $k \in \mathbb{N}_0$. Let x be a point in $\mathbb{R}^n$ with coordinates x_i, $i \in \mathbb{N}$, $i \le n$. Throughout this section, these are cartesian coordinates.

The partial derivative $D^\gamma u$ is defined by

$$(D^\gamma u)(x) = \left(\partial_{x_1}^{\gamma_1} \dots \partial_{x_n}^{\gamma_n} u\right)(x) = \left(\frac{\partial^{|\gamma|} u}{\partial x_1^{\gamma_1} \dots \partial x_n^{\gamma_n}}\right)(x), \quad x \in \mathcal{G}\,, \tag{45}$$

with $|\gamma| = \sum_{i=1}^n \gamma_i$ being the order of the derivative. The set of all derivatives of u of order k at point x is denoted by $(D^k u)(x) = \{(D^\gamma u)(x) : \ |\gamma| = k\}$.

Definition 3.2 (Region). A subset $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$ is called a region, if it is open and connected. Here, connected means that for any two points $x, y \in \mathcal{G}$, there exists a continuous function $f : [0,1] \to \mathcal{G}$ such that $f(0) = x$ and $f(1) = y$.

Remark 3.3.

(i) If u is defined as a function on $\mathbb{R}^n$, $n \in \mathbb{N}$, such that $u : \ \mathbb{R}^n \to \mathbb{R}$, the restriction of u to $\mathcal{G} \subset \mathbb{R}^n$ is denoted by $u|_{\mathcal{G}}$.

(ii) The gradient of a differentiable function $u : \ \mathbb{R}^n \to \mathbb{R}$ with respect to the variable x is defined as the vector of all first derivatives and denoted by

$$\nabla_x u = (\partial_{x_1} u, \dots, \partial_{x_n} u)^T\,. \tag{46}$$

We omit the index x if it is clear with respect to which variable the differentiation has to be carried out.

(iii) The directional derivative of u with respect to a unit vector e is given by $(\nabla_x u) \cdot e$. The directional derivative with respect to the outer unit normal vector of a bounded region $\mathcal{G}$ is denoted by $\partial_n u$.

(iv) The divergence $\nabla_x \cdot$ of a differentiable vector field $h : \mathcal{G} \to \mathbb{R}^n$, $n \in \mathbb{N}$, with respect to the variable x is the scalar value

$$\nabla_x \cdot h = \sum_{k=1}^n \frac{\partial h_k}{\partial x_k}. \tag{47}$$

(v) If $\nabla_x f$ is differentiable, the Laplace operator $\Delta_x f$ with respect to the variable x is given by

$$\Delta_x f = \nabla_x \cdot (\nabla_x f) = \sum_{k=1}^n \frac{\partial^2 f}{\partial x_k^2}. \tag{48}$$

For vector-valued functions, the Laplace operator is defined component-wise.

As we now have introduced the notation for strong derivatives, we can define function spaces of continuously differentiable functions (see, e.g., [1]).

Definition 3.4 (Spaces of Continuously Differentiable Functions). Let $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, be a region. For $k \in \mathbb{N}_0$, we denote by $\mathrm{C}^k(\mathcal{G})$ the vector space of all functions $u : \mathcal{G} \to \mathbb{R}$ which together with all their derivatives $D^\gamma u$ of order $|\gamma| \le k$ are continuous on $\mathcal{G}$. We write $\mathrm{C}(\mathcal{G})$ for $\mathrm{C}^0(\mathcal{G})$. The space of infinitely continuously differentiable functions is given by $\mathrm{C}^\infty(\mathcal{G}) = \bigcap_{k=0}^\infty \mathrm{C}^k(\mathcal{G})$.

The subspace of functions in $\mathrm{C}^k(\mathcal{G})$ that have compact support in $\mathcal{G}$ is denoted by $\mathrm{C}_c^k(\mathcal{G})$. A function u has compact support in $\mathcal{G}$ if there is a compact set $K \subset \mathcal{G}$ such that

$$\mathrm{supp}(u) = \overline{\{x \in \mathcal{G} : \ u(x) \neq 0\}} \subset K\,. \tag{49}$$

The spaces $\mathrm{C}^k(\overline{\mathcal{G}})$ contain all functions $u \in \mathrm{C}^k(\mathcal{G})$ for which $D^\gamma u$ is bounded and uniformly continuous for all $\gamma \in \mathbb{N}_0^k$ with $0 \leq |\gamma| \leq k$, i.e., it possesses a unique, bounded, continuous extension to the closure $\overline{\mathcal{G}}$. These spaces are Banach spaces when equipped with the norm

$$\|u\|_{\mathrm{C}^k(\overline{\mathcal{G}})} = \max_{0 \leq |\gamma| \leq k} \sup_{x \in \mathcal{G}} |(D^\gamma u)(x)| \ . \tag{50}$$

In some cases, functions are required to be more regular than just being continuous, but requiring them to be continuously differentiable would be too much. Thus, we introduce the spaces of Hölder-continuous functions ([1]).

Definition 3.5 (Hölder-Continuous Functions). Let $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, be a region, $\gamma \in \mathbb{N}_0^k$, and $k \in \mathbb{N}_0$. The space $\mathrm{C}^{k,s}(\overline{\mathcal{G}})$, $0 < s \leq 1$, is the subspace of functions $u \in \mathrm{C}^k(\overline{\mathcal{G}})$ whose derivatives $D^\gamma u$ of order k satisfy

$$|(D^\gamma u)(x) - (D^\gamma u)(y)| \leq C \, \|x - y\|^s \quad \text{for all } x, y \in \mathcal{G} \tag{51}$$

with a constant $C \in \mathbb{R}^+$. We say u has Hölder-continuous derivatives of order k with Hölder exponent s or, in the special case $s = 1$, u has Lipschitz-continuous derivatives of order k. $\mathrm{C}^{k,s}(\overline{\mathcal{G}})$ is a Banach space if equipped with the norm

$$\|u\|_{\mathrm{C}^{k,s}(\overline{\mathcal{G}})} = \|u\|_{\mathrm{C}^k(\overline{\mathcal{G}})} + \max_{0 \leq |\gamma| \leq k} \sup_{\substack{x,y \in \mathcal{G} \\ x \neq y}} \frac{|(D^\gamma u)(x) - (D^\gamma u)(y)|}{\|x - y\|^s} \ . \tag{52}$$

For $r \geq s > 0$, the inclusion $\mathrm{C}^{k,r}(\overline{\mathcal{G}}) \subset \mathrm{C}^{k,s}(\overline{\mathcal{G}})$ is valid.

If a function u satisfies

$$\lim_{\delta \to 0} \sup_{\substack{x,y \in \mathcal{G} \\ 0 < \|x-y\| < \delta}} \frac{|(D^\gamma u)(x) - (D^\gamma u)(y)|}{\|x - y\|^s} < \infty \tag{53}$$

for each $|\gamma| = k$, the derivatives of order k of u are called uniformly Hölder-continuous and u is an element of the space $\mathrm{C}_u^{k,s}(\overline{\mathcal{G}})$.

When looking for functions whose values on the boundary of a region are prescribed, it is often useful, if not even necessary, to restrict the kind of region under consideration to answer questions of existence and uniqueness. The region is required to have some kind of regularity. In order to give different kinds of regularity properties, we need another definition ([1]).

Definition 3.6 (m-smooth Transformation). Let Φ be a one-to-one transformation of a region $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, onto a region $G \subset \mathbb{R}^n$ with $\Psi = \Phi^{-1}$. We call Φ m-smooth if, writing $y = \Phi(x)$ and

$$\begin{aligned} y_1 &= \phi_1(x_1, \ldots, x_n), & x_1 &= \psi_1(y_1, \ldots, y_n), \\ y_2 &= \phi_2(x_1, \ldots, x_n), & x_2 &= \psi_2(y_1, \ldots, y_n), \\ &\vdots & &\vdots \\ y_n &= \phi_n(x_1, \ldots, x_n), & x_n &= \psi_n(y_1, \ldots, y_n), \end{aligned} \tag{54}$$

the functions $\phi_1, \ldots, \phi_n$ belong to $\mathrm{C}^m(\overline{\mathcal{G}})$ and the functions $\psi_1, \ldots, \psi_n$ belong to $\mathrm{C}^m(\overline{G})$.

The following summary of regularity conditions is taken from the book by Adams [1] and is just slightly adapted. Here, a finite cone $\mathscr{C}_x$ with vertex $x \in \mathbb{R}^n$ is defined as the set

$$\mathscr{C}_x = B_{r_1}(x) \cap \{x + C(y - x) : \ y \in B_{r_2}(z)\} \tag{55}$$

with $B_{r_2}(z)$ being a ball around $z \in \mathbb{R}^n$ such that $x \notin B_{r_2}(z)$ and $r_1, r_2, C \in \mathbb{R}^+$. Two cones $\mathscr{C}_x$ and $\mathscr{C}_{\widetilde{x}}$ are called congruent if there exists an isometry $f : \mathbb{R}^n \to \mathbb{R}^n$ with $f(\mathscr{C}_x) = f(\mathscr{C}_{\widetilde{x}})$.

Given an index set J, a collection of sets $\{U_j : \ j \in J\}$ is called a cover of a set $\mathcal{G}$ if

$$\mathcal{G} \subset \bigcup_{j \in J} U_j. \tag{56}$$

A cover is called open cover if all sets $\{U_j : \ j \in J\}$ are open. An open cover is said to be locally finite if any compact set in $\mathbb{R}^n$ can intersect at most finitely many elements of $\{U_j : \ j \in J\}$ ([1]). Locally finite collections of sets are countable. Thus, we can assume $J \subset \mathbb{N}$.

Definition 3.7 (Regularity of Domains). Let $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, be a region. $\mathcal{G}$ has

(i) the segment property if there exists a locally finite open cover $\{U_j : \ j \in J\}$ of $\partial\mathcal{G}$ and a corresponding set $\{y_j : \ j \in J\}$ of non-zero vectors such that if $x \in \overline{\mathcal{G}} \cap U_j$ for some j, then $x + \varepsilon y_j \in \mathcal{G}$ for $0 < \varepsilon < 1$;

(ii) the cone property if there exists a finite cone $\mathscr{C}$ such that each point $x \in \mathcal{G}$ is the vertex of a finite cone $\mathscr{C}_x$ contained in $\mathcal{G}$ and congruent to $\mathscr{C}$;

(iii) the uniform cone property if there exists a locally finite open cover $\{U_j : j \in J\}$ of $\partial\mathcal{G}$ and a corresponding set $\{\mathscr{C}_j : j \in J\}$ of finite cones, each congruent to some fixed finite cone $\mathscr{C}$, such that
   (a) for some finite $M \in \mathbb{R}^+$, every U_j has a diameter less than M,
   (b) for some $\delta > 0$, $\bigcup_{j=1}^{\infty} U_j \supset \{x \in \mathcal{G} : \ \mathrm{dist}(x, \partial\mathcal{G}) < \delta\}$,
   (c) for every j, $Q_j = \bigcup_{x \in \mathcal{G} \cap U_j} (x + \mathscr{C}_j) \subset \mathcal{G}$,
   (d) for some finite $N \in \mathbb{N}$, every collection of $N + 1$ of the sets Q_j has an empty intersection;

(iv) the strong local Lipschitz property if there exist positive numbers δ and M, a locally finite open cover $\{U_j : \ j \in J\}$ of $\partial\mathcal{G}$, and for each U_j a real-valued function f_j of $n - 1$ real variables, such that
   (a) for some finite $N \in \mathbb{N}$, every collection of $N + 1$ of the sets U_j has an empty intersection,
   (b) for every pair of points $x, y \in \{z \in \mathcal{G} : \ \mathrm{dist}(z, \partial\mathcal{G}) < \delta\}$ such that $\|x - y\| < \delta$, there exists j such that $x, y \in \{z \in U_j : \ \mathrm{dist}(z, \partial U_j) > \delta\}$,
   (c) each function f_j satisfies a Lipschitz condition with constant M,
   (d) for some cartesian coordinate system $(x_{j,l})_{l=1}^n$ in U_j, the set $\mathcal{G} \cap U_j$ is represented by the inequality $x_{j,n} < f_j(x_{j,1}, \ldots, x_{j,n-1})$;

(v) the uniform C^m-regularity property if there exists a locally finite open cover $\{U_j : j \in J\}$ of $\partial\mathcal{G}$ and a corresponding set $\{\Phi_j : j \in J\}$ of m-smooth one-to-one transformations with Φ_j taking U_j onto $B_1(0) \subset \mathbb{R}^n$, such that
   (a) for some $\delta > 0$, $\bigcup_{j=1}^{\infty} \Psi_j(B_{0.5}(0)) \supset \{x \in \mathcal{G} : \operatorname{dist}(x, \partial\mathcal{G}) < \delta\}$, where $\Psi = \Phi^{-1}$,
   (b) for some finite $N \in \mathbb{N}$, every collection of $N+1$ of the sets U_j has an empty intersection,
   (c) for each j, $\Phi_j(U_j \cap \mathcal{G}) = \{y \in B_1(0) : y_n > 0\}$,
   (d) if $(\phi_{j,1}, \ldots, \phi_{j,n})$ and $(\psi_{j,1}, \ldots, \psi_{j,n})$ denote the components of Φ_j and Ψ_j, respectively, then there exists a finite M such that for all $\gamma \in \mathbb{N}_0^n$, $|\gamma| \leq m$, for every $1 \leq i \leq n$, and for every j, we have $|D^\gamma \phi_{j,i}(x)| \leq M$, $x \in U_j$, and $|D^\gamma \psi_{j,i}(y)| \leq M$, $y \in B_1(0)$.

For the different kinds of regularity, we have (v) $\overset{m \geq 1}{\Longrightarrow}$ (iv) $\Longrightarrow$ (iii) $\Longrightarrow$ (i). These regularity properties require $\mathcal{G}$ to lie on only one side of its boundary, whereas the cone property does not impose this condition.

Remark 3.8.

(i) If $\mathcal{G}$ is bounded, the requirements for $\mathcal{G}$ being strong local Lipschitz reduce to the condition that for each point $x \in \partial\mathcal{G}$, there exists a neighborhood U of x such that $U \cap \partial\mathcal{G}$ is the graph of a Lipschitz-continuous function.

(ii) In some cases it is necessary to require that the parts of the one-to-one transformation mentioned in the definition of the C^m-regularity property have not only bounded derivatives, but Hölder-continuous ones. This yields the $\mathrm{C}^{m,s}$-regularity property.

As already mentioned, the above introduced definition of strong differentiability with continuous or even Hölder-continuous derivatives is often too restrictive. Therefore, we need some other, weaker definition of derivatives. To define these weak derivatives, we need a definition of convergence in $\mathrm{C}_c^\infty(\mathcal{G})$ first (see, e.g., [1]).

Definition 3.9 (Convergence in $\mathrm{C}_c^\infty(\mathcal{G})$). Let $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, be a bounded region, $(\phi_l)_{l \in \mathbb{N}}$ in $\mathrm{C}_c^\infty(\mathcal{G})$, and $\phi_0 \in \mathrm{C}_c^\infty(\mathcal{G})$. The sequence $(\phi_l)_{l \in \mathbb{N}}$ is said to converge towards ϕ_0 in $\mathrm{C}_c^\infty(\mathcal{G})$ for $l \to \infty$ if there is a compact subset $K \subset \mathcal{G}$ such that

$$\operatorname{supp}(\phi_l) \subset K \quad \text{for all } l \in \mathbb{N}\,, \tag{57}$$

$$\operatorname{supp}(\phi_0) \subset K\,, \tag{58}$$

as well as all partial derivatives of ϕ_l of arbitrary order converge uniformly to those of ϕ_0, i.e.,

$$\sup_{x \in \mathcal{G}} |(D^\gamma \phi_l)(x) - D^\gamma \phi_0(x)| \xrightarrow{l \to \infty} 0 \quad \text{for all } \gamma \in \mathbb{N}_0^n\,. \tag{59}$$

Remark 3.10.

(i) $\mathrm{C}_c^\infty(\mathcal{G})$ is often denoted by $\mathscr{D}(\mathcal{G})$ and called the space of test functions, although the latter identification is not unique. It is a topological vector space, but not normable ([1, 17]).

(ii) A function ϕ defined on $\mathbb{R}^n$ is called finite if it vanishes outside a bounded set. The space $C_c^\infty(\mathbb{R}^n)$ consist of all finite, arbitrary often continuously differentiable functions on $\mathbb{R}^n$.

The above definition allows us to define distributions (see [16]).

Definition 3.11 (Distribution). Let $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, be a bounded region. A distribution (or generalized function) is a linear functional $f : \mathrm{C}_c^\infty(\mathcal{G}) \to \mathbb{R}$ which is continuous in the following sense: If a sequence $(\phi_l)_{l\in\mathbb{N}_0}$ in $\mathrm{C}_c^\infty(\mathcal{G})$ converges for $l \to \infty$ towards $\phi \in \mathrm{C}_c^\infty(\mathcal{G})$, then $f(\phi_l) = \langle f, \phi_l\rangle$ converges for $l \to \infty$ towards $f(\phi)$. The set of all distributions is denoted by $(\mathrm{C}_c^\infty(\mathcal{G}))'$.
The space of vector-valued distributions is defined accordingly.

Remark 3.12. The space of distributions is often denoted by $\mathscr{D}'(\mathcal{G})$. If we consider $\mathrm{C}_c^\infty(\mathcal{G})$ as a topological vector space, $(\mathrm{C}_c^\infty(\mathcal{G}))'$ is its topological dual ([1, 17]).

There is another way to characterize functions that is useful to present here in anticipation of a more general concept that we introduce later on. For this, we have to explain our interpretation of the integral of a function.

Within this chapter, all integrals are understood in the sense of Lebesgue integrals. In the following, we denote by V_n the Lebesgue measure on a given measurable (e.g., open or closed) subset of $\mathbb{R}^n$. If there is no confusion, we omit the index n in $\mathrm{d}V_n$. A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, \infty\}$ is called Lebesgue-measurable if the set $\{x : x \in \mathbb{R}^n, f(x) > c\}$ is measurable for arbitrary $c \in \mathbb{R}$ (see [4]). We also have to define when a function is integrable. Following [4], we consider a measurable set $A \in \mathbb{R}^n$ and introduce step functions $s = \sum_{k=1}^m a_k \chi_{A_k}$ with coefficients $a_k \in \mathbb{R}$ and χ_{A_k} the indicator function of $A_k \subset A$, given by

$$\chi_{A_k}(x) = \begin{cases} 1, x \in A_k, \\ 0, x \notin A_k. \end{cases} \tag{60}$$

For step functions, we can define

$$\int_A s(x)\, \mathrm{d}V_n(x) = \sum_{k=1}^m a_k V_n(A_k). \tag{61}$$

This allows us to define the integral for measurable positive functions $f : \mathbb{R}^n \to \mathbb{R}^+ \cup \{\infty\}$ as

$$\int_A f(x)\, \mathrm{d}V_n(x) = \sup \int_A s(x)\, \mathrm{d}V_n(x) \tag{62}$$

whereas the supremum is taken over all step functions s which vanish outside A and satisfy $0 \le s \le f$ inside A. For measurable functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, \infty\}$, we introduce

$$f^+(x) = \max(f(x), 0), \tag{63}$$

$$f^-(x) = \max(-f(x), 0). \tag{64}$$

If at least one of the integrals $\int_A f^+(x)\ \mathrm{d}V_n(x), \int_A f^-(x)\ \mathrm{d}V_n(x)$ is finite, the integral of f exists and is given by

$$\int_A f(x)\ \mathrm{d}V_n(x) = \int_A f^+(x)\ \mathrm{d}V_n(x) - \int_A f^-(x)\ \mathrm{d}V_n(x) \tag{65}$$

with values in $\mathbb{R} \cup \{-\infty, \infty\}$. A measurable function is called integrable if its integral exists and takes values in $\mathbb{R}$.

Now, we can define ([1])

Definition 3.13 (Locally Integrable Functions). Let $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, be a region. A function u is called locally integrable on $\mathcal{G}$ if for every compact subset $K \subset \mathcal{G}$ we have

$$\int_K |f(x)|\ \mathrm{d}V_n(x) < \infty. \tag{66}$$

For every locally integrable function u, we can define a corresponding distribution $T_u \in (\mathrm{C}_c^\infty(\mathcal{G}))'$ simply by

$$T_u(\phi) = \int_{\mathcal{G}} u(x)\phi(x)\ \mathrm{d}V(x), \quad \phi \in \mathrm{C}_c^\infty(\mathcal{G})\ . \tag{67}$$

Usually, notation is a little bit abused by also using u instead of T_u to denote the corresponding distribution. Distributions that correspond in that way to a locally integrable function are called regular.

There are many distributions for which no corresponding locally integrable function can be found. The most prominent example is the evaluation of a function ϕ at a certain point x, known as Dirac's delta distribution. If $0 \in \mathcal{G}$, the evaluation of a function $\phi \in \mathrm{C}_c^\infty(\mathcal{G})$ is given by

$$\delta(\phi) = \phi(0). \tag{68}$$

It is easy to prove that there is no locally integrable function for which

$$\int_{\mathcal{G}} \delta(x)\phi(x)\ \mathrm{d}V(x) = \phi(0), \quad \phi \in \mathrm{C}^\infty(\mathcal{G})\ . \tag{69}$$

However, δ satisfies Definition 3.11.

It is obvious how addition of two distributions and multiplication with a constant should be defined on distributions. Distributions may even be multiplied by smooth functions ([1]). For $T \in (\mathrm{C}_c^\infty(\mathcal{G}))'$ and $u \in \mathrm{C}^\infty(\mathcal{G})$, the product $uT \in (\mathrm{C}_c^\infty(\mathcal{G}))'$ is defined by

$$(uT)(\phi) = T(u\phi), \quad \phi \in \mathrm{C}_c^\infty(\mathcal{G})\ . \tag{70}$$

The support of a distribution is defined as follows ([4]).

Definition 3.14 (Support of a Distribution). Suppose $T \in (\mathrm{C}_c^\infty(\mathcal{G}))'$ for an open bounded region $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$. The support of T is defined as

$$\operatorname{supp}(T) \coloneqq \left\{x \in \overline{\mathcal{G}} :\ \forall \delta > 0 :\ T|_{\mathcal{G} \cap B_\delta(x)} \neq 0\right\}. \tag{71}$$

The support of a distribution as defined in Definition 3.14 is closed (see [4]) and the concept of compact support is directly transfered to distributions.

Another operation which we need not only on distributions is convolution ([17]).

Definition 3.15 (Convolution). Let u be a function defined on $\mathbb{R}^n$, $n \in \mathbb{N}$, and $x, y \in \mathbb{R}^n$. We define

$$(\mathsf{T}_x u)(y) = u(y - x) \,, \tag{72}$$

$$\breve{u}(y) = u(-y) \,, \tag{73}$$

$$(\mathsf{T}_x \breve{u})(y) = u(x - y) \,. \tag{74}$$

Let v be another function on $\mathbb{R}^n$. The convolution $u * v$ is defined as

$$(u * v)(x) = \int_{\mathbb{R}^n} u(y) v(x - y) \, \mathrm{d}V(y) = \int_{\mathbb{R}^n} u(y) (\mathsf{T}_x \breve{v})(y) \, \mathrm{d}V(y) \tag{75}$$

if the integral exists for V-almost all $x \in \mathbb{R}^n$, i.e., the set N of all points x for which the integral does not exist has vanishing measure, i.e., $V(N) = 0$.

For a distribution $u \in (\mathrm{C}_c^\infty(\mathbb{R}^n))'$ and $\phi \in \mathrm{C}_c^\infty(\mathbb{R}^n)$, the function $u * \phi$ is defined by

$$(u * \phi)(x) \coloneqq u(\mathsf{T}_x \breve{\phi}). \tag{76}$$

Theorem 3.16 (Properties of Convolutions). *Let $u \in (\mathrm{C}_c^\infty(\mathbb{R}^n))'$, $\phi, \psi \in \mathrm{C}_c^\infty(\mathbb{R}^n)$. Then the following holds:*

(i) $\mathsf{T}_x(u * \phi) = (\mathsf{T}_x u) * \phi = u * (\mathsf{T}_x \phi)$ *for all* $x \in \mathbb{R}^n$*;*
(ii) $u * \phi \in \mathrm{C}^\infty(\mathbb{R}^n)$ *and* $u * (\phi * \psi) = (u * \phi) * \psi$*;*
(iii) *the operator L, defined by*

$$L\phi = u * \phi, \quad \phi \in \mathrm{C}_c^\infty(\mathbb{R}^n) \,, \tag{77}$$

is a linear mapping of $\mathrm{C}_c^\infty(\mathbb{R}^n)$ into $\mathrm{C}^\infty(\mathbb{R}^n)$ which satisfies $\mathsf{T}_x L = L \mathsf{T}_x$, $x \in \mathbb{R}^n$.

Until now, convolutions for distributions are only declared if a distribution is convolved with an element of $\mathrm{C}_c^\infty(\mathbb{R}^n)$. The next lemma extends convolution to elements of $\mathrm{C}^\infty(\mathbb{R}^n)$ ([17]).

Lemma 3.17. *Let $u \in (\mathrm{C}_c^\infty(\mathbb{R}^n))'$ have compact support, $\phi \in \mathrm{C}^\infty(\mathbb{R}^n)$, $\psi \in \mathrm{C}_c^\infty(\mathbb{R}^n)$. The convolution $u * \phi \in \mathrm{C}^\infty(\mathbb{R}^n)$ is well defined. Moreover,*

(i) $\mathsf{T}_x(u * \phi) = (\mathsf{T}_x u) * \phi = u * (\mathsf{T}_x \phi)$ *for all* $x \in \mathbb{R}^n$*,*
(ii) $u * \psi \in \mathrm{C}_c^\infty(\mathbb{R}^n)$*,*
(iii) $u * (\phi * \psi) = (u * \phi) * \psi = (u * \psi) * \phi$*.*

Convolutions may also be defined between distributions ([17]).

Lemma 3.18 (Convolutions between Distributions). *Let $u, v, w \in (\mathrm{C}_c^\infty(\mathbb{R}^n))'$, $n \in \mathbb{N}$.*

(i) *If at least one of u, v has compact support, the convolution $u * v$ is defined by $(u * v) * \phi = u * (v * \phi)$ for all $\phi \in \mathrm{C}_c^\infty(\mathbb{R}^n)$ and $u * v = v * u$.*

(ii) *If at least one of the supports* $\operatorname{supp}(u)$, $\operatorname{supp}(v)$ *is compact, we have* $\operatorname{supp}(u * v) \subset \operatorname{supp}(u) + \operatorname{supp}(v)$.
(iii) *If at least two of the supports* $\operatorname{supp}(u)$, $\operatorname{supp}(v)$, $\operatorname{supp}(w)$ *are compact, we have* $(u * v) * w = u * (v * w)$.

We can now define weak derivatives by defining derivatives of distributions ([1, 17]).

Definition 3.19 (Weak Derivative). Let $\mathcal{G}$ be a bounded region in $\mathbb{R}^n$, $n \in \mathbb{N}$, and $u \in (\mathrm{C}_c^\infty(\mathcal{G}))'$. The weak derivative of u with respect to x_i, $i \in \{1, \ldots, n\}$, is defined by

$$\langle \partial_{x_i} u, \phi \rangle = - \langle u, \partial_{x_i} \phi \rangle , \quad \phi \in \mathrm{C}_c^\infty(\mathcal{G}) . \tag{78}$$

For a multi-index $\gamma \in \mathbb{N}_0^n$, we have the generalization

$$\langle D^\gamma u, \phi \rangle = (-1)^{|\gamma|} \langle u, D^\gamma \phi \rangle , \quad \phi \in \mathrm{C}_c^\infty(\mathcal{G}) . \tag{79}$$

Remark 3.20. We use the same notation for weak derivatives and classical (strong) partial derivatives (based on the limit of difference quotients). If a continuous strong derivative exists, it coincides with the weak derivative as can be seen by integration by parts ([17]).

Theorem 3.21 (Weak Derivatives and Convolution).

(i) *Suppose* $u \in (\mathrm{C}_c^\infty(\mathbb{R}^n))'$ *and* $\phi \in \mathrm{C}_c^\infty(\mathbb{R}^n)$, $n \in \mathbb{N}$, *or* $u \in (\mathrm{C}_c^\infty(\mathbb{R}^n))'$ *with compact support and* $\phi \in \mathrm{C}^\infty(\mathbb{R}^n)$, *then* $D^\gamma(u * \phi) = (D^\gamma u) * \phi = u * (D^\gamma \phi)$ *for all* $\gamma \in \mathbb{N}_0^n$.
(ii) *Suppose* $u \in (\mathrm{C}_c^\infty(\mathbb{R}^n))'$ *and* δ *is the delta distribution, then* $D^\gamma u = (D^\gamma \delta) * u$ *for all* $\gamma \in \mathbb{N}_0^n$. *In particular,* $u = \delta * u$.
(iii) *Suppose* $u, v \in (\mathrm{C}_c^\infty(\mathbb{R}^n))'$ *and at least one of them has compact support, then* $D^\gamma(u * v) = (D^\gamma u) * v = u * (D^\gamma v)$ *for all* $\gamma \in \mathbb{N}_0^n$.

A consequence of Theorem 3.21 is that it allows to give an informal integral expression for Dirac's delta distribution and its derivatives, with a slight abuse of notation, by

$$\int_{\mathbb{R}^n} D^\gamma \delta(x - y) \phi(y) \, \mathrm{d}V(y) = (-1)^{|\gamma|} (D^\gamma \phi)(x) . \tag{80}$$

It is easy to prove that the weak derivative of a distribution is also a distribution. Thus, for every distribution, there exist weak derivatives of arbitrary order. Nevertheless, classes of distributions and their derivatives can be distinguished if we introduce a new concept of regularity based on integrability. We begin by defining Lebesgue spaces.

Definition 3.22 (Lebesgue Spaces). Let $\mathcal{G}$ be a bounded region in $\mathbb{R}^n$, $n \in \mathbb{N}$, and $p \in \mathbb{R}^+$. The Lebesgue space $\mathrm{L}^p(\mathcal{G})$ consists of all equivalence classes with respect to the Lebesgue measure V of V-almost everywhere identical functions on $\mathcal{G}$, whose representatives u satisfy

$$\int_{\mathcal{G}} |u(x)|^p \, \mathrm{d}V(x) < \infty . \tag{81}$$

Moreover, the space $\mathrm{L}^\infty(\mathcal{G})$ contains all such equivalence classes whose representatives are measurable, essentially bounded functions $u : \mathcal{G} \to \mathbb{R}$, i.e.,

$$\operatorname*{ess\,sup}_{x\in\mathcal{G}} |u(x)| \coloneqq \inf_{V(N)=0} \sup_{x\in\mathcal{G}\setminus N} |u(x)| < \infty \,. \tag{82}$$

Remark 3.23. It is convenient to identify a function with its respective equivalence class. However, we remind the reader that it is, in general, not possible to evaluate a function from $\mathrm{L}^p(\mathcal{G})$ at a point $x \in \mathcal{G}$.

We summarize a few properties of the Lebesgue spaces ([1]).

Lemma 3.24 (Properties of Lebesgue Spaces). *Let $\mathcal{G}$ be a bounded region in $\mathbb{R}^n$, $n \in \mathbb{N}$, and $1 \le p < \infty$.*

(i) *The Lebesgue space $\mathrm{L}^p(\mathcal{G})$ is a Banach space with respect to the norm*

$$\|u\|_{\mathrm{L}^p(\mathcal{G})} = \left(\int_{\mathcal{G}} |u(x)|^p \, \mathrm{d}V(x) \right)^{\frac{1}{p}} \,. \tag{83}$$

$\mathrm{L}^\infty(\mathcal{G})$ is a Banach space with respect to the norm

$$\|u\|_{\mathrm{L}^\infty(\mathcal{G})} = \operatorname*{ess\,sup}_{x\in\mathcal{G}} |u(x)| \,. \tag{84}$$

(ii) *The space $\mathrm{L}^2(\mathcal{G})$ is a Hilbert space if equipped with the scalar product*

$$(u, v)_{\mathrm{L}^2(\mathcal{G})} = \int_{\mathcal{G}} u(x)v(x) \, \mathrm{d}V(x) \,. \tag{85}$$

(iii) *For arbitrary $1 \le p_1, p_2 < \infty$ with $p_1 \ge p_2$, we have $\mathrm{L}^{p_1}(\mathcal{G}) \subset \mathrm{L}^{p_2}(\mathcal{G})$ and $\mathrm{L}^{p_1}(\mathcal{G}) \subset \mathrm{L}^1_{\mathrm{loc}}(\mathcal{G})$, with $\mathrm{L}^1_{\mathrm{loc}}(\mathcal{G})$ being the space of locally integrable functions.*

(iv) *Let $1 < p_1 < \infty$ and p_2 such that $\frac{1}{p_1} + \frac{1}{p_2} = 1$. For $u \in \mathrm{L}^{p_1}(\mathcal{G})$, $v \in \mathrm{L}^{p_2}(\mathcal{G})$, we have $uv \in \mathrm{L}^1(\mathcal{G})$ and*

$$\|uv\|_{\mathrm{L}^1(\mathcal{G})} \le \|u\|_{\mathrm{L}^{p_1}(\mathcal{G})} \, \|v\|_{\mathrm{L}^{p_2}(\mathcal{G})} \,. \tag{86}$$

This is known as Hölder's inequality. It also holds for $u \in \mathrm{L}^1(\mathcal{G})$ and $v \in \mathrm{L}^\infty(\mathcal{G})$. Then we have $uv \in \mathrm{L}^1(\mathcal{G})$.

(v) *Let $1 < p_1 < \infty$ and p_2 such that $\frac{1}{p_1} + \frac{1}{p_2} = 1$. Then the dual space $(\mathrm{L}^{p_1}(\mathcal{G}))'$ of $(\mathrm{L}^{p_1}(\mathcal{G}))$ is isometrically isomorph to $\mathrm{L}^{p_2}(\mathcal{G})$. Moreover, $\left(\mathrm{L}^1(\mathcal{G})\right)'$ is isometrically isomorph to $\mathrm{L}^\infty(\mathcal{G})$, but $(\mathrm{L}^\infty(\mathcal{G}))'$ is not isometrically isomorph to $\mathrm{L}^1(\mathcal{G})$.*

(vi) *$\mathrm{C}_c(\mathcal{G})$ and $\mathrm{C}_c^\infty(\mathcal{G})$ are dense subspaces of $\mathrm{L}^p(\mathcal{G})$ for all $1 \le p \le \infty$.*

The definition of Lebesgue spaces allows us to evaluate the regularity of a distribution by asking if it is also an element of some Lebesgue space. It comes naturally to extend this to a distribution's derivative. This gives rise to the definition of Sobolev spaces ([1, 14, 21]).

Definition 3.25 (Sobolev Spaces). Let $\mathcal{G}$ be a bounded region in $\mathbb{R}^n$, $n \in \mathbb{N}$, and $1 \leq p \leq \infty$. The Sobolev space $\mathrm{W}^{k,p}(\mathcal{G})$, $k \in \mathbb{N}_0$, is defined as the subspace of $\mathrm{L}^p(\mathcal{G})$ with

$$\mathrm{W}^{k,p}(\mathcal{G}) = \{u \in \mathrm{L}^p(\mathcal{G}) : \ D^\gamma u \in \mathrm{L}^p(\mathcal{G}) \quad \text{for all } \gamma \in \mathbb{N}_0^n, |\gamma| \leq k\} \ . \tag{87}$$

$\mathrm{W}^{k,p}(\mathcal{G})$ is a separable Banach space with respect to the norm

$$\|u\|_{\mathrm{W}^{k,p}(\mathcal{G})} = \left(\sum_{|\gamma| \leq k} \int_{\mathcal{G}} |D^\gamma u(x)|^p \ \mathrm{d}V(x) \right)^{\frac{1}{p}} \ . \tag{88}$$

For $p = 2$, we denote $\mathrm{H}^k(\mathcal{G}) = \mathrm{W}^{k,2}(\mathcal{G})$. These spaces are separable Hilbert spaces with inner product

$$(u, v)_{\mathrm{H}^k(\mathcal{G})} = \sum_{|\gamma| \leq k} (D^\gamma u, D^\gamma v)_{\mathrm{L}^2(\mathcal{G})} \ . \tag{89}$$

Moreover, the concept of Hölder-continuity can also be transferred to weak derivatives in the following sense ([21]).

Definition 3.26 (Sobolev–Slobodeckij Spaces). Let $\mathcal{G}$ be a bounded region in $\mathbb{R}^n$, $n \in \mathbb{N}$, and $1 \leq p \leq \infty$. The Sobolev–Slobodeckij space $\mathrm{W}^{k,p}(\mathcal{G})$ of fractional order $k = r + s$ with $r \in \mathbb{N}_0$ and $0 < s < 1$ is defined as the subspace of $\mathrm{W}^{r,p}(\mathcal{G})$ with

$$\mathrm{W}^{k,p}(\mathcal{G}) = \left\{ u \in \mathrm{W}^{r,p}(\mathcal{G}) : \ |D^\gamma u|_{s,p,\mathcal{G}} < \infty \quad \text{for all } \gamma \in \mathbb{N}_0^n, |\gamma| = k \right\} \ , \tag{90}$$

where the semi-norm $|u|_{s,p,\mathcal{G}}$ is given by

$$|u|_{s,p,\mathcal{G}} = \left(\int_{\mathcal{G}} \int_{\mathcal{G}} \frac{|u(x) - u(y)|^p}{\|x - y\|^{n+ps}} \, \mathrm{d}V(x) \, \mathrm{d}V(y) \right)^{\frac{1}{p}} \ . \tag{91}$$

$\mathrm{W}^{k,p}(\mathcal{G})$ is a Banach space if equipped with the norm

$$\|u\|_{\mathrm{W}^{k,p}(\mathcal{G})} = \left(\|u\|^p_{\mathrm{W}^{r,p}(\mathcal{G})} + \sum_{|\gamma| = k} |D^\gamma u|^p_{s,p,\mathcal{G}} \right)^{\frac{1}{p}} \ . \tag{92}$$

As before, for $p = 2$, we denote $\mathrm{H}^k(\mathcal{G}) = \mathrm{W}^{k,2}(\mathcal{G})$.

In what follows, we formulate results only for scalar-valued functions, although similar results are valid for vector-valued functions.

The following relations of Sobolev spaces and spaces of continuous differentiable functions is particularly useful when considering numerical solution schemes (see [1]).

Lemma 3.27. *Let $\mathcal{G}$ be a bounded region in $\mathbb{R}^n$, $n \in \mathbb{N}$, and $1 \leq p < \infty$. The Sobolev space $\mathrm{W}^{k,p}(\mathcal{G})$, $k \in \mathbb{N}_0$, is the completion of $\mathrm{C}^\infty(\mathcal{G})$ with respect to the norm $\|\cdot\|_{\mathrm{W}^{k,p}(\mathcal{G})}$.*

If $\mathcal{G}$ has the segment property, then the set of restrictions to $\mathcal{G}$ of functions in $\mathrm{C}_c^\infty(\mathbb{R}^n)$ is dense in $\mathrm{W}^{k,p}(\mathcal{G})$.

Lemma 3.27 suggests the definition of another class of Sobolev spaces.

Definition 3.28. Let $\mathcal{G}$ be a bounded region in $\mathbb{R}^n$, $n \in \mathbb{N}$, and $1 \leq p < \infty$. The Sobolev space $\mathrm{W}_0^{k,p}(\mathcal{G})$, $k \in \mathbb{N}_0$, is defined as the completion of $\mathrm{C}_c^\infty(\mathcal{G})$ with respect to the norm $\|\cdot\|_{\mathrm{W}^{k,p}(\mathcal{G})}$.

The above definition is motivated by the theory of partial differential equations. In general, if we are looking for a solution to a given partial differential equation on $\mathcal{G}$, it is necessary to prescribe boundary conditions to achieve uniqueness. The simplest way to do this is by assuming that the solution shall vanish at the boundary (so-called homogeneous Dirichlet boundary condition). We add more details to these concepts later.

Remark 3.29. In general, $\mathrm{W}_0^{k,p}(\mathcal{G}) \neq \mathrm{W}^{k,p}(\mathcal{G})$. For conditions on $\mathcal{G}$ under which those spaces are equal, the reader is referred to, e.g., [1].

With this definition, we can characterize Sobolev spaces with negative index ([1]).

Definition 3.30 (Sobolev Spaces with Negative Index). Let $\mathcal{G}$ be a bounded region in $\mathbb{R}^n$, $n \in \mathbb{N}$, and $1 < p_1 < \infty$, p_2 such that $\frac{1}{p_1} + \frac{1}{p_2} = 1$. The Sobolev space $\mathrm{W}^{-k,p_2}(\mathcal{G})$, $k \in \mathbb{R}^+$, is defined as

$$\mathrm{W}^{-k,p_2}(\mathcal{G}) = \left\{ f \in (\mathrm{C}_c^\infty(\mathcal{G}))' : \ \|f\|_{\mathrm{W}^{-k,p_2}(\mathcal{G})} < \infty \right\} , \tag{93}$$

with

$$\|f\|_{\mathrm{W}^{-k,p_2}(\mathcal{G})} = \sup_{0 \neq u \in \mathrm{C}_c^\infty(\mathcal{G})} \frac{|f(u)|}{\|u\|_{\mathrm{W}^{k,p_1}(\mathcal{G})}} . \tag{94}$$

$\mathrm{W}^{-k,p_2}(\mathcal{G})$ is the dual space of $\mathrm{W}_0^{k,p_1}(\mathcal{G})$.

An essential property of Sobolev spaces is the existence of the following embeddings ([1]).

Theorem 3.31 (Sobolev Embedding Theorem). *Let $\mathcal{G}$ be a bounded region in $\mathbb{R}^n$, $n \in \mathbb{N}$, $j, k \in \mathbb{N}_0$, $1 \leq p_1, p_2 < \infty$.*

(i) *If $\mathcal{G}$ has the cone property, the following embeddings, marked by $\hookrightarrow$, exist:*

(a) *Suppose $kp_1 < n$ and $p_1 \leq p_2 \leq \frac{np_1}{n-kp_1}$. Then*

$$\mathrm{W}^{j+k,p_1}(\mathcal{G}) \hookrightarrow \mathrm{W}^{j,p_2}(\mathcal{G}) . \tag{95}$$

(b) *Suppose $kp_1 = n$, $p_1 \leq p_2 < \infty$. Then*

$$\mathrm{W}^{j+k,p_1}(\mathcal{G}) \hookrightarrow \mathrm{W}^{j,p_2}(\mathcal{G}) . \tag{96}$$

Moreover, if $p_1 = 1$ and, thus, $k = n$, this also holds for $p_2 = \infty$.

(ii) *If $\mathcal{G}$ has the strong local Lipschitz property, additional embeddings hold:*

(a) *Suppose $kp_1 > n > (k-1)p_1$. Then*

$$\mathrm{W}^{j+k,p_1}(\mathcal{G}) \hookrightarrow \mathrm{C}^{j,s}(\overline{\mathcal{G}}), \quad 0 < s < k - \frac{n}{p_1} . \tag{97}$$

(b) *Suppose* $n = (k-1)p_1$. *Then*

$$\mathrm{W}^{j+k,p_1}(\mathcal{G}) \hookrightarrow \mathrm{C}^{j,s}(\overline{\mathcal{G}}), \quad 0 < s < 1\,. \tag{98}$$

The last embedding holds for $s = 1$ *if* $n = k-1$ *and* $p_1 = 1$.

As mentioned above, the discussion of partial differential equations involves the necessity to specify in some sense the values an element of a Sobolev space takes on the boundary of a region $\mathcal{G}$. This is not a trivial problem as elements of Sobolev spaces are equivalence classes like the elements of Lebesgue spaces on which the definition of Sobolev spaces is based. Lemma 3.27 allows us to find a solution to this dilemma by introducing the trace operator ([13, 21]). However, as this involves Sobolev spaces defined on the boundary $\partial\mathcal{G}$ of a bounded region, we first have to define proper measures on the boundary, which are given by so-called Hausdorff measures ([2]).

Definition 3.32 (Hausdorff Measure). Let S be a smooth surface in $\mathbb{R}^n$, $n \in \mathbb{N}$ given by

$$S = \{(x, g(x)) \in \mathbb{R}^n : x \in D\} \tag{99}$$

with $D \subset \mathbb{R}^{n-1}$ open and bounded and $g \in C^1(\overline{D})$.

For any subset $E \subset S$ for which the set $\widetilde{E} \coloneqq \{x \in D : (x, g(x)) \in E\}$ is measurable with respect to the $(n-1)$-dimensional Lebesgue measure V_{n-1}, we define the Hausdorff measure S_{n-1} of E by

$$S_{n-1}(E) \coloneqq \int_{\widetilde{E}} \sqrt{1 + |\nabla g(x)|^2}\, \mathrm{d}V_{n-1}(x). \tag{100}$$

The norms of all Sobolev spaces defined on the boundary $\partial\mathcal{G}$ of a region $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{R}$ are understood with respect to the corresponding Hausdorff measure S_{n-1}.

Remark 3.33. It is possible to extend the above definition in several ways. For example, we can combine several surfaces which satisfy Definition 3.32 by gluing them. This is necessary for closed surfaces, which do not directly satisfy Definition 3.32, e.g, the unit sphere which can be seen as the combination of two hemispheres. For an even more general definition see, e.g., [19].

Theorem 3.34 (Trace Operator). *Let* $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, *be a bounded region with the uniform* $\mathrm{C}^{m,s}$*-regularity property.*

(i) *Let* $\frac{1}{2} < k \leq m + s$, *whereas for* $k \in \mathbb{N}$, $k = m-1$, $s = 1$ *is allowed. There is a continuous linear operator* $T_0 : \mathrm{H}^k(\mathcal{G}) \to \mathrm{H}^{k-\frac{1}{2}}(\partial\mathcal{G})$, *called trace operator, such that*

$$T_0 u = u|_{\partial\mathcal{G}} \quad \text{for all } u \in \mathrm{C}^{\lfloor k \rfloor + 1}(\overline{\mathcal{G}})\,. \tag{101}$$

If $k \in \mathbb{N}$, *we have* $u \in \mathrm{C}^k(\overline{\mathcal{G}})$.

(ii) *Let $k+1 \leq m+s$, whereas for $k \in \mathbb{N}$, $m = k$ and $s = 1$ is allowed and $l \in \mathbb{N}$ such that $k - l > \frac{1}{2}$. There is another continuous linear trace operator $T_l : \mathrm{H}^k(\mathcal{G}) \to \bigotimes_{i=0}^{l} \mathrm{H}^{k-i-\frac{1}{2}}(\partial\mathcal{G})$ such that*

$$T_l u = \left(u|_{\partial\mathcal{G}}, \partial_{-n(x)} u|_{\partial\mathcal{G}}, \ldots, \partial^l_{-n(x)} u|_{\partial\mathcal{G}} \right) \quad \text{for all } u \in \mathrm{C}^{\lfloor k \rfloor + l + 1}(\overline{\mathcal{G}}) \,. \tag{102}$$

If $k \in \mathbb{N}$, we have $u \in \mathrm{C}^{k+l}(\overline{\mathcal{G}})$. Here, $\partial_{-n(x)} u$ is the directional derivative with respect to the inner normal on $\partial\mathcal{G}$.

Remark 3.35. Definition 3.28 and Theorem 3.34 are compatible, as we have

$$T_0 u = 0 \quad \text{on } \partial\mathcal{G} \quad \text{for all } u \in \mathrm{W}_0^{k,p}(\mathcal{G}) \,, \tag{103}$$

i.e., the elements of $\mathrm{W}_0^{k,p}(\mathcal{G})$ satisfy a homogeneous Dirichlet boundary condition. For $k \geq 2$, it is possible to show that

$$T_0 D^\gamma u = 0 \quad \text{on } \partial\mathcal{G} \quad \text{for all } u \in \mathrm{W}_0^{k,p}(\mathcal{G}) \text{ with } |\gamma| \leq k-1 \,. \tag{104}$$

The definition of all the above spaces can be generalized to functions which take values in a separable Banach space V ([16]). Let $I \subset \mathbb{R}$ be a bounded open interval and V be a separable Banach space with (topological) dual V'. We start by defining $\mathrm{C}(I;V)$ to be the space of all bounded continuous functions $u : I \to V$, $t \mapsto u(t)$ and equip it with the norm

$$\|u\|_{\mathrm{C}(I;V)} = \sup_{t \in I} \|u(t)\|_V \,. \tag{105}$$

Analogously, $\mathrm{C}^k(I;V)$, $k \in \mathbb{N}$, is defined as the space of all functions $u : I \to V$ whose derivatives in I, i.e., with respect to t, up to order k are of class $\mathrm{C}(I;V)$.

Moreover, the Lebesgue spaces $\mathrm{L}^p(\mathcal{G})$, $1 \leq p < \infty$, can be generalized to $\mathrm{L}^p(I;V)$ by substituting the absolute value in their definition and the definition of their norms by the norm on V, thus yielding the norm

$$\|u\|_{\mathrm{L}^p(I;V)} = \left(\int_I \|u(t)\|_V^p \; dV_1(t) \right)^{\frac{1}{p}} \,. \tag{106}$$

If V is a separable Hilbert space, $\mathrm{L}^2(I;V)$ is also a Hilbert space. The space $\mathrm{L}^\infty(I;V)$ consists of all measurable, essentially bounded functions $u : I \to V$. It is a Banach space with respect to the norm

$$\|u\|_{\mathrm{L}^\infty(I;V)} = \operatorname*{ess\,sup}_{t \in I} \|u(t)\|_V \,. \tag{107}$$

The spaces $\mathrm{C}(\overline{I};V)$ and $\mathrm{L}^p(\overline{I};V)$, $1 \leq p \leq \infty$, are defined accordingly.

The generalization of Sobolev spaces to V-valued functions is straightforward. We show how this is done for $\mathrm{H}^1(I;\mathrm{L}^2(\mathcal{G}))$, where $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, is an open bounded region. The corresponding norm is given by

$$\|u\|_{\mathrm{H}^1(I;\mathrm{L}^2(\mathcal{G}))} = \left(\int_I \left(\int_{\mathcal{G}} \left(|u(x,t)|^2 + |\partial_t u(x,t)|^2 \right) \mathrm{d}V(x) \right) dV_1(t) \right)^{\frac{1}{2}} \,. \tag{108}$$

For separable Hilbert spaces V, the following embedding theorem can be established:

Lemma 3.36 (Sobolev Lemma for Hilbert Space-valued Functions). *Let $I \subset \mathbb{R}$ be a bounded open interval and V be a separable Hilbert space. Then any function $u \in \mathrm{H}^1(I;V)$ has a continuous representative in $\mathrm{C}(\overline{I};V)$.*

Remark 3.37. See, e.g., [18] for a proof in a more general setting with Banach spaces instead of Hilbert spaces.

A well-known tool in geodesy is Fourier transformation, which can also be used to define Sobolev spaces. However, it is not possible to define Fourier transformation for all distributions. Thus, we need the concept of rapidly decreasing functions ([4]).

Definition 3.38 (Schwartz Space of Rapidly Decreasing Functions).
For $\phi \in \mathrm{C}^\infty(\mathbb{R}^n)$, $k, l, n \in \mathbb{N}_0$, $n \neq 0$, $\alpha \in \mathbb{N}_0^n$, we define

$$p_{k,l}(\phi) \coloneqq \sup_{x \in \mathbb{R}^n} \left(|x|^k + 1\right) \sum_{|\alpha| \leq l} |D^\alpha \phi(x)| \, . \tag{109}$$

ϕ is called rapidly decreasing if $p_{k,l}(\phi) < \infty$ for all $k, l \in \mathbb{N}_0$. The space $\mathscr{S}(\mathbb{R}^n)$ of all rapidly decreasing functions is called Schwartz space.

Convergence of a sequence $(\phi_j)_{j \in \mathbb{N}}$ in $\mathscr{S}(\mathbb{R}^n)$ is defined by

$$\phi_j \xrightarrow{\mathscr{S}} \phi \quad \Longleftrightarrow \quad \lim_{j \to \infty} p_{k,l}(\phi_j - \phi) = 0 \ \forall \ k, l \in \mathbb{N}_0 \, . \tag{110}$$

From the definition, it is clear that $\mathrm{C}_c^\infty(\mathbb{R}^n) \subset \mathscr{S}(\mathbb{R}^n)$.

We can now define Fourier transformation for rapidly decreasing functions ([4]):

Definition 3.39 (Fourier Transformation in Schwartz Space). Let $\phi \in \mathscr{S}(\mathbb{R}^n)$. The Fourier transform $\mathscr{F}\phi$ of ϕ is defined by

$$\mathscr{F}\phi(\xi) \coloneqq (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} e^{-ix \cdot \xi} \phi(x) \mathrm{d}V(x), \quad \xi \in \mathbb{R}^n \, . \tag{111}$$

We summarize some properties of the Fourier transformation $\mathscr{F}$ for rapidly decreasing functions ([4]):

Theorem 3.40 (Properties of Fourier Transformation on $\mathscr{S}(\mathbb{R}^n)$). *Let $\phi \in \mathscr{S}(\mathbb{R}^n)$, $\alpha \in \mathbb{N}_0^n$:*

(i) $x^\alpha \phi, D^\alpha \phi, \mathscr{F}\phi, D^\alpha \mathscr{F}\phi, \mathscr{F} D^\alpha \phi \in \mathscr{S}(\mathbb{R}^n)$;
(ii) $D^\alpha \mathscr{F}\phi = (-i)^{|\alpha|} \mathscr{F}(x^\alpha \phi)$, $\xi^\alpha \mathscr{F}\phi = (-i)^{|\alpha|} \mathscr{F}(D^\alpha \phi)$;
(iii) $\mathscr{F}\left(e^{-\frac{|\cdot|^2}{2}}\right)(\xi) = e^{-\frac{|\xi|^2}{2}}$;
(iv) $\mathscr{F} : \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{R}^n)$ *is bijective, periodic with period 4, and bicontinuous with inverse*

$$\left(\mathscr{F}^{-1}\phi\right)(x) = (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} e^{ix \cdot \xi} \phi(\xi) \mathrm{d}V(\xi) \, . \tag{112}$$

As mentioned earlier, Fourier transformation cannot be extended to the space of distributions. However, we can consider the dual of the Schwartz space, $\mathscr{S}'(\mathbb{R}^n)$. Since $\mathrm{C}_c^\infty(\mathbb{R}^n) \subset \mathscr{S}(\mathbb{R}^n)$, we have $\mathscr{S}'(\mathbb{R}^n) \subset (\mathrm{C}_c^\infty(\mathbb{R}^n))'$.

The elements of $\mathscr{S}'(\mathbb{R}^n)$ are called tempered distributions. Any distribution with compact support can be extended to be a tempered distribution. Regular distributions are tempered distributions if their corresponding locally integrable function satisfies $u \in \mathrm{L}^p(\mathbb{R}^n)$ for some $1 \leq p \leq \infty$. A counter-example would be the distribution corresponding to $u : x \mapsto u(x) = e^{|x|^2}$.

We can now extend the Fourier transformation to $\mathscr{S}'(\mathbb{R}^n)$ ([4]):

Definition 3.41 (Fourier Transformation of Tempered Distributions). Let $T \in \mathscr{S}'(\mathbb{R}^n)$. The Fourier transform $\mathscr{F}T$ of T is defined by

$$\mathscr{F}T(\phi) \coloneqq T\left(\mathscr{F}\phi\right), \quad \phi \in \mathscr{S}(\mathbb{R}^n)\,. \tag{113}$$

Again, we summarize some properties of the Fourier transformation for tempered distributions ([4]):

Theorem 3.42 (Properties of Fourier Transformation on $\mathscr{S}'(\mathbb{R}^n)$).

(i) $T \in \mathscr{S}'(\mathbb{R}^n) \;\Rightarrow\; \mathscr{F}T \in \mathscr{S}'(\mathbb{R}^n)$;
(ii) *for regular distributions* $T_u \in \mathscr{S}'(\mathbb{R}^n)$, *we have* $\mathscr{F}T_u = T_{\mathscr{F}u}$;
(iii) $\mathscr{F}, \mathscr{F}^{-1} : \mathscr{S}'(\mathbb{R}^n) \to \mathscr{S}'(\mathbb{R}^n)$ *are bijective, periodic with period* 4, *and bicontinuous with*

$$\mathscr{F}\mathscr{F}^{-1}T = \mathscr{F}^{-1}\mathscr{F}T = T\,. \tag{114}$$

(iv) $T \in \mathscr{S}'(\mathbb{R}^n), \phi \in \mathscr{S}(\mathbb{R}^n) \;\Rightarrow\; \mathscr{F}^{-1}T\left(\phi(\xi)\right) = \mathscr{F}T\left(\phi(-\xi)\right)$.

An important feature of the Fourier transformation is that it can be extended to $\mathrm{L}^2(\mathbb{R}^n)$ ([4]). As an operator from $\mathrm{L}^2(\mathbb{R}^n)$ to $\mathrm{L}^2(\mathbb{R}^n)$, the Fourier transformation is an isometric isomorphism ([4]):

Theorem 3.43 (Fourier Transformation on $\mathrm{L}^2(\mathbb{R}^n)$). *The Fourier transformation has a unique extension to* $\mathrm{L}^2(\mathbb{R}^n)$ *and for all* $\phi, \psi \in \mathrm{L}^2(\mathbb{R}^n)$, *we have*

$$(u, v)_{\mathrm{L}^2(\mathbb{R}^n)} = (\mathscr{F}u, \mathscr{F}v)_{\mathrm{L}^2(\mathbb{R}^n)}\,, \tag{115}$$

i.e., $\mathscr{F} : \mathrm{L}^2(\mathbb{R}^n) \to \mathrm{L}^2(\mathbb{R}^n)$ *is an isometric isomorphism.*

Fourier transformation can also be used to characterize Sobolev spaces and is especially useful when discussing Sobolev spaces with non-integer index ([4])

Theorem 3.44 (Characterization of $\mathrm{H}^k(\mathbb{R}^n)$ and $\mathrm{H}^k(\mathcal{G})$ via Fourier Transformation). *Let* $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$ *be a bounded region,* $k \in \mathbb{R}_0^+$.

(i) *For* $u \in \mathrm{H}^k(\mathbb{R}^n)$ *exist* $c_1, c_2 \in \mathbb{R}^+$ *such that*

$$c_1 \left\|u\right\|^2_{\mathrm{H}^k(\mathbb{R}^n)} \leq \int_{\mathbb{R}^n} \left(1 + |\xi|\right)^{2k} \left|\mathscr{F}u\right|^2 \leq c_2 \left\|u\right\|^2_{\mathrm{H}^k(\mathbb{R}^n)}.$$

(ii) *If there exists a continuous operator* $E : \mathrm{H}^k(\mathcal{G}) \to \mathrm{H}^k(\mathbb{R}^n)$ *with* $Eu|_{\mathcal{G}} = u$ *for all* $u \in \mathrm{H}^k(\mathcal{G})$*, the space* $\mathrm{H}^k(\mathcal{G})$ *coincides with the restriction of functions in* $\mathrm{H}^k(\mathbb{R}^n)$ *on* $\mathcal{G}$ *and the norm*

$$\|u\|'_{\mathrm{H}^k(\mathcal{G})} = \inf_{\substack{\widetilde{u}\in \mathrm{H}^k(\mathbb{R}^n) \\ \widetilde{u}|_{\mathcal{G}}}} \left\| (1+|\cdot|)^k \, \mathscr{F}\widetilde{u} \right\|_{\mathrm{L}^2(\mathbb{R}^n)} \tag{116}$$

is equivalent to the norm in $\mathrm{H}^k(\mathcal{G})$.

Remark 3.45.

(a) For a domain $\mathcal{G}$ to satisfy an extension condition as in Theorem 3.44, it is sufficient that $\mathcal{G}$ is bounded and has the uniform cone property. This result is due to Calderón and Zygmund (see, e.g., [21] and the references therein).

(b) It is also possible to define spaces of Sobolev-type via Fourier transformation based on other Lebesgue spaces $\mathrm{L}^p(\mathbb{R}^n)$. Those are called Bessel potential spaces. However, they are in general not identical to the Sobolev or Sobolev–Slobodeckij spaces as given in this chapter. For details, the reader is referred to [1] and the references therein.

4. Differential equations

Assume we have an open bounded region $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, $n > 1$, and a map

$$\mathcal{F} : \mathbb{R}^{n^k} \times \mathbb{R}^{n^{k-1}} \times \cdots \times \mathbb{R}^n \times \mathbb{R} \times \mathcal{G} \to \mathbb{R}, \quad k \in \mathbb{N} \,. \tag{117}$$

Then for $u \in \mathrm{C}^k(\mathcal{G})$

$$\mathcal{F}\big(\big(D^k u\big)(x), \ldots, u(x), x\big) = 0 \quad \text{for all } x \in \mathcal{G} \tag{118}$$

is a partial differential equation (PDE) of order k if at least one derivative of order k is actually a part of the equation and no derivative of higher order than k is present. This can be done analogously for systems of differential equations. We only deal with linear PDEs here that can be written as

$$\sum_{|\gamma|\le k} a_\gamma(x) \left(D^\gamma u\right)(x) = f(x), \ \gamma \in \mathbb{N}_0^n \quad \text{for all } x \in \mathcal{G} \,, \tag{119}$$

with given coefficient functions a_γ and right-hand side f. If $f = 0$, the PDE is called homogeneous.

Many important systems of linear PDEs consist of second-order PDEs. There are three main classes of these PDEs. We start by defining an elliptic differential operator (cf., e.g., [6]).

Definition 4.1 (Elliptic PDEs). Let $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, $u \in \mathrm{C}^k(\mathcal{G})$. Let L be a linear differential operator of second order such that

$$u \mapsto Lu = \sum_{i,j=1}^{n} a_{ij} \partial_{x_i} \partial_{x_j} u + \sum_{i=1}^{n} b_i \partial_{x_i} u + cu \tag{120}$$

with given functions $a_{ij} : \mathcal{G} \to \mathbb{R}$, $b_i : \mathcal{G} \to \mathbb{R}$, and $c : \mathcal{G} \to \mathbb{R}$ sufficiently smooth.

L is called uniformly elliptic if there exists a constant $C \in \mathbb{R}^+$ such that

$$\sum_{i,j=1}^{n} a_{ij}(x) y_i y_j \geq C \|y\|^2 \tag{121}$$

for almost every $x \in \mathcal{G}$ and all $y \in \mathbb{R}^n$.

With this definition, we can now define the two other main classes (cf., e.g., [6, 9]). For both, there is one distinguished variable, denoted by t rather than as a component of a vector x, which is usually the time being distinct from spatial variables summarized in x.

Definition 4.2 (Parabolic PDE). Let $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, $(0, t_{\text{end}}) \subset \mathbb{R}$, $t_{\text{end}} \in \mathbb{R}^+$, $u \in \mathrm{C}^k(\mathcal{G} \times (0, t_{\text{end}}))$. Let L be a linear differential operator of second order such that

$$u \mapsto Lu = \sum_{i,j=1}^{n} a_{ij} \partial_{x_i} \partial_{x_j} u + \sum_{i=1}^{n} b_i \partial_{x_i} u + cu - \partial_t u \tag{122}$$

with given functions $a_{ij} : \mathcal{G} \times (0, t_{\text{end}}) \to \mathbb{R}$, $b_i : \mathcal{G} \times (0, t_{\text{end}}) \to \mathbb{R}$, and $c : \mathcal{G} \times (0, t_{\text{end}}) \to \mathbb{R}$ sufficiently smooth.

L is called uniformly parabolic if there exist constants C_0, $C_1 > 0$ such that

$$C_0 \|y\|^2 \leq \sum_{i,j=1}^{n} a_{ij}(x,t) y_i y_j \leq C_1 \|y\|^2 \tag{123}$$

for all $(x,t) \in \mathcal{G} \times (0, t_{\text{end}})$ and all $y \in \mathbb{R}^n$.

Definition 4.3 (Hyperbolic PDE). Let $\mathcal{G} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, $(0, t_{\text{end}}) \subset \mathbb{R}$, $t_{\text{end}} \in \mathbb{R}^+$, $u \in \mathrm{C}^k(\mathcal{G} \times (0, t_{\text{end}}))$. Let L be a linear differential operator of second order such that

$$u \mapsto Lu = \sum_{i,j=1}^{n} a_{ij} \partial_{x_i} \partial_{x_j} u + \sum_{i=1}^{n} b_i \partial_{x_i} u + cu - \partial_t^2 u \tag{124}$$

with given functions $a_{ij} : \mathcal{G} \times (0, t_{\text{end}}) \to \mathbb{R}$, $b_i : \mathcal{G} \times (0, t_{\text{end}}) \to \mathbb{R}$, and $c : \mathcal{G} \times (0, t_{\text{end}}) \to \mathbb{R}$ sufficiently smooth.

L is called uniformly hyperbolic if there exists a constant $C \in \mathbb{R}^+$ such that

$$\sum_{i,j=1}^{n} a_{ij}(x,t) y_i y_j \geq C \|y\|^2 \tag{125}$$

for all $(x,t) \in \mathcal{G} \times (0, t_{\text{end}})$ and all $y \in \mathbb{R}^n$.

Remark 4.4.

(i) It is possible that the character of a differential operator changes with x, e.g., when there is a function of x as coefficient of the time derivative term. Such equations can be locally elliptic, parabolic, or hyperbolic instead of uniformly, i.e., they are of one of these types on certain subregions of $\mathcal{G}$.

(ii) Not all linear second-order PDEs are of one of the above classes for $n > 2$.

(iii) Conditions on the actual smoothness of coefficients depends on the specific problem under considerations. For example, the so-called Cauchy–Kovalev-

skaya Theorem (see, e.g., [15]) needs analytic coefficients to guarantee the existence and uniqueness of analytic solutions. If we want the operator L in (119) to be continuous from $H^{l+k}(\mathcal{G})$ to $H^l(\mathcal{G})$, this can be achieved by requiring $a_\gamma \in C^l\left(\overline{\mathcal{G}}\right)$ (see, e.g., [21]).

As already pointed out, the formulations of differential equations as given above with strong partial derivatives is often not suited to find answers to the questions of solvability, uniqueness of solutions, or their regularity. Instead, we would like to have a formulation based on weak derivatives.

Let us assume that a linear second-order differential equation is given in its strong form by

$$Lu(x) = f(x) \quad \text{for all } x \in \mathcal{G} \,. \tag{126}$$

For simplicity, we equip this equation with the homogeneous Dirichlet boundary condition

$$u(x) = 0 \quad \text{for all } x \in \partial\mathcal{G} \,. \tag{127}$$

In the context of differential equations, we often use the abbreviation $\Gamma = \partial\mathcal{G}$.

A classical strong solution of this PDE has to be in $\mathrm{C}^2(\mathcal{G})$ which is a rather restrictive requirement. We can relax this requirement in two points. First, we can change over to a weakly differentiable solution. For this purpose, suppose v is an arbitrary function belonging to $\mathrm{C}_c^\infty(\mathcal{G})$, multiply the differential equation by v and integrate over $\mathcal{G}$. We obtain

$$(Lu, v)_{L^2(\mathcal{G})} = (f, v)_{L^2(\mathcal{G})} \quad \text{for all } v \in \mathrm{C}_c^\infty(\mathcal{G}) \,. \tag{128}$$

Moreover, we can relax the requirements on differentiability of u by performing an integration by parts on the left-hand side, which gives us a bilinear form $a(u, v)$. Thus, a useful assumption on u is $u \in \mathrm{H}_0^1(\mathcal{G})$. Additionally, as $\mathrm{C}_c^\infty(\mathcal{G})$ is a dense subspace of $\mathrm{H}_0^1(\mathcal{G})$, we can extend the space of functions with which we multiply to $\mathrm{H}_0^1(\mathcal{G})$. This yields

$$a(u, v) = f(v) \quad \text{for all } v \in \mathrm{H}_0^1(\mathcal{G}) \,. \tag{129}$$

Here, we interpreted the right-hand side as a linear functional on $\mathrm{H}_0^1(\mathcal{G})$. It is easy to see that every solution of the strong formulation is also a solution of the weak formulation. However, the opposite may not be true.

For other kinds of boundary conditions, the above procedure is changed in two points. On the one hand, if the values of u on the boundary are given and different from zero, u has to belong to another subspace of $\mathrm{H}^1(\mathcal{G})$. On the other hand, if normal derivatives of u are specified in a Neumann boundary condition, integration by parts yields some integrals over (parts of) the boundary Γ of $\mathcal{G}$. These are usually incorporated into the linear form f on the right-hand side. Other modifications may be necessary for other boundary conditions.

To answer the question whether a unique solution to a PDE in its weak form (129) exists, we can use Corollary 2.36 or directly the Theorems by Riesz (2.31) or Lax–Milgram (2.35) for elliptic PDEs. There are also general results for parabolic or hyperbolic systems, but those are out of the scope of this chapter.

5. Reproducing kernel functions

Let $\mathcal{S} \subset \mathbb{K}^n$, $n \in \mathbb{N}$, $x, y \in \mathcal{S}$ and $(V, (\cdot,\cdot)_V)$ a Hilbert space of functions defined on $\mathcal{S}$ or a superset of $\mathcal{S}$. Following [3], we define

Definition 5.1 (Reproducing Kernel). A function $K : \mathcal{S} \times \mathcal{S} \mapsto \mathbb{K}$, is called a reproducing kernel if

(a) for each fixed $y \in \mathcal{S}$, we have $K(\cdot, y) \in V$ and
(b) for every $f \in V$ and for every point $y \in \mathcal{S}$, the reproducing property

$$f(y) = (f, K(\cdot, y))_V \tag{130}$$

holds.

The question whether a given Hilbert space V has a reproducing kernel is answered by Davis [3].

Theorem 5.2 (Aronszajn, Existence of a Reproducing Kernel). *A necessary and sufficient condition that V has a reproducing kernel function is that for each fixed $y \in \mathcal{S}$, the linear functional given by $\delta(y)$, i.e.,*

$$V \ni f \mapsto \delta(y)f = f(y) \in \mathbb{K} \tag{131}$$

is bounded.

Reproducing kernel functions have the following properties ([3]):

Theorem 5.3 (Uniqueness and Symmetry of a Reproducing Kernel). *If a Hilbert space V possesses a reproducing kernel K, the kernel is unique and for all $x, y \in \mathcal{S}$, we have*

$$K(x, y) = \overline{K(y, x)}, \tag{132}$$

whereas the overline marks complex conjugation.

Theorem 5.4. *Let V have a reproducing kernel and let $\lim_{n\to\infty}\|f - f_n\|_V = 0$. Then, for each $x \in \mathcal{S}$, we have*

$$\lim_{n\to\infty} f_n(x) = f(x). \tag{133}$$

The convergence holds uniformly in every subset $\mathcal{U}$ of $\mathcal{S}$ for which $K : \mathcal{S} \to \mathbb{K}$, $y \mapsto K(y, y)$ is bounded.

As a consequence, we obtain the following corollary ([3]):

Corollary 5.5. *If V has a reproducing kernel, then the expansion of a function with respect to an orthonormal basis converges pointwise to the function and uniformly in every subset $\mathcal{U}$ of $\mathcal{S}$ for which $K : \mathcal{S} \to \mathbb{K}$, $y \mapsto K(y, y)$ is bounded.*

Reproducing kernels can also be used to find representatives for linear functionals on V. Let L be a linear functional on the reproducing kernel Hilbert space V. According to the Riesz representation Theorem 2.31, L has a representative h. As h is an element of V, we have

$$h(y) = (h, K(\cdot, y))_V = \overline{(K(\cdot, y), h)_V}. \tag{134}$$

On the other hand, $K(\cdot, y)$ is an element of V and we can apply the functional L to $K(\cdot, y)$ to get

$$LK(\cdot, y) = (K(\cdot, y), h)_V . \tag{135}$$

By comparing Eqs. (134) and (135), we get

Theorem 5.6. *Let V have a reproducing kernel $K(\cdot, \cdot)$, and let L be a bounded linear functional defined on V. Then the function*

$$h(y) = \overline{LK(\cdot, y)} \tag{136}$$

is in V and for all $f \in V$,

$$L(f) = (f, h)_V . \tag{137}$$

Moreover, we have

$$\|L\|_{V'}^2 = L_y \overline{L_x K(x, y)}. \tag{138}$$

Here, the indices x and y on L are used to clarify that we first regard $K(x, y)$ as a function of x with parameter y and apply L with respect to x and then regard $\overline{L_x K(x, y)}$ as a function of y with parameter x and apply L with respect to y.

An important example for a reproducing kernel Hilbert space is given in the following definition.

Definition 5.7. Let $\mathcal{G} \subset \mathbb{C}$ be a bounded region. The space of all functions $f \in \mathrm{L}^2(\mathcal{G})$ which are analytic is denoted by $\mathrm{L}^2_{\mathrm{A}}(\mathcal{G})$.

Theorem 5.8. *Let $\{h_n : n \in \mathbb{N}\}$ be an orthonormal basis in $\mathrm{L}^2_{\mathrm{A}}(\mathcal{G})$. Then $K(x, y) = \sum_{n=1}^{\infty} h_n(x)\overline{h_n(y)}$ is a (and hence the) reproducing kernel for $\mathrm{L}^2_{\mathrm{A}}(\mathcal{G})$; that is, for all $f \in \mathrm{L}^2_{\mathrm{A}}(\mathcal{G})$ we have*

$$f(y) = (f(\cdot), K(\cdot, y))_V = \int_{\mathcal{G}} f(x)\overline{K(x, y)}\,\mathrm{d}x. \tag{139}$$

$K(\cdot, \cdot)$ is known as the Bergman kernel.

Combining Theorems 5.6 and 5.8 yields

Corollary 5.9. *Let $\{h_n : n \in \mathbb{N}\}$ be an orthonormal basis in $\mathrm{L}^2_{\mathrm{A}}(\mathcal{G})$ and L a bounded linear functional defined on $\mathrm{L}^2_{\mathrm{A}}(\mathcal{G})$. Then*

$$f(y) = \sum_{n=1}^{\infty} h_n(y)\overline{L\, h_n(x)} \tag{140}$$

is the representative of L, and

$$\|L\|^2 = \sum_{n=1}^{\infty} |L\, h_n|^2 = L_y \overline{L_x K(x, y)}. \tag{141}$$

The above theorems show how an orthonormal basis can be used to find a reproducing kernel. However, we can also use a reproducing kernel to find an orthonormal basis using Theorem 5.6 ([3]).

Theorem 5.10. *Let V be a Hilbert space of functions that has the reproducing kernel $K(\cdot,\cdot)$. If $(L_n)_{n\in\mathbb{N}}$ is a sequence of bounded linear functionals on V such that from $L_n f = 0$, for all $n \in \mathbb{N}$, follows that $f = 0$, then the functions*

$$h_n(y) = \overline{L_n\, K(\cdot, y)}, \quad n \in \mathbb{N} \tag{142}$$

form a basis for V.

Moreover, Theorem 5.8 can be used the other way around to examine whether an orthonormal system is complete ([3]).

Theorem 5.11. *Let $K(\cdot,\cdot)$ be the reproducing kernel of $\mathrm{L}^2_{\mathrm{A}}(\mathcal{G})$ and $\{h_n : n \in \mathbb{N}\}$ be an orthonormal system. This system is complete if and only if*

$$K(x,x) = \sum_{n=1}^{\infty} |h_n(x)|^2 \quad \text{for all } x \in \mathcal{G}. \tag{143}$$

Another example with special relevance in geodesy is the space of spherical harmonics of a fixed degree.

Let $\Omega = \partial B_1(0) \subset \mathbb{R}^3$ be the unit sphere in $\mathbb{R}^3$. Let $H_n : \mathbb{R}^3 \to \mathbb{R}$ be a homogeneous harmonic polynomial of degree $n \in \mathbb{N}$, i.e., $H_n(\alpha x) = \alpha^3 H_n(x)$ for all $\alpha \in \mathbb{R}$ for all $x \in \mathbb{R}^3$ and $\Delta_x H_n$ vanishes on $\mathbb{R}^3$. Then the restriction $Y_n = H_n|_\Omega$ is a spherical harmonic of degree n and the following theorem holds ([8]):

Theorem 5.12 (Reproducing Kernel for Spherical Harmonics).
The space $\mathrm{Harm}_n(\Omega)$ *of all spherical harmonics of degree $n \in \mathbb{N}$ is a reproducing kernel Hilbert space with the kernel given by*

$$K(\zeta,\eta) = \frac{2n+1}{4\pi} P_n(\zeta \cdot \eta), \qquad \zeta, \eta \in \Omega \tag{144}$$

with $P_n : [-1,+1] \to \mathbb{R}$ the Legendre polynomial of degree n.

As the space of spherical harmonics in the above example is only finite dimensional, let us consider another example with an infinite-dimensional space.

Consider the operator of the Newton potential

$$A : \mathrm{L}^2(B_1(0)) \to R(A) \tag{145}$$

$$f \mapsto Af = \int_\Omega \frac{f(y)}{4\pi\,|\cdot - y|} \mathrm{d}V(y)\Big|_{\mathbb{R}^3 \setminus \overline{B_1(0)}}. \tag{146}$$

The null space of A consists of all functions in $\mathrm{L}^2(B_1(0))$ that are orthogonal to harmonic functions in $B_1(0)$. We denote the space of functions which are harmonic in $B_1(0)$ by $\mathrm{Pot}(B_1(0))$. It can be shown (see, e.g., [7]) that $A|_{\mathrm{Pot}(B_1(0))}$ is a linear bijective operator. We can define an inner product on $R(A)$ by

$$(h_1, h_2)_{R(A)} \coloneqq (A^{-1}h_1, A^{-1}h_2)_{\mathrm{L}^2(B_1(0))}. \tag{147}$$

Moreover, $R(A)$ equipped with this inner product is a Hilbert space which satisfies Theorem 5.2.

To find the reproducing kernel, we observe that for any $x \in \mathbb{R}^3 \setminus \overline{B_1(0)}$, $f \in \mathrm{Pot}(B_1(0))$ and $h = Af$, we have

$$h(x) = \left(f, \frac{1}{4\pi\,|x-\cdot|}\right)_{\mathrm{L}^2(B_1(0))} = \left(Af, A\left[\frac{1}{4\pi\,|x-\cdot|}\right]\right)_{R(A)} = \left(h, A\left[\frac{1}{4\pi\,|x-\cdot|}\right]\right)_{R(A)}. \tag{148}$$

Thus, the reproducing kernel is given by

$$K_{R(A)}(x,y) = A\left[\frac{1}{4\pi\,|x-\cdot|}\right] = \frac{1}{(4\pi)^2}\int_\Omega \frac{1}{|x-z|\,|y-z|}\mathrm{d}V(z). \tag{149}$$

6. Summary

As we have seen in this chapter, functional analysis provides many useful concepts to tackle geodetic problems. The structures and results which we discussed here are the very foundations for the solution of such problems as the (stochastic) oblique derivative problem ([11]), the inverse problem of determining the density distribution in the Earth's crust from gravity measurements ([7]) or the very successful and still expanding applications of wavelets to deal with local data concentration and data refinement ([7, 8]). They are also used in [12] to derive limit formulae and jump relations of potential theory in Sobolev spaces.

As can be deduced from their success so far, the application of functional analytic results and concepts has a key role in facing future challenges in geodesy. Among those are the incorporation of heterogeneous data, i.e., measurements of different quantities related to the gravity potential of the Earth to determine said potential. Moreover, the unequal distribution of those measurements, which may provide a high data density in some regions but show gaps in others, calls for locally oriented methods as opposed to the classical, globally oriented methods.

References

[1] R. Adams. *Sobolev Spaces*. Academic Press, New York, 1975.

[2] H.W. Alt. *Lineare Funktionalanalysis*. Springer, Berlin, Heidelberg, 2012.

[3] P. Davis. *Interpolation and Approximation*. Blaisdell Publishing Company, Waltham, MA, 1963.

[4] M. Dobrowolski. *Angewandte Funktionalanalysis*. Springer, Berlin, Heidelberg, 2010.

[5] P. Enflo. A counterexample to the approximation property in Banach spaces. *Acta Math.*, 130:309–317, 1973.

[6] L. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 1998.

[7] W. Freeden and C. Gerhards. *Geomathematically Oriented Potential Theory*. Chapman & Hall / CRC Press, Boca Raton, 2012.

[8] W. Freeden and V. Michel. *Multiscale Potential Theory with Applications to Geoscience*. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, 2004.

[9] A. Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Inc., Englewood Cliffs, 1964.

[10] G. Grosche, V. Ziegler, D. Ziegler, and E. Zeidler, editors. *Teubner Taschenbuch der Mathematik – Teil II*. B.G. Teubner, Stuttgart, Leipzig, 1995.

[11] M. Grothaus and T. Raskop. The Outer Oblique Boundary Problem of Potential Theory. *Numer Funct Anal Optim*, 30:711–750, 2009.

[12] M. Grothaus and T. Raskop. Limit Formulae and Jump Relations of Potential Theory in Sobolev Spaces. *Int. J. Geomath.*, 1:51–100, 2010.

[13] J.-L. Lions. *Equations Différentielles Operationelles et Problèmes aux Limites*, volume 111 of *Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen*. Springer, Berlin, 1961.

[14] W. McLean. *Strongly Elliptic Systems and Boundary Integral Equations*. Cambridge University Press, Cambridge, 2000.

[15] V.P. Mikhailov. *Partial Differential Equations*. MIR Publishers, Moscow, 1978.

[16] M. Renardy and R.C. Rogers. *An Introduction to Partial Differential Equations*. Texts in Applied Mathematics. Springer, New York, 1993.

[17] W. Rudin. *Functional Analysis*. McGraw-Hill, New York, 1973.

[18] R.E. Showalter. *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, volume 49 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, 1996.

[19] L. Simon. *Lectures on Geometric Measure Theory*, volume 3 of *Proceedings of the Centre for Mathematical Analysis*. Centre for Mathematical Analysis, The Australian National University, Canberra, 1984.

[20] G.G. Stokes. On the variation of gravity on the surface of the Earth. *Trans Camb Phil Soc*, 8:672–695, 1849.

[21] J. Wloka. *Partial Differential Equations*. Cambridge University Press, Cambridge, 1992.

Matthias Augustin
Mathematical Image Analysis Group
Fakultät 6
Saarland University
D-66123 Saarbrücken, Germany

Sarah Eberle
Numerical Analysis Group
Mathematical Institute
University of Tübingen
D-72076 Tübingen, Germany

Martin Grothaus
Functional Analysis and Stochastic Analysis Group
Fachbereich Mathematik
University of Kaiserslautern
D-67663 Kaiserslautern, Germany

Handbook of Mathematical Geodesy
Geosystems Mathematics, 201–314

Ill-Posed Problems: Operator Methodologies of Resolution and Regularization

Willi Freeden and M. Zuhair Nashed

Abstract. A general framework of regularization and approximation methods for ill-posed problems is developed. Three levels in the resolution processes are distinguished and emphasized: *philosophy of resolution, regularization-approximation schema*, and *regularization algorithms*. Dilemmas and methodologies of resolution of ill-posed problems and their numerical implementations are examined with particular reference to the problem of finding numerically minimum weighted-norm least squares solutions of first kind integral equations (and more generally of linear operator equations with non-closed range). An emphasis is placed on the role of constraints, function space methods, the role of generalized inverses, and reproducing kernels in the regularization and stable computational resolution of these problems. The thrust of the contribution is devoted to the interdisciplinary character of operator-theoretic and regularization methods for ill-posed problems, in particular in mathematical geoscience.

Keywords. Ill-posed problems, inverse problems, regularization.

1. Introduction

This contribution on operator-theoretic approaches to ill-posed problems (IPP's) develops a general framework for regularization and approximation methods for ill-posed problems. Most inverse problems are ill-posed. For example, gravimetric and downward continuation problems of geodesy are ill-posed. Three levels in the resolution processes are distinguished and analyzed in this expository research paper: *philosophy of resolution, regularization-approximation schema, and regularization algorithms*. Our essential objective is to provide an outlook within

Ill-posed and inverse problems represent classical topics in the research of mathematical geodesy. Hence, the necessity of such a contribution comes naturally. This is the reason why the editors and the publisher have decided to include this chapter here despite the fact that its content has been extracted from W. Freeden, M.Z. Nashed, Operator-Theoretic and Regularization Approaches to Ill-Posed Problems, GEM Int. J. Geomath., Springer, 2017 (https://doi.org/10.1007/s13137-017-0100-0).

which discretization and other approximation methods can be better motivated, interpreted, and understood. Our development will be fairly general in scope and theory, and it is applicable to a wide range of ill-posed problems. Each method for resolution, whether regularized in the classical sense or non-regularized, involves a critical "parameter" whose "optimal value" is crucial to the amenability and numerical implementation of the method. For example, in Tikhonov-type regularization it is the regularization parameter, or more generally the choice of the regularization operator. In projection and other discrete methods, it is the optimal dimension of the approximating subspaces. In discretization methods it is the choice of the mesh size beyond which a further refinement will lead to instability. In multiscale methods it is the scale parameter to determine the scale space in which multiresolution is realizable relative to the data width. In iterative methods it is the level at which one should terminate the iteration (i.e., it is the stopping rule for the iterates). In filteration-truncation methods it is the number of terms to be included, etc. This paper delineates unifying principles that quantify the choice of the parameter, the type of estimates, and a priori information that are needed to arrive at an "optimal" value for this parameter.

Methodologies and dilemmas of the resolution of ill-posed problems and their numerical implementations are examined with particular reference to the problem of finding minimum weighted-norm least squares solutions of linear operator equations with non-closed range. A common problem in all these methods is delineated: Each method reduces the problem of resolution to a "non-standard" minimization problem involving an unknown critical "parameter" whose "optimal" value is crucial to the numerical realization and amenability of the method. The "non-standardness" results from the fact that one does not have explicitly, or a priori, the function to be minimized: It has to built up using additional information, convergence rate estimates, noise characteristics and robustness conditions, etc. Several results are discussed that represent and complement advances in regularization of inverse and ill-posed problems. An emphasis is placed on the role of constraints, function space methods, the role of generalized inverses, and reproducing kernels in the regularization and stable computational resolution of these problems. The thrust of the work is devoted to the interdisciplinary character of operator-theoretic methods for ill-posed problems. It is hoped that the viewpoints and approaches developed in this work for geodetically relevant obligations would be found useful in connection with other ill-posed problems in diverse areas of application. In fact, our purpose is to provide an outlook within which technical results can be better motivated and understood. Within this framework, criteria can be given relative to which the scope and limitations of the various methods can be assessed. This is important both in theory and practice since there is no cure-all method for ill-posed problems; therefore it is imperative to be able to clarify why a certain method works in some context as well as when not to use that method. The work discusses at length the intuitive principles that underlie the various methods and establishes some results within this framework, thereby omitting technicalities of the proofs.

Within the framework of the paper we are mainly interested in pointing out those aspects that are related to generalized inverses in (reproducing kernel) Hilbert spaces and those which are not. Only standard notation from functional analysis is used; reference may be made to any introductory book on functional analysis (e.g., [140, 143]). In a number of ill-posed problems (for example, in mathematical geodesy, the gravimetry problem, the gradiometry problem, etc.), the operator A is an integral operator, and the problem $Ax = y$ is essentially one of "solving" a *Fredholm integral equation of the first kind.* If y belongs to the range of the operator A, we may ask for an exact solution, while in the case in which y fails to belong to the range of A (the case more typically met in applied problems of the type described above), we must confront the fundamental issue of deciding what should be meant by a "solution" and, only then, seek appropriate techniques for the resolution of the problem.

During the past three decades a substantial amount of machinery from functional analysis, theory of special functions, optimization as well as approximation theory and numerical analysis has been brought to bear on the resolution and understanding of IPPs, and the interdisciplinary character of many inverse and ill-posed problems has emerged very clearly. The interdisciplinary character of IPP's in Applied Sciences is also stressed in many survey papers, which also give excellent account of the state of the art for various problems in practice and contain extensive bibliographies. Three problems are essentially treated in the literature:

(1) the *identification* problem,
(2) the *synthesis* (or controllability) problem, and
(3) best *approximate synthesis.*

These problems are all subsumed in the general problem of studying an operator equation of the form $Ax = y$, where A usually is assumed to be an operator with non-continuous inverse.

The vivid research activity in the field of ill-posed and inverse problems has led to a vast literature on inverse and ill-posed problems. We list only a selection of contributions, where the reader is also referred to the literature therein.

textbooks: [14, 22, 60, 63, 105, 106, 124, 132, 137, 142, 148, 169, 170, 173, 177, 178, 248, 250, 254],

conference reports, handbooks: [7, 12, 25, 65, 82, 83, 107, 128, 141, 173, 177, 182, 183, 185, 187–189, 192, 223, 233].

In addition to the books we mention the journals:

Inverse Problems,
Inverse Problems in Science and Engineering,
Inverse Problems and Imaging,
Journal on Inverse and Ill-Posed Problems,
Journal of Mathematical Imaging and Vision,
Mathematical Inverse Problems.

Many authors contributed results in different areas of ill-posed and inverse problems (note that he following list is rather incomplete, for more details the reader is referred to the references in the aforementioned textbooks or the below listed journal and handbook publications):

generalized inverse and least squares problems: [11, 39, 47, 50, 51, 57, 58, 60, 66, 95, 99, 114, 125, 151, 153, 157, 160, 164–166, 173, 175, 178, 182–186, 194, 197, 208, 209, 219, 220, 226, 240, 261],

truncated singular value, Tikhonov regularization, and discrepancy principles: [24–28, 38, 52, 53, 60, 64, 68, 69, 92, 93, 105, 107, 118, 126, 146, 147, 152, 154, 155, 168, 176, 178, 179, 181, 184, 201, 202, 210, 216, 218, 225, 235, 243, 247, 249, 250, 256],

inversion in (*reproducing kernel*) *Hilbert spaces*: [67, 122, 144, 184, 195, 196, 232, 236, 258],

projection methods, moment problems: [6, 7, 34, 35, 52, 53, 61, 137, 180, 184, 199, 241, 262],

iterative methods, finite element methods, other computational methods: [3, 6, 11, 20, 32, 33, 38, 40, 54, 80, 92, 102, 108, 112, 113, 115–117, 133, 134, 138, 172, 184, 200, 204, 214, 242, 246, 251, 253, 255, 263, 266–268],

mollifier methods: [63, 148, 149],

variational methods, implicit function theorems, ill-posed problems in differential equations: [1, 5, 9, 10, 14, 36, 42, 43, 48, 50, 56, 61, 62, 145, 189–191, 206, 215, 224, 230, 252, 257],

multiscale methods: [41, 44, 45, 74–77, 81, 91, 94, 141, 150, 152, 156, 161–163].

The *GEM–International Journal on Geomathematics* is a forum in which geoscientifically relevant ill-posed problems gain appropriate recognition. Many further references will be given in due course.

It is also hoped that the viewpoints and approaches developed in this paper would be found useful in connection with other inverse problems of various (not necessarily geoscientific) research areas.

2. Solvability of ill-posed operator equations

Schematically, a *direct* (*forward*) *problem* can be formulated as follows:

$$\text{object} \longrightarrow \text{data information of the object.}$$

The *inverse problem* is considered the "inverse" to the forward problem which relates the object (sub)information to the object:

$$\text{data information of the object} \longrightarrow \text{object.}$$

An object may be understood to be the systematic relationship of all data subinformation, object parameters, and other auxiliary information. It may be linear or non-linear, deterministic or random, etc.

In standard functional analytic nomenclature (see, e.g., [120, 135, 229, 245, 265]) we are usually confronted with the following operator equation: Given spaces X, Y equipped with the settings of norm and inner product, respectively. Consider a mapping A from X to Y, i.e., $(A; X, Y)$ with

$$A : X \to Y. \tag{2.1}$$

The *Direct Problem* (*DP*) is as follows: Given $x \in X$, find $y = Ax \in Y$. The *Inverse Problem* (*IP*) is as follows: Given an observed output y, find an input x that produces it, i.e., $Ax = y \in Y$, or given a desired output z, find an input x that produces an output $y = Ax \in Y$ that is as "close" to z as possible.

A *Well-Posed* (*Properly-Posed*) *Problem in the sense of Hadamard* is as follows: For each "data" $y \in Y$, the operator equation $X \ni x \mapsto Ax = y \in Y$ has one and only one solution, and the solution depends continuously on y. In more detail, a mathematical problem is well posed in the sense of Hadamard (cf. [109, 110]), if it satisfies the following properties:

(H1) *Existence*: For all (suitable) data, there exists a solution of the problem (in an appropriate sense).
(H2) *Uniqueness*: For all (suitable) data, the solution is unique.
(H3) *Stability*: The solution depends continuously on the data.

According to this definition, *a problem is ill posed or improperly posed in the sense of Hadamard if one of these three conditions is violated.*

As already mentioned, ill-posed problems arise in many branches of science, engineering, and mathematics, including computer vision, natural language processing, machine learning, statistics, statistical inference, medical imaging, remote sensing, non-destructive testing, astronomy, geodesy and geophysics, exploration and prospection, and many other fields. It should be noted that J. Hardamard (1865–1963) dismissed ill-posed problems as irrelevant to physics or real world applications, but he was proven wrong four decades after his declaration. In fact, it turned out that Hadamard's classification had a tremendous influence on the development of mathematics. Some years ago, starting from Hadamard's properties a more relevant understanding of ill-posedness was provided by a more detailed functional analytical background (cf. [185]) that will be explained later on.

2.1. Finite-dimensional matrix equations and generalized inverse

Since any numerical approximation procedure usually leads to finite-dimensional problems involving a singular functional analytic context, we begin with the recapitulation of finite systems of linear equations (see, e.g., [23, 99, 178, 260] and the list of references therein for more details). After having treated the finite-dimensional situation, we turn to the analogous theory in infinite-dimensional operator framework.

Spectral matrix representation. We start with a linear matrix equation of the form

$$Ax = y, \tag{2.2}$$

with $A \in \mathbb{K}^{n\times n}$ ($\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$) being a Hermitian positive definite matrix with n rows and n columns. From the spectral theory it is well known that there exist eigenvalues $0 < \lambda_1 \le \cdots \le \lambda_n$ and a corresponding unitary matrix $U = (u_1, \ldots, u_n)$ of eigenvectors $u_i \in \mathbb{C}^n \backslash \{0\}$ (i.e., $u_i^H u_j = \delta_{i,j}$) such that A has a representation of the form

$$U^H A U \;=\; \underbrace{\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ & & \ddots \\ 0 & & \lambda_n \end{pmatrix}}_{n\times n} \;=\; \operatorname{diag}(\lambda_1, \ldots, \lambda_n). \tag{2.3}$$

The *condition number* of A is given by the quotient of the largest and smallest eigenvalue, i.e., $\kappa = \frac{\lambda_n}{\lambda_1}$ (note that $\lambda_1 > 0$). For the sake of simplicity and coherence with the analysis for the infinite-dimensional case below, we shall assume here that the scaling is such that $\lambda_n = 1$, so that $\kappa = \lambda_1^{-1}$. The condition number is a measure for the stable solvability of the problem (2.2).

Ill-conditioned matrix equations. Assume that we have *noisy data* y^ε instead of y, which satisfy

$$\|y^\varepsilon - y\| \le \varepsilon \tag{2.4}$$

in the Euclidean norm on $\mathbb{K}^n$. Let x^ε denote the solution with right-hand side y^ε. Then it follows from the spectral representation that

$$x^\varepsilon - x \;=\; U \operatorname{diag}\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_n}\right) U^H \,(y^\varepsilon - y). \tag{2.5}$$

Hence, observing the orthogonality of eigenvectors we are led to the estimate

$$\|x^\varepsilon - x\|^2 = \sum_{i=1}^{n} \lambda_i^{-2} |u_i^H (y^\varepsilon - y)|^2 \le \lambda_1^{-2} \|y^\varepsilon - y\|^2. \tag{2.6}$$

In other words, we have

$$\|x^\varepsilon - x\| \le \kappa \|y^\varepsilon - y\| \le \kappa\, \varepsilon. \tag{2.7}$$

The sharpness of this estimate can be immediately seen for $y^\varepsilon - y = \varepsilon u_1$. It is clear that with increasing condition number of the matrix A, the noise amplification increases in the worst case. For large κ one therefore speaks of an *"ill-conditioned problem"* (ICP). But it should be remarked that a finite-dimensional linear problem is never ill posed (in the sense that the third condition in Hadamard's classification is violated), but for κ large one certainly comes close to this case.

We also observe that errors in low frequencies (i.e., corresponding to eigenvectors with large eigenvalues) are amplified less. Following our nomenclature we see that an error in the lowest frequency, i.e., $y^\varepsilon - y = \varepsilon\, u_n$ is not amplified at all. In fact, we just obtain $\|x^\varepsilon - x\| = \varepsilon$ from the spectral representation. This is a typical effect for inverse problems. It means that not all possible versions of noise of the same size are equally bad, high-frequency noise corresponding to low

eigenvalues is always worse than low-frequency noise. However, in practice, we are able to make any assumption on the noise only in rare exceptions, so that a regularization method has to deal with arbitrary noise.

Until now, we have assumed that the matrix A is Hermitian positive definite, i.e., the minimal eigenvalue is positive. If this is not the case, the matrix has a non-trivial null space. If λ_r denotes the minimal non-zero eigenvalue, then the solution formula becomes $x = \sum_{i=r}^{n} \lambda_i^{-1} u_i u_i^H y$, and the problem is solvable if and only if $u_i^H y = 0$ for $i < r$. If the data set is noisy, i.e., instead of y we have y^ε, we are led to use the projection Py^ε onto the range of A. In doing so we obtain for the corresponding solution x^ε with data Py^ε that $x^\varepsilon - x = \sum_{i=r}^{n} \lambda_i^{-1} u_i u_i^H (Py^\varepsilon - y)$. Since $u_i^H P y^\varepsilon = u_i^H y^\varepsilon$ for $i \geq r$ we thus can estimate similarly as described above $\|x^\varepsilon - x\| \leq \lambda_r \varepsilon$. Consequently, there is no error propagation in the null space components and the noise amplification is actually determined by the minimal non-zero eigenvalue.

Matricial generalized inverse (pseudoinverse, Moore–Penrose inverse). Let $A \in \mathbb{K}^{n\times m}$ ($\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$) be a matrix with n rows and m columns, $y \in \mathbb{K}^m$. Note that A is not required to be square, no rank assumptions are made at this stage. Then the linear system

$$Ax = y, \quad x \in \mathbb{K}^n, \tag{2.8}$$

need not have a (unique) solution. If the system (2.8) is unsolvable, a reasonable generalized notion of a solution is a "least squares solution", which minimizes the residual $Ax - y$ in the Euclidean norm (note that $\|\ \|$ stands for the Euclidean norm in this subsection): A vector $x \in \mathbb{K}^n$ is a

(1) *least squares solution of* (2.8) if and only if

$$\|Ax - y\| = \inf\{\|Az - y\| : z \in \mathbb{K}^n\}, \tag{2.9}$$

(2) *best-approximate solution* (*or minimal norm solution*) *of* (2.8) if and only if x is a least squares solution and satisfies

$$\|x\| = \inf\{\|z\| : z \text{ is a least squares solution}\}. \tag{2.10}$$

The following results are well known from classical linear algebra: A vector x^* is a solution of (2.9) if and only if the *normal equations*

$$A^H A x^* = A^H y \tag{2.11}$$

are satisfied. The problem (2.9) possesses a unique solution if and only if A has full rank. If $A^\dagger \in \mathbb{K}^{m\times n}$ may be understood as the matrix which assigns to each $y \in \mathbb{K}^n$ the best-approximate solution of (2.8), then it is called the *generalized inverse* (*also designated, Moore–Penrose inverse or pseudoinverse*) *of* A.

In order to construct $A^\dagger$ and, hence, best-approximate solutions via the so-called *singular value decomposition* (*SVD*) *of* A we recall the definition of *singular values of a matrix* A:

Let $\sigma_1, \ldots, \sigma_r > 0$ be such that $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0$ are the positive eigenvalues of the matrix $A^H A$ (each one written down as often as its multiplicity is). Then $\sigma_1, \ldots, \sigma_r$ are denoted the (non-zero) singular values of A.

This setting makes sense, since $A^H A$ is positive semidefinite. Obviously, $r \leq \min\{n, m\}$, where r is the rank of A. We know that a Hermitian matrix can be diagonalized, where the diagonal elements are its eigenvalues. The following theorem generalizes this result to the non-Hermitian case.

Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ be the singular values of A. Then there exist unitary matrices $U \in \mathbb{K}^{m\times m}$ and $V \in \mathbb{K}^{n\times n}$ such that

$$V^H A U = \underbrace{\begin{pmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{pmatrix}}_{m\times n} \tag{2.12}$$

holds true. The columns of U and V are eigenvectors of $A^H A$ and AA^H, respectively. The expression (2.12) *is the so-called singular value decomposition (SVD) of A.*

The singular value decomposition (SVD) is not unique, since the unitary matrices U and V are not. Obviously, from (2.12), we obtain

$$A = V \begin{pmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{pmatrix} U^H, \tag{2.13}$$

since V and U are invertible, the rank of A is r, the number of non-zero singular values (counted with multiplicity). Note that with $U = (u_1, \ldots, u_m)$, $V = (v_1, \ldots v_n)$, we have for $i \in \{1, \ldots, r\}$

$$Au_i = \sigma_i v_i \tag{2.14}$$

and

$$A^H v_i = \sigma_i u_i, \tag{2.15}$$

which follows from the singular value decomposition (2.12) via multiplication by V and U^H, respectively. The system $\{(\sigma_i; u_i, v_i) : i \in \{1, \ldots, r\}\}$ is the so-called *singular system* for A. The system $\{v_1, \ldots, v_r\}$ is an orthonormal basis for the

range $\mathcal{R}(A) = \{y : Ax = y\}$, for any $x \in \mathbb{K}^n$,

$$Ax = \sum_{i=1}^{r} \langle Ax, v_i \rangle v_i = \sum_{i=1}^{r} \langle x, A^H v_i \rangle v_i, \tag{2.16}$$

which implies in connection with (2.15) that

$$Ax = \sum_{i=1}^{r} \sigma_i \langle x, u_i \rangle v_i \tag{2.17}$$

holds true. Analogously, for all $y \in \mathbb{K}^m$,

$$A^H y = \sum_{i=1}^{r} \sigma_i \langle y, v_i \rangle u_i. \tag{2.18}$$

Note that if A has real entries, so U and V have. The notion of a singular system and the expansions (2.17) and (2.18) generalize to compact operators on infinite-dimensional spaces, e.g., integral operators, as we will see later on.

Let A have the SVD (2.12). *Then*

$$A^\dagger = U \underbrace{\begin{pmatrix} \frac{1}{\sigma_1} & & & & 0 \\ & \ddots & & & \\ & & \frac{1}{\sigma_r} & & \\ & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{pmatrix}}_{n \times m} V^H. \tag{2.19}$$

This also implies the existence and uniqueness of a best-approximate solution. Since $A^\dagger b$ is the least squares solution of minimal norm, we obtain that $A^\dagger b$ is a solution of the normal equations $A^H Ax = A^H b$ with minimal norm, i.e.,

$$A^\dagger y = (A^H A)^\dagger A^H y. \tag{2.20}$$

This means that in order to approximate $A^\dagger b$ we may as well compute an approximation to the minimal-norm solution in the normal equations, a fact we will heavily use in the construction of regularization methods, later on.

Historical remarks. It should be mentioned that during the last century, the concept of a pseudoinverse (generalized inverse) has rated considerable attention in the mathematical as well as geodetic literature (a bibliography, for example, listing over 1700 references on the subject is due to [178]). One of the most significant applications of generalized inverses is to problems of best fit. Therefore one might seek such evidence in the writings of those who laid the foundations of the method of least squares. C.F. Gauss developed the method of least squares in 1794, but he did not publish his results until several years later (see, e.g., [55, 97, 217] for a review of the subject). Gauss's interest in the subject may be dated back to his

considerations of problems in geodesy. One should point out that Gauss [87, 88] did not formally display $A^\dagger$. However, following [226], the ingredients for the construction of a generalized inverse were essentially available to him, but he did not use them. Indeed, there appears to be no evidence that he was inclined to proceed in that direction. On the other hand, his approach to the problem of determining best estimates is certainly in the spirit of generalized inverses. Early interest in the first half of the last century in the subject of generalized inverses was initiated by a paper on matrices by R. Penrose [208]. Indeed, basic elements of this concept had been considered somewhat earlier. For example, E.H. Moore [165] presented a development of the notion (see also R. Baer [17], A. Bjerhammar [29, 30], K. Friedrichs [84], F.Helmert [119], E.H. Moore [166], C.R. Rao, S.K. Mitra [220], C.L. Siegel [238], and H. Wolf [261]). Moreover, in the setting of integral and differential operators the concept was considered even earlier by I. Fredholm [70] and W.A. Hurwitz [127], and by D. Hilbert [121] (see [223] for a discussion of generalized inverses in classical analysis, and see also [23, 31, 178] for brief historical sketches of the subject).

Truncated singular value regularization. The decomposition (2.19), more concretely, the identity

$$A^\dagger y = \sum_{i=1}^{r} \frac{\langle y, v_i \rangle}{\sigma_i}\, u_i \tag{2.21}$$

also shows how errors in y affect the result $A^\dagger y$: Errors in components of y corresponding to small singular values are amplified by the large factor of the singular value, so that such data errors are dangerous. This explains the numerical instability of (2.21), if A has small singular values. Although the problem of computing the best-approximate solution is well posed, it is then numerically unstable. The first idea to reduce this instability is to replace (2.21) by

$$x_\alpha = \sum_{\substack{i=1 \\ \sigma_i^2 \geq a}}^{r} \frac{\langle y, v_i \rangle}{\sigma_i}\, u_i \tag{2.22}$$

with an appropriately chosen value $\alpha > 0$; this truncation is the first example of a regularization, where the original problem is replaced by a neighboring one, which is more stable. However, the choice of the "regularization parameter" α is quite crucial.

If we use (2.22) with perturbed data y^ε (with $\|y - y^\varepsilon\| \leq \varepsilon$), we obtain as the "regularized solution"

$$x_\alpha^\varepsilon = \sum_{\substack{i=1 \\ \sigma_i^2 \geq \alpha}}^{r} \frac{\langle y^\varepsilon, v_i \rangle}{\sigma_i} u_i. \tag{2.23}$$

We estimate the total error between x_α^ε and the sought-for quantity $A^\dagger y$:

$$\begin{aligned}\|x_\alpha^\varepsilon - A^\dagger y\| &= \left\| \sum_{\substack{i=1\\ \sigma_i^2 \geq \alpha}}^{r} \frac{\langle y^\varepsilon, v_i\rangle}{\sigma_i} u_i - \sum_{i=1}^{r} \frac{\langle y, v_i\rangle}{\sigma_i} u_i \right\| \\ &= \left\| \sum_{i=1}^{r} \frac{\langle y, v_i\rangle}{\sigma_i} u_i - \sum_{\substack{i=1\\ \sigma_i^2 \geq \alpha}}^{r} \frac{\langle y, v_i\rangle}{\sigma_i} u_i \right\| + \left\| \sum_{\substack{i=1\\ \sigma_i^2 \geq \alpha}}^{r} \frac{\langle y, v_i\rangle - \langle y^\varepsilon, v_i\rangle}{\sigma_i} u_i \right\| .\end{aligned} \tag{2.24}$$

Since the elements u_i are orthonormal, we have

$$\left\| \sum_{i=1}^{r} \frac{\langle y, v_i\rangle}{\sigma_i} u_i - \sum_{\substack{i=1\\ \sigma_i^2 \geq \alpha}}^{r} \frac{\langle y, v_i\rangle}{\sigma_i} u_i \right\| = \left\| \sum_{\substack{i=1\\ \sigma_i^2 < \alpha}}^{r} \frac{\langle y, v_i\rangle}{\sigma_i} u_i \right\|^2 = \sum_{\substack{i=1\\ \sigma_i^2 < \alpha}}^{r} \frac{|\langle y, v_i\rangle|^2}{\sigma_i^2} \to 0 \tag{2.25}$$

for $\alpha \to 0$. Hence, for sufficiently small α, the last sum is empty such that

$$\left\| \sum_{\substack{i=1\\ \sigma_i^2 \geq \alpha}}^{r} \frac{\langle y, v_i\rangle - \langle y^\varepsilon, v_i\rangle}{\sigma_i} u_i \right\|^2 = \sum_{\substack{i=1\\ \sigma_i^2 \geq \alpha}}^{r} \frac{|\langle y - y^\varepsilon, v_i\rangle|^2}{\sigma_i^2} \leq \frac{1}{\alpha} \sum_{i=1}^{r} |\langle y - y^\varepsilon, v_i\rangle|^2 \leq \frac{\varepsilon^2}{\alpha}. \tag{2.26}$$

The second error term does not blow up as $\alpha \to 0$, since the sum has always at most r terms, hence, it can be estimated by $\varepsilon^2\ (\min\{\sigma_i^2 : i \in \{1, \ldots, r\}\})^{-1}$. The sum (2.23) is called *truncated singular value expansion*. It can be interpreted as applying a low-pass filter to the data.

Tikhonov regularization. Another way of making (2.21) $\to$ (2.22) more stable would be to replace it by the sum

$$x_\alpha^\varepsilon = \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 + \alpha} \langle y^\varepsilon, v_i\rangle u_i. \tag{2.27}$$

This is a classical variant of the famous *Tikhonov regularization method* for matrix equations, which we shall consider in more detail in infinite dimensions. It is helpful to characterize it in a different way: Let x_α^ε be defined by (2.27). Then, by the orthonormality of the u_i, we have, for all $j \in \{1, \ldots, r\}$,

$$\langle x_\alpha^\varepsilon, u_j\rangle = \frac{\sigma_j}{\sigma_j^2 + \alpha} \langle y^\varepsilon, v_j\rangle \tag{2.28}$$

and

$$\sigma_j^2 \langle x_\alpha^\varepsilon, u_j\rangle + \alpha \langle x_\alpha^\varepsilon, u_j\rangle = \sigma_j \langle y^\varepsilon, v_j\rangle. \tag{2.29}$$

Now, because of (2.17) and (2.18), it follows that

$$\sum_{j=1}^{r} \sigma_j \langle y^\varepsilon, v_j \rangle u_j = \sum_{j=1}^{r} \langle y^\varepsilon, Au_j \rangle u_j = \sum_{j=1}^{r} \langle A^H y^\varepsilon, u_j \rangle u_j = A^H y^\varepsilon \tag{2.30}$$

and

$$\begin{aligned} \sum_{j=1}^{r} \left(\sigma_j^2 \langle x_\alpha^\varepsilon, u_j \rangle + \alpha \langle x_\alpha^\varepsilon, u_j \rangle \right) u_j &= \sum_{j=1}^{r} \left(\langle x_\alpha^\varepsilon, A^H A u_j \rangle + \alpha \langle x_\alpha^\varepsilon, u_j \rangle u_j \right) \\ &= A^H A x_\alpha^\varepsilon + \alpha x_\alpha^\varepsilon, \end{aligned} \tag{2.31}$$

note that x_α^ε is in the linear span of $\{u_1, \ldots, u_r\}$ which follows from (2.27). Now, the identity (2.29) implies in vector nomenclature that

$$\left(A^H A + \alpha I \right) x_\alpha^\varepsilon = A^H y^\varepsilon, \tag{2.32}$$

which is an alternative characterization of the *Tikhonov regularization*. From this "*regularized normal equation*", we can obtain still another characterization of x_α^ε, namely as the unique minimizer of the so-called *Tikhonov functional*

$$x \mapsto \|Ax - y^\varepsilon\|^2 + \alpha \|x\|^2, \tag{2.33}$$

which can be seen by putting the first derivative of the functional in (2.33) to 0, resulting exactly in the linear equation (2.32). The minimization of (2.33) can be seen as a combination of the two minimizations that appear in the definition of a best-approximate solution. It has also interpretations as a penalty method, e.g., via Lagrange multipliers.

The computation explained above can be formally performed for $\alpha = 0$, too. In this case it shows that x is the solution of minimal norm of the normal equation $A^H Ax = A^H y$ which was already attacked by C.F. Gauss [87] (see also the contributions by R.L. Plackett [217] and D.W. Robinson [226], and for a deeper insight [178]). The monograph [178] also contains a brief historical sketch of this subject. The book [260] (see also the references therein) presents an overview about numerical methods and procedures. If $A^H A$ is invertible (and hence positive definite), the normal equation $A^H Ax = A^H y$ can be solved by standard Cholesky decomposition, which leads to an alternative method for computing best-approximation solutions, for which no SVD is needed.

2.2. Least squares problems and generalized inverses

As usual, the domain, range, and null space (kernel) of any operator $A : \mathcal{D}(A) \to \mathcal{R}(A)$ are denoted by $\mathcal{D}(A), \mathcal{R}(A)$, and $\mathcal{N}(A)$, respectively.

We start with solvability conditions of an operator equation in Hilbert spaces, where the concepts of "distance" and "angle" are at the disposal for an applicant. More concretely, let X and Y be Hilbert spaces and let $A : X \to Y$ be a bounded

linear operator whose range $\mathcal{R}(A)$ is not necessarily closed. Then we have the orthogonal decompositions

$$X = \mathcal{N}(A) \oplus \mathcal{N}(A)^{\perp}, \tag{2.34}$$

$$Y = \overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^{\perp}, \tag{2.35}$$

and

$$\mathcal{N}(A^*) = \mathcal{R}(A)^{\perp}, \tag{2.36}$$

where $\mathcal{N}(A)^{\perp}$ is the orthogonal complement of $\mathcal{N}(A)$, $\overline{\mathcal{R}(A)}$ is the closure of the range of A, and A^* is the *adjoint operator of* A, i.e., $\langle Ax, y\rangle = \langle x, A^*y\rangle$ for all $x \in X$ and $y \in Y$.

We consider the operator equation

$$Ax = y. \tag{2.37}$$

Four (mutually exclusive) situations arise (cf. [184]):

1) $\mathcal{R}(A)$ is dense in Y, (hence, $\mathcal{N}(A^*) = \{0\}$), and $y \in \mathcal{R}(A)$;
2) $\mathcal{R}(A)$ is dense in Y, and $y \notin \mathcal{R}(A)$;
3) $\overline{\mathcal{R}(A)}$ is a proper subspace of Y, and $y \in \mathcal{R}(A) + \mathcal{R}(A)^{\perp}$;
4) $\overline{\mathcal{R}(A)} \neq Y$, and $y \notin \mathcal{R}(A) + \mathcal{R}(A)^{\perp}$.

In case 1) one has, of course, a solution in the classical sense; in case 2) and 4) a classical solution does not exist, while in case 3) a solution need not exist.

We say x is a *"least squares solution" of* (2.37) if

$$\inf\{\|Au - y\| : u \in X\} = \|Ax - y\|. \tag{2.38}$$

Since

$$\|Au - y\|^2 = \|Au - Qy\|^2 + \|y - Qy\|^2, \tag{2.39}$$

where Q is the orthogonal projector of Y onto $\overline{R(A)}$, it is clear that a least squares solution exists if and only if

$$y \in \mathcal{R}(A) + \mathcal{R}(A)^{\perp}, \tag{2.40}$$

where $\mathcal{R}(A) + \mathcal{R}(A)^{\perp}$ is a dense set in Y. For such y the set of all least squares solutions of (2.37), denoted by $\mathcal{L}(y)$, is a non-empty closed convex set (indeed $\mathcal{L}(y)$ is the translate of $\mathcal{N}(A)$ by a fixed element of $\mathcal{N}(y)$), hence, it has a unique element of minimal norm, denoted by $A^{\dagger}y$.

The *generalized inverse* (*or pseudoinverse*) $A^{\dagger}$ is the linear operator which assigns to each $y \in \mathcal{D}(A^{\dagger}) := \mathcal{R}(A) + \mathcal{R}(A)^{\perp}$, the unique element in $\mathcal{L}(y) \cap \mathcal{N}(A)^{\perp}$, so that $\mathcal{L}(y) = A^{\dagger}y + \mathcal{N}(A)$. It is easy to show that $A^{\dagger}y$ is the minimal norm solution (equivalently the unique solution in $\mathcal{N}(A)^{\perp}$) of the *normal equation*

$$A^*Ax = A^*y \tag{2.41}$$

(the equation obtained by setting the first variation of $\|Ax - y\|^2$ equal to zero). It also follows that $A^{\dagger} = (A/\mathcal{N}(A)^{\perp})^{-1}Q$ so that $A^{\dagger}$ can be characterized as the linear operator with the function-theoretic properties:

$$\mathcal{D}(A^{\dagger}) = \mathcal{R}(A) + \mathcal{R}(A)^{\perp}, \quad \mathcal{N}(A^{\dagger}) = \mathcal{R}(A)^{\perp} = \mathcal{N}(A^*) \tag{2.42}$$

and

$$\mathcal{R}(A^\dagger) = \mathcal{N}(A)^\perp. \tag{2.43}$$

The equivalence of these characterizations of $A^\dagger$ is established in [173] (see also [104, 185] for a lucid exposition and [185, 194] for generalization to unbounded operators).

In case 1) above, $A^\dagger$ gives the minimal-norm solution of 3). In case 3), Equation (2.37) has a least squares solution (which is unique if and only if $\mathcal{N}(A) = \{0\}$). In both cases the infimum in (2.38) is attained and is equal to zero and $\|y - Qy\|$, respectively. Case 2) and 4) are pathological and usually are not under discussion in generalized inverse theory, since in both cases $y \notin \mathcal{D}(A^\dagger)$, and the infimum in (2.38) is not attained.

As canonical evolution of Hadamard's classification, M.Z. Nashed [178, 184] called the operator equation (2.37) *well posed in the least squares* (*relative to* X *and* Y) if for each $y \in Y$ the equation has a unique least squares solution (of minimal norm), which depends continuously on y; otherwise the problem is ill posed. The advantage of adopting this notion of well-posedness is that it focuses on infinite-dimensional problems (e.g., an inconsistent finite system of linear algebraic equations will not be ill posed in above sense, while it is ill-posed in the sense of Hadamard). It follows immediately from the open mapping theorem in functional analysis (see, e.g., [245]) that the following statements are equivalent:

a) the problem (2.37) is well posed;
b) $\mathcal{R}(A)$ is closed;
c) $A^\dagger$ is bounded.

Summarizing we are led to the following conclusion (see [16, 184, 185]): The problem $(A; X, Y)$ is called *well posed in the sense of Nashed*, if $\mathcal{R}(A)$ is closed in Y. If $\mathcal{R}(A)$ is not closed in Y, the problem $(A; X, Y)$ is called *ill posed in the sense of Nashed*.

2.3. Weighted least squares problems

Very often we are interested in weighted minimal-norm least squares solutions. Let $\mathcal{L}_A(y)$ be the set of all least squares solutions of $Ax = y$, where A is a bounded linear operator from X into Y. Let Z be a Hilbert space and $L : \mathcal{D}_L \subset X \to Z$ be a closed linear operator with dense domain and closed range. For $y \in \mathcal{D}(A^\dagger)$, we consider the following problem: find $w \in \mathcal{L}_A(y)$ such that

$$\|Lw\| \leq \|Lu\| \quad \text{for all } u \in \mathcal{L}_A(y). \tag{2.44}$$

If $L(\mathcal{N}(A))$ is closed and $\mathcal{N}(A) \cap \mathcal{N}(L) = \{0\}$, then (2.44) has a unique solution $w(y)$. We denote by $A_L^\dagger$ the linear map induced by $y \to w(y)$ and call it the *weighted generalized inverse* of A We define a new inner product and norm on $\mathcal{D}(L)$ by

$$[u, v]_L := (Au, Av) + (Lu, Lv) \tag{2.45}$$

with $\|u\|_L := \sqrt{[u,u]_L}$. We denote the space $\mathcal{D}(L)$ with this new inner product by X_L. It is easy to show that $A_L^\dagger y$ is the least squares solution of X_L-minimal norm of $Ax = y$. Let

$$\mathcal{M} := \{x \in X : L^*Lx \in \mathcal{N}(A)^\perp\}. \tag{2.46}$$

Then $\mathcal{M}$ is the orthogonal complement of $\mathcal{N}(A)$ with respect to $[\cdot,\cdot]$, and $\mathcal{R}(A_L^\dagger) = \mathcal{M}$.

2.4. Singular value decomposition for compact operators

Next we discuss a certain set of operators, viz. compact operators, that turn out to be specific prototypes for generating a large class of ill-posed problems.

Let X, Y be normed spaces. An operator $A : X \to Y$ is called compact*, if one of the following equivalent conditions is fulfilled:*

(1) *Every bounded subset $U \subset X$ possesses an image in Y, which is relatively compact, i.e., $\overline{A(U)}$ is a compact set.*
(2) *For every bounded sequence $\{x_n\}_n \subset X$ the sequence $\{Ax_n\}_n$ possesses a subsequence that converges in Y.*

By convention, we introduce the following settings:

$$K(X,Y) = \{A : X \to Y : A \text{ is linear and compact}\} \tag{2.47}$$

and

$$K(X) = K(X,X). \tag{2.48}$$

Let X, Y, Z be normed spaces. Then the following statements hold true:

(1) $K(X,Y) \subset L(X,Y)$.
(2) If $A \in L(X,Y)$ with $\dim \mathcal{R}(A) < \infty$, then A is compact.
(3) If $A : X \to Y$ is compact and $B : Y \to Z$ is continuous or $A : X \to Y$ is continuous and $B : Y \to Z$ is compact, then $AB : X \to Z$ is compact.
(4) The identity operator $I : X \to X$ is compact if and only if X is finite-dimensional.
(5) Let Y be a Banach space. Then $K(X,Y)$ is closed, i.e., a sequence of compact operators $\{A_n\}_n \subset K(X,Y)$ $\lim_{n\to\infty} \|A_n - A\|_{X\to Y} = 0$ has a compact limit, i.e., the limit operator A is compact.
(4) If A is compact and invertible and X is not finite-dimensional, then A^{-1} is *not* continuous.

Example 1. Let $\mathcal{G}$ be a *regular region in* $\mathbb{R}^q$, i.e., a bounded region $\mathcal{G}$ dividing $\mathbb{R}^q$ uniquely into the inner space $\mathcal{G}$ and the outer space $\mathcal{G}^c = \mathbb{R}^3 \backslash \overline{\mathcal{G}}$, $\overline{\mathcal{G}} = \mathcal{G} \cup \partial\mathcal{G}$, such that the boundary $\partial\mathcal{G}$ is an orientable smooth Lipschitzian manifold of dimension $q-1$, and suppose that K is of class $C^{(0)}\left(\overline{\mathcal{G}} \times \overline{\mathcal{G}}\right)$. We introduce the integral operator $A : C^{(0)}\left(\overline{\mathcal{G}}\right) \to C^{(0)}\left(\overline{\mathcal{G}}\right)$ by letting

$$(AF)(x) = \int_{\mathcal{G}} K(x,y) F(y)\, dy, \quad F \in C^{(0)}\left(\overline{\mathcal{G}}\right). \tag{2.49}$$

The proof of the compactness of A can be based on a well-known theorem of classical analysis, namely the *Theorem of Arzelà–Ascoli*. This theorem provides two equivalent properties to the relative compactness of a subset $U \subset C^{(0)}\left(\overline{\mathcal{G}}\right)$:

Let $\emptyset \neq \mathcal{G}$ be regular. A subset $U \subset C^{(0)}\left(\overline{\mathcal{G}}\right)$ is relatively compact if and only if the following two statements are valid:

(1) *U is equicontinuous, i.e., for every $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$, such that for all $F \in U$*

$$|F(x) - F(y)| < \varepsilon \tag{2.50}$$

for all $x, y \in G$ with $\|x - y\| < \delta(\varepsilon)$.

(2) *U is bounded, i.e., there exists an $M > 0$ with $\|F\|_{C^{(0)}\left(\overline{\mathcal{G}}\right)} < M$ for all $F \in U$.*

In accordance with the theorem of Arzelà–Ascoli we are now interested in applying the conditions (1) and (2) to the context of the integral operator introduced in (2.49):

(1) K is uniformly continuous in $\overline{\mathcal{G}} \times \overline{\mathcal{G}}$. Therefore, AU is equicontinuous.

(2) Suppose that $U \subset C^{(0)}(\overline{\mathcal{G}})$ is *bounded*. Assume that F is of class U. Then we have

$$|AF(x)| = \left| \int_{\mathcal{G}} K(x,y) F(y)\, dy \right| \leq M \sup_{x,y \in \overline{\mathcal{G}}} |K(x,y)| \, \|\mathcal{G}\| < \infty. \tag{2.51}$$

In other words, AU is bounded.

As a consequence, by virtue of the theorem of Arzelà–Ascoli, we are able to conclude that AU is relatively compact, so that the operator A defined by (2.49) is compact.

Example 2. Let $\mathcal{G}$ be a regular region in $\mathbb{R}^q$, and K be of class $L^2\left(\overline{\mathcal{G}} \times \overline{\mathcal{G}}\right)$, then $A : L^2\left(\overline{\mathcal{G}}\right) \to L^2\left(\overline{\mathcal{G}}\right)$ given by

$$AF = \int_{\mathcal{G}} K(\cdot, y) F(y)\, dy, \quad F \in L^2\left(\overline{\mathcal{G}}\right) \tag{2.52}$$

is compact (the proof can be found in, e.g., [120, 139]).

Example 3. Let $\mathcal{G}$ be a regular region in $\mathbb{R}^q$. We introduce the operator $A : L^2\left(\overline{\mathcal{G}}\right) \to L^2\left(\overline{\mathcal{G}}\right)$ given by

$$AF(x) = \int_{\mathcal{G}} K(x,y) F(y)\, dy, \quad F \in L^2(\mathcal{G}). \tag{2.53}$$

If K is continuous for $x \neq y$ and *weakly singular*, i.e., there exist a value $\alpha \in (0, q)$ and a constant $C > 0$ such that

$$|K(x,y)| \leq C \frac{1}{|x-y|^{q-\alpha}}, \tag{2.54}$$

then A is compact (for the proof see, e.g., [135]). As a consequence, the *Newton volume integral*

$$V(x) = AF(x) = \frac{1}{4\pi} \int_{\mathcal{G}} \frac{1}{|x-y|} F(y)\, dy, \quad F \in L^2(\mathcal{G}), \tag{2.55}$$

occurring in the so-called *inverse gravimetry problem* of determining the geologic density distributions F inside the Earth from the Earth's gravitational potential V in $\overline{\mathcal{G}^c} \subset \mathbb{R}^3$ forms a compact operator A.

Singular value decomposition. Next we are concerned with the introduction of eigenvalues and eigenfunctions corresponding to an operator $A \in L(X)$. Let X be a normed space. Suppose that A is of class $L(X)$.

(1) $\lambda \in \mathbb{C}$ is called a *regular value* of A if and only if $\lambda I - A$ is continuously invertible. $\rho(A) = \{\lambda \in \mathbb{C} : \lambda I - A$ continuously invertible$\}$ is called the *resolvent set*.
(2) $\sigma(A) = \mathbb{C} \setminus \rho(A)$ is called the *spectrum* of A.
(3) $\lambda \in \sigma(A)$ is called an *eigenvalue* of A if $\mathcal{N}(\lambda I - A) \neq \{0\}$. The elements of $\mathcal{N}(\lambda I - A) \setminus \{0\}$ are called *eigenvectors* of A corresponding to the eigenvalue λ.

The following results are standard for a Banach space X (see, e.g., [120]):

(1) If $\lambda \in \sigma(A)$, then $|\lambda| \leq \|A\|$, i.e., the spectrum is bounded.
(2) $\sigma(A) \subset \mathbb{C}$ is compact.

Let X be a normed space. Suppose that A is a compact operator on X (i.e., $A \in K(X)$).

(1) If $\lambda \in \sigma(A) \setminus \{0\}$, then λ is an eigenvalue of A.
(2) If λ is an eigenvalue of A, then $\dim \mathcal{N}(\lambda I - A) < \infty$.
(3) $\sigma(A)$ is at most countable. Furthermore, $0 \in \sigma(A)$.
(4) 0 is the only accumulation point of $\sigma(A)$.

Central in our considerations about compact operator is the following *spectral theorem for compact self-adjoint operators* that can be seen in parallel to the finite-dimensional case of matrix operators:

Let X be a Hilbert space. Assume that A is of class $K(X)$ and that A is self-adjoint, i.e., $A^* = A$. Then there exists an orthonormal system $\{x_i\}_{i \in \mathbb{N}} \subset X$ and a sequence $\{\mu_i\}_{i \in \mathbb{N}} \subset \mathbb{R}$ (finite or countable) with $|\mu_1| \geq |\mu_2| \geq \cdots > 0$, such that

$$Ax = \sum_{i=1}^{\infty} \mu_i \langle x, x_i \rangle x_i \tag{2.56}$$

holds true for all $x \in X$.

Suppose that X and Y are Hilbert spaces. Furthermore, let A be of class $K(X, Y)$. Then A^*A is also compact and obviously self-adjoint. Due to the spectral theorem there exist a sequence $\{\lambda_i\}_{i \in \mathbb{N}} \subset \mathbb{R}$ and an orthonormal system $\{x_i\}_{i \in \mathbb{N}} \subset X$ such that

$$A^*Ax = \sum_{i=1}^{\infty} \lambda_i \langle x, x_i \rangle x_i, \quad x \in X. \tag{2.57}$$

Suppose that $\lambda_i \in \sigma(A^*A) \setminus \{0\}$ and denote, as usual, by x_i its corresponding eigenvector. It follows that

$$\lambda_i \|x_i\|^2 = \lambda_i \langle x_i, x_i \rangle = \langle \lambda_i x_i, x_i \rangle_X = \langle A^*Ax_i, x_i \rangle_X = \langle Ax_i, Ax_i \rangle_Y = \|Ax_i\|_Y^2 . \tag{2.58}$$

Therefore we are able to conclude that $\lambda_i > 0$.

Singular values. In the sequel, we assume that the eigenvalues are listed in the chronological order as follows:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_i \geq \lambda_{i+1} \geq \cdots \geq 0. \tag{2.59}$$

Set $\sigma_j = \sqrt{\lambda_j}$. Moreover, let $y_i = \frac{1}{\sigma_i} Ax_i$ i.e., $Ax_i = \sigma_i y_i$, $i \in \mathbb{N}$, and

$$A^*y_i = A^* \left(\frac{1}{\sigma_i} Ax_i \right) = \frac{1}{\sigma_i} A^*Ax_i = \frac{1}{\sigma_i} \lambda_i x_i = \sigma_i x_i. \tag{2.60}$$

It is not hard to see that

$$\begin{aligned} \langle y_i, y_k \rangle_Y &= \frac{1}{\sigma_i \sigma_k} \langle Ax_i, Ax_k \rangle_Y = \frac{1}{\sigma_i \sigma_k} \langle A^*Ax_i, x_k \rangle_X = \frac{1}{\sigma_i \sigma_k} \langle \lambda_i x_i, x_k \rangle_X \\ &= \frac{\lambda_i}{\sigma_i \sigma_k} \langle x_i, x_k \rangle_X = \frac{\sigma_i}{\sigma_k} \delta_{i,k} = \delta_{i,k}. \end{aligned} \tag{2.61}$$

Thus, $\{y_i\}_{i \in \mathbb{N}} \subset Y$ forms a complete orthonormal system (ONS) in $\overline{\mathcal{R}(A)}$, so that the system $\{x_i\}_{i \in \mathbb{N}}$ is a complete ONS in $\mathcal{N}(A)^\perp$. Now, assume that x is a member of $\mathcal{N}(A)^\perp$. Then it follows that

$$x = \sum_{i=1}^{\infty} \langle x, x_i \rangle_X \, x_i \tag{2.62}$$

and

$$Ax = \sum_{i=1}^{\infty} \langle x, x_i \rangle_X \, Ax_i = \sum_{i=1}^{\infty} \sigma_i \langle x, x_i \rangle_X \, y_i \tag{2.63}$$

for all $x \in \mathcal{N}(A)^\perp$.

Let X, Y be Hilbert spaces. The set $\{\sigma_i; x_i, y_i\}_{i \in \mathbb{N}} \subset (0, \infty) \times X \times Y$ is called the singular system of an operator $A \in K(X, Y)$. The values σ_i are called the singular values of A. The elements x_i, y_i are called the singular vectors. Furthermore, the series

$$Ax = \sum_{i=1}^{\infty} \sigma_i \langle x, x_i \rangle_X \, y_i, \;\; x \in X \tag{2.64}$$

is called the singular value decomposition (SVD) of A.

Picard condition. The following condition plays an essential role in the solvability of inverse problems.

If $A : X \to Y$ is compact with singular value decomposition (SVD)

$$\{\sigma_i; x_i, y_i\}_{i \in \mathbb{N}},$$

then $y \in \overline{\mathcal{R}(A)}$ *is an element of* $\mathcal{R}(A)$ *if and only if*

$$\sum_{i=1}^{\infty} \frac{|\langle y, y_i\rangle|^2}{\sigma_i^2} \tag{2.65}$$

is convergent.

Obviously, from the Picard condition, it follows that $\sigma_i^{-2}|\langle y, y_i\rangle|^2 \to 0$ if $i \to \infty$ so that information about the decay of the Fourier coefficients of an element y becomes available.

Let A *be a compact operator* (*i.e.,* $A \in K(X,Y)$) *with SVD* $\{\sigma_i; x_i, y_i\}_{i\in\mathbb{N}}$. *Then the generalized inverse* (*or pseudoinverse*) *of a compact operator is representable in the form*

$$A^{\dagger}y = \sum_{i=1}^{\infty} \frac{1}{\sigma_i} \langle y, y_i\rangle\, x_i \tag{2.66}$$

for all $y \in \mathcal{D}\left(A^{\dagger}\right)$.

If $\mathcal{R}(A)$ is finite dimensional (i.e., there exist only finitely many element y_i), then $\mathcal{R}\left(A^{\dagger}\right) < \infty$. Therefore, $A^{\dagger}$ is compact. In particular, $A^{\dagger}$ is continuous.

The representation of the generalized inverse in terms of the singular value decomposition (2.66) opens the perspective to classify ill-posed problems. Indeed, the summands $\sigma_i^{-1}\langle y, y_i\rangle x_i$ occurring in the series (2.66) depend closely on the singular values. If the values σ_i are small, then the contribution by the series (2.66) becomes large. The existence of SVD can be guaranteed for all compact operators. Nevertheless, the concrete knowledge of SVD is critical. Only in rare exceptions, SVD is explicitly known in practice.

Regularization methods. Obviously, the first two criteria (H1) and (H2) determining a well-posed problem in the sense of Hadamard, can always be enforced by considering the generalized inverse $A^{\dagger}$. A violating of the third criterion, i.e., instability arises if the spectrum of the operator A is not bounded away from zero. Thus, it seems to be natural to construct regularizing approximations via modifying the smallest singular values. In accordance with the singular value decomposition of the generalized inverse, it follows that such a modification of small values and, hence, a construction of regularization operators can be obtained in the form

$$x_\alpha = R_\alpha y = \sum_{i=1}^{\infty} \sigma_i F_\alpha(\sigma_i^2)\langle y, y_i\rangle x_i \quad y \in Y, \tag{2.67}$$

with some function $F_\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$F_\alpha(\lambda) \to \frac{1}{\lambda}, \quad \lambda > 0,\ \alpha \to 0. \tag{2.68}$$

Such an operator R_α as defined by (2.67) and (2.68) may be understood as a *regularization operator* if

$$\lambda|F_\alpha(\lambda)| \le C_{F_\alpha} < \infty, \quad \lambda > 0. \tag{2.69}$$

If (2.69) is satisfied, then we are able to see that

$$\|R_\alpha y\|^2 = \sum_{i=1}^{\infty} \sigma_i^2 (F_\alpha(\sigma_i))^2 |\langle y, y_i\rangle|^2 \le C_{F_\alpha}^2 \sum_{i=1}^{\infty} |\langle y, y_i\rangle|^2 \le C_{F_\alpha}^2 \|y\|^2, \tag{2.70}$$

where C_{F_α} is a bound for the norm of R_α. Note that the pointwise convergence of F_α immediately implies the pointwise convergence of R_α to $A^\dagger$.

Truncated singular value regularization. Let A be a compact operator (i.e., $A \in K(X,Y)$) with SVD $\{\sigma_i; x_i, y_i\}_{i\in\mathbb{N}}$. The main idea of truncated singular value decomposition is to ignore all singular values below a certain threshold value, which we can identify with the regularization parameter α, hence, the representation of the regularized solution is given by

$$F_\alpha(\lambda) = \begin{cases} \frac{1}{\lambda}, & \lambda \ge \alpha \\ 0, & \lambda < \alpha \end{cases} \tag{2.71}$$

i.e.,

$$x_\alpha = R_\alpha y = \sum_{\sigma_i \ge \alpha} \frac{1}{\sigma_i} \langle y, y_i\rangle \; x_i, \quad y \in Y, \tag{2.72}$$

which explains the name truncated singular value decomposition, since all terms in the sum corresponding to small singular values are truncated. Since 0 is the only accumulation point of the singular values of a compact operator, the sum in (2.72) is always finite for $\alpha > 0$. In particular, only a finite number of singular values and singular vectors has to be computed in order to realize this method. On the other hand it should be mentioned that, for α being sufficiently small, the number of singular values that need to be computed can increase strongly. Obviously, $C_{F_\alpha} = \alpha^{-1}$.

Lavrentiev regularization. The main idea of this regularization method is to shift all singular values by α, i.e.,

$$F_\alpha(\lambda) = \frac{1}{\sqrt{\lambda}(\sqrt{\lambda} + \alpha)}, \quad \lambda > 0, \; \alpha > 0, \tag{2.73}$$

so that

$$x_\alpha = R_\alpha y = \sum_{i=1}^{\infty} \frac{1}{\sigma_i + \alpha} \langle y, y_i\rangle \; x_i, \quad y \in Y. \tag{2.74}$$

In this case, the sum is really infinite and the full singular system is needed in order to compute the solution. However, if A is a positive semidefinite operator (and, thus, $x_i = y_i$), we obtain

$$(A + \alpha I) \; x_\alpha = \sum_{i=1}^{\infty} \langle y, y_i\rangle \; x_i = y. \tag{2.75}$$

Hence, the regularized solution can also be obtained in this case without any knowledge of the singular system as the solution of the linear equation $(A + \alpha I) \; x_\alpha = y$. Clearly, $(\sigma + \alpha)^{-1} \le \alpha^{-1}$, hence, $C_{F_\alpha} = \alpha^{-1}$.

Tikhonov regularization. The regularized solution is given by

$$F_\alpha(\lambda) = \frac{1}{\lambda + \alpha}, \quad \lambda > 0,\ \alpha > 0, \tag{2.76}$$

so that

$$x_\alpha = R_\alpha y = \sum_{i=1}^{\infty} \frac{\sigma_i}{\sigma_i^2 + \alpha} \langle y, y_i \rangle\ x_i, \quad y \in Y. \tag{2.77}$$

As in the case of Lavrentiev regularization, we can compute x_α defined by (2.77) without any knowledge of the singular system. In fact, it is easy to see that

$$(A^*A + \alpha I)\ x_\alpha = A^* y \tag{2.78}$$

and, hence, we can solve a well-posed linear system to obtain x_α. From this representation it also follows that Tikhonov regularization is just Lavrentiev regularization applied to the normal equation. It is not hard to see that $\lambda^2 + \alpha \geq 2\lambda\sqrt{\alpha}$, hence, C_{F_α} can be chosen as $2\alpha^{-1/2}$.

Asymptotic regularization. Asymptotic regularization is usually constructed from the solution x of the initial value problem

$$x'(t) = -A^*(Ax(t) - y), \quad t > 0, \tag{2.79}$$

$$x(0) = 0, \tag{2.80}$$

as $x_\alpha = x(\frac{1}{\alpha}\ \cdot)$. By representing x in terms of the singular vectors x_i in the form

$$x(t) = \sum_{i=1}^{\infty}\ \alpha_i(t)\ x_i \tag{2.81}$$

with $\alpha_i(0) = 0$, we obtain from the singular value decomposition

$$\alpha_i'(t) = -\sigma_i^2 \alpha_i(t) + \sigma_i\ \langle y_i, y \rangle. \tag{2.82}$$

This ordinary differential equation can be solved analytically by

$$\alpha_i(t) = \left(1 - \exp(-\sigma_i^2 t)\right) \frac{1}{\sigma_i} \langle\ y_i, y \rangle. \tag{2.83}$$

Hence, the regularized solution is given by

$$F_\alpha(\lambda) = \left(1 - \exp\left(-\frac{\lambda}{\alpha}\right)\right) \frac{1}{\lambda} \tag{2.84}$$

i.e.,

$$x_\alpha = \sum_{i=1}^{\infty} \left(1 - \exp\left(-\frac{\sigma_i^2}{\alpha}\right)\right) \frac{1}{\sigma_i} \langle y, y_i \rangle\ x_i, \quad y \in Y. \tag{2.85}$$

Error estimates. For the error between $x^\dagger$ and x_α^ε in the case of noisy data y^ε, we are able to write (with $x_\alpha^\varepsilon = R_\alpha y^\varepsilon$)

$$x^\dagger - x_\alpha^\varepsilon = (x^\dagger - x_\alpha) + (x_\alpha - x_\alpha^\varepsilon). \tag{2.86}$$

The first term $x^\dagger - x_\alpha$ is the *approximation error of the regularization method*, which is independent of the noise. The second term $x_\alpha - x_\alpha^\varepsilon$ corresponding to the

propagation of data noise in the regularized case. By aid of the triangle inequality it follows that

$$\|x^\dagger - x_\alpha^\varepsilon\| \le \|x^\dagger - x_\alpha\| + \|x_\alpha - x_\alpha^\varepsilon\|. \tag{2.87}$$

Thus, the estimate of the error between the regularized solution and the exact solution, can be handled by two error terms separately. It is clear that such an estimation provides a guideline for the parameter choice, namely by choosing α such that the terms on the right-hand side are balanced.

Next we deal with an estimate of the approximation error, which is independent of the noise level ε:

Let $F_\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ be a piecewise continuous function satisfying the assumptions

$$F_\alpha(\lambda) \to \frac{1}{\lambda}, \quad \lambda > 0, \quad \alpha \to 0,$$
$$|F_\alpha(\lambda)| \le M_\alpha < \infty, \quad \lambda > 0,$$

and

$$\sup_{\alpha,\lambda}(\lambda F_\alpha(\lambda)) \le C_F < \infty \tag{2.88}$$

for some constant $C_F > 0$. Moreover, let the regularization operator be defined by (2.67). *Then, for all $y \in \mathcal{D}(A^\dagger)$, we have*

$$R_\alpha y \to A^\dagger y, \quad \alpha \to 0. \tag{2.89}$$

The function $t \mapsto F_\alpha(t)$, $t \in \mathbb{R}^+$, converges pointwise to the function

$$F(t) = \begin{cases} 0 & , \quad t > 0 \\ 1 & , \quad t = 0. \end{cases} \tag{2.90}$$

Due to the discontinuity at zero, the convergence of $tF_\alpha(t) - 1$ to zero is becoming slower and slower as t decreases to zero. Since it is allowed to specify an arbitrarily small singular value σ_i and the minimal norm solution $x^\dagger = x_i$, the convergence of regularized solutions is arbitrarily slow. On the other hand, we observe that there is a possibly faster convergence if the components $\langle x^\dagger, x_i\rangle$ decay sufficiently fast compared to the eigenvalues. For example, if we have $|\langle x^\dagger, x_i\rangle| \le c\sigma_i^\mu$ for some constant $c > 0$ and $\mu > 0$, then it follows

$$\begin{aligned} \limsup_{\alpha\to 0} \|R_\alpha y - A^\dagger y\|^2 &\le \limsup_{\alpha\to 0} c^2 \sum_{n=1}^{\infty} (\sigma_i F_\alpha(\sigma_i) - 1)^2 \sigma_i^{2\mu} \\ &\le c^2 \sum_{n=1}^{\infty} \lim_{\alpha} (\sigma_i^{1+\mu} F_\alpha(\sigma_i) - \sigma_i^\mu)^2. \end{aligned} \tag{2.91}$$

In other words, one has to consider the limit of the function $t \mapsto |t^{1+\mu} F_\alpha(t) - t^\mu|$ as $t \to \infty$ instead, which is usually much faster. For example, in case of the truncated singular value decomposition, we obtain

$$|t^{1+\mu} F_\alpha(t) - t^\mu| = \begin{cases} 0 & , \quad t \ge \alpha \\ t^\mu & , \quad t < \alpha. \end{cases} \tag{2.92}$$

If the singular values of the operator decay sufficiently fast (which is the typical case for ill-posed problems), e.g., $\sum_{n=1}^{\infty} \sigma_i^{\mu} < \infty$, we are confronted with the situation

$$\|R_\alpha y - A^\dagger y\|^2 \le c^2 \sum_{\sigma_i < \alpha} \sigma_i^{2\mu} \le c^2 \alpha^\mu \sum_{i=1}^{\infty} \sigma_i^\mu \tag{2.93}$$

so that the term $\|R_\alpha y - A_y^\dagger\|$ is of order $\alpha^{\mu/2}$. Consequently, we somehow need smoothness of the solution (in terms of the smoothing properties of the operator) in order to obtain a convergence rate in terms of α. We shall pursue this idea by introducing spaces of smoothness involving the absolute value of a compact operator.

Next we are concerned with the propagation of the data error through the regularization. Let F_α and C_F be as given above, and let $x_\alpha = R_\alpha y$, $x_\alpha^\varepsilon = R_\alpha y^\varepsilon$. From the singular value decomposition it follows directly

$$\begin{aligned} \|Ax_\alpha - Ax_\alpha^\varepsilon\|^2 &\le \sum_{i=1}^{\infty} \left(\sigma_i^2 F_\alpha(\sigma_i^2)\right)^2 |\langle y - y^\varepsilon, y_i\rangle|^2 \\ &\le C_F^2 \sum_{n=1}^{\infty} |\langle y - y^\varepsilon, y_i\rangle|^2 = C_F^2 \|y - y^\varepsilon\|^2 \le (C_F \varepsilon)^2, \end{aligned} \tag{2.94}$$

so that

$$\|Ax_\alpha - Ax_\alpha^\varepsilon\| \le C_F \varepsilon \tag{2.95}$$

is valid. In the same way we obtain

$$\begin{aligned} \|x_\alpha - x_\alpha^\varepsilon\|^2 &\le \sum_{i=1}^{\infty} (F_\alpha(\sigma_i))^2 |\langle y - y^\varepsilon, y_i\rangle|^2 \\ &\le M_\alpha^2 \sum_{i=1}^{\infty} |\langle y - y^\varepsilon, y_i\rangle|^2 = M_\alpha^2 \|y - y^\varepsilon\|^2 \le (M_\alpha \varepsilon)^2, \end{aligned} \tag{2.96}$$

so that

$$\|x_\alpha - x_\alpha^\alpha\| \le M_\alpha \varepsilon \tag{2.97}$$

is implied (note that (2.97) estimates the norm of R_α by C_F).

3. Operator methodologies of resolution

Numerous methods have been proposed for treating and regularizing various types of ill-posed problems (IPP's). The rationale in most methods for resolution (approximate solvability) of IPP is to construct a "solution" that is acceptable physically as a meaningful approximation and is sufficiently stable from the computational standpoint, hence, an emphasis is put on the distinction between "solution" and "resolution". As already mentioned, the main dilemma of modeling of ill-posed problems is that the closer the mathematical model describes the IPP, the worse is the "condition number" of the associated computational problem (i.e.,

the more sensitive to errors). For ill-posed problems, the difficulty is to bring additional information about the desired solution, compromises, or new outlooks as aids to the resolution of IPP. It is conventional to use the phrase "regularization of an ill-posed problem" to refer to various approaches to circumvent the lack of continuous dependence (as well as to bring about existence and uniqueness if necessary). Roughly speaking, this entails an analysis of an IPP via an analysis of an associated well-posed problems, i.e., a system (usually a sequence or a family) of well-posed problems, yielding meaningful answers to the IPP. We distinguish three aspects of regularization:

(a) *strategy of resolution and reconstruction*,
(b) *regularization-approximation schema*,
(c) *regularization algorithms*.

One of the purposes of our work is to dramatize this delineation with reference to specific methods and results.

The *strategy of resolution and reconstruction of ill-posed problems* involves one or more of the following intuitive ideas (cf. [184]):

(α) change the notion of what is meant by a solution (e.g., ε-approximate solution: $\|Au - y\| \leq \varepsilon$, where $\varepsilon > 0$ is prescribed; quasi-solution: $\|Au - y\| \leq \|Ax - y\|$ for all $x \in \mathcal{M}$, a prescribed subset of the domain of A; least squares solution of minimal norm, etc.),
(β) modify the operator equation or the problem itself,
(γ) change the spaces and/or topologies,
(δ) specify the type of involved noise ("strong" or "weak" noise).

The philosophy of resolution leads to the use of algebraic methods versus function space methods, statistical versus deterministic approaches, strong versus weak noise (see [50, 51, 53], where the concept of weakly bounded noise was first introduced), etc.

By a *regularization-approximation scheme* we refer to a variety of methods such as Tikhonov's regularization, projection methods, multiscale methods, iterative approximation, etc., that can be applied to ill-posed problems. These schemes turn into algorithms once a resolution strategy can be effectively implemented. Unfortunately, this requires a determination of a suitable value of a certain parameter associated with the scheme (e.g., regularization parameter, mesh size, dimension of subspace in the projection scheme, specification of the level of a scale space, classification of noise, etc.). This is not a trivial problem since it involves a trade-off between accuracy and numerical stability, a situation that does not usually arise in well-posed problems.

From the standpoint of mathematical and numerical analysis one can roughly group "regularization methods" into three categories (cf. [184]):

(a) Regularization methods in function spaces is one category. This includes Tikhonov-type regularization, the method of quasi-reversibility, the use for certain function spaces such as scale spaces in multi-resolutions, the method

of generalized inverses (pseudoinverses) in reproducing kernel Hilbert spaces, and multiscale wavelet regularization.

(b) Resolution of ill-posed problems by "control of dimensionality" is another category. This includes projection methods, discretization and moment-discretization schemes. The success of these methods hinges on the possibility of obtaining approximate solutions while keeping the dimensionality of the finite-dimensional problem within the "range of numerical stability". It also hinges on deriving error estimates for the approximate solutions that is crucial to the control of the dimensionality.

(c) A third category are iterative and filtration methods which can be applied either to the problem in function spaces or to a discrete version of it. The crucial ingredient in iterative methods is to stop the iteration before instability creeps into the process. Thus iterative methods have to be modified or accelerated so as to provide a desirable accuracy by the time a stopping rule is applied. Filtration methods refer to procedures where, for example, singular functions and values producing highly oscillatory solutions are eliminated. Various "low pass" filters can, of course, be used. The last sentence in (b) is also crucial for the determination of a stopping rule.

3.1. Concept of regularization revisited

The concept of a *regularizer* plays an important role in obtaining an approximate solution of an IPP in the presence of contamination in the data. We shall explain show that it is, indeed, a generic concept that can be used to unify some of the principles occurring in various regularization-approximation schemes (Tikhonov's regularization, truncated or filtered singular-value expansions, projection methods, multiscale techniques, iterative methods, etc).

More explicitly, let X, Y be normed spaces and let $A : X \to Y$ be a one-to-one mapping (not necessarily linear). Note that the assumption that A is one-to-one is imposed for the convenience of the linear case. In fact, it can be dropped if one uses generalized inverses, as is done in several papers of Nashed [176, 178]. This assumption will also be dropped in this contribution after we will have motivated and explained the auxiliary procedure for the case when A is assumed to be one-to-one.

A *regularizer to the operator equation* $Ax = y$ *is a one-parameter family of operators* $\{R_t : t \in \Gamma\}$ *where* Γ *is an index set of real numbers with* $0 \in \overline{\Gamma}$ (*the closure of* Γ) *satisfying the following conditions:*

(C1) *For each* $t \in \Gamma, R_t$ *is a continuous operator on all of* Y *into* X.
(C2) *For each* $x \in X$, $\lim_{t\to 0} \|R_t Ax - x\| = 0$.

Strong noise. Ill-posed problems of mathematical practice are usually regularizable in following sense: For $y \in \mathcal{R}(A)$, let y^ε be known with "noise level", $\|y^\varepsilon - y\| \leq \varepsilon$ (*"strong noise condition"*). Consider the operator equation between normed spaces X, Y

$$Ax = y^\varepsilon. \tag{3.1}$$

The existence of a regularizer enables us to calculate an "approximate solution" $\|x^\varepsilon - A^{-1}y\| \to 0$ and $\|Ax^\varepsilon - y\| \to 0$. We explain this fact in more detail. Let x^ε be given by $x^\varepsilon := R_t y^\varepsilon$. Then

$$\|x^\varepsilon - A^{-1}y\| \le \|R_t y - A^{-1}y\| + \|R_t y^\varepsilon - R_t y\|, \tag{3.2}$$

where we assume that both norms on the right side of (3.2) are known. Note that $\|R_t y - A^{-1}y\|$ gives a *rate of convergence* of $R_t Ax$ to x and $\|R_t y^\varepsilon - R_t y\|$ is the *modulus of continuity of the operator* R_t at y. Observe that $\|R_t y^\varepsilon - R_t y\| \to 0$ as $\varepsilon \to 0$ for any fixed t. For given $\varepsilon > 0$ we choose $t = t(\varepsilon)$ to minimize the right side of (3.2). Then $x^\varepsilon = R_{t(\varepsilon)} y^\varepsilon$ has the claimed property since $\|R_{t(\varepsilon)} y^\varepsilon - A^{-1}y\| \to 0$ as $\varepsilon \to 0$.

At this stage we shall be primarily interested in the case when A will be assumed to be linear. For simplicity, R_t will be required to be linear, too. We then have

$$\|x^\varepsilon - A^{-1}y\| \le \|R_t y - A^{-1}y\| + M(t)\,\varepsilon \tag{3.3}$$

where $\|R_t\| \le M(t)$ (note that the operators R_t are not uniformly bounded since A^{-1} is unbounded, so $M(t) \to \infty$ as $t \to 0$).

The notion of a regularizer can be easily extended to weighted least squares problems (see Subsection 2.3): We say that the problem (3.1), or equivalently $A_L^\dagger$, is regularizable if there exists a one-parameter family of linear operators $\{R_t : t \in \Gamma\}$ with $\mathcal{R}(R_t) \subset \mathcal{M}$ such that $\lim_{t\to 0} \|R_t Ax - x\| = 0$ for $x \in \mathcal{M}$ and for each $t > 0$, R_t is bounded. Again, the family $\{R_t\}$ is not uniformly bounded in t since $\mathcal{R}(A)$ is non-closed. As before, the existence of a regularizer provides us with a family of approximate solutions determined by a well-posed problem. In the presence of contamination in y, say $\|y^\varepsilon - y\| \le \varepsilon$, the error $\|R_t y^\varepsilon - A_L^\dagger y\|_X \to 0$ as $t \to 0$ (in fact, it blows up). The criterion then is to choose t to minimize the error:

$$\begin{aligned} \|R_t y^\varepsilon - A_L^\dagger y\| &\le \|R_t y - A_L^\dagger y\| + \|R_t(y^\varepsilon - y)\| \\ &\le \|R_t y - A_L^\dagger y\| + \|R_t\|\varepsilon. \end{aligned} \tag{3.4}$$

The first term (*regularization error*) tends to zero as $t \to 0$, while the second term (*magnification of contamination error* due to ill-posedness) tends to ∞:

$$\boxed{\begin{array}{lll} & \left\| A_L^\dagger y - R_t y \right\| & \longrightarrow 0 \\ t \to 0 \quad \nearrow & & \\ \quad\quad \searrow & \|R_t\,(y - y^\varepsilon)\| \le \|R_t\|\,\varepsilon & \longrightarrow \infty \end{array}} \tag{3.5}$$

$$\begin{array}{lll} & \left\| A_L^\dagger y - R_t y \right\| & \longrightarrow \infty \\ \nearrow & & \\ t \to \infty & & \\ \searrow & & \\ & \| R_t \left(y - y^\varepsilon \right) \| \leq \| R_t \| \, \varepsilon & \longrightarrow 0 \end{array} \tag{3.6}$$

If we know an *error estimate* for the first term and a growth estimate for C_t, a suitable t can be determined. Such estimates can be obtained for particular regularizers R_t using additional information on the solution $A_L^\dagger y$, e.g., smoothness, and some robustness condition on C_t, e.g., one might estimate that $C_t \leq d(t)$ is a known function which tends to ∞ as $t \to 0$, and $\| R_t y - A_L^\dagger y \| \leq b(t)$, where $b(t) \to 0$ as $t \to 0$. Then, an optimal $t(\varepsilon)$ can be easily calculated, and for this $t(\varepsilon)$, $R_{t(z)} y^\varepsilon \to A_L^\dagger y$, as $\varepsilon \to 0$. Concrete realizations of regularizers abound in regularization methods, projection and iterative methods, etc., as we shall see in the remaining work. In the case of an iterative scheme, $t = 1/n$, $x_n = R_t y$ represents the nth iterate of a process which converges to $A_L^\dagger$ (or A^{-1}) in the absence of contamination in y. In projection methods $t = 1/n$, where n is the dimension of the approximating subspace. In finite differences, t represents the mesh size h. The preceding results then show how to obtain stable approximate solutions in the presence of error in y it the needed estimates in (3.4) are available.

Weak noise. Let $K : X \to Y$ be a linear compact operator between the Hilbert spaces X and Y. The inner products and norms of X and Y are denoted by $\langle \cdot, \cdot \rangle_X$, $\langle \cdot, \cdot \rangle_Y$ and $\| \cdot \|_X, \| \cdot \|_Y$ (note that we do not use subscripts if they are clear from the context, here and elsewhere). Consider the data $y \in Y$ according to the equation

$$y = K x_0 + \eta \ , \tag{3.7}$$

where $\eta \in Y$ is the unknown noise and $x_0 \in X$ is an unknown element one wishes to recover from the data y. The following model is imposed on the noise. Let $A : Y \to Y$ be linear, compact, Hermitian, and positive-definite (i.e., $\langle y, Ay \rangle > 0$ for all $y \in Y$, $y \neq 0$), and let

$$\varepsilon^2 := \langle \eta, A\eta \rangle. \tag{3.8}$$

We assume that ε is "small" and investigate what happens when $\varepsilon \to 0$. The operator A introduced above is not meant to be arbitrary. In fact, it must be connected with K in the sense that, for some $m \geq 1$ (not necessarily integer), the range of K is continuously embedded into the range of A^m, so that

$$A^{-m} K : X \to Y \quad \text{is continuous} \ . \tag{3.9}$$

If η satisfies (3.8), (3.9), it is referred to as *weakly bounded noise*.

Some comments should be made: In a deterministic setting, a reasonable model for the noise is that it is "high-frequency", and we would like to investigate what happens when the frequency tends to ∞, but without the noise tending to 0 strongly, that is without assuming that $\| \eta \|_Y \to 0$. Thus, $\eta \to 0$ weakly begins capturing the essence of "noise". Then, for any linear compact operator

$S : Y \to Y$, we would have $\|S\eta\|_Y \to 0$. So, in this sense, there is nothing unusual about (3.8) and (3.9). Moreover, we would like (3.8) to capture the whole truth, i.e., the relations

$$\langle \eta, A^p \eta \rangle = o(\varepsilon^2) \quad \text{and} \quad \langle \eta, A^q \eta \rangle = O(\varepsilon^2) \tag{3.10}$$

fail for $p > 1$ and $q < 1$ as $\varepsilon \to 0$. This may be a tall order, although examples of operators A and noises η satisfying (3.8)–(3.10) are easily constructed (for more details see [50]). At the same time A is supposed to capture the smoothing effect of K in the sense of (3.9). Ideally, one would like $A^{-m}K$ to be continuous with a continuous inverse. The natural choice $A = (KK^*)^{1/2m}$ would achieve this, but would have to be reconciled with (3.8) and possibly (3.10). The condition (3.9) is not unreasonable.

Eggermont et al. [53] show how the weak noise model leads to simple bounds on expressions like $\langle \eta, y \rangle_Y$ for $y \in A^m(Y)$, the range of A^m. For $\beta > 0$, they introduce the inner product on $A^m(Y)$, by letting

$$\langle y, z \rangle_{m,\beta} = \langle y, z \rangle_Y + \beta^{2m} \langle A^{-m}, y, A^{-m} z \rangle_Y, \quad z \in A^m(Y), \tag{3.11}$$

and denote the associated norm by $\| \cdot \|_{m,\beta}$. The following result is of interest in itself, but it also later on plays a crucial role in the context of Tikhonov regularization with weakly bounded noise:

Suppose that $m \geq 1$. *Under the assumptions* (3.8), (3.9) *on the weakly bounded noise, for all* $y \in A^m(Y)$ *and all* $\beta > 0$

$$|\langle \eta, y \rangle_Y| \leq \beta^{-1/2} \varepsilon \, \|y\|_{m,\beta} \tag{3.12}$$

(*note that the factor* $\beta^{-1/2}$ *stays the same, regardless of* m).

3.2. Use of compactness and a priori bounds

The use of *a priori bounds* (more generally, a priori information) about the solution of an ill-posed problem has long been recognized to play a significant role in bringing about continuous dependence (i.e., of providing a regularizing effect). Early in the study of ill-posed problems, a fundamental observation was made by Tikhonov (cf. [246, 249] for original references) that the restriction to a compact set insures well-posedness. More precisely, suppose X and Y are metric spaces and $F : X \to Y$ is a continuous injection, and let $C \subset X$ be compact. Then $F^{-1} : F(C) \to C$ is continuous: To show this, let $W \subset C$ be open in the relative topology, then the complement of W relative to C, denoted by W^c, is closed and, hence, compact since C is compact. Continuity of F implies that $F(W^c)$ is compact and, therefore, closed. From this it follows that $F(W)$ is open, for F is injective (and hence $F(W^c) \cap F(W)$ is empty).

Remark. The use of differential operators as smoothing conditions often leads to a setting in which the restriction of the domain of the operator to a compact

set automatically prevails. For example, consider the simple situation treated by Tikhonov (see [250]), where

$$J_\alpha(f) = \|Af - g\|^2 + \alpha\ \Omega(f), \quad \alpha > 0, \tag{3.13}$$

with

$$\Omega(f) = \int_0^1 \{p(x)[f'(x)]^2 + q(x)[f(x)]^2\}\ dx, \tag{3.14}$$

when p and q are positive, q is continuous, and q has a continuous derivative. Then the set

$$C_r := \{f \in L^2[0,1] : \Omega(f) \le r\} \tag{3.15}$$

is compact in X for each $r > 0$. The smoothing function $\Omega(f)$ can be considered to be induced by a differential operator L, i.e., $\Omega(f) = \|Lf\|^2$, where $L^*Lf = -(pf')' + qf$ on $[0,1]$ with the boundary conditions $f'(0) = f'(1) = 0$. In particular, taking $p = q = 1$, it follows that C_r is a ball in the Sobolev space $W_2^1[0,1]$. Thus by restricting solutions to lie in a ball in $W_2^1[0,1]$ the problem is no longer ill-posed. Similar results can be obtained using more general differential operators and related spaces that are compactly embedded in X. The success of the aforementioned approach hinges on the fact that the unit ball of $W_2^1[0,1]$ is a compact set in the topology of $L^2[0,1]$ (note that it is, of course, not compact in the topology of $W_1^2[0,1]$, for more details the reader is referred, e.g., to [2, 184, 185]).

3.3. Tikhonov's regularization

Let $L : \mathcal{D}(L) \subset X \to Z$, where Z is a Hilbert space, be a closed linear operator with dense domain and closed range. We first assume (see also [176, 247, 249, 250]) that $\mathcal{N}(L)$ is finite dimensional and that $\mathcal{N}(L) \cap \mathcal{N}(A) = \{0\}$. We endow $\mathcal{D}(L)$ with the topology induced by the graph norm $|u| := (\|u\|^2 + \|Lu\|^2)^{1/2}$. Then L becomes a bounded operator on $\mathcal{D}(L)$. We define a new inner product on $\mathcal{D}(L)$ by

$$[u, v]_L = (Au, Av)_Y + (Lu, Lv)_Z \tag{3.16}$$

Then the induced norm $\|u\|_L := \sqrt{[u,u]}$ is equivalent to the graph norm of u. Thus, both A and L are bounded operators on $\mathcal{D}(L)$ equipped with the inner product (3.16); we denote this Hilbert space by X_L. The discussion shows that, under the hypotheses listed above, without loss of generality we may restrict ourselves to the case in which $A : X \to Y$ and $L : X \to Z$ are both bounded. For each $y \in D(A^\dagger)$, there is a unique element $x \in \mathcal{L}(y)$ which minimizes $\|Lu\|$. Let $A_L^\dagger y := x$, and define

$$\mathcal{M} := \{x \in X : L^*Lu \in \mathcal{N}(A)^\perp\}. \tag{3.17}$$

Then it is not difficult to show that $\mathcal{M}$ is the orthogonal complement of $\mathcal{N}(A)$ with respect to the inner product (3.16), so that $A_L^\dagger$ is the generalized inverse relative to the decompositions:

$$X : \mathcal{N}(A) \oplus \mathcal{M}, \quad Y = \overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^\perp. \tag{3.18}$$

Under the above assumptions for each $\alpha > 0$, there exists a unique x_α which minimizes

$$J_\alpha(x) := \|Ax - y\|_Y^2 + \alpha \|Lx\|_Z^2. \tag{3.19}$$

Furthermore,

$$x_\alpha = (A^*A + \alpha L^*L)^{-1} A^* y \tag{3.20}$$

and x_α converges to $A_L^\dagger y$ for $y \in \mathcal{D}(A^\dagger)$ as $\alpha \to 0$, and diverges otherwise. Thus it follows that in the presence of contamination, say y is replaced by y_ε, where $\|y_\varepsilon - y\| \leq \varepsilon$ for some $\varepsilon > 0$, the norm of the corresponding x_α^ε may well diverge. It should be noted that $(A^*A + \alpha L^*L)^{-1} A^*$ does not converge in the uniform operator topology as $\alpha \to 0$; in fact, $\|(A^*A + \alpha L^*L)^{-1} A^*\| \to \infty$. Furthermore, the equation system $(A^*A + \alpha L^*L)x = A^*y$ is poorly conditioned for small α, and hence numerically unstable. Thus, both contamination and numerical approximation dictate that a choice for α has to be made which would be a suitable compromise between accuracy and stability. Several procedures for the choice of "optimal" α are available, some of which take into consideration various a priori information. A simple choice that works would be $\alpha = \varepsilon^2$.

Strongly bounded noise in Tikhonov's regularization. We come back to the situation where $K : X \to Y$ is a linear compact operator between the Hilbert spaces X and Y. Consider the data $y \in Y$ according to the equation $y = Kx_0 + \eta$, where $\eta \in Y$ is the unknown noise and $x_0 \in X$ is an unknown element one wishes to recover from the data y. We study Tikhonov's regularization as a scheme to recover x_0 from the data y in the *strong noise model*

$$y = Kx_0 + \eta \quad \text{with} \quad \|\eta\|_Y \leq \varepsilon \ . \tag{3.21}$$

The interest is in what happens when $\varepsilon \to 0$. It should be noted that, in the Tikhonov regularization scheme, the unknown x_0 is estimated by specifying the solution $x = x^{\alpha,\varepsilon}$ of the problem

$$\text{minimize} \ \ \|Kx - y\|_Y^2 + \alpha \|x\|_X^2 \quad \text{over} \ \ x \in X \tag{3.22}$$

for some regularization parameter α, $\alpha > 0$, yet to be specified. This procedure dates back to [216, 249]. Its minimizer exists and is unique. Moreover, it is well-known (see, e.g., [106]) that convergence rates on the error $\|x^{\alpha,\delta} - x_0\|_X$ can be obtained from a source condition. For simplicity, it is assumed there that there exists a $z_0 \in X$ such that the "source condition"

$$x_0 = (K^*K)^{\nu/2} z_0 \quad \text{for some} \quad 0 < \nu \leq 2 \tag{3.23}$$

holds true. Precise necessary and sufficient conditions are given in [202]. In the study of convergence rates under the source condition (3.23), it is assumed here that ν is known and that α is chosen accordingly. Clearly, one wants to obtain bounds on the error $\|x^{\alpha,\varepsilon} - x_0\|_X$. As usual, this is broken up into two parts

$$\|x^{\alpha,\varepsilon} - x_0\|_X \leq \|x^{\alpha,\varepsilon} - x^{\alpha,0}\|_X + \|x^{\alpha,0} - x_0\|_X, \tag{3.24}$$

where $x^{\alpha,\varepsilon}$ is the "noiseless" estimator, i.e., the minimizer of $\|Kx-y\|_Y^2+\alpha\|x\|_X^2$. Thus, $x^{\alpha,\varepsilon}-x^{\alpha,0}$ is the noise part of the error and $x^{\alpha,0}-x_0$ is the error introduced by the regularization.

The following results (see, e.g., [50, 106]) are well known:

(1) There exists a constant c such that for all α, $0<\alpha\leq 1$,

$$\|x^{\alpha,\varepsilon}-x^{\alpha,0}\|_X \leq c\,\alpha^{-\frac{1}{2}}\|\eta\|_Y. \tag{3.25}$$

(2) Under the source condition (3.23), there exists a constant such that for all α, $0<\alpha\leq 1$,

$$\|x^{\alpha,\varepsilon}-x_0\|_X \leq c\,\alpha^{\nu/2}. \tag{3.26}$$

As a consequence, the two results (3.25), (3.26) above then provide the following convergence rates:

Assuming the source condition (3.23) *and the condition* (3.21) *on the noise for* $\alpha\to 0$ *we have*

$$\|x^{\alpha,\varepsilon}-x_0\|_X = O\left(\alpha^{-1/2}\,\varepsilon+\alpha^{\nu/2}\right). \tag{3.27}$$

Moreover, if $\alpha\asymp\varepsilon^{2/(\nu+1)}$ *then*

$$\|x^{\alpha,\varepsilon}-x_0\|_X = O\left(\varepsilon^{\nu/(\nu+1)}\right)\ . \tag{3.28}$$

Weak noise in Tikhonov's regularization. Tikhonov's regularization may also be considered (cf. [49, 52, 53]) as the scheme to recover x_0 from the data y in the *weak noise model*

$$y = Kx_0+\eta\ . \tag{3.29}$$

Thus, we assume that there is a smoothing operator A such that the noise η and A satisfy (3.8) and (3.9). In particular, $\langle\eta, A\eta\rangle_Y=\varepsilon^2$, and the discussion (cf. [53]) focusses on what happens when $\varepsilon\to 0$. Formally, Tikhonov regularization does not depend on the noise being strongly or weakly bounded. Thus x_0 is estimated by the solution $x=x^{\alpha,\varepsilon}$ of the problem (3.22) for some positive regularization parameter α yet to be specified. Again we want to obtain bounds on the error $\|x^{\alpha,\varepsilon}-x_0\|_X$, and it is broken up as $\|x^{\alpha,\varepsilon}-x_0\|_X \leq \|x^{\alpha,\varepsilon}-x^{\alpha,0}\|_X+\|x^{\alpha,0}-x_0\|_X$, where $x^{\alpha,0}$ is the "noiseless" estimator, i.e., the minimizer of $\|Kx-y\|_Y^2+\alpha\|x\|_X^2$. Thus, $x^{\alpha,\varepsilon}-x^{\alpha,0}$ is the noise part of the error and $x^{\alpha,0}-x_0$ is the error caused by the regularization. It is useful to introduce a new norm on X by way of

$$\|x\|_{\alpha,X}^2 = \|Kx\|_Y^2+\alpha\,\|x\|_X^2\ . \tag{3.30}$$

Assuming again the source condition (3.23) we see that the noiseless part $x^{\alpha,0}-x_0$ can be covered as before, but the treatment of the noise part is markedly different from the case of strong noise (see [51]):

(1) *Under the conditions* (4.216), (3.9) *on the noise* η, *there exists a constant* C *depending on* A *only such that for* $\alpha\to 0$

$$\|x^{\alpha,\varepsilon}-x^{\alpha,0}\|_{\alpha,X}^2 \leq C\,\alpha^{-\frac{1}{4m}}\varepsilon\ . \tag{3.31}$$

This leads to the convergence rates (3.32) *and* (3.33) *specified below, where it is shown in* [50] *that they are optimal, following arguments from* [201]*, but assuming, in addition, that* $A^{-m}K$ *has a continuous inverse.*

(2) *Assuming the source condition* (3.23) *and the conditions* (4.216) *and* (3.9) *on the noise for* $\alpha \to 0$*, we have*

$$\|x^{\alpha,\varepsilon} - x_0\|_X = O\left(\alpha^{-\frac{1}{2}-\frac{1}{4m}}\varepsilon + \alpha^{\nu/2}\right). \tag{3.32}$$

Moreover, if $\alpha \asymp \varepsilon^{4m/(2m\nu+2m+1)}$ *then*

$$\|x^{\alpha,\varepsilon} - x_0\|_X = O\left(\varepsilon^{2m\nu/(2m\nu+2m+1)}\right). \tag{3.33}$$

3.4. Characterization of regularizers

Let A be of class $L(X,Y)$ with non-closed range. In what follows we introduce concepts of regularizing families for the ill-posed problem $Ax = y$ based on bounded outer inverses of the operator A:

A linear operator $B : Y \to X$ *is called an* inner inverse of A *if* $ABA = A$. *A* (*nonzero*) *linear operator* $B : Y \to X$ *is called an* outer inverse of A *if* $BAB = B$.

In what follows, B is always taken to be a non-zero operator. In the case of Hilbert spaces, the regularizers will approximate least squares solutions of $Ax = y$. In the case of Banach spaces, we assume that A is injective, and $\mathcal{R}(A)$ is dense in Y (otherwise the regularizers would apply to the equation $Ax = Qy$, where Q is a continuous projector of Y onto $\mathcal{R}(A)$, whose existence has to be assumed).

We will classify ill-posed in Banach spaces according to the type of regularizing families that they admit (cf. [185] [186]). In the case of Hilbert spaces this classification will particularly distinguish the set of all compact operators with infinite-dimensional range within the set of all bounded operators with non-closed range.

Many of the operator-theoretic aspects of ill-posed linear equation (including regularization and stabilization methods) are really problems in operator ranges and operator factorizations. In particular, ranges of outer inverses play several roles. First, it should be noted that very often the constructed approximation solution is in the range of some outer inverse (or an approximate outer inverse). Second, the problem of finding an "optimal" outer inverse with a prescribed rank can be solved for several classes of operators. Third, outer inverses with a prescribed finite-dimensional range can be easily constructed; this cannot always be done in case the prescribed range is infinite dimensional and the outer inverse is required to be bounded. Fourth, outer inverses have desirable "stability/continuous dependence" properties which inner inverses or the generalized inverse lack.

Our purpose is to introduce notions of regularizers in form of bounded outer inverses with infinite-dimensional range (within this class, convergent regularizers can be selected to provide "optimal" resolution), approximate outer inverses and approximate right inverses in scales of norms.

Remark. The set of all operators in $L(X,Y)$ that have bounded outer inverses with infinite-dimensional range and the set of full-rank $m \times n$ matrices share several common properties: Each of them is both *open* and *dense*, and all elements of each of the sets have outer inverses with the maximal possible rank (namely, the same as the rank of A). These properties and other results to be analyzed next indicate that, in Hilbert space, an equation involving a bounded non-compact operator with non-closed range is "less" ill-posed than an equation with a compact operator with infinite-dimensional range. In comparison with least squares or generalized inverse problems for $m \times n$ matrices, one may say that for operators with non-closed range, the case of a non-compact operator corresponds to the full-rank case for matrices, while the case of a (nondegenerate) compact operator is the infinite-dimensional analog of the rank-deficient case for matrices.

Outer inverses in "solvability" and "regularization" of ill-posed problems. If X is of dimension m and Y is of dimension n, it follows from the property $ABA = A$ that the rank of any inner inverse of A cannot be less than the rank of A. Similarly it follows from the property $BAB = B$ that the rank of any outer inverse of A cannot exceed the rank of A. Moreover, if $r := \text{rank } A$, then one can construct outer inverses of rank s for any $s \leq r$ and inner inverses of rank t for any $r \leq t \leq \min(m,n)$ (see, e.g., [178]). Even in the finite-dimensional case the possibility of approximating the generalized inverse (or the least squares solution of minimal norm in Hilbert space settings) by an outer inverse of rank $s < r$ is an attractive feature of outer inverses. A similar approximation by inner inverses is of course not possible.

The situation is more drastic in the infinite-dimensional case as can be seen from the following known result: *If the range of $A \in L(X,Y)$ is non-closed, then A has no bounded inner inverse.* In fact, if B is any inner inverse of A, then AB and BA are linear idempotents, with $\mathcal{N}(BA) = \mathcal{N}(A)$ and $\mathcal{R}(A) = \mathcal{R}(AB)$. Thus the following algebraic decompositions hold:

$$X = \mathcal{N}(A) \dot{+} \mathcal{R}(BA), \tag{3.34}$$

$$Y = \mathcal{R}(A) \dot{+} \mathcal{N}(AB), \tag{3.35}$$

where $\dot{+}$ denotes the algebraic direct sum. Now, suppose B is bounded, then the projectors BA and AB are continuous (equivalently, the decomposition in (3.34) and (3.35) are topological) and so $\mathcal{R}(A)$ is closed, which contradicts the assumption. Thus, we are led to the statement:

No regularizer can be an inner inverse.

The non-boundedness of the inner inverse of A in case of a non-closed range of $A \in L(X,Y)$ is actually a part of the following known statement (see [194]): *Let $A \in L(X,Y)$, where X and Y are Banach spaces. Then A has a bounded inner inverse B if and only if the decompositions* (3.34) *and* (3.35) *are topological. Equivalently, A has a bounded inner inverse if and only if $\mathcal{N}(A)$ and $\mathcal{R}(A)$ have*

topological complements in X and Y, respectively, (i.e., the projectors on $\mathcal{N}(A)$ and $\mathcal{R}(A)$ are continuous).

If X and Y are Hilbert spaces, then $A \in L(X,Y)$ has a bounded inner inverse if and only if $\mathcal{R}(A)$ is closed. It should be noted in all these cases that it does not mean that all inner inverses are bounded, unless A is invertible.

Truncated singular value expansions as outer inverses (cf. [185]): Let H_1 and H_2 be Hilbert spaces and let $K : H_1 \to H_2$ be a (nonzero) compact linear operator. Let K^* denote the adjoint of K. Since K^*K is a non-negative symmetric compact linear operator on H_1 we have in terms of the singular system $\{\sigma_k; x_k, y_k\}$ for K the following spectral representation $K^*Kx = \sum_{k=1}^{\infty} \sigma_k^2 \langle x, x_k\rangle x_k$, where $\{x_k\}$ is an orthonormal set of eigenvectors of K^*K with $K^*Kx_k = \sigma_k^2 x_k$ with $\sigma_1 \geq \sigma_2 \geq \cdots > 0$. Set $y_k := \sigma_k^{-1} K x_k$. Then the y_k's form an orthonormal set in H_2, and it is easy to show that

$$Kx = \sum_{k=1}^{\infty} \sigma_k \langle x, x_k\rangle y_k. \tag{3.36}$$

Obviously, the series (3.36) is the singular value expansion (SVD) of K (note that, if $\operatorname{rank}(K) = r$, then the number of non-zero singular values is r and the summation in (3.36) extends from $k = 1$ to r).

From here on, we assume, unless stated otherwise, that the range K is infinite dimensional. Then we have an infinite number of non-zero singular values with $\sigma_n \to 0$ as $n \to \infty$.

Let m be a fixed positive integer and define the operator $B_m : H_2 \to H_1$ by

$$B_m y := \sum_{k=1}^{m} \sigma_k^{-1} \langle y,\ y_k\rangle x_k. \tag{3.37}$$

It then follows that

$$KB_m\ y = \sum_{k=1}^{m} \sigma_k^{-1} \langle y, y_k\rangle K x_k = \sum_{k=1}^{m} \langle y, y_k\rangle y_k \tag{3.38}$$

and

$$\begin{aligned} B_m\ K\ B_m\ y &= \sum_{k=1}^{m} \sigma_k^{-1} \left\langle \sum_{i=1}^{m} \langle y, y_i\rangle y_i, y_k \right\rangle\ x_k \\ &= \sum_{k=1}^{m} \sigma_k^{-1} \langle y, y_k\rangle x_k \\ &= B_m\ y. \end{aligned} \tag{3.39}$$

Thus, for each $m \in \mathbb{N}$, B_m is an outer inverse of rank m. For each $y \in \mathcal{D}(K^\dagger) := \mathcal{R}(K) + \mathcal{R}(K)^\perp$, $\|B_m\ y - K^\dagger y\| \to 0$ as $m \to \infty$, where $K^\dagger y = \sum_{i=1}^{\infty} \sigma_i^{-1} \langle y, y_i\rangle x_i$, but the convergence is not uniform and the operators B_m are not uniformly bounded.

The SVD is particularly useful because it permits a quantification of the notion of near rank deficiency. It is well known from linear algebra that for any $m \times n$ real or complex matrix A of rank r and any $k < r$,

$$\inf \{\|A - B\|_{\text{Frob}} : \text{rank}(B) = k\} = \|A - A_k\|_{\text{Frob}}, \tag{3.40}$$

where A_k is the truncated singular value decomposition (with k terms) of A and $\|\cdot\|_{\text{Frob}}$ is the Frobenius norm (see, e.g., [99]). In view of the above observation that the TSVD is an outer inverse and by use of the well-known relation between the singular values and singular vectors of A and those of the Moore–Penrose of A, it follows that

$$\inf \left\{\|A^\dagger - B\|_{\text{Frob}} : BAB = B, \text{rank}(B) = k\right\} = \|A^\dagger - B_k\|_{\text{Frob}}, \tag{3.41}$$

where

$$B_k\, y = \sum_{i=1}^{k} \sigma_i^{-1} \langle y, y_i \rangle x_i. \tag{3.42}$$

The same analysis and properties of outer inverses can be easily extended to linear inverse problems with discrete data (such as those that arise from moment discretization of ill-posed linear integral and operator equations in Hilbert space (see, e.g., [24, 178]) or when projection methods on finite-dimensional subspaces are used (see, e.g., [105, 181]).

Outer inverses are not "equation solvers" (cf. [185]), i.e., if B is an outer inverse for A, which is not also an inner inverse, then for $y \in \mathcal{R}(A)$, $x := By$ is not a solution to $Ax = y$, and for $y \notin \mathcal{R}(A)$, $x := By$ is not a least squares solution (in the case of a Hilbert space). One finds in some books statements like: "Since almost every application of various generalized inverses involves subsets of 1-inverses (inner inverses), we will mainly consider inner inverses that satisfy additional conditions...". For ill-posed problems (see [185]), the situation is precisely the opposite. We are not interested in a generalized or inner inverse that would be an "equations solver", since such an "inverse" will be unbounded. Rather we seek a bounded operator that has some "inverse-like" properties, and that can serve simultaneously as an "approximate inverse" and stabilizer to the inverse problem. Indeed, outer inverses possess these qualities:

(a) If B is an outer inverse of A, then B is also an inner inverse of $\tilde{A} := A|\mathcal{R}(B)$.
(b) For all $y \in \mathcal{R}(AB)$, $x := By$ is the unique solution in $\mathcal{M} = \mathcal{R}(B)$ of the equation $Ax = y$.

From these properties (see [185] for more details), it follows that the unique solution of $Ax = y$ in $\mathcal{R}(B)$ is a "regularized" solution and can be constructed in a stable way. If $\mathcal{R}(BA)$ or $\mathcal{R}(B)$ is infinite-dimensional, then we have in a sense the possibility of "infinite resolution", and the equation with bounded outer inverses of infinite rank are not as ill posed as those for which an outer inverse with infinite rank does not exist.

Approximate outer and approximate right inverses. The concept of a regularizer plays an important role in obtaining an approximate solution of an ill-posed problem. Let $A : X \to Y$ be a one-to-one mapping (not necessarily linear). In the already known definition by Tikhonov (see e.g. [170, 249]) a regularizer is a bounded "approximate" inverse or "approximate" generalized inverse. Any regularizer must also satisfy the defining equations of an inner as well as outer inverse approximately, but it cannot satisfy the defining equation of an inner inverse exactly. This motivates considering regularizers (cf. [185]) that satisfy the defining equation of an outer inverse and, in addition, have "maximum" rank.

A family $\mathcal{F}$ of regularizers by outer inverses for the ill-posed problem $Ax = y$ is said to be of type I if each $B \in \mathcal{F}$ satisfies the following conditions:

(i) $B \in L(Y, X)$
(ii) $BAB = B$,
(iii) *the range of B is infinite-dimensional.*

An ill-posed problem that does not admit a family of regularizers of type I is said to be of type II.

Any $B \in \mathcal{F}$ is called a regularizer. Of course, $u := By$ is not necessarily a good approximation to the "solution" x of $Ax = y$ for each $B \in \mathcal{F}$, just like, say, $(A^*A + \alpha I)^{-1}A^*y$ is not necessarily a good approximation for each $\alpha > 0$. As already mentioned, every "regularization method" involves a critical "parameter", whose optimal value – or at least a suitable choice of it – is crucial to the approximation of the solution. Thus criteria or strategies have to be developed for selecting a suitable or "optimal" outer inverse from our class of regularizers. We will not address this problem here. Instead we focus only on the operator-theoretic aspects of outer inverses as a class of regularizers.

As an example, we note that *Tikhonov's regularization provides an approximate outer inverse, but not an outer inverse.* Let

$$B_\alpha := (K^*K + \alpha I)^{-1}K^*, \qquad \alpha > 0. \tag{3.43}$$

Then

$$B_\alpha - B_\alpha K B_\alpha = \alpha(K^*K + \alpha I)^{-1}B_\alpha, \quad \alpha > 0. \tag{3.44}$$

Note that

$$B_\alpha - B_\alpha K B_\alpha \to 0, \quad \alpha \to 0. \tag{3.45}$$

Regularization operators obtained via spectral families (such as those in [19, 59, 104, 179]) are usually not outer inverses. The same is true for the "regularized" truncated SVD.

Tikhonov's regularization operator and other regularizers obtained by spectral families are approximate outer inverses in the following sense:

$A \in L(X, Y)$ is *approximately outer-invertible* if, for each $\mu \in (0, 1)$, there exists a $B_\mu \in L(Y, X)$ with the following properties:

$$\|(B_\mu A B_\mu - B_\mu)y\| \le C(\mu)\|B_\mu y\| \tag{3.46}$$

and

$$\|B_\mu y\| \le \Gamma_\mu \|y\| \quad \text{for all} \quad y \in Y. \tag{3.47}$$

Each such B_μ is called an approximate outer inverse of A. Properties of $C(\mu)$, Γ_μ and B_μ are to be prescribed for convergence analysis.

For a compact linear operator $K : H_1 \to H_2$, *the filtered truncated SVD*

$$B_r y = \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 + \alpha} \langle y, y_i \rangle \; x_i, \quad \alpha > 0 \tag{3.48}$$

is an approximate outer inverse, but not an outer inverse (choose $\mu = r^{-1}$).

It is also useful to introduce a notion of "approximate right-invertibility", which abstracts some characteristics of regularization methods:

$A \in L(X, Y)$ is called *approximately right-invertible* if, for each $\mu \in (0, 1)$, there exists a norm $\|\cdot\|_\mu$ on X and a $B_\mu : Y \to X$ such that for all $y \in Y$ and all $x \in X$,

$$\|AB_\mu y - y\| \le \mu \; \|y\| \tag{3.49}$$

$$\|B_\mu y\|_\mu \le \Gamma(\mu) \; \|y\|, \tag{3.50}$$

and

$$\|x\|_\mu \to \|x\| \quad \text{as} \quad \mu \to 0^+. \tag{3.51}$$

Each such B_μ is called an *approximate right inverse* of A. Properties of $C(\mu)$ and Γ_μ and V_μ are to be prescribed for convergence analysis.

Each such B_μ is called an *approximate right inverse* of A with a bound $\Gamma(\mu)$ (note that B_μ need not be linear). For regularization one requires $\Gamma(\mu) = O(\mu^{-\gamma})$ or a similar behaviour. Again $\mu \; \|y\|$ in (3.49) may be replaced by $C(\mu) \; \|y\|$.

For regularizers of type I we have bounded outer inverses with infinite-dimensional range. For ill-posed problems for which such regularizers do not exist, we may use approximate outer inverses as regularizers. Again, these can only be approximate inner inverses or approximate right inverses.

Characterizations of ill-posed problems. Next we deal with characterizations of ill-posed problems of so-called type I and II: *Let* Out(L) *denote that set of all* (*nonzero*) *outer inverses to a* (*nonzero*) *linear transformation $L : V \to W$*, where V and W are vector spaces over the same field. The following proposition is immediate:

Let $L : V \to W$ be a (nonzero) linear transformation. Then the following statements are equivalent for any (nonzero) linear transformation $M : W \to V$ (where $\dotplus$ again denotes algebraic direct sum):

(a) $M \in \mathrm{Out}(L)$.
(b) ML is idempotent and $V = \mathcal{R}(M) \dotplus \mathcal{N}(ML)$.
(c) LM is idempotent and $W = N(M) \dotplus \mathcal{R}(LM)$.
(d) LM is idempotent and $\mathcal{R}(M) \cap \mathcal{N}(L) = \{0\}$.

The next proposition establishes the existence of (algebraic) outer inverses with a prescribed range:

Let V_1 be a subspace of V such that $V_1 \cap \mathcal{N}(L) = \{0\}$ and W_2 be an algebraic complement of LV_1. Then there exists an outer inverse M to L such that $\mathcal{R}(M) = V_1$ and $\mathcal{N}(M) = W_2$. Under these conditions $M|LV_1 = (L|V_1)^{-1}$.

We now consider bounded outer inverses of $A \in L(X, Y)$, $A \neq 0$, and their connections with certain topological complements. A topological direct sum will be denoted by $\oplus$. We are again interested in conditions under which there exists a bounded outer inverse B with a range and a null space prescribed.

Let $A : X \to Y$ be a bounded linear operator with non-closed range. Then the ill-posed problem $Ax = y$ is of type I if and only if the following two conditions hold:

(a) $\mathcal{R}(A)$ *contains a closed infinite-dimensional subspace, say M.*
(b) $\mathcal{N}(A)$ *is (topologically) complemented in the subspace $A^{-1}(M)$, the inverse image of M under A.*

In the case of Hilbert spaces, these characterizing conditions take an explicit and simple form.

Let A be a bounded linear operator on a Hilbert space H_1 into a Hilbert space H_2, and let the range of A be non-closed. Then the following statements are equivalent:

(i) *The ill-posed problem $Ax = y$ is of type I,*
(ii) $\mathcal{R}(A)$ *contains a closed infinite-dimensional subspace,*
(iii) *A is not compact.*

An operator $A \in L(X, Y)$ is *strictly singular* if the subspaces $Z \subset X$ for with the restriction $A|Z$ has a bounded inverse on AZ, the image of Z under A, are necessarily finite dimensional (see, e.g., [136]). This notion of a strictly singular operator is clearly the right generalization of a compact operator in Hilbert space. It is natural to ask if it is possible to characterize ill-posed problems of type I in Banach spaces by the condition that "the operator is not strictly singular". However, this is not possible (cf. [185, 186]).

Remarks. Finally some comments should be made (following Nashed [185]):

(i) For various aspects of operator factorizations and operator ranges, [184] has initiated the study of bounded (or densely defined closed) linear operators which have bounded outer inverses of infinite rank within the framework of operator factorization and operator ranges.
(ii) The classification of ill-posed linear problems as proposed here provides also a classification of ill-posed non-linear problems based on properties of outer inverses of the Fréchet or Hadamard derivative of the non-linear operator (cf. [174]). Approximate outer inverses have been used by B.D. Craven and M.Z. Nashed [36] in the context of inverse function theorems when the derivative does not have a bounded inverse of a bounded generalized inverse.

(iii) Examples of operator equations with a non-compact bounded linear operator with non-closed range arise from various integral and convolution operators on the half-line and for certain generalized Wiener–Hopf operators. They also arise if zero belongs to the continuous spectrum of a linear, bounded selfadjoint and non-negative operator on a Hilbert space.

(iv) For various aspects of construction and computation of outer inverses of a prescribed rank $s \leq r$ for a matrix of rank r, see [23] and several papers and the annotated bibliography in [177]. For constructions of outer inverses of bounded linear operators in Banach spaces the reader is referred to [165, 166, 179]. For stability properties for outer inverses, see [184–186]. For perturbation analysis of outer inverses, see [166, 179, 185, 186]. Convergence analysis of regularization operators hinges upon perturbation and stability properties of outer inverses (see [185, 186]).

4. Reconstruction methods and regularizing filters

Assume that A is of class $L(X,Y)$, X, Y Hilbert spaces. Let $\{R_t\}_{t>0}$ be a family of continuous operators (regularizers) from Y to X with $R_t 0 = 0$. If there exists a mapping $\alpha : (0,\infty) \times Y \to (0,\infty)$, such that, for all elements $y \in \mathcal{R}(A)$ and regularization parameter $t = \alpha(\varepsilon, y^\varepsilon)$,

$$\sup\left\{\left\|A^\dagger y - R_{\alpha(\varepsilon,y^\varepsilon)} y^\varepsilon\right\| : y^\varepsilon \in Y \text{ with } \|y - y^\varepsilon\| \leq \varepsilon\right\} \overset{\varepsilon\to 0}{\longrightarrow} 0, \tag{4.1}$$

then the pair $\left(\{R_t\}_{t>0}, \alpha\right)$ is called a *regularization method* (or simply *a regularization*) *of* $A^\dagger$.

If the sequence $\{R_t\}_{t>0}$ is a subset of $L(Y,X)$, then the regularization $\left(\{R_t\}_{t>0}, \alpha\right)$ is called linear. The mapping α is known as the *parameter choice* provided that

$$\sup\left\{\alpha(\varepsilon, y^\varepsilon) : y^\varepsilon \in Y \text{ with } \|y - y^\varepsilon\| \leq \varepsilon\right\} \overset{\varepsilon\to 0}{\longrightarrow} 0. \tag{4.2}$$

If α is only dependent on ε, i.e., $\alpha(\varepsilon, y^\varepsilon) = \alpha(\varepsilon)$, we speak of an *a priori parameter choice*. If α is dependent on ε and y^ε, i.e., $\alpha = \alpha(\varepsilon, y^\varepsilon)$, it is called an *a posteriori parameter choice*.

The limit relation (4.1) can be equivalently written in the form

$$\sup\left\{\left\|x - R_{\alpha(\varepsilon,y^\varepsilon)} y^\varepsilon\right\| : y^\varepsilon \in Y \text{ with } \|Ax - y^\varepsilon\| \leq \varepsilon\right\} \overset{\varepsilon\to 0}{\longrightarrow} 0 \tag{4.3}$$

for all $x \in \mathcal{N}(A)^\perp$. A direct consequence of a regularization method is that the limit relation

$$\lim_{\varepsilon\to 0} \left\|A^\dagger y - R_{\alpha(\varepsilon,y)} y\right\| = 0 \tag{4.4}$$

holds true for all $y \in \mathcal{R}(A)$ (note that, in Equation (4.4), the regularization is applied to y instead of y^ε). It is usual to collect all regularization parameters that are relevant in the limit relation (4.4) in a set $\Gamma = \{\alpha(\varepsilon, y) : \varepsilon > 0, y \in \mathcal{R}(A)\}$. Because of the fact that $\lim_{\varepsilon\to 0} \alpha(\varepsilon, y) = 0$ the set Γ possesses an accumulation

point at 0. Moreover, $\lim_{\Gamma \ni \lambda \to 0} \left\| A^\dagger y - R_\lambda y \right\| = 0$ for all $y \in \mathcal{R}(A)$. All in all, we are led to the following result:

Assume that A is of class $L(X,Y)$. Let $\left(\{R_t\}_{t>0}, \alpha\right)$ be a regularization method of $A^\dagger$. Then, the subfamily $\{R_\lambda\}_{\lambda \in \Gamma}$ converges pointwise in $\mathcal{R}(A)$ to $A^\dagger$ for $\lambda \to 0$.

As a direct consequence we obtain the statement (see, e.g., Louis [148], Rieder [227]):

Assume that A is of class $L(X,Y)$. Let $\left(\{R_t\}_{t>0}, \alpha\right)$ be a regularization of $A^\dagger$. If $\mathcal{R}(A) \neq \overline{\mathcal{R}(A)}$, then $\{\|R_t\|\}_{t>0}$ is unbounded.

As already known, for a linear regularization, the *reconstruction error* $\|A^\dagger y - R_t y^\varepsilon\|$ can be split by use of a family $\{R_t\}_{t>0}$ in the following way:

$$\begin{aligned} \left\| A^\dagger y - R_t y^\varepsilon \right\| &\leq \underbrace{\left\| A^\dagger y - R_t y \right\|}_{\text{approximation error}} + \underbrace{\| R_t y - R_t y^\varepsilon \|}_{\text{data error}} \\ &\leq \|R_t\| \, \|y - y^\varepsilon\| \leq \|R_t\| \cdot \varepsilon. \end{aligned} \tag{4.5}$$

Once again, it should be pointed out, that both parts of the reconstruction error, i.e., the approximation error and the data error, exhibit an opposed behavior in limit considerations for $t \to 0$ and $t \to \infty$.

Hence, an essential task is to search for a value t_{opt} that balances both errors. In fact, the regularization parameter should be chosen in such a way that $\alpha(\varepsilon, y^\varepsilon) \approx t_{\text{opt}}$.

Classification of regularization methods. The question (cf. [148]) arises how to classify regularization methods. Since all methods need to converge as $\varepsilon \to 0$, an obvious criterion of specifying their characteristics is the speed of the convergence, i.e., we make the attempt to introduce a classification with respect to the speed of convergence of the total error:

$$\sup \left\{ \left\| A^\dagger y - R_{\alpha(\varepsilon, y^\varepsilon)} y^\varepsilon \right\| : y \in \mathcal{R}(A), y^\varepsilon \in Y \text{ with } \|y - y^\varepsilon\| \leq \varepsilon \right\} \overset{\varepsilon \to 0}{\longrightarrow} 0. \tag{4.6}$$

Unfortunately, it turns out that the speed of the convergence is arbitrarily slow for <u>all</u> regularization methods.

Let A be of class $L(X,Y)$. Assume that $\mathcal{R}(A) \neq \overline{\mathcal{R}(A)}$. Let $(\{R_t\}_{t>0}, \alpha)$ be a regularization of $A^\dagger$. Then there exists <u>no</u> function $h : [0,\infty) \to [0,\infty)$ with $\lim_{\varepsilon \to 0} h(\varepsilon) = 0$ such that

$$\sup \left\{ \left\| A^\dagger y - R_{\alpha(\varepsilon, y^\varepsilon)} y^\varepsilon \right\| : y \in \mathcal{R}(A), \|y\| \leq 1, y^\varepsilon \in Y \text{ with } \|y - y^\varepsilon\| \leq \varepsilon \right\} \leq h(\varepsilon). \tag{4.7}$$

Powers of absolute values and smooth Hilbert spaces. Next our interest is to show that the concept of smoothness can be used for the classification of regularization methods. An auxiliary tool is the *absolute value $|A|$ of the operator A*. In order to motivate the setting $|A|$ we start with the explanation of a *functional calculus for compact operators* which also helps us to introduce filters for purposes of

regularization in the next subsection. Even better, the functional calculus for compact operators enables us to reduce spectral features to a study in terms of real functions.

Let A be of class $K(X,Y)$ with the singular system $\{\sigma_j; x_j, y_j\}_{j\in\mathbb{N}}$. Let $\Phi : [0,\infty) \to \mathbb{R}$ be a piecewise continuous function defined on the interval $\left[0, \|A\|^2\right]$. Then we understand the operator $\Phi(A^*A)$ to be given as

$$\Phi(A^*A)\, x = \sum_{j=1}^{\infty} \Phi\left(\sigma_j^2\right) \langle x, x_j\rangle\, x_j + \Phi(0)\, P_{\mathcal{N}(A)} x. \tag{4.8}$$

Note that the series on the right side of (4.8) is convergent on the interval $\left[0, \|A\|^2\right]$.

Let $A \in K(X,Y)$ be a compact operator with the singular system

$$\{\sigma_j; x_j, y_j\}_{j\in\mathbb{N}}.$$

Suppose that $\Phi : [0,\infty) \to \mathbb{R}$ is piecewise continuous. Then the following properties hold true:

(a) $\|A\| = \sigma_1$, *i.e., the norm coincides with the largest singular value of A,*

(b) $$\Phi(A^*A)A^* = A^*\Phi(AA^*), \tag{4.9}$$

(c) $$\|\Phi(A^*A)\| = \sup_{j\in\mathbb{N}} |\Phi(\sigma_j^2)| \le \sup_{0\le\lambda\le\|A\|^2} |\Phi(\lambda)|, \tag{4.10}$$

(d) $$\|\Phi(A^*A)A^*\| = \|A^*\Phi(AA^*)\| = \sup_{j\in\mathbb{N}} \left(\sigma_j |\Phi(\sigma_j^2)|\right) \le \sup_{0\le\lambda\le\|A\|^2} \left(\sqrt{\lambda}\, |\Phi(\lambda)|\right). \tag{4.11}$$

Example. If $\Phi = 1$, then it is clear that

$$\Phi(A^*A)\, x = \sum_{j=1}^{\infty} \langle x, x_j\rangle\, x_j + P_{\mathcal{N}(A)} x = P_{\overline{\mathcal{R}(A^*)}} x + P_{\mathcal{N}(A)} x = P_X x = x. \tag{4.12}$$

This explains the occurrence of the term $\Phi(0)\, P_{\mathcal{N}(A)} x$ in (4.8).

Example. If $\Phi(t) = t^{1/2}, t \ge 0$, then

$$(A^*A)^{1/2}\, x = \sum_{j=1}^{\infty} \sigma_j \langle x, x_j\rangle\, x_j + \underbrace{\Phi(0)}_{=0} \left(P_{\mathcal{N}(A)} x\right) = \sum_{j=1}^{\infty} \sigma_j \langle x, x_j\rangle\, x_j, \tag{4.13}$$

holds true for all $x \in X$.

The operator $|A| = (A^*A)^{1/2}$ is called *the absolute value of A*. Analogously, $|A^*|$ is given by

$$|A^*|\, y = (AA^*)^{1/2}\, y = \sum_{j=1}^{\infty} \sigma_j \langle y, y_j\rangle\, y_j, \quad y \in Y. \tag{4.14}$$

It is easy to see that

$$\|\, |A|\, x\|^2 = \langle |A|x, |A|x\rangle = \langle A^*Ax, x\rangle = \langle Ax, Ax\rangle = \|Ax\|^2. \tag{4.15}$$

Remark. Our notation (4.13) can be used to introduce fractional powers $(A^*A)^\mu$, $\mu \geq 0$, of A^*A:

$$|A|^{2\mu}x = (A^*A)^\mu x = \sum_{j=1}^{\infty} \sigma_j^{2\mu} \langle x, x_j \rangle x_j. \tag{4.16}$$

Let X, Y be Hilbert spaces. Suppose that A is of class $K(X,Y)$. Then

(1) $\mathcal{R}(A^*) = \mathcal{R}(|A|) = \mathcal{R}\left((A^*A)^{1/2}\right)$,
(2) $\mathcal{R}(A) = \mathcal{R}(|A^*|) = \mathcal{R}\left((AA^*)^{1/2}\right)$.

Next we introduce subspaces of the Hilbert space X involving on the concept of powers of the absolute value $|A|$ of the operator A: For $\nu \geq 0$, let X_ν be defined by

$$X_\nu = \mathcal{R}\left(|A|^\nu\right) = \left\{ |A|^\nu x : x \in \mathcal{N}(A)^\perp \right\}. \tag{4.17}$$

Then the following properties can be verified by straightforward arguments:

(1) $X_\nu \subset \mathcal{N}(A)^\perp \subset X$ for all $\nu \geq 0$,
(2) $X_\nu \subset X_\mu$ for $\nu > \mu \geq 0$,
(3) $X_0 = \mathcal{N}(A)^\perp$.

In connection with

$$x = |A|^\nu z = \sum_{k=1}^{\infty} \sigma_k^\nu \langle z, x_k \rangle x_k \tag{4.18}$$

we are able to impose the following norm on X_ν :

$$\|x\|_\nu^2 = \|z\|^2 = \sum_{k=1}^{\infty} \frac{|\langle z, x_k \rangle|^2}{\sigma_k^{2\nu}} \sigma_k^{2\nu} = \sum_{k=1}^{\infty} \frac{|\langle x, x_k \rangle|^2}{\sigma_k^{2\nu}}. \tag{4.19}$$

Alternatively, the spaces X_ν can be characterized by the norms (4.19), i.e.,

$$X_\nu = \left\{ x \in \mathcal{N}(A)^\perp : \|x\|_\nu < \infty \right\}. \tag{4.20}$$

The spaces X_ν impose conditions on the smoothness of the elements $x \in X$.

After these preliminaries about powers of absolute values we come back to the discussion of the speed of the convergence of a regularization method involving the concept of smoothness, i.e., the solution of an operator equation will be assumed to be a member of a subspace X_ν, $\nu > 0$, of X.

Indeed, in a large number of ill-posed problems $(A; X, Y)$ the operator A shows the property that the image Ax is smoother than x. In concrete situations this leads us to functions x_k in the singular system with strongly growing oscillations for increasing k. Looking at the higher frequency parts of an element $x \in X$, i.e., the inner products $\langle x, x_k \rangle$ for large k, we notice that there is a damping effect on Ax by the factor σ_k. Hence, the norm $\|x\|_\nu$ can be interpreted in the sense that $\langle x, x_k \rangle / \sigma_k^\nu \to 0$ for $k \to \infty$ is demanded. In addition, the larger the value of ν is chosen, the faster $\langle x, x_k \rangle$ has to converge to 0. Therefore it can be concluded that the element x does not contain relevant high frequency components. In other words, x can be regarded as "smooth".

Optimality of reconstruction methods. In the following, a stable *reconstruction method* for the solution of the operator equation involving $A \in L(X,Y)$ is understood to be a continuous (not-necessarily linear) mapping $T : Y \to X$ with $T\,0 = 0$. The question is which reconstruction error does occur in the *best worst case*, if the data are noisy.

The *worst case error of a reconstruction method* T for A corresponding to the noise level ε and the additional information $\|A^\dagger y\|_\nu \le \rho$ is defined as

$$E_\nu(\varepsilon,\rho,T) = \sup\left\{\|Ty^\varepsilon - A^\dagger y\| :\ y \in \mathcal{R}(A), y^\varepsilon \in Y, \|y - y^\varepsilon\| \le \varepsilon, \|A^\dagger y\|_\nu \le \rho\right\}. \tag{4.21}$$

Another expression (avoiding the occurrence of the operator $A^\dagger$) can be formulated by

$$E_\nu(\varepsilon,\rho,T) = \sup\left\{\|Ty^\varepsilon - x\| :\ x \in X_\nu, y^\varepsilon \in Y, \|Ax - y^\varepsilon\| \le \varepsilon, \|x\|_\nu \le \rho\right\}. \tag{4.22}$$

Clearly, the smaller the worst case error, the better the reconstruction method.

The *best worst case error* for A corresponding to the noise level ε and the additional restriction $\|A^\dagger y\|_\nu \le \rho$ is understood to be

$$E_\nu(\varepsilon,\rho) = \inf\left\{E_\nu(\varepsilon,\rho,T) :\ T : Y \to X \text{ continuous},\ T0 = 0\right\}. \tag{4.23}$$

Note that the best worst case error, i.e., $E_\nu(\varepsilon,\rho)$, depends on the problem (i.e., on the operator A), but not on the reconstruction method.

The next result tells us about the quantity of the best worst case error.

Let A be of class $L(X,Y)$. Then we have

$$E_\nu(\varepsilon,\rho) = e_\nu(\varepsilon,\rho), \tag{4.24}$$

where

$$e_\nu(\varepsilon,\rho) = \sup\left\{\|x\| :\ x \in X_\nu, \|Ax\| \le \varepsilon, \|x\|_\nu \le \rho\right\}. \tag{4.25}$$

Clearly, our results characterize the best worst case error independently of the knowledge of a specific reconstruction method.

Let A be of class $L(X,Y)$. Then, for $\nu > 0$,

$$e_\nu(\varepsilon,\rho) \le \rho^{\frac{1}{\nu+1}}\varepsilon^{\frac{\nu}{\nu+1}}. \tag{4.26}$$

Furthermore, there exists a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ with $\varepsilon_k \to 0$ for $k \to \infty$ such that

$$e_\nu(\varepsilon,\rho) = \rho^{\frac{1}{\nu+1}}\varepsilon^{\frac{\nu}{\nu+1}}. \tag{4.27}$$

In other words, the estimate (4.26) is sharp, i.e., it cannot be improved.

Regularizing filters. If A is an injective operator of class $K(X,Y)$, then $A^\dagger$ can be expressed in the form $(A^*A)^{-1}A^*$. The non-continuity is caused by the term $(A^*A)^{-1}$, that has to be stabilized. In connection with the functional calculus for compact operators we are therefore led to filters as appropriate tools for regularization.

Let $\{F_t\}_{t>0}$, $F_t : \left[0, \|A\|^2\right] \to \mathbb{R}$ be a family of piecewise continuous functions satisfying the conditions

(F1) $\lim_{t\to 0} F_t(\lambda) = \frac{1}{\lambda}$ for all $\lambda \in \left(0, \|A\|^2\right]$,

(F2) $\lambda |F_t(\lambda)| \le C_F$ for all $\lambda \in \left[0, \|A\|^2\right]$ and $t > 0$.

Then the family $\{F_t\}_{t>0}$ is called a *filter relative to A*.

By virtue of Condition (F1), $F_t(A^*A)$ becomes a continuous operator, which converges in pointwise sense to $(A^*A)^{-1}$ as $t \to 0$. This is the reason why we let

$$R_t y = F_t(A^*A)\, A^* y, \quad y \in Y. \tag{4.28}$$

As a consequence, $\{F_t\}_{t>0}$ filters the influence of small singular values of A on the operator R_t in (4.28).

In terms of the singular system $\{\sigma_k; x_k, y_k\}_{k\in\mathbb{N}}$ we are able to write

$$\begin{aligned} F_t(A^*A)\, A^* y &= \sum_{k=1}^{\infty} F_t\left(\sigma_k^2\right) \sigma_k \langle y, y_k\rangle x_k + F_t(0)\, P_{\mathcal{N}(A)} A^* y \\ &= \sum_{k=1}^{\infty} F_t\left(\sigma_k^2\right) \sigma_k \langle y, y_k\rangle x_k \end{aligned} \tag{4.29}$$

due to fact that $P_{\mathcal{N}(A)} A^* y = 0$ (note that $A^* y \in \mathcal{R}(A^*) = \mathcal{N}(A)^{\perp}$). Considering the approximation error we obtain, for $y \in \mathcal{R}(A)$, that

$$\begin{aligned} A^\dagger y - R_t y &= A^\dagger y - F_t(A^*A)\, A^* y \\ &= A^\dagger y - F_t(A^*A)\, A^* A A^\dagger y \\ &= (I - F_t(A^*A)\, A^* A)\, A^\dagger y \\ &= p_t(A^*A)\, A^\dagger y, \end{aligned} \tag{4.30}$$

where the function $p_t : \lambda \mapsto p_t(\lambda), \lambda \in \left[0, \|A\|^2\right]$ is given by

$$p_t(\lambda) = 1 - \lambda F_t(\lambda), \quad \lambda \in \left[0, \|A\|^2\right]. \tag{4.31}$$

The identity (4.30) leads us to the formulation of the following result:

Assume that $A \in K(X,Y)$. Let $\{F_t\}_{t>0}$ be a filter. Then

$$\lim_{t\to 0} \quad R_t y = \begin{cases} A^\dagger y & y \in \mathcal{D}\left(A^\dagger\right), \\ \infty & y \notin \mathcal{D}\left(A^\dagger\right), \end{cases} \tag{4.32}$$

where

$$R_t y = F_t(A^*A)\, A^* y$$

for $y \in Y$.

The next result concretizes the stability of $R_t y^\varepsilon$ under the noise level ε in more detail:

Let $\{F_t\}_{t>0}$ *be a filter so that* (F1), (F2) *hold true. For* $y, y^\varepsilon \in Y$ *with* $\|y - y^\varepsilon\| \leq \varepsilon$, *set* $x_t = R_t y$ *and* $x_t^\varepsilon = R_t y^\varepsilon$. *Then, for the residual term, we have*

$$\|Ax_t - Ax_t^\varepsilon\| \leq C_F \varepsilon, \tag{4.33}$$

while, for the error term, it follows that

$$\|x_t - x_t^\varepsilon\| \leq \varepsilon \sqrt{C_F M(t)}$$

where we have used the abbreviation

$$M(t) = \sup_{0 \leq \lambda \leq \|A\|^2} |F_t(\lambda)|. \tag{4.34}$$

Let us continue with the *estimate of the total error*, thereby using $x^\dagger = A^\dagger y$, $y \in D\left(A^\dagger\right)$. We base our considerations on the usual splitting into the approximation error and the data error in the form

$$\begin{aligned} \left\|A^\dagger y - R_t y^\varepsilon\right\| &\leq \left\|A^\dagger y - R_t y\right\| + \|R_t y - R_t y^\varepsilon\| \\ &= \left\|x^\dagger - x_t\right\| + \|x_t - x_t^\varepsilon\| \\ &\overset{t \to 0}{\leq} \underbrace{\left\|x^\dagger - x_t\right\|}_{\to 0} + \varepsilon \sqrt{C_F M(t)}. \end{aligned} \tag{4.35}$$

Obviously, the approximation error $\|x^\dagger - R_t y\|$ tends to 0. The discussion of the data error is much more problematic than the approximation error: From the limit relation $\lim_{t \to 0} F_t(\lambda) = \frac{1}{\lambda}$ imposed on a filter within the interval $[0, \|A\|^2]$ it follows that $M(t) \overset{t \to 0}{\longrightarrow} \infty$. Therefore, for the total error, we are confronted with a divergent behavior as $t \to 0$. Nevertheless, convergence properties can be forced by a suitable coupling of t and ε. In fact, it can be deduced that, under the a priori parameter choice indicated above, the approximation error as well as the data error converge to 0, if the noise level ε tends to zero (cf. [227]):

Let $\{F_t\}_{t>0}$ *be a filter. If we choose* $\alpha : (0, \infty) \to (0, \infty), \varepsilon \mapsto \alpha(\varepsilon)$, *such that* $\alpha(\varepsilon) \overset{\varepsilon \to 0}{\longrightarrow} 0$ *as well as* $\varepsilon \sqrt{M(\alpha(\varepsilon))} \overset{\varepsilon \to 0}{\longrightarrow} 0$, *then* $\left(\{R_t\}_{t>0}, \alpha\right)$ *with* $R_t = F_t(A^*A)A^*$ *is a regularization of* $A^\dagger$ (*by convention,* $\{F_t\}_{t>0}$ *is called a regularizing filter*).

Next we mention which additional requirement imposed on filters $R_t = F_t(A^*A)A^*$ will be necessary to guarantee the order optimality.

Let $\{F_t\}_{t>0}$ *be a regularizing filter for* $A \in L(X, Y)$. *Assume there are* $t_0 > 0$, $\mu > 0$, *and a function* $\omega_\mu : (0, t_0] \to \mathbb{R}$ *such that*

$$\sup_{0 \leq \lambda \leq \|A\|^2} \lambda^{\mu/2} |p_t(\lambda)| \leq \omega_\mu(t) \tag{4.36}$$

for all $t \in (0, t_0]$ (*with* $p_t(\lambda) = 1 - \lambda F_t(\lambda)$). *Let* $y \in \mathcal{R}(A)$ *and let* $x^\dagger = A^\dagger y$ *in* X_μ *with* $\|x^\dagger\|_\mu \leq \varrho$. *Then the following estimates*

(a) $\|x^\dagger - x_t\|_X \leq \omega_\mu(t)\varrho$,
(b) $\|Ax^\dagger - Ax_t\|_Y \leq \omega_{\mu+1}(t)\varrho$

hold true for $x_t = R_t y = F_t(A^*A)A^*y$ *and* $0 < t < t_0$. *Moreover, let*

(1) $\omega_\mu(t) \leq C_p t^{\frac{\mu}{2}}$ *for* $t \to 0$,
(2) $M(t) = \sup_{0 \leq \lambda \leq \|A\|^2} |F_t(\lambda)| \leq C_M t^{-1}$ *for* $t \to 0$,

where $\mu > 0$, $C_p, C_M > 0$ *are constants. Let the a priori parameter choice* $\alpha : (0,\infty) \to (0,\infty)$ *fulfill*

$$C_1 \left(\frac{\varepsilon}{\rho}\right)^{\frac{2}{\mu+1}} \leq \alpha(\varepsilon) \leq C_2 \left(\frac{\varepsilon}{\rho}\right)^{\frac{2}{\mu+1}}, \quad \varepsilon \to 0, \tag{4.37}$$

where C_1, C_2 *are positive constants. Then,* $(\{R_t\}_{t>0}, \alpha)$, $R_t = F_t(A^*A)A^*$, *is an order optimal regularization for* $A^\dagger$ *with respect to* X_μ.

Of course, we need to know the values ρ *and* μ *to guarantee the order optimality. Without the availability of* ρ *but based on the knowledge of* μ, *we are led to* $\alpha(\varepsilon) = C\varepsilon^{\frac{2}{\mu+1}}$ *with* C *being a positive constant to find an order optimal regularization. Without any information of both parameters* ρ *as well as* μ *we have to deal with a posteriori parameter choices.*

An asymptotic behavior of ω_μ determines the speed of convergence for the reconstruction error. An important feature is the so-called *qualification*.

Let $\{F_t\}_{t\geq 0}$ be a regularizing filter for $A^\dagger$, where $A \in L(X,Y)$ satisfying the asymptotic relation

$$M(t) \leq \frac{C_M}{t}, \quad t \to 0. \tag{4.38}$$

where C_M is a positive constant. The maximal value μ_0, such that there exists, for all values $\mu \in (0, \mu_0]$, a constant $C_p > 0$ satisfying

$$\sup \lambda^{\mu/2} |p_t(\lambda)| \leq C_p\, t^{\mu/2}, \quad t \to 0, \tag{4.39}$$

is called the *qualification of the filter* (remember $p_t(\lambda) = 1 - \lambda F_t(\lambda)$).

In other words, the qualification can be regarded as the maximal rate of decay. If the qualification is finite, there exists a parameter choice $\alpha(\varepsilon)$ such that

$$\left\|A^\dagger y - R_{\alpha(\varepsilon)} y^\varepsilon\right\| = O\left(\varepsilon^{\mu_0/(\mu_0+1)}\right), \ \varepsilon \to 0, \tag{4.40}$$

holds true for $A^\dagger y \in X_{\mu_0}$. If the qualification is infinite, there exists a parameter choice $\alpha(\varepsilon)$ for which the error decay comes arbitrarily close to $O(\varepsilon)$. As a consequence, filters with infinite qualification are more advantageous than others.

Order optimality of special regularizing filters. In the following we recover important examples of regularization methods constituted by filters, i.e., the truncated singular value decomposition SVD and the Tikhonov–Phillips regularization.

Truncated singular value decomposition revisited: The SVD of $A^\dagger$ is

$$A^\dagger y = \sum_{k=1}^{\infty} \frac{1}{\sigma_k} \langle y, y_k \rangle\, x_k.$$

We choose the filter

$$F_t(\lambda) = \begin{cases} \frac{1}{\lambda}, & \lambda \geq t, \\ 0, & \lambda < t. \end{cases} \tag{4.41}$$

Then

$$R_t y = F_t(A^*A)\, A^* y = \sum_{k=1}^{\infty} F_t\left(\sigma_k^2\right) \sigma_k \langle y, y_k \rangle x_k = \sum_{\sigma_k \geq \sqrt{t}} \frac{1}{\sigma_k} \langle y, y_k \rangle x_k \tag{4.42}$$

is the truncated SVD of $A^\dagger$ (with finitely many summands).

Clearly we have

(1) $\lim_{t\to 0} F_t(\lambda) = \frac{1}{\lambda}$ for all $\lambda > 0$,
(2) $C_F = \sup_{0\leq\lambda\leq\|A\|^2} \lambda |F_t(\lambda)| = 1$ for all $t \leq \|A\|^2$,
(3) $M(t) = \sup_{0\leq\lambda\leq\|A\|^2} |F_t(\lambda)| = \frac{1}{t}$ for all $t > 0$.

The total error can be described as follows:

$$\begin{aligned} \left\|A^\dagger y - R_t y^\varepsilon\right\| &\leq \left\|A^\dagger y - R_t y\right\| + \left\|R_t y - R_t y^\varepsilon\right\| \\ &= \left\| \sum_{\sigma_k < \sqrt{t}} \frac{1}{\sigma_k} \langle y, y_k \rangle x_k \right\| + \varepsilon \sqrt{C_F M(t)} \\ &\overset{\text{Parseval}}{=} \left(\sum_{\sigma_k < \sqrt{t}} \frac{1}{\sigma_k^2} |\langle y, y_k \rangle|^2 \right)^{1/2} + \frac{\varepsilon}{\sqrt{t}}. \end{aligned} \tag{4.43}$$

Moreover, for all $\mu > 0$ and $0 \leq t \leq \|A\|^2$, we have

$$\sup_{0\leq\lambda\leq\|A\|^2} \lambda^{\mu/2} |p_t(\lambda)| = \sup_{0\leq\lambda\leq\|A\|^2} \lambda^{\mu/2} |1 - \lambda F_t(\lambda)| = \sup_{0\leq\lambda\leq t} \lambda^{\mu/2} = t^{\mu/2}.$$

Together with $M(t) = \frac{1}{t}$ we find that this filter possesses an infinite qualification, and it is order optimal for all $\mu > 0$. However, one can show that the TSVD is *not* optimal for *any* $\mu > 0$.

Asymptotic regularization revisited. We choose the following filter

$$F_t(\lambda) = \begin{cases} \frac{1-\exp\left(-\frac{\lambda}{t}\right)}{\lambda} & \lambda > 0 \\ \frac{1}{t} & \lambda = 0. \end{cases} \tag{4.44}$$

Then we obtain

$$R_t y = \sum_{k=1}^{\infty} F_t\left(\sigma_k^2\right) \sigma_k \langle y, y_k \rangle x_k = \sum_{k=1}^{\infty} \frac{1 - \exp\left(-\frac{\sigma_k^2}{t}\right)}{\sigma_k} \langle y, y_k \rangle \; x_k.$$

It is not difficult to prove the following properties:

(1) $\lim_{t\to 0} F_t(\lambda) = \frac{1}{\lambda}$ for $\lambda > 0$
(2) $\lambda |F_t(\lambda)| = \left\{ \begin{array}{c} 1 - \exp\left(-\frac{\lambda}{t}\right), \quad \lambda > 0 \\ \frac{\lambda}{t}, \quad \lambda = 0 \end{array} \right\} . \leq 1 \quad , \quad t > 0,$

(3) $\sup_{0\le\lambda\le\|A\|^2} \lambda\,|F_t(\lambda)| = 1 = C_F$,
(4) $M(t) = \sup_{0\le\lambda\le\|A\|^2} |F_t(\lambda)| = \frac{1}{t}$, $t>0$ (note that $F_t(\lambda)$ is monotonously decreasing in λ and $\lim_{\lambda\to 0} F_t(\lambda) = \frac{1}{t}$).

The qualification of the filter again is infinite.

Tikhonov's regularization revisited. Using the filter

$$F_t(\lambda) = \frac{1}{\lambda + t}, \quad t>0, \tag{4.45}$$

we find that

$$R_t y = F_t(A^*A)\,A^*y = \sum_{k=1}^{\infty} \frac{\sigma_k}{\sigma_k^2 + t}\,\langle y, y_k\rangle\, x_k \tag{4.46}$$

and

$$\begin{aligned}(A^*A + tI)\,R_t y &= \sum_{k=1}^{\infty} \frac{\sigma_k}{\sigma_k^2+t}\,\langle y, y_k\rangle\,\left(\sigma_k^2 + t\right) x_k = \sum_{k=1}^{\infty} \sigma_k\,\langle y, y_k\rangle\, x_k \\ &= \sum_{k=1}^{\infty} \langle y, \sigma_k y_k\rangle\, x_k = \sum_{k=1}^{\infty} \langle y, Ax_k\rangle\, x_k = \sum_{k=1}^{\infty} \langle A^*y, x_k\rangle\, x_k = A^*y,\end{aligned} \tag{4.47}$$

i.e., $x_t = R_t y$ is the unique solution of the equations $(A^*A + tI)\,x_t = A^*y$. Such equations are called *regularized normal equations*. It can be easily seen that

(1) $\lim_{t\to 0} F_t(\lambda) = \frac{1}{\lambda}$, $\lambda > 0$,
(2) $\lambda\,|F_t(\lambda)| = \frac{\lambda}{\lambda+t} \le 1 = C_F$ for all $t>0$,
(3) $M(t) = \sup_{0\le\lambda\le\|A\|^2} |F_t(\lambda)| = \sup_{0\le\lambda\le\|A\|^2} \frac{1}{\lambda+t} = \frac{1}{t}$, $t>0$.

In order to compute the qualification for Tikhonov's regularization we take a look at the term

$$\sup_{0\le\lambda\le\|A\|^2} \lambda^{\frac{\mu}{2}} |p_t(\lambda)| = \sup_{0\le\lambda\le\|A\|^2} \underbrace{t^{\frac{\mu}{2}} \frac{\left(\frac{\lambda}{t}\right)^{\frac{\mu}{2}}}{\frac{\lambda}{t}+1}}_{=h_\mu(\lambda,t)}. \tag{4.48}$$

In fact, we are led to distinguish two cases:

Case 1: For $\mu > 2$: $h_\mu(\lambda, t)$ is strictly monotonously increasing in λ.

Case 2: For $\mu \le 2$: we find the estimate

$$\sup_{0\le\lambda\le\|A\|^2} h_\mu(\lambda,t) \le t^{\frac{\mu}{2}} \underbrace{\sup_{0\le z<\infty} \frac{z^{\frac{\mu}{2}}}{z+1}}_{=C_p<\infty}. \tag{4.49}$$

Summarizing our considerations we obtain

$$\sup_{0\le\lambda\le\|A\|^2} \lambda^{\frac{\mu}{2}} |p_t(\lambda)| \le \begin{cases} C_p t^{\frac{\mu}{2}}, & : \quad 0<\mu\le 2, \\ \|A\|^{\mu-2} t, & : \quad \mu > 2. \end{cases} \tag{4.50}$$

In other words, the qualification of the Tikhonov filter is $\mu_0 = 2$, i.e., we arrive at the order optimality and even at the optimality for $0 < \mu \leq 2$, if we use the parameter choice $\alpha(\varepsilon) = \frac{1}{\mu}\left(\frac{\varepsilon}{\rho}\right)^{\frac{2}{\mu+1}}$.

Morozov's discrepancy principle. Let $y \in \mathcal{D}\left(A^{\dagger}\right)$ and $y^{\varepsilon} \in Y$ with $\|y - y^{\varepsilon}\| < \varepsilon$ and $x_t^{\varepsilon} = F_t\left(A^*A\right)A^*y^{\varepsilon}$. The *idea of the discrepancy principle* can be explained as follows: Choose the parameter $\alpha = \alpha\left(\varepsilon, y^{\varepsilon}\right)$ such that

$$\|Ax_t^{\varepsilon} - y^{\varepsilon}\| \approx \varepsilon. \tag{4.51}$$

In other words, the *residual* or *discrepancy* of y^{ε} is assumed to have the order of the data error.

In order to realize the assumption (4.51) we consider the defect function

$$d : t \mapsto d(t) = \|Ax_t^{\varepsilon} - y^{\varepsilon}\|. \tag{4.52}$$

It is not hard to see that

$$\begin{aligned} d(t) &= \left\|AF_t\left(A^*A\right)A^*y^{\varepsilon} - y^{\varepsilon}\right\| \\ &= \left\|p_t\left(AA^*\right)y^{\varepsilon}\right\| \\ &= \left(\sum_{k=1}^{\infty} p_t^2\left(\sigma_k^2\right)\left|\langle y^{\varepsilon}, y_k\rangle\right|^2 + \underbrace{p_t(0)}_{=1}\left\|P_{\mathcal{N}(A^*)}y^{\varepsilon}\right\|^2\right)^{1/2}, \end{aligned} \tag{4.53}$$

so that

$$\lim_{t\to 0, t>0} d(t) = \left\|P_{\mathcal{N}(A^*)}y^{\varepsilon}\right\| = \left\|P_{\mathcal{R}(A)^{\perp}}y^{\varepsilon}\right\|. \tag{4.54}$$

If $y \notin \mathcal{R}(A)$, then $\left\|P_{\mathcal{R}(A)^{\perp}}y^{\varepsilon}\right\|_Y$ can be arbitrarily large. However, if $y \in \mathcal{R}(A)$, then we have

$$\left\|P_{\mathcal{R}(A)^{\perp}}y^{\varepsilon}\right\| = \left\|P_{\mathcal{R}(A)^{\perp}}\left(y - y^{\varepsilon}\right)\right\| \leq \|y - y^{\varepsilon}\| \leq \varepsilon \tag{4.55}$$

and, therefore,

$$\lim_{t\to 0, t>0} d(t) \leq \varepsilon \tag{4.56}$$

In other words, for all $\tau > 1$ exists a t_0 with $d(t) < \tau\varepsilon$ for all $t \leq t_0$.

Let $\tau > 1$ be chosen (fixed) and $\{t_k\}$ be a strictly monotonously decreasing sequence with limit zero. Determine k^*, such that

$$d\left(t_{k^*}\right) \leq \tau\varepsilon \leq d\left(t_i\right), \quad i = 1, \ldots, k^* - 1 \tag{4.57}$$

Set $\alpha\left(\varepsilon, y^{\varepsilon}\right) = t_{k^*}$.

Let A be of class $L(X, Y)$. Suppose that $\{F_t\}_{t>0}$ is a regularizing filter with qualification $\mu_0 > 1$. Moreover, assume that $M(t) \leq \frac{C_M}{t}$ for $t \to 0$. Let the parameter choice $\alpha : (0, \infty) \times Y \to (0, \infty)$ be taken in accordance with the discrepancy principle, such that the sequence $\{t_k\}_k$ satisfies $t_k = \theta_k t_{k-1}$, where $0 < \vartheta \leq \theta_k < 1$ for all k. Furthermore, suppose that $\tau > \sup\{|p_t(\lambda)| : t > 0, 0 \leq \lambda \leq \|A\|^2\} \geq p_t(0) = 1$.

Then $\left(\{R_t\}_{t>0}, \gamma\right)$ with $R_t = F_t\left(A^*A\right)A^*$ is an order optimal regularization of $A^{\dagger}$ with respect to X_{μ} for all $(0, \mu_0 - 1]$.

Note that, for the values $\mu \in (\mu_0 - 1, \mu_0]$ we obtain nothing, whereas, the Tikhonov case leads to the order of convergence $O(\varepsilon^{\frac{1}{2}})$ instead of $O(\varepsilon^{\frac{2}{3}})$. This is the reason why the discrepancy principle may produce suboptimal rates of convergence in case of a finite qualification. However, in connection with an infinite qualification, no such problems arise.

Generalized discrepancy principle: Next we are interested in a generalization of the discrepancy principle. To this end we make a reformulation: Let $\{t_k\}$ be a strictly monotonously decreasing sequence with limit zero. Then, t_{k^*} is chosen as follows:

$$\begin{aligned} t_{k^*} &= \sup\left\{t_k : \left\|Ax_{t_k}^{\varepsilon} - y^{\varepsilon}\right\|^2 \leq \tau\varepsilon^2\right\} \\ &= \sup\left\{t_k : \left\|p_{t_k}(AA^*)\, y^{\varepsilon}\right\|^2 \leq \tau\varepsilon^2\right\} \\ &= \sup\left\{t_k : \left\langle y^{\varepsilon}, p_{t_k}^2(AA^*)\, y^{\varepsilon}\right\rangle \leq \tau\varepsilon^2\right\}. \end{aligned} \tag{4.58}$$

In doing so we have used the function $s_t = p_t^2$.

For a generalization we allow arbitrary functions s_t in the following sense: Let $\tau > 1$ be chosen (fixed) and $\{t_k\}$ as before. Determine k^* such that

$$t_{k^*} = \sup\left\{t_k : \langle y^{\varepsilon}, s_{t_k}(AA^*)\, y^{\varepsilon}\rangle \leq \tau\varepsilon^2\right\}. \tag{4.59}$$

Set

$$\alpha(\varepsilon, y^{\varepsilon}) = t_{k^*}\ . \tag{4.60}$$

We have to look for functions s_t which yield order optimal methods for the whole parameter domain $(0, \mu_0]$. An answer is given by the example:

$$s_t(\lambda) = p_t^{2+\frac{2}{\mu_0}}(\lambda) = p_t^3(\lambda) = \left(\frac{t}{t+\lambda}\right)^3. \tag{4.61}$$

The generalized discrepancy principle with this family of functions s_t together with the Tikhonov regularization is an order optimal method with respect to X_μ for $\mu \in (0, 2]$.

In the previous considerations we have studied a number of *a posteriori parameter choice rules* which all depend in one way or the other on the computed approximation – and on the given data error level ε. A perfect example to illustrate this general reasoning is the discrepancy principle where reconstructions are discarded unless their data fit has the order to the noise level ε.

In practical examples such noise level information is not always available (or reliable). For instance, a given discrete data vector may consist of a finite number of measurements, for each of which we may or may not know the standard deviation and/or a worst-case error bound. Typically, the worst-case bound will be a severe overestimation, while the standard deviation might underestimate the true error. both estimates may therefore lead to a significant loss of accuracy when used in these parameter choice rules. Another uncertainty problem arises if we are going to embed the discrete data into a continuous model by some interpolation or approximation process. Then we have to estimate the L^2-norm of the difference

between the constructed function and the true data function from the discrete noise information, and from a priori assumed smoothness properties of the data.

Often it is necessary to consider alternative (a posteriori) parameter choice rules that avoid knowledge of the noise level, and to determine some realistic regularization parameter on the basis of the actual performance of the regularization method under consideration. Such heuristic parameter choice rules will be called *error free*. A good reference to these strategies is [63]. It must be emphasized, however, that error free parameter choice rules cannot provide a convergent regularization method in the strict sense. Still, there are examples where an error free rule leads to better reconstructions than some sophisticated order-optimal rule, cf., e.g., [113] for some numerical comparisons.

Another heuristic parameter choice rule which can be interpreted via some kind of error estimation is the method of generalized cross-validation introduced by Wahba (cf. [256] for the history of this method and a more detailed exposition). It applies to problems where A is an operator into a finite-dimensional data space, e.g., a generalized moment problem.

Another very popular error-free parameter choice rule has been advocated by Hansen [116]. This method is based on an inspection of the residual norms of the computed approximations, this time by relating them to the norms of the approximations themselves (cf. [113]). In spite of its use in several applications, for example, in satellite to satellite tracking, satellite gravity gradiometry (see, e.g., [72, 79, 79, 94] [231], there still lacks a sound mathematical foundation of the L-curve method.

Tikhonov–Phillips regularization. The Tikhonov–Phillips filter is given by

$$F_t(\lambda) = 1/(\lambda + t), \quad t > 0, \quad \lambda \in \left[0, \|A\|^2\right]. \tag{4.62}$$

Let A be of class $L(X, Y)$. Without loss of generality, suppose that A is injective (otherwise we have to replace X by $\mathcal{N}(A)^{\perp}$). Let Z be a Hilbert space and $B \in L(X, Z)$ be continuously invertible, i.e., there exists a value $\beta > 0$ such that

$$\beta \|x\|_X \leq \|Bx\|_Z \quad \text{for all } x \in X. \tag{4.63}$$

Before we deal with the generalization of the Tikhonov–Phillips regularization we mention some preparatory results:

(1) *Lax–Milgram Lemma* Suppose that $L \in L(X)$. Assume there exists a value $\lambda > 0$ such that

$$\langle Lx, x\rangle \geq \lambda \|x\|^2 \tag{4.64}$$

holds for all $x \in X$. Then L is continuously invertible and

$$\left\|L^{-1}\right\| \leq 1/\lambda. \tag{4.65}$$

(2) The stabilized normal equation

$$(A^*A + tB^*B)\, x = A^*y \tag{4.66}$$

with $y \in Y$ possesses a unique solution for all $t > 0$ which continuously depends on y.

Our aim is to show that the solution of (4.66) can be equivalently obtained by minimizing the argument of the *Tikhonov–Phillips functional* given by

$$J_{t,y}(x) = \|Ax - y\|^2 + t\|Bx\|^2. \tag{4.67}$$

Note that the second term on the right-hand side of (4.67) is called the *penalty term of the Tikhonov–Phillips functional*:

Let A and B be given as indicated above. Moreover, assume that $y \in Y$ and $t > 0$. Then the following statements are equivalent:

(1) $(A^*A + tB^*B)\, x_t = A^*y$,
(2) x_t *minimizes the functional*

$$J_{t,y}(x) = \|Ax - y\|^2 + t\|Bx\|^2, \tag{4.68}$$

i.e., $x_t = \arg\min\{J_{t,y}(x) : x \in X\}$.

Our purpose is to comment on this result in more detail: For $y \in Y$ and the family of generalized Tikhonov–Phillips regularizations R_t given by

$$\begin{aligned} x_t = R_t y &= (A^*A + tB^*B)^{-1} A^*y \\ &= \arg\min\{J_{t,y}(x) : x \in X\}, \end{aligned} \tag{4.69}$$

the penalty term satisfies the estimate

$$\|Bx_t\| \le \frac{1}{\sqrt{t}}\|y\|. \tag{4.70}$$

In fact, the property (4.70) explains the role of the penalty term. If t is large, then $\|Bx_t\|$ is small compared to $\|Ax_t - y\|_Y$. If $t \ll 1$, then $\|Bx_t\|$ becomes large in comparison with the residual term $\|Ax_t - y\|$. All in all, the choice of the operator influences the character of x_t. Some features can be strengthened, where others can be weakened.

Let A and B be given as before. Furthermore, suppose that $y \in \mathcal{D}(A^\dagger)$ and $r > 0$. Set

$$\delta_r = \inf\left\{\left\|\frac{1}{\beta^2}B^*BA^\dagger y - A^*y'\right\| : y' \in Y, \|y'\| \le r\right\}. \tag{4.71}$$

Then the following statements hold true for $x_t = (A^*A + tB^*B)^{-1}A^*y$:

(1) $\left\|x_t - A^\dagger y\right\|^2 \le \delta_r^2 + t\beta^2 r^2, \quad r, t > 0,$
(2) $\lim_{t\to 0} x_t = A^\dagger y$.

Some additional effort is needed to formulate regularizations under a priori parameter choice.

Let A, B given as before. If we choose $\gamma : (0,\infty) \to (0,\infty)$ such that

$$\lim_{\varepsilon\to 0}\gamma(\varepsilon) = 0 \quad \textit{and} \quad \lim_{\varepsilon\to 0}\frac{\varepsilon}{\sqrt{\gamma(\varepsilon)}} = 0, \tag{4.72}$$

then $\left(\{R_t\}_{t>0}, \gamma\right)$ with $R_t = (A^*A + tB^*B)^{-1}A^*$ is a regularization of $A^\dagger$.

In order to derive further convergence results we impose further assumptions on B: If $\{(\sigma_n; x_n, y_n)\}$ is the singular system of A and if, for $B \in L(X, Z)$, we set

$$B^*Bx = \sum_{k=1}^{\infty} \beta_k^2 \langle x, x_k \rangle x_k, \quad \beta_k \in \left[\beta, \|B\|^2\right] \tag{4.73}$$

(this is a particular specialization), then we get

$$\begin{aligned} R_t y &= (A^*A + tB^*B)^{-1} A^* y \\ &= \sum_{k=1}^{\infty} \frac{\sigma_k}{\sigma_k^2 + t\beta_k^2} \langle y, y_k \rangle x_k. \end{aligned} \tag{4.74}$$

Note that the classical Tikhonov–Phillips regularization uses $B = I$, i.e., $\beta_k = 1$ for all k. By β_k we control which singular value gets damped and how strongly it is regularized.

In general, R_t cannot be written as a filter, i.e., in the form $F_t(A^*A)A^*$. However, the following helpful estimates can be easily verified (see [227]):

$$\left\|F_{\|B\|^2 t}(A^*A)A^*y\right\| \le \left\|R_t y\right\| \le \left\|F_{\beta^2 t}(A^*A)A^*y\right\| \quad \text{for all } y \in Y, \tag{4.75}$$

and

$$\left\|p_{\beta^2 t}(A^*A)x\right\| \le \left\|(I - R_t A)x\right\| \le \left\|p_{\|B\|^2 t}(A^*A)x\right\| \quad \text{for all } x \in X, \tag{4.76}$$

where $p_t(\lambda) = 1 - \lambda F_t(\lambda) = \frac{t}{\lambda + t}$ and $F_t(\lambda) = \frac{1}{\lambda + t}$ (as for the classical Tikhonov filter).

The family $\{R_t\}_{t>0}$ possesses the same asymptotic behavior for $t \to 0$ as the classical Tikhonov filter, the behavior for $\varepsilon \to 0$ is independent of B.

Suppose that $A \in K(X, Y)$ and $B \in L(X, Z)$. Assume that the representation (4.73) *holds true. Let R_t be given in the form $R_t = (A^*A + tB^*B)^{-1}A^*$, $t > 0$.*

(a) *If the a priori parameter γ is chosen such that*

$$C_\gamma \left(\frac{\varepsilon}{\rho}\right)^{\frac{2}{\mu+1}} \le \gamma(\varepsilon) \le C_\Gamma \left(\frac{\varepsilon}{\rho}\right)^{\frac{2}{\mu+1}} \quad \text{for } \varepsilon \to 0, \tag{4.77}$$

(*as in* (4.37)), *where C_γ and C_Γ are positive constants, then the method $(\{R_t\}_{t>0}, \gamma)$ is an order optimal regularization of $A^\dagger$ with respect to X_μ, $\mu \in (0, 2]$.*

(b) *If we choose γ according to the discrepancy principle, then $(\{R_t\}_{t>0}, \gamma)$ is an order optimal regularization of $A^\dagger$ with respect to X_μ, $\mu \in (0, 1]$. The order of decay of the error $O(\varepsilon^{\frac{1}{2}})$ is maximal.*

(c) *If we choose γ in accordance with the generalized discrepancy principle with $t_k^* = \sup\{t_k \mid \eta(t_k) \le \tau\varepsilon^2\}$, where*

$$\eta(t) = \|Ax_t^\varepsilon - y^\varepsilon\|^2 - \langle A^*(Ax_t^\varepsilon - y^\varepsilon), (A^*A + tB^*B)^{-1}A^*(Ax_t^\varepsilon - y^\varepsilon)\rangle, \tag{4.78}$$

with $y^\varepsilon \in Y$ and $x_t^\varepsilon = R_t y^\varepsilon$, then $(\{R_t\}_{t>0}, \gamma)$ is an order optimal regularization of $A^\dagger$ with respect to X_μ, $\mu \in (0, 2]$.

Once again, it is possible to prove that $O(\varepsilon^{\frac{2}{3}})$ is the maximal order of decay for the error. In other words, the qualification does not increase based on this generalization. However, it should be remarked that the reconstructions using $B \neq I$ may lead to strongly different results. In particular, it is possible to adapt β_k to the spectrum of the noise, if this information is known or it can be estimated by other methods.

4.1. Generalized inverses in reproducing kernel Hilbert spaces

Within the L^2-context the range of a compact linear operator K with infinite-dimensional range is always non-closed. In [195–197] we are confronted with the question: Can one endow $\mathcal{R}(K)$ with a new inner product that would make $\mathcal{R}(K)$ a Hilbert space and that would have additional useful properties?

Reproducing kernel Hilbert space (RKHS) framework. A Hilbert space H of complex-valued functions on a (bounded) set $\mathcal{G}$ (e.g., a regular region) is called a *reproducing kernel Hilbert space* (*RKHS*) if all the evaluation functional $H \ni x \mapsto x(t) \in \mathbb{C}$ are continuous (bounded) for each fixed $t \in \mathcal{G}$, i.e., there exists a positive constant C_t for each $t \in \mathcal{G}$ such that $|x(t)| \leq C_t \, \|x\|_H$ for all $x \in H$. By the *Riesz Representation Theorem*, for each $t \in \mathcal{G}$, there exists a unique element Q_t such that $x(t) = \langle x, Q_t \rangle_H$ for all $x \in H$. The reproducing kernel $Q(\cdot, \cdot) : \mathcal{G} \times \mathcal{G} \mapsto \mathbb{C}$ of a RKHS H is defined by $Q(s,t) = \langle Q_s, Q_t \rangle_H, \;\; s, t \in \mathcal{G}$.

We list some basic properties of RKHS's that are particularly relevant in approximation and estimation theory:

- $Q(s,t) = \overline{Q(t,s)}$ for all $t, s \in \mathcal{G}$.
- $Q(s,s) \geq 0$ for all $s \in \mathcal{G}$.
- $|Q(s,t)| \leq \sqrt{Q(s,s)}\sqrt{Q(t,t)}$ for all $s, t \in \mathcal{G}$.
- The reproducing kernel $Q(s,t)$ on $\mathcal{G} \times \mathcal{G}$ is a non-negative definite Hermitian kernel. Conversely by the *Aronszajn–Moore Theorem*, every non-negative definite Hermitian function $Q(\cdot, \cdot)$ on $\mathcal{G} \times \mathcal{G}$ determines a unique Hilbert space H_Q for which $Q(\cdot, \cdot)$ is a reproducing kernel ([15]) (note that a complex-valued kernel F on $\mathcal{G} \times \mathcal{G}$ is said to be positive definite if, for any n points $t_1, \ldots, t_n \in \mathcal{G}$, the matrix $A = (F(t_i, t_j))_{1 \leq i,j \leq n}$ is non-negative definite, i.e.,

$$u^H A u = \sum_{i,j=1}^{n} \overline{u_i} \; F(t_i, t_j) \; u_j \geq 0 \tag{4.79}$$

  for all $u = (u_1, \ldots, u_n) \in \mathbb{C}^n$).
- A closed subspace $\tilde{H}$ of a RKHS H is also a RKHS. Moreover, the orthogonal projector P of H onto $\tilde{H}$ and the reproducing kernel $\tilde{Q}(s,t)$ of the RKHS $\tilde{H}$ are related by $Pf(s) = \langle f, \tilde{Q}_s \rangle, s \in \mathcal{G}$ for all $f \in H$ where $\tilde{Q}_k = PQ$.
- In a RKHS, the element representing a given bounded linear functional L can be expressed by means of the reproducing kernel: $L(f) = \langle f, h \rangle_H$, where $h = L(Q)$.

  Similarly, for a bounded linear operator L on H to H, we have that $Lf(t) = \langle Lf, h \rangle = \langle f, L^* h \rangle$.

- If $\mathcal{G}$ is a bounded domain or if $\mathcal{G}$ is an unbounded domain but

$$\int_{\mathcal{G}\times\mathcal{G}} |Q(t,s)|^2 dt\ ds < \infty, \tag{4.80}$$

and $Q(s,t)$ is continuous on $\mathcal{G}\times\mathcal{G}$, then H_Q is a space of continuous functions.
- Every finite-dimensional function space is a RKHS H with reproducing kernel

$$Q(s,t) = \sum_{i=1}^{n} u_i(s)\ \overline{u_i(t)}, \tag{4.81}$$

where $\{u_i\}_{i=1}^n$ is an orthonormal basis for H (notice that the sum in the above definition of the kernel Q is invariant under the choice of an orthonormal basis).
- If the integral relation

$$\int_{\mathcal{G}\times\mathcal{G}} |Q(s,t)|^2\ ds\ dt < \infty, \tag{4.82}$$

holds true, then $Q(\cdot,\cdot)$ has a countable sequence of eigenvalues and eigenfunctions (*Theorem of Mercer*).
- Let $\{\varphi_n\}_{n\in\mathbb{N}}$ be a sequence of complex functions defined on $\mathcal{G}$ such that, for every $t \in \mathcal{G}$,

$$\sum_{n=1}^{\infty} |\varphi_n(t)|^2 < \infty. \tag{4.83}$$

For every sequence $\{c_n\}_{n\in\mathbb{N}}$ with $\sum_{n=1}^{\infty} |c_n|^2 < \infty$, the series $\sum_{n=1}^{\infty} c_n\varphi_n(t)$ is then convergent in $\mathbb{C}$ for every $t \in \mathcal{G}$. The functions which are the sums of such series form a linear subspace H, on which we are able to define the structure of a separable Hilbert space by taking as scalar product, for

$$f = \sum_{n=1}^{\infty} c_n\varphi_n, \quad g = \sum_{n=1}^{\infty} d_n\varphi_n, \tag{4.84}$$

the number

$$\langle f, g\rangle_H = \sum_{n=1}^{\infty} c_n\overline{d_n}. \tag{4.85}$$

This space has a reproducing kernel, namely

$$Q(x,y) = \sum_{n=1}^{\infty} \varphi_n(t)\overline{\varphi_n(s)},\ t,s \in \mathcal{G}\times\mathcal{G}. \tag{4.86}$$

- Let H be a separable RKHS, then its reproducing kernel $Q(\cdot,\cdot)$ has the expansion

$$Q(s,t) = \sum_{n=1}^{\infty} \varphi_n(t)\ \overline{\varphi_n(s)}, \tag{4.87}$$

where $\{\varphi_n\}_{n=1}^{\infty}$ is an orthonormal basis for H (we remark that for a general separable Hilbert space H, $\sum_{n=1}^{\infty} \varphi_n(t)\ \overline{\varphi_n(s)}$ is not a reproducing kernel

(note that $L^2(\mathcal{G})$ is not an $RKHS$) and also that ϕ_n's do not generally correspond to sampling expansions. If they do, i.e., if $\varphi_n(t) = Q(t_n, t)$ for some sequence $\{t_n\}$, then we have that $f(t) = \sum_{n=1}^{\infty} f(t_n)\ \varphi_n(t)$, this constitutes a *sampling theorem*.)

- If the reproducing kernel $Q(s,t)$ of a RKHS H is continuous on $\mathcal{G}\times\mathcal{G}$, then H is a space of continuous functions (being uniformly continuous on a bounded $\mathcal{G}$). This follows from

$$|x(t) - x(s)| = |\langle x, Q_t - Q_s\rangle_H| \le \|x\|_H\ \|Q_t - Q_s\|_H \tag{4.88}$$

and

$$\|Q_t - Q_s\|^2 = Q(t,t) - 2Q(t,s) + Q(s,s) \tag{4.89}$$

for all $s, t \in \mathcal{G}$.

- Strong convergence in a RKHS H implies pointwise convergence and uniform convergence on compact sets, because of the fact

$$|x(t) - x_n(t)| = |\langle x - x_n, Q_t\rangle_H| \le \sqrt{Q(t,t)}\ \|x - x_n\|_H. \tag{4.90}$$

- Let H_Q denote the RKHS with reproducing kernel Q, and denote the inner product and norm in H_Q by $\langle\cdot,\cdot\rangle_{H_Q}$ and $\|\cdot\|_{H_Q}$, respectively. Note that $Q(s,s')(= Q_s(s'))$ is a non-negative definite Hermitian kernel on $\mathcal{G}\times\mathcal{G}$, and that $\{Q_s, s\in\mathcal{G}\}$ spans H_Q since $\langle Q_s, x\rangle_{H_Q} = 0$, $s\in\mathcal{G}$, implies $x(s) = 0$. For more properties of reproducing kernel spaces the reader is referred to, e.g., [15, 37, 131] and the references therein.
- For every positive definite kernel $Q(\cdot,\cdot)$ on $\mathcal{G}\times\mathcal{G}$, there exist a zero mean Gaussian process with $Q(\cdot,\cdot)$ as its covariance, giving rise to the relation between Bayes estimates, Gaussian processes, and optimization processes in RHKS (for more details the reader is referred to the geodetic literature, see, e.g., [100, 159, 160, 167], and the monographs [148, 256]).

Interest in reproducing kernel Hilbert spaces have increased in recent years, as the computer capacity has made solutions of ever larger and more complex problems practicable. Indeed, new reproducing kernel representations and new applications (in particular in physical geodesy and geophysics) are being contributed at a rapid rate. For example, a certain RHKS in terms of outer harmonics allows the adequate determination of the Earth's gravitational potential (see, e.g., [71, 237] for early spline approaches) in consistency with gravitational observables of heterogeneous type (that are interpretable as (bounded) linear functionals on the RKHS under consideration).

Hilbert–Schmidt operator theory. An RKHS H_Q with RK Q determines a self-adjoint *Hilbert–Schmidt operator* (also denoted by Q) on $L^2(\mathcal{G})$ to $L^2(\mathcal{G})$ by letting

$$(Qx)(s) = \int_{\mathcal{G}} Q(s,s')\ x(s')\ ds', \quad x\in L^2(\mathcal{G}). \tag{4.91}$$

Since Q is assumed to be continuous, then by the Theorem of Mercer (see, e.g., [139]), the operator Q has an $L^2(\mathcal{G})$-complete orthonormal system of eigenfunctions $\{\phi_i\}_{i=1}^{\infty}$ and corresponding eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ with $\lambda_i \ge 0$ and $\sum_{i=1}^{\infty}\lambda_i < \infty$.

Thus Q is a *trace-class* operator (see, e.g., [46]) so that $Q(\cdot,\cdot)$ has the uniformly convergent Fourier expansions

$$Q(s,s') = \sum_{i=1}^{\infty} \lambda_i \ \phi_i(s) \ \phi_i(s') \tag{4.92}$$

and

$$Qx = \sum_{i=1}^{\infty} \lambda_i \langle x, \phi_i \rangle_{L^2(\mathcal{G})} \ \phi_i, \tag{4.93}$$

where $\langle \cdot,\cdot \rangle_{L^2(\mathcal{G})}$ is the inner product in $L^2(\mathcal{G})$. It is well known (see, for example, [249]) that the space H_Q consists of all $x \in L^2(\mathcal{G})$ satisfying the condition

$$\sum_{i=1}^{\infty} {\lambda_i}^{-1} \left(\langle x, \phi_i \rangle_{L^2(\mathcal{G})} \right)^2 < \infty, \tag{4.94}$$

(note that the notational convention $0/0 = 0$ is being adopted) with inner product $\langle \cdot,\cdot \rangle_{H_Q}$ given by

$$\langle x_1, x_2 \rangle_{H_Q} = \sum_{i=1}^{\infty} \lambda_i^{-1} \langle x_1, \phi_i \rangle_{L^2(\mathcal{G})} \langle x_2, \phi_i \rangle_{L^2(\mathcal{G})}. \tag{4.95}$$

The operator Q has a well-defined symmetric square root $Q^{1/2}$ which is a Hilbert–Schmidt operator (see, e.g., [195–197]):

$$Q^{1/2}x = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \langle x, \phi_i \rangle_{L^2(\mathcal{G})} \ \phi_i. \tag{4.96}$$

Thus, since $\mathcal{N}(Q) = \mathcal{N}(Q^{1/2})$, we have

$$H_Q = Q^{1/2}(L^2(\mathcal{G})) = Q^{1/2} \left(L^2(\mathcal{G}) \ominus \mathcal{N}(Q) \right). \tag{4.97}$$

$(Q^{1/2})^\dagger$ has the representation

$$(Q^{1/2})^\dagger x = \sum_{i=1}^{\infty} (\sqrt{\lambda_i})^\dagger \langle x, \phi_i \rangle_{L^2(\mathcal{G})} \ \phi_i \tag{4.98}$$

on $H_Q \oplus H_Q^\perp$ ($\perp$ in $L^2(\mathcal{G})$), where, for Θ a real number, $\Theta^\dagger = \Theta^{-1}$ for $\Theta \neq 0; \Theta^\dagger = 0$ for $\Theta = 0$. Similarly, $Q^\dagger$ has the representation

$$Q^\dagger x = \sum_{i=1}^{\infty} \lambda_i^\dagger \langle x, \phi_i \rangle_{L^2(\mathcal{G})} \ \phi_i. \tag{4.99}$$

For any operator Q on $L^2(\mathcal{G})$, induced by an RK $Q(s,s')$ as defined in (4.91), we shall adopt the notational conventions (cf. [195–197])

$$Q^{-1} := Q^\dagger, \tag{4.100}$$

$$Q^{-1/2} := (Q^{1/2})^\dagger. \tag{4.101}$$

This leads to the relations

$$\|x\|_{H_Q} = \inf\left\{\|p\|_{L^2(\mathcal{G})} : p \in L^2(\mathcal{G}),\ x = Q^{1/2}p\right\}, \quad x \in H_Q. \tag{4.102}$$

and

$$\langle x_1, x_2\rangle_{H_Q} = \langle Q^{1/2}x_1, Q^{1/2}x_2\rangle_{L^2(\mathcal{G})}, \quad x_1, x_2 \in H_Q, \tag{4.103}$$

and, if $x_1 \in H_Q$ and $x_2 \in H_Q$ with $x_2 = Q\rho$ for some $\rho \in L^2(\mathcal{G})$, then

$$\langle x_1, x_2\rangle_{H_Q} = \langle x_1, \rho\rangle_{L^2(\mathcal{G})}\ . \tag{4.104}$$

Relationship between generalized inverses and L^2-topology. We are now ready to explore properties of the generalized inverse of a linear operator between two RK spaces (cf. [195–197]). To this end we let $X = L^2(\mathcal{G})$ and $Y = L^2(\mathcal{H})$ denote the Hilbert spaces of square-integrable real-valued functions on the closed, bounded domains $\mathcal{G}$ and $\mathcal{H}$, respectively. Let A be a linear operator from X into Y. Let $\subset$ denote point set inclusion only, and suppose that A has the following properties:

$$H_Q \subset \mathcal{D}(A) \subset X, \tag{4.105}$$

where H_Q is an RKHS with continuous RK on $\mathcal{G} \times \mathcal{G}$;

$$A(H_Q) = H_{\overline{R}} \subset H_R \subset Y, \tag{4.106}$$

where $H_{\overline{R}}$ and H_R are RKHS's with continuous RK's on $\mathcal{H} \times \mathcal{H}$; so that

$$\mathcal{N}(A) \text{ in } H_Q \text{ is closed in } H_Q. \tag{4.107}$$

We emphasize, in particular, that the space $H_{\overline{R}}$ is not necessarily closed in the topology of H_R.

Let $A^\dagger_{(X,Y)}$ denote the generalized inverse of A, when A is considered as a mapping from X into Y, and let $A^\dagger_{(Q,R)}$ denote the generalized inverse of A when A is considered as a mapping from H_Q into H_R. Now, the topologies in (X, Y) are not the same as the topologies in (H_Q, H_R). Thus, the generalized inverses $A^\dagger_{(X,Y)}$ and $A^\dagger_{(Q,R)}$ show distinct continuity properties, in general. We shall develop the relation between $A^\dagger_{(Q,R)}$ and certain (X, Y) and (Y, Y) generalized inverses. In the sequel, the operators $R : Y \to Y$ and $R^{1/2} : Y \to Y$ are defined from the RK of H_R analogous to Q and $Q^{1/2}$ (see (4.93) and (4.96)). We continue the notational convention of (4.100), i.e., $R^{-1} = R^\dagger = R^\dagger_{(Y,Y)}$ and $R^{-1/2} = (R^{1/2})^\dagger_{(Y,Y)}$.

From [195–197] we are able to deduce the following result: *Under assumptions* (4.105)–(4.107)*, let* $y \in \mathcal{D}(A^\dagger_{(Q,R)})$*, i.e.,* $y \in H_{\overline{R}} \oplus H^\perp_{\overline{R}}$ ($\perp$ *in* H_R)*. Then*

$$y \in \mathcal{D}(Q^{1/2}(R^{-1/2}AQ^{1/2})^\dagger_{(X,Y)}R^{-1/2}) \tag{4.108}$$

and

$$A^\dagger_{(Q,R)}y = Q^{1/2}(R^{-1/2}AQ^{1/2})^\dagger_{(X,Y)}R^{-1/2}\ y. \tag{4.109}$$

Moreover, it follows that $A^\dagger_{(Q,R)}$ is bounded, provided that $A(H_Q) = H_R$.

It should be noted that an operator A may satisfy the assumption $A(H_Q) = H_R$ while failing to have a closed range in the space Y. This is, for example, the case if A is a Hilbert–Schmidt linear integral operator (with non-degenerate kernel) on X. It is this observation which makes RKHS useful in the context of regularization and approximation of ill-posed linear operator equations.

Explicit representation of minimal norm solutions. We assume that H_Q is chosen so that the linear functionals $\{\mathcal{E}_t : t \in \mathcal{H}\}$ defined by

$$\mathcal{E}_t x = (Ax)(t) \tag{4.110}$$

are continuous in H_Q. Then, by the Riesz representation theorem, there exists $\{\eta_t, t \in \mathcal{H}\} \in H_Q$ such that

$$(Ax)(t) = \langle \eta_t, x \rangle_{H_Q}, \quad t \in \mathcal{H}, \quad x \in H_Q, \tag{4.111}$$

where η_t is explicitly given by

$$\eta_t(s) = \langle y_r, Q_s \rangle_{H_Q} = (AQ_s)(t) \tag{4.112}$$

($\eta_t(s)$ is readily obtained in a more explicit form from (4.112) if A is a differential or integral operator).

Let $R(t, t')$ be the non-negative definite kernel on $\mathcal{H} \times \mathcal{H}$ given by

$$R(t, t') = \langle \eta_t, \eta_i \rangle_{H_Q}, \quad t, t' \in \mathcal{H}. \tag{4.113}$$

Assume that H_R is the RKHS with RK R given by (4.113). Let R_t be the element of H_R defined by $R_t(t') = R(t, t')$, and let $\langle \cdot, \cdot \rangle_{H_R}$ be the inner product in H_R. Suppose that V is the closure of the span of $\{\eta_t, t \in \mathcal{H}\}$ in H_Q. Now, $\{R_t, t \in \mathcal{H}\}$ spans H_R, and by the properties of RKHS, we have

$$\langle \eta_t, \eta_{t'} \rangle_{H_Q} = R(t, t') = \langle R_t, R_{t'} \rangle_{H_R}. \tag{4.114}$$

Thus there is an isometric isomorphism between the subspace V and H_R, generated by the correspondence

$$\eta_t \in V \sim R_t \in H_R. \tag{4.115}$$

Then, $x \in V \sim y \in H_R$ if and only if $\langle \eta_t, x \rangle_{H_Q} = y(t) = \langle R_t, g \rangle_{H_R}$, $t \in \mathcal{H}$, i.e., if and only if $y(t) = (Ax)(t), t \in \mathcal{H}$. Thus, $A(H_Q) = A(V) = H_R$. The null space of A in H_Q is $\{x : x \in H_Q, \|Ax\|_{H_R} = 0\}$. Since

$$\langle \eta_t, x \rangle_{H_Q} = 0, \quad t \in \mathcal{H} \quad \text{and} \quad x \in H_Q \Rightarrow x \in V^{\perp}, \tag{4.116}$$

and $x \in V$ implies $\|x\|_Q = \|Ax\|_V$, it follows that the null space of A in H_Q is $V^{\perp}$ ($\perp$ in H_Q). Hence, (4.110) entails that the null space of $A : H_Q \to H_R$ in H_Q is always closed, irrespective of the topological properties of $A : X \to Y$.

We list the following table of corresponding sets and elements, under the correspondence "$\sim$" of (4.115), where the entries on the left are in H_Q:

$$V \sim H_R, \tag{4.117}$$

$$x \sim y, \tag{4.118}$$

$$\eta_t \sim R_t, \tag{4.119}$$

$$P_V Q_s \sim \eta_s^*. \tag{4.120}$$

Here, P_V is the projector from H_Q onto the (closed) subspace V, $y(t) = \langle \eta_t, x \rangle_{H_Q}$, $t \in \mathcal{H}$, and $\eta_s^* = AQ_s = A(P_V Q_s)$, i.e.,

$$\eta_s^*(t) = \langle \eta_i, P_V Q_s \rangle_{H_Q} = \eta_t(s). \tag{4.121}$$

This leads to the following result (the proof is given in [196, 197]):

Let A and H_Q satisfy (4.110)*, and let R be given by* (4.114)*, where η_i is defined by* (4.111)*. Let $\eta_s^* = AQ_s$. Then, for $y \in H_R$,*

$$(A^\dagger_{(Q,R)} y)(x) = \langle \eta_s^*, y \rangle_{H_R}, \quad s \in \mathcal{G}. \tag{4.122}$$

We also obtain another operator representation of $A^\dagger_{(Q,R)}$:

Under the assumptions

(i) *$\mathcal{D}(A^*)$ is dense in Y, where A^* is the adjoint of A considered as an operator from X to Y,*
(ii) *A and H_Q satisfy* (4.111)*,*
(iii) *H_Q and $H_R = A(H_Q)$ possess continuous reproducing kernels,*

we have, for $y \in H_R$,

$$(A^\dagger_{(Q,R)} y)(s) = (QA^*(AQA^*)^\dagger_{(Y,Y)} y)(s), \quad s \in \mathcal{G}. \tag{4.123}$$

Poorly conditioned operator equations. For $A : X \to Y$, the *pseudocondition number* of A (relative to the norms of X and Y) is given by

$$\gamma(A; X, Y) = \sup_{\substack{x \neq 0 \\ x \in \mathcal{D}(A)}} \frac{\|Ax\|_Y}{\|x\|_X} \sup_{\substack{y \neq 0 \\ y \in \mathcal{D}(A^\dagger)}} \frac{\|A^\dagger y\|_X}{\|y\|_Y}. \tag{4.124}$$

The equation $Ax = y$ is said to be *poorly conditioned in the spaces X, Y* if the number $\gamma(A; X, Y)$ is much greater than 1 (note that $1 \leq \gamma(A; X, Y)$; for ill-posed problems, γ is not finite).

Suppose H_Q is an RKHS with $H_Q \subset \mathcal{D}(A)$, and A and H_Q satisfy (4.110) with $A(H_Q) = H_R$, R given by (4.114). Then $\gamma(A; H_Q, H_R) = 1$. To see this, write $x \in H_Q$ in the form $x = x_1 + x_2$, where $x_2 \in V^\perp$. Then $Ax = Ax_1 = y_1$ and $\|y_1\|_{H_R} = \|x_1\|_{H_Q}$. Thus

$$\gamma(A; H_Q, H_R) = \sup_{x \neq 0} \frac{\|y_1\|_{H_R}}{\|x\|_{H_Q}} \sup_{y_1 \neq 0} \frac{\|x_1\|_{H_Q}}{\|y_1\|_{H_R}} = 1. \tag{4.125}$$

On the other hand, the number $\gamma(A; X, Y)$ may be large. Thus, the casting of the operator equation $Ax = y$ in the reproducing kernel spaces H_Q, H_R always leads to a well-conditioned (indeed, optimally-conditioned) problem.

Regularization of pseudosolutions in reproducing kernel spaces. We study properties of *regularized pseudosolutions* (in RKHS) x_α of the operator equation $Ax = y$, where y is not necessarily in the range of the operator A. By a *regularized pseudosolution* we mean a solution to the variational problem: Find x_α in H_Q to minimize

$$J_y(x) = \|y - Ax\|^2_{H_P} + \alpha\|x\|^2_{H_Q}, \qquad \alpha > 0, \tag{4.126}$$

where H_Q is an RKHS in the domain of A, $\|\cdot\|_{H_P}$ denotes the norm in an RKHS H_P with RK P, $H_P \subset Y$, $J_y \subset Y$, $\phi_y(x)$ is assigned the value $+\infty$ if $y - Ax \notin H_P$, and $\alpha > 0$. We suppose A and H_Q satisfy (4.110), hence, $A(H_Q) = H_R$ possesses an RK. As before, A may be unbounded, invertible, or compact considered as an operator from $X(= L^2(\mathcal{G}))$ to $Y(= L^2(\mathcal{H}))$. It is assumed that y possesses a (not necessarily unique) representation $y = y_0 + \xi$, for some $y_0 \in A(H_Q)$ and $\xi \in H_P$, where ξ may be thought of as a "disturbance".

For $\alpha > 0$, let $H_{\alpha P}$ be the RKHS with RK $\alpha P(t, t')$, where $P(t, t')$ is the RK on $\mathcal{H} \times \mathcal{H}$ associated with H_P. We have $H_P = H_{\alpha P}$ and

$$\|\cdot\|^2_{H_P} = \alpha\|\cdot\|^2_{H_{\alpha P}}. \tag{4.127}$$

Let $R(\alpha) = R + \alpha P$, and let $H_{R(\alpha)}$ be the RKHS with RK $R(\alpha) = R(\alpha, t, t')$. According to [15], $H_{R(\alpha)}$ is the Hilbert space of functions of the form

$$y = y_0 + \xi, \tag{4.128}$$

where $y_0 \in H_R$ and $\xi \in H_P$. Following [15], we note that this decomposition is not unique unless H_R and H_P have no element in common except the zero element. The norm in $H_{R(\alpha)}$ is given by

$$\|y\|^2_{R(\alpha)} = \min\{\|y_0\|^2_{H_R} + \|\xi\|^2_{H_{\alpha R}} : y_0 \in H_R, \xi \in H_P, y_0 + \xi = y\}, \tag{4.129}$$

where, however, the y_0 and ξ attaining the minimum in (4.129) are easily shown to be unique by the strict convexity of the norm.

Consider the problem of finding $x_\alpha \in H_Q$ to minimize $J_y(x)$ in (4.126) for $y \in H_{R(\alpha)}$. Then $y - Ax_\alpha$ must be in H_P and it is obvious that $x_\alpha \in V$, the orthogonal complement of the null space of A in H_Q. For any $x \in V$, $\|x\|_{H_Q} = \|Ax\|_{H_R}$ by the isometric isomorphism between V and H_R, and (4.126) may be written in the equivalent form: Find $x_\alpha \in V$ to minimize

$$\alpha\|Ax\|^2_{H_R} + \|y - Ax\|^2_{H_P}. \tag{4.130}$$

Comparing (4.129) and (4.130) with the aid of (4.127), we see that y_0 and ξ attaining the minimum on the right-hand side of (4.129) are related to the solution x_α of the minimization problem (4.130), by

$$y_0 = Ax_\alpha \quad \text{and} \quad \xi = y - Ax_\alpha. \tag{4.131}$$

A representation of the solution x_α is given (see [195, 196]) as follows:

Suppose $\mathcal{D}(A^*)$ *is dense in* $Y, H_Q \subset \mathcal{D}(A)$ *and* A *and* H_Q *satisfy* (4.110). *Suppose* $H_Q, H_R(= A(H_Q))$ *and* $H_P \subset Y$ *all have continuous RK's. Then, for* $y \in H_{R(\alpha)}$, *the unique minimizing element* $x_\alpha \in H_Q$ *of the functional* $J_y(x)$ *is given by*

$$\langle \eta_s^*, y\rangle_{R(\alpha)} = x_\alpha(s) = (QA^*(AQA^* + \alpha P)^\dagger_{(Y,Y)}y)(s) \quad s \in \mathcal{G}, \tag{4.132}$$

where $\eta_s^* = AQ_s$. We call the (linear) mapping which assigns to each $y \in H_{R(\alpha)}$ the unique minimizing element x_α the *regularization operator* of the equation $Ax = y$.

The most useful situations occur, of course, when H_R is strictly contained in $H_{R(\alpha)}$. For example, H_R may be a dense subset of Y in the Y-topology and $H_{R(\alpha)}$ a bigger dense subset. We deal with this situation later. If $H_R^\perp$ (in Y) is not empty, then P may be chosen so that the closure of H_P in the Y-topology equals $H_R^\perp$ in Y. Then $H_P \cap H_R = \{0\}$, $H_{\alpha P}$ and H_R are orthogonal subspaces of $H_{R(\alpha)}$ (see, e.g., [15]), and the decomposition (4.128) is unique. In this case we have the following theorem which shows that the regularization operator is indeed a generalized inverse in an appropriate RKHS: *If* $H_P \cap H_R = \{0\}$, *then the minimizing element* x_α *of* (4.126) *is the solution to the problem: Find* $x \in \mathcal{L}(y)$ *to minimize*

$$\|x\|_{H_Q}, \tag{4.133}$$

where

$$\mathcal{L}(y) = \{x : x \in H_Q, \|y - Ax\|_{H_{R(\alpha)}} = \inf_{z \in H_Q} \|y - Az\|_{H_{R(\alpha)}}\}. \tag{4.134}$$

It should be remarked that, in our approach, we have

$$A(H_Q) = H_R \subset H_{R(\alpha)} \subset Y. \tag{4.135}$$

Replacing $H_{\overline{R}}$ and H_R in (4.106) by H_R and $H_{R(\alpha)}$, respectively, we get from (4.109)

$$A^\dagger_{(Q,R(\alpha))}y = Q^{1/2}[(R + \alpha P)^{-1/2}AQ^{1/2}]^\dagger_{(X,Y)}(R + \alpha P)^{-1/2}y \tag{4.136}$$

for $y \in \mathcal{D}(A^\dagger_{(Q,R(\alpha))})$.

It is helpful to remember that the topology on H_R is not, in general, the restriction of the topology of $H_{R(\alpha)}$, with the notable exception of the case $H_R \cap H_P = \{0\}$. In [129] a concrete example is provided arising in the approximate solution of boundary value problems, where H_R is not a closed subspace of $H_{R(\alpha)}$. If $H_R \cap H_P = \{0\}$, then H_R is a closed subspace of $H_{R(\alpha)}$, and we have

$$A^\dagger_{(Q,R(\alpha))} = QA^*(R + \alpha P)^{-1}. \tag{4.137}$$

Note that in this case, the generalized inverse and the regularization operator coincide. If $H_R = A(H_Q)$ is not closed in $H_{R(\alpha)}$, then the regularization operator and the generalized inverse are different. Also, the right-hand side of (4.136) and

(4.137) are not the same: (4.137) has maximal domain $H_{R(\alpha)}$, while (4.136) has maximal domain $\mathcal{H}_R \oplus H_R^{\perp}$ ($\perp$ in $H_{R(\alpha)}$).

Rates of convergence to the generalized inverse. We note some properties of x_α as $\alpha \to 0$ when $H_R \subset H_P$. If $y \in H_R = A(H_Q)$, then we have $x_\alpha \to A^{\dagger}_{(Q,R)}y$ as $\alpha \to 0$; here we may say something about the rate of convergence if certain additional conditions are satisfied (compare also with [129]). However, y may not be in the domain of $A^{\dagger}_{(Q,R)}$. This situation can occur if, for example, H_R is dense in $H_{R(1)}$. In this case, $\lim_{\alpha\to 0}\|x_\alpha\|_Q = \infty$: *Assume that* $y = Ax_0 + \xi_0$, *where* $x_0 \in V$, $\xi_0 \in H_P$ *and suppose that* $H_R \subset H_P$. Then the following properties hold true:

(i) $B = P^{-1/2}R^{1/2}$ *is a bounded operator on* $Y = L^2(\mathcal{H})$.

(ii) *If* $\xi_0 = 0$ *and* $\|(B^*B)^{-1}R^{-1/2}(Ax_0)\|_{L^2(\mathcal{H})} < \infty$, *then*

$$\|A^{\dagger}_{(Q,R)}y - x_\alpha\|^2_{H_Q} = O(\alpha^2). \tag{4.138}$$

(iii) *If* $\xi_0 = 0$ *and* $\|(B^*B)^{-1/2}R^{-1/2}(Ax_0)\|_{L^2(\mathcal{H})} < \infty$, *then*

$$\|A^{\dagger}_{(Q,R)}y - x_\alpha\|^2_{H_Q} = O(\alpha). \tag{4.139}$$

(iv) *If* $\xi_0 \notin H_R$, *then* $\lim_{\alpha\to 0}\|x_\alpha\|_{H_Q} = \infty$.

Here inverses indicated by "– " are the generalized inverses in the topology of L^2-spaces.

4.2. Projection methods

First, we recall the definition of a projection operator known from functional analysis: Let X be a normed space over the field $\mathbb{K}$ ($\mathbb{K}$ is equal to $\mathbb{R}$ or $\mathbb{C}$). Let $U \subset X$ be a closed subspace. A linear bounded operator $P : X \to X$ is called a *projection operator* on U if $Px \in U$ for all $x \in X$ and $Px = x$ for all $x \in U$. Every non-trivial projection operator satisfies $P^2 = P$ and $\|P\| \geq 1$.

The following two examples provide some important projection operators:

(a) (*Orthogonal Projection.*) Let X be a pre-Hilbert space. Suppose that $U \subset X$ is a complete subspace. Let $Px \in U$ be the *best approximation* to x in U, i.e., Px satisfies the relation

$$\|Px - x\| \leq \|u - x\| \quad \text{for all} \quad u \in U. \tag{4.140}$$

$P : X \to U$ is linear and $Px \in U$ is characterized by $\langle x - Px, u\rangle = 0$ for all $u \in U$, i.e., $x - Px \in U^{\perp}$. Therefore,

$$\begin{aligned}\|x\|^2 &= \|Px + (x - Px)\|^2 \\ &= \|Px\|^2 + \|x - Px\|^2 + 2\underbrace{Re\langle x - Px, Px\rangle}_{=0} \geq \|Px\|^2,\end{aligned}$$

i.e., $\|P\| = 1$.

(b) (*Interpolation Operator.*) Let $X = C^{(0)}([a, b])$ be the space of real-valued continuous functions on $[a, b]$ supplied with the supremum norm $\|\cdot\|_{C^{(0)}[a,b]}$. Then

X is a normed space over $\mathbb{R}$. Let $U = \text{span}\{u_1, \ldots, u_n\}$ be an n-dimensional subspace and $t_1, \ldots, t_n \in [a, b]$ such that the interpolation problem in U is uniquely solvable, i.e., $\det(u_j(t_k)) \neq 0$. We define $Px \in U$ by the interpolant of $x \in C^{(0)}([a, b])$ in U, i.e., $u = Px \in U$ satisfies $u(t_i) = x(t_i)$ for all $i = 1, \ldots, n$. Then $P : X \to U$ is a projection operator.

Examples for U in (b) are spaces of algebraic or trigonometric polynomials. As a drawback of these choices, we note that from the results of Faber (see, for example, [198]) the interpolating polynomials of continuous functions x do not, in general, converge to x as the degree of the polynomials tends to infinity. Nonetheless, trigonometric interpolation at equidistant points converges with optimal order of convergence.

Next we are concerned with a certain class of projection methods in Banach spaces, essentially following the monograph of A. Kirsch [137]: Let X and Y be Banach spaces and $A : X \to Y$ be bounded and one-to-one. Furthermore, let $X_n \subset X$ and $Y_n \subset Y$ be finite-dimensional subspaces of dimension n and $Q_n : Y \to Y_n$ be a projection operator. For $y \in Y$, the projection method for solving the equations $Ax = y$ is to solve the equations

$$Q_n A x_n = Q_n y \quad \text{for} \quad x_n \in X_n. \tag{4.141}$$

Assume that $\{\hat{x}_1, \ldots, \hat{x}_n\}$ and $\{\hat{y}_1, \ldots, \hat{y}_n\}$ are bases of X_n and Y_n, respectively. Then $Q_n y$ and every $Q_n A\hat{x}_j$, $j = 1, \ldots, n$, are representable in the forms

$$Q_n y = \sum_{i=1}^{n} \beta_i \hat{y}_i \quad \text{and} \quad Q_n A\hat{x}_j = \sum_{i=1}^{n} B_{ij}\hat{y}_i, \quad j = 1, \ldots, n, \tag{4.142}$$

with $\beta_i, B_{ij} \in \mathbb{K}$. The linear combination $x_n = \sum_{j=1}^{n} \alpha_j \hat{x}_j$ solves (4.141) if and only if $\alpha = (\alpha_1, \ldots, \alpha_n)^T \in \mathbb{K}^n$ solves the finite system of linear equations

$$\sum_{i=1}^{n} B_{ij}\alpha_j = \beta_i, i = 1, \ldots, n. \tag{4.143}$$

We are led to the following important classes of projection methods for $A : X \to Y$ being a bounded and one-to-one operator.

Let X and Y be pre-Hilbert spaces and $X_n \subset X$ and $Y_n \subset Y$ be finite-dimensional subspaces with $\dim X_n = \dim Y_n = n$. Let $Q_n : Y \to Y_n$ be the *orthogonal projection.* Then the projected equation $Q_n A x_n = Q_n y$ is equivalent to

$$\langle Ax_n, z_n\rangle = \langle y, z_n\rangle \quad \text{for all} \quad z_n \in Y_n. \tag{4.144}$$

We let $X_n = \text{span}\{\hat{x}_1, \ldots, \hat{x}_n\}$ and $Y_n = \text{span}\{\hat{y}_1, \ldots, \hat{y}_n\}$. Looking for a solution of (4.144) in the form of a linear combination $x_n = \sum_{j=1}^{n} \alpha_j \hat{x}_j$ we arrive at

$$\sum_{j=1}^{n} \alpha_j \langle A\hat{x}_j, \hat{y}_i\rangle = \langle y, \hat{y}_i\rangle \quad \text{for} \quad i = 1, \ldots, n, \tag{4.145}$$

or in matrix-vector $A\alpha = \beta$, where $A_{ij} = \langle A\hat{x}_j,\ \hat{y}_i\rangle$ and $\beta_i = \langle \hat{y},\ \hat{y}_i\rangle,\ i = 1, \ldots, n$.

A priori assumption. For the remaining part of this subsection about projection methods, it is helpful to make the following *a priori assumption* (*APA*):

(i) *Let $A : X \to Y$ be a linear, bounded, and injective operator between Banach spaces, $X_n \subset X$ and $Y_n \subset Y$ be finite-dimensional subspaces of dimension n and $Q_n : Y \to Y_n$ be a projection operator. We assume that $\bigcup_{n\in\mathbb{N}} X_n$ is dense in X and the $Q_n A|_{X_n} : X_n \to Y_n$ is one-to-one and, thus, invertible.*

(ii) *Let $x \in X$ be the solution of*

$$Ax = y. \tag{4.146}$$

By $x_n \in X_n$, we understand the unique solutions of the equations

$$Q_n A x_n = Q_n y, \ n \in \mathbb{N}. \tag{4.147}$$

As a consequence of (APA) we are allowed to represent the solutions x_n of (4.147) in the form $x_n = R_n y$, where $R_n : Y \to X_n \subset X$ is defined by

$$R_n := (Q_n A|_{X_n})^{-1} Q_n : Y \to X_n \subset X. \tag{4.148}$$

Suppose that (APA) is valid throughout this section. The *projection method is called convergent* if the approximate solutions $x_n \in X_n$ of (4.147) converge to the exact solution $x \in X$ of (4.146) for every $y \in A(X)$, i.e., if the limit relation

$$R_n A x = (Q_n A|_{X_n})^{-1} Q_n A x \to x, \quad n \to \infty, \tag{4.149}$$

holds true for every $x \in X$.

Obviously, this definition of convergence coincides with the definition of a regularization strategy for the equation $Ax = y$. Therefore, the projection method converges if and only if R_n is a regularization strategy for the equation $Ax = y$.

Convergence can only be expected if we require that $\bigcup_{n\in\mathbb{N}} X_n$ is dense in X and $Q_n y \to y$ for all $y \in A(X)$. For a compact operator A, however, this property is not sufficient for the convergence. In fact we have to assume an additional boundedness condition:

The solution $x_n = R_n y \in X_n$ of (4.147) *converges to x for every $y = Ax$ if and only if there exists $c > 0$ such that*

$$\|R_n A\| \le c \quad \text{for all} \quad n \in \mathbb{N}. \tag{4.150}$$

If (4.150) *is satisfied, the following error estimate can be shown to be valid*

$$\|x_n - x\| \le (1 + c) \min_{z_n \in X_n} \|z_n - x\| \tag{4.151}$$

with the same constant c as in (4.150).

The estimates (4.150) and (4.151) can be verified in straightforward way.

Suppose that $\|R_n A\|$ is bounded. The operator $R_n A$ is a projection operator onto X_n since for $z_n \in X_n$ we have $R_n A z_n = (Q_n A|X_n)^{-1} Q_n A z_n = z_n$. Thus we are able to deduce that

$$x_n - x = (R_n A - I)x = (R_n A - I)(x - z_n) \quad \text{for all} \quad z_n \in X_n. \tag{4.152}$$

It follows that

$$\|x_n - x\| \leq (c+1)\|x - z_n\| \quad \textit{for all} \quad z_n \in X_n \tag{4.153}$$

such that (4.151) *is valid. Convergence* $x_n \to x$ *follows from the fact that* $\bigcup_{n\in\mathbb{N}} X_n$ *is dense in* X.

Thus far, we were concerned with the case, where the right-hand side y is exactly known. Next we consider the case where the right-hand side is known only approximately, i.e., we start from an element $y^\varepsilon \in Y$ with $\|y^\varepsilon - y\| \leq \varepsilon$. To this end we understand the operator R_n from (4.148) as a regularization operator in the usual sense so that we are led to distinguish two kinds of errors for the right-hand side. A straightforward application of the triangle inequality yields with $x_n^\varepsilon := R_n y^\varepsilon$ the inequality

$$\begin{aligned} \|x_n^\varepsilon - x\| &\leq \|x_n^\varepsilon - R_n y\| + \|R_n y - x\| \\ &\leq \|R_n\| \, \|y^\varepsilon - y\| + \|R_n A x - x\|. \end{aligned} \tag{4.154}$$

As usual, we are confronted with the dilemma of IP: The error ε of the right-hand side is multiplied by the norm of R_n. The second term describes the discretization error against the exact data.

In practice one solves the discrete system (4.143) where the vector β is replaced by a perturbed vector $\beta^\varepsilon \in \mathbb{K}^n$ with

$$|\beta^\varepsilon - \beta|^2 = \sum_{j=1}^{n} |\beta_j^\varepsilon - \beta_j|^2 \leq \varepsilon^2. \tag{4.155}$$

We will call this fact the *discrete perturbation* of the right-hand side. Instead of (4.143) one solves $B\alpha^\varepsilon = \beta$ and defines $x_n^\varepsilon \in X_n$ by $x_n^\varepsilon = \sum_{j=1}^n \alpha_j^\varepsilon \hat{x}_j$. Note that the choice of the basis functions $\hat{x}_j \in X_n$ and $\hat{y}_j \in Y_n$ are essential rather than the norm of Y. Unfortunately, it turns out, that the condition number of B reflects the ill-conditioning of the equation $Ax = y$. In this respect it should be mentioned that it suffices to study the question of convergence for the "principal part" of the operator A under discussion. More concretely, if the projection method converges for an operator A, then convergence and error estimates also hold true for $A + C$, where C is compact relative to A (i.e., $A^{-1}C$ is compact).

The proof of the following result is, e.g., given in the monograph [137]:

Let $C : X \to Y$ *be a linear operator with* $C(X) \subset A(X)$ *such that* $A + C$ *is one-to-one and* $A^{-1}C$ *is compact in* X. *Assume, furthermore, that the projection method converges for* A, *i.,e., that* $R_n A x \to x$, $n \to \infty$, *for every* $x \in X$, *where*

$$R_n = (Q_n A|_{X_n})^{-1} Q_n.$$

Then it also converges for $A + C$ *i.e.,*

$$(Q_n(A+C)|_{X_n})^{-1} Q_n (A+C)x \to x, \; n \to \infty, \quad \textit{for all} \quad x \in X.$$

Let $x \in X$ be the solution of $(A + C)x = y$ and $x^\varepsilon \in X_n$ be the solution of the corresponding projected equation $Q_n(A+C)x_n^\varepsilon = y_n^\varepsilon$ for some $y_n^\varepsilon \in Y_n$. Then there exists a constant $c > 0$ such that

$$\|x - x_n^\varepsilon\| \le c\left[\|A^{-1}Cx - R_nCx\| + \|A^{-1}y - R_ny_n^\varepsilon\|\right] \tag{4.156}$$

for all sufficiently large n and $\varepsilon > 0$.

It should be mentioned that the first term on the right-hand side of (4.156) is just the error of the projection methods for the equations $Ax = Cx$ without perturbation of the right-hand side. This allows to assure the estimate

$$\|A^{-1}Cx - R_nCx\| \le (1 + c) \min_{z_n \in X_N} \|A^{-1}Cx - z_n\|.$$

The second term on the right-hand side of (4.156) is the error for the equation $Ax = y$. Hence, our results include both the continuous and the discrete perturbations of the right-hand side. For the continuous case we set $y_n^\varepsilon := Q_ny^\varepsilon$, while in the discrete case, we set $y_n^\varepsilon = \sum_{i=1}^n \beta_i^\varepsilon y_i$.

All in all, a framework for reduction of inverse and identification problems to finite-dimensional problems exists within the concept of projection methods (cf. [184]). Projection methods (e.g., spline or finite-element functions) can be either applied directly to IPP or to the regularized problem, i.e., to the problem of minimizing a Tikhonov type functional (see, e.g., [158, 199, 241]). The moment-discretization (or semidiscretization) method for integral equations of the first kind and for IPP in reproducing kernel Hilbert spaces (see [178, 195]), viewed as a projection method, is both quasi-optimal and robust. It is also particularly suited when values of the data function $y(s)$ are known only at a finite number of points. These properties, together with the convergence and commutativity properties established in [178] may account for the favorable behavior of computer implementations of this method [16], which the authors [16] call Nashed's method.

Galerkin methods. We deal with the situation that X and Y are (real or complex) Hilbert spaces. Moreover, $A : X \to Y$ is assumed to be linear, bounded, and one-to-one; $X_n \subset X$ and $Y_n \subset Y$ are assumed to be finite-dimensional subspaces with $\dim X_n = \dim Y_n = n$; and $Q_n : Y \to Y_n$ is the orthogonal projection operator onto Y_n. Then, $Q_nAx_n = Q_ny$ reduces to the so-called *Galerkin equations*

$$\langle Ax_n, z_n\rangle = \langle y, z_n\rangle \quad \text{for all} \quad z_n \in Y_n. \tag{4.157}$$

Choosing bases $\{\hat{x}_1, \ldots, \hat{x}_n\}$ and $\{\hat{y}_1, \ldots, \hat{y}_n\}$ of X_n and Y_n, respectively, we are led to a finite system in the coefficients of $x_n = \sum_{j=1}^n \alpha_j\hat{x}_j$ (see (4.145)):

$$\sum_{i=1}^n B_{ij}\alpha_j = \beta_i, \quad i = 1, \ldots, n, \tag{4.158}$$

where we have used the abbreviations $B_{ij} = \langle A\hat{x}_j, \hat{y}_i\rangle_Y$ and $\beta_i = \langle y, \hat{y}_i\rangle_Y$. We observe that B_{ij} and β_i coincide with the settings in (4.142) only if the set $\{\hat{y}_j : j = 1, \ldots, n\}$ forms an orthonormal basis of Y_n.

It should be remarked that the Galerkin method is also known as the *Petrov–Galerkin method* (see [215]) since Petrov was the first to consider the general situation of (4.157). The special case $X = Y$ and $X_n = Y_n$ was studied by Bubnov in 1914 and later by Galerkin in 1915 (see [86]). For this reason, this special case is also known as the *Bubnov–Galerkin method*. In the case when the operator A is self-adjoint and positive definite, we will see that the Bubnow–Galerkin method coincides with the *Rayleigh–Ritz method* (see [221, 228]).

Error estimates. The following error estimates for the Galerkin method of the form (4.154) (see, e.g., [137]) differ only in the first term, which corresponds to the perturbation of the right-hand side. The second term bounds the error for the exact right-hand side and tends to zero, provided the boundedness assumption (4.150) is satisfied.

Assume that the Galerkin equations (4.157) *are uniquely solvable for every right-hand side of the equation* $Ax = y$.

(a) *Let* $y^\varepsilon \in Y$ *with* $\|y - y^\varepsilon\| \le \varepsilon$ *be given and* $x_n^\varepsilon \in X_n$ *be the solution of*

$$\langle Ax_n^\varepsilon, z_n\rangle = \langle y^\varepsilon, z_n\rangle \quad \text{for all} \quad z_n \in Y_n. \tag{4.159}$$

Then the following error estimate holds true:

$$\|x_n^\varepsilon - x\| \le \varepsilon\|R_n\| + \|R_nAx - x\|. \tag{4.160}$$

(b) *Let* B *and* β *be given by* (4.158) *and* $\beta^\varepsilon \in \mathbb{K}^n$ *with* $|\beta - \beta^\varepsilon| \le \varepsilon$, *where* $|\cdot|$ *denotes the Euclidean norm in* $\mathbb{K}^n$. *Let* $\alpha^\varepsilon \in \mathbb{K}^n$ *be the solution of* $B\alpha^\varepsilon := \beta^\varepsilon$. *Set* $x_n^\varepsilon := \sum_{j=1}^n \alpha_j^\varepsilon \hat{x}_j \in X_n$. *Then the following error estimate holds true:*

$$\|x_n^\varepsilon - x\| \le \frac{a_n}{\lambda_n}\,\varepsilon + \|R_nAx - x\|, \tag{4.161}$$

$$\|x_n^\varepsilon - x\| \le b_n\|R_n\|\varepsilon + \|R_nAx - x\|, \tag{4.162}$$

where

$$a_n = \max\left\{\left\|\sum_{j=1}^n \rho_j\hat{x}_j\right\|_X : \sum_{j=1}^n |\rho_j|^2 = 1\right\}, \tag{4.163}$$

$$b_n = \max\left\{\sqrt{\sum_{j=1}^n |\rho_j|^2} : \left\|\sum_{j=1}^n \rho_j\hat{y}_j\right\| = 1\right\}, \tag{4.164}$$

and $\lambda_n > 0$ *denotes the smallest singular value of the matrix* B.

Next we are interested in deriving error estimates for three particularly interesting choices for the finite-dimensional subspaces X_n and Y_n (see, e.g., [137]). The cases, where X_n and Y_n are coupled by $Y_n = A(X_n)$ or $X_n = A^*(Y_n)$ will lead to the least squares method or the dual least squares method, respectively. In addition we will study the Bubnov–Galerkin method for the case where A additionally satisfies the so-called Garding inequality. In all cases, we formulate the

Galerkin equations for the perturbed cases first without using particular bases and then with respect to given bases in X_n and Y_n.

Least squares method. For a finite-dimensional subspace $X_n \subset X$, determine $x_n \in X_n$ such that

$$\|Ax_n - y\| \leq \|Az_n - y\| \quad \text{for all} \quad z_n \in X_n. \tag{4.165}$$

Clearly, existence and uniqueness of $x_n \in X_n$ can be guaranteed easily since X_n is finite-dimensional and A is assumed to be one-to-one. The solution $x_n \in X_n$ of the least squares problem is characterized by

$$\langle Ax_n, Az_n \rangle = \langle y, Az_n \rangle \quad \text{for all} \quad z_n \in X_n. \tag{4.166}$$

We notice that this method is a special case of the Galerkin method, where we have $Y_n = A(X_n)$. Choosing a basis $\{\hat{x}_j, j = 1, \ldots, n\}$ of X_n leads to the finite linear system

$$\sum_{j=1}^{n} \alpha_j \langle A\hat{x}_j, Ax_i \rangle = \beta_i = \langle y, A\hat{x}_i \rangle \quad \text{for all} \quad i = 1, \ldots, n, \tag{4.167}$$

i.e., in matrix-vector nomenclature $B\alpha = \beta$. The corresponding matrix $B \in \mathbb{K}^{n \times n}$ with $B_{ij} = \langle A\hat{x}_j, A\hat{x}_j \rangle_Y$ is Hermitian and positive definite, since A is assumed to be one-to-one.

Of practical interest is the case where the right-hand side is perturbed by an error. Let $x_n^\varepsilon \in X_N$ solve the equations

$$\langle Ax_n^\varepsilon, Az_n \rangle = \langle y^\varepsilon, Az_n \rangle \quad \text{for all} \quad z_n \in X_n, \tag{4.168}$$

where $y^\varepsilon \in Y$ is the perturbed right-hand side satisfying $\|y^\varepsilon - y\|_Y \leq \varepsilon$. For the discrete perturbation, we assume that $\beta \in \mathbb{K}^n$ is replaced by $\beta^\varepsilon \in \mathbb{K}^n$ with $|\beta^\varepsilon - \beta| \leq \varepsilon$, where $|\cdot|$ denotes the Euclidean norm in $\mathbb{K}^n$. This leads to the following finite system of equations in the unknown coefficients of $x_n^\varepsilon = \sum_{j=1}^n a_j^\varepsilon \hat{x}_j$:

$$\sum_{j=1}^{n} \alpha_j^\varepsilon \langle A\hat{x}_j, A\hat{x}_i \rangle = \beta_i^\varepsilon \quad \text{for all} \quad i = 1, \ldots, n. \tag{4.169}$$

The system (4.169) is uniquely solvable, since the matrix B is positive definite.

Obviously, for least squares methods, the boundedness condition (4.150) is not satisfied without imposing additional assumptions (for more details we refer, e.g., to [139, 234]):

Let $A : X \to Y$ be a linear, bounded, and injective operator between Hilbert spaces. Suppose that $X_n \subset X$ form finite-dimensional subspaces such that $\bigcup_{n \in \mathbb{N}} X_n$ is dense in X. Let $x \in X$ be the solution of $Ax = y$ and $x_n^\varepsilon \in X_n$ be the least squares solution from (4.168) *or* (4.169). *Denote by σ_n the quantity*

$$\sigma_n = \max \{\|z_n\| : z_n \in X_n, \|Az_n\| = 1\}. \tag{4.170}$$

Suppose that there exists $c > 0$ independent of n, such that

$$\min_{z_n \in X_n} \{\|x - z_n\| + \sigma_n \|A(x - z_n)\|\} \le c\|x\| \quad \text{for all} \quad x \in X. \tag{4.171}$$

Then, the least squares method is convergent, and we have $\|R_n\| \le \sigma_n$. Moreover, we have the error estimate

$$\|x - x_n^\varepsilon\| \le r_n \sigma_n \, \varepsilon + \overline{c} \, \min\{\|x - z_n\| : z_n \in X_n\} \tag{4.172}$$

for some $\overline{c} > 0$. Here, $r_n = 1$, if $x_n^\varepsilon \in X_n$ solves (4.168)*, i.e., ε measures the continuous perturbation $\|y^\varepsilon - y\|_Y$. If ε measures the discrete error $|\beta^\varepsilon - \beta|$ in the Euclidean norm and $x_n^\varepsilon = \sum_{j=1}^n \alpha_j^\varepsilon \hat{x}_j \in X_n$, where the vector $\alpha^\varepsilon = (\alpha_1^\varepsilon, \ldots, \alpha_n^\varepsilon)^T$ solves* (4.169)*, then r_n is given by*

$$r_n = \max\left\{ \sqrt{\sum_{j=1}^n |\rho_j|^2} : \left\| A\left(\sum_{j=1}^n \rho_j \hat{x}_j \right) \right\| = 1 \right\}. \tag{4.173}$$

For further numerical aspects of least squares method, we refer, e.g., to [57, 58, 134, 157, 173, 178].

Dual least squares method. As another variant of the Galerkin method, we come to the dual least squares method. In this case the boundedness condition (4.150) is indeed always satisfied: Given some finite-dimensional subspaces $Y_n \subset Y$, determine $u_n \in Y_n$ such that

$$\langle A^* u_n, z_n \rangle = \langle y, z_n \rangle \quad \text{for all} \quad z_n \in Y_n, \tag{4.174}$$

where, as always, $A^* : Y \to X$ denotes the adjoint of A. Then $x_n = A^* z_n$ is called the *dual least squares solution*. It is a special case of the Galerkin method, where $X_n = A^*(Y_n)$. Writing (4.174) for $y = Ax$ in the form

$$\langle A^* u_n, A^* z_n \rangle = \langle x, A^* z_n \rangle \quad \text{for all} \quad z_n \in Y_n, \tag{4.175}$$

we see that the dual least squares method is just the least squares method for the equation $A^* u = x$. This explains the standard terminology in the literature.

Suppose now that the right-hand side is perturbed. Let $y^\varepsilon \in Y$ be given such that $\|y^\varepsilon - y\| \le \varepsilon$. Instead of the linear equation (4.174), we determine $x_n^\varepsilon := A^* u_n^\varepsilon \in X_n$ via

$$\langle A^* u_n^\varepsilon, A^* z_n \rangle = \langle y^\varepsilon, z_n \rangle \quad \text{for all} \quad z_n \in Y_n. \tag{4.176}$$

For discrete perturbations, we specify a basis $\{\hat{y}_j, j = 1, \ldots, n\}$ of Y_n and assume that the right-hand side $\beta_i = \langle y, \hat{y}_i \rangle_Y, i = 1, \ldots, n$, of the resulting linear equations are perturbed by a vector $\beta^\varepsilon \in \mathbb{K}^n$ with $|\beta^\varepsilon - \beta| \le \varepsilon$, where $|\cdot|$ denotes the Euclidean norm in $\mathbb{K}^n$. Instead of (4.174) we are then led to

$$x_n^\varepsilon = A^* u_n^\varepsilon = \sum_{j=1}^n \alpha_j^\varepsilon A^* \hat{y}_j, \tag{4.177}$$

where $\alpha^\varepsilon \in \mathbb{K}^n$ solves the linear equation

$$\sum_{j=1}^{n} \alpha_j^\varepsilon \langle A^* \hat{y}_j, A^* \hat{y}_i \rangle = \beta_i^\varepsilon, \quad i = 1, \ldots, n. \tag{4.178}$$

Results on convergence and error estimates are listed, e.g., in the textbook [137].

Let X and Y be Hilbert spaces. Suppose that $A : X \to Y$ is linear, bounded, an one-to-one such that the range $A(X)$ is dense in Y. Let $Y_n \subset Y$ form finite-dimensional subspaces such that $\bigcup_{n\in\mathbb{N}} Y_n$ is dense in Y. Assume that $x \in X$ is the solution of $Ax = y$. Then the linear equations (4.176) *and* (4.178) *are uniquely solvable for every right-hand side and every $n \in \mathbb{N}$. Furthermore, the dual least squares method is convergent, and we have*

$$\|R_n\| \le \sigma_n = \max\{\|z_n\| : z_n \in Y_n, \|A^* z_n\| = 1\}. \tag{4.179}$$

Moreover, we have the error estimate

$$\|x - x_n^\varepsilon\| \le r_n \sigma_n \, \varepsilon + c \, \min\{\|x - z_n\| : z_n \in A^*(Y_n)\} \tag{4.180}$$

for some $c > 0$. Here, $r_n = 1$ if $r_n^\varepsilon \in X_n$ solves (4.176)*, i.e., ε measures the norm $\|y^\varepsilon - y\|$ in Y. If ε measures the discrete error $|\beta^\varepsilon - \beta|$ and $x_n^\delta = \sum_{j=1}^n \alpha_j^\varepsilon A^* \hat{y}_j \in X_n$, where α^ε solves* (4.178)*, then r_n is given by*

$$r_n = \max\left\{ \sqrt{\sum_{j=1}^{n} |\rho_j|^2} : \left\| \sum_{j=1}^{n} \rho_j \hat{y}_j \right\| = 1 \right\} \tag{4.181}$$

(note that $r_n = 1$ if $\{\hat{y}_j, j = 1, \ldots, n\}$ constitutes an orthonormal system in Y).

Bubnov–Galerkin method. We assume that $A : X \to X$ is a linear and bounded operator and $X_n, n \in \mathbb{N}$, are finite-dimensional subspaces. The Galerkin method amounts to the problem of determining $x_n \in X_n$ such that

$$\langle Ax_n, z_n \rangle = \langle y, z_n \rangle \quad \text{for all} \quad z_n \in X_n. \tag{4.182}$$

This special case is called the *Bubnov–Galerkin method*. Again, we consider the perturbation of the right-hand side. If $y^\varepsilon \in Y$ is chosen such that $\|y^\varepsilon - y\| \le \varepsilon$ represents a perturbed right-hand side, then instead of (4.182) we study the equation system

$$\langle Ax_n^\varepsilon, z_n \rangle = \langle y^\varepsilon, z_n \rangle \quad \text{for all} \quad z_n \in X_n. \tag{4.183}$$

An alternative is to choose a basis $\{\hat{x}_j, j = 1, \ldots, n\}$ of X_n. We assume that the right-hand side $\beta_i = (y, \hat{y}_i)$, $i = 1, \ldots, n$ of the Galerkin equations are perturbed by a vector $\beta^\varepsilon \in \mathbb{K}^n$ with $|\beta^\varepsilon - \beta| \le \varepsilon$, where $|\cdot|$ denotes again the Euclidean norm in $\mathbb{K}^n$. In this case, instead of (4.182), we have to solve

$$\sum_{j=1}^{n} a_j^\varepsilon \langle A\hat{x}_j, \hat{x}_i \rangle = \beta_i^\varepsilon \quad \text{for} \quad i = 1, \ldots, n. \tag{4.184}$$

For $\alpha^\varepsilon \in \mathbb{K}^n$ we set $x_n^\varepsilon = \sum_{j=1}^n \alpha_j^\varepsilon \hat{x}_j$.

Next we show that the *Rayleigh–Ritz method*, in fact, is a special case of the Bubnov–Galerkin method.

Rayleigh–Ritz method. Let $A : X \to X$ be also self-adjoint and positive definite, so that $\langle Ax, y\rangle_X = \langle x, Ay\rangle_X$ and $\langle Ax, x\rangle_X > 0$ for all $x, y \in X$ with $x \neq 0$. We introduce the functional

$$J(z) = \langle Az, z\rangle - 2Re\langle y, z\rangle \quad \text{for} \quad z \in X. \tag{4.185}$$

The identity

$$J(z) - J(x) = 2Re\langle Ax - y, z - x\rangle + \langle A(z - x), z - x\rangle \tag{4.186}$$

and the positivity of A tells us that $x \in X$ is the unique minimum of J if and only if x solves $Ax = y$. The Rayleigh–Ritz method is to minimize J over the finite-dimensional subspace X_n. From (4.186), we see that if $x_n \in X_n$ minimizes J on X_n, then , for $z_n = x_n \pm \varepsilon u_n$ with $u_n \in X_n$ and $\varepsilon > 0$, it follows that

$$0 \leq J(z_n) - J(x_n) = \pm\varepsilon \quad 2Re\langle Ax_n - y, u_n\rangle + \varepsilon^2\langle Au_n, u_n\rangle$$

for all $u_n \in X_n$. By dividing $\varepsilon > 0$ and afterwards letting $\varepsilon \to 0$ we find that $x_n \in X_n$ satisfies the equation system (4.182). If, on the other hand, $x_n \in X_n$ solves (4.182), then we get from (4.186),

$$J(z_n) - J(x_n) = \langle A(z_n - x_n), z_n - x_n\rangle \geq 0$$

for all $z_n \in X_n$. Therefore, the *Rayleigh–Ritz method coincides with the Bubnov–Galerkin method.*

Finally we are interested in the Bubnov–Galerkin method for the important class of coercive operators. As preparatory material we briefly recapitulate some settings (see, e.g., [137]):

(i) A *Gelfand triple* (V, X, V^*) consists of a reflexive Banach space V, an Hilbert space X, and the dual space V^* of V such that
   (a) V is a dense subspace of X,
   (b) the imbedding $J : V \to X$ is bounded.

   It is conventional to write (see, e.g., [137]) $V \subset X \subset V^*$ since we can identify X with a dense subspace of V^*. This identifications given by the dual operator $J^* : X \to V^*$ of J, where we identify the dual of the Hilbert space X by itself. From $(x, y) = \langle J^*x, y\rangle$, for all $x \in X$ and $y \in V$ we see that with this identification the dual pairing $\langle\cdot,\cdot\rangle$ in (V^*, V) is an extension of the inner product $(\cdot,\cdot)$ in X, i.e., we write

$$\langle x, y\rangle = (x, y) \quad \text{for all} \quad x \in Y \quad \text{and} \quad y \in V.$$

   Furthermore, we have the estimates

$$|\langle x, y\rangle| \leq \|x\|_{V^*}\|y\|_V \quad \text{for all} \quad x \in V^*, y \in V,$$

   thus,

$$|\langle x, y\rangle| \leq \|x\|_{V^*}\|y\|_V \quad \text{for all} \quad x \in X, y \in V.$$

   It is well known that J^* is one-to-one and has a dense range.

(ii) Let V be a reflexive Banach space with dual space V^*. We denote the norms in V and V^* by $\|\cdot\|_V$ and $\|\cdot\|_{V^*}$, respectively. A linear bounded operator $A : V^* \to V$ is called *coercive* if there exists $\gamma > 0$ with

$$Re\langle x, Ax\rangle \geq \gamma \, \|x\|_V^2 \quad \text{for all} \quad x \in V^*, \tag{4.187}$$

where $\langle\cdot,\cdot\rangle$ denotes the dual pairing in (V^*, V).

(iii) The operator A satisfies *Garding's inequality* if there exists a linear compact operator $C : V^* \to V$ such that $A + C$ is coercive, i.e.,

$$Re(x, Ax) \geq \gamma \, \|x\|_{V^*}^2 - Re\langle x, Cx\rangle \quad \text{for all} \quad x \in V^*.$$

Note that, by the same argument as in the Lax–Milgram theorem, it can be shown that every coercive operator is an isomorphism from V^* onto V. Coercive operators play an important role in the study of partial differential equations and integral equations by variational methods. In the conventional definition, the roles of V and V^* are interchanged. For integral operators that are "smoothing", our definition seems to be more appropriate. However, both definitions are equivalent in the sense that the inverse operator $A^{-1} : V \to V^*$ is coercive in the usual sense with γ replaced by $\gamma/\|A\|^2$.

Convergence of the Bubnov–Galerkin method. After these preparations we are in the position to formulate convergence of the Bubnov–Galerkin method for coercive operators (see [137]).

Let (V, X, V^*) be a Gelfand triple, and $X_n \subset V$ be finite-dimensional subspaces such that $\bigcup_{n\in\mathbb{N}} X_n$ is dense in X. Let $K : V^* \to V$ be coercive with constant $\gamma > 0$. Let $x \in X$ be the solution of $Ax = y$. Then we have the following results:

(a) *There exist unique solutions of the Galerkin equations* (4.182)–(4.184)*, and the Bubnov–Galerkin method converges in V^* with*

$$\|x - x_n\|_{V^*} \leq c \, \min\{\|x - z_n\|_{V^*} : z_n \in X_n\} \tag{4.188}$$

for some $c > 0$.

(b) *Define the quantity $\rho_n > 0$ by*

$$\rho_n = \max\{\|u\| : u \in X_n, \ \|u\|_{V^*} = 1\} \tag{4.189}$$

and the orthogonal projection operator P_n from X onto X_n. The Bubnov–Galerkin method converges in X if there exists $c > 0$ with

$$\|u - P_n u\|_{V^*} \leq \frac{c}{\rho_n}\|u\| \quad \textit{for all} \quad u \in X. \tag{4.190}$$

In this case, we have the estimates

$$\|R_n\| \leq \frac{1}{\gamma}\rho_n^2 \tag{4.191}$$

and

$$\|x - x_n^\varepsilon\| \leq c[r_n\rho_n^2 + \min\{\|x - z_n\| : z_n \in X_n\}] \tag{4.192}$$

for some $c > 0$. Here $r_n = 1$ if $x_n^\varepsilon \in X_n$ solves (4.183)*, i.e., ε measures the norm $\|y^\varepsilon - y\|$ in X. If ε measures the discrete error $|\beta^\varepsilon - \beta|$ in the Euclidean norm and $x_n^\varepsilon = \sum_{j=1}^n \alpha_j^\varepsilon \hat{x}_j \in X_n$, where α^ε solves* (4.184)*, then r_n is given by*

$$r_n = \max\left\{ \sqrt{\sum_{j=1}^n |\rho_j|^2} : \| \sum_{j=1}^n \rho_j \hat{x}_j \| = 1 \right\}. \tag{4.193}$$

Again, we note that $r_n = 1$ if $\{\hat{x}_j, j = 1, \ldots, n\}$ forms an orthonormal system in X. For further details, we refer to [200] and the monographs [22, 137, 139, 148].

4.3. Multiscale methods as regularization schemes

Next a compact operator equation is dealt within regularization methods, based on filtering techniques by means of wavelets. In a general setup a singular integral approach to regularization is established, decomposition/reconstruction regularization wavelets are introduced in the frequency space which allow the regularization in form of a multiresolution analysis. Two different types of regularization wavelets are discussed in more detail, namely (non-locally supported) Tikhonov–Phillips regularization wavelets and (bandlimited) truncated singular value decomposition wavelets. Our considerations closely follow Freeden, Schneider [77] about regularization and multiresolution.

Let $(X, \langle\cdot,\cdot\rangle)$ and $(Y, \langle\cdot,\cdot\rangle)$ be separable real functional Hilbert spaces over domains $\mathcal{G}_X$ and $\mathcal{G}_Y$, respectively, i.e., X, respectively, Y consists of functions $x : \mathcal{G}_X \to \mathbb{R}$, respectively, $y : \mathcal{G}_Y \to \mathbb{R}$. We consider a linear, compact operator $A : X \to Y$ satisfying $\mathcal{N}(A) = \{0\}$, $\mathcal{R}(A) \subsetneq Y$, $\overline{\mathcal{R}(A)} = Y$, so that $A^*A : X \to X$ is self-adjoint.

In the sequel, we denote by $\{x_n\}_{n=0,1,\ldots}$ a complete orthonormal system in $(X, \langle\cdot,\cdot\rangle_X)$ and by $\{y_n\}_{n=0,1,\ldots}$ a complete orthonormal system in $(Y, \langle\cdot,\cdot\rangle_Y)$ such that the singular values $\{\sigma_n\}_{n=0,1,\ldots}$ of A satisfy $Ax_n = \sigma_n y_n$, $A^*y_n = \sigma_n x_n$, $n \in \mathbb{N}_0$. Since A is supposed to be injective it follows that $\sigma_n > 0$ for all $n \in \mathbb{N}_0$. Any $p \in P$, $P \in \{X, Y\}$, can be represented in terms of an orthonormal (Fourier) expansion with respect to $\{p_n\}_{n=0,1,\ldots}$

$$p = \sum_{n=0}^\infty p_P^\wedge(n) p_n \tag{4.194}$$

with

$$p_P^\wedge(n) = \langle p, p_n \rangle_P, \tag{4.195}$$

where the equality in (4.194) is understood in the $\|\cdot\|_P$-sense. In conclusion, any element of $\mathcal{R}(A)$ admits an expansion of the form

$$Ax = \sum_{n=0}^\infty \sigma_n x_X^\wedge(n) y_n. \tag{4.196}$$

As is well known, the Picard condition tells us that the problem

$$Ax = y, \quad x \in X,\ y \in Y \tag{4.197}$$

has a solution if and only if $y \in Y$ satisfies

$$\sum_{n=0}^{\infty} (\sigma_n^{-1}\ y_Y^{\wedge}(n))^2 < \infty. \tag{4.198}$$

In this case it is known that the solution of (4.197) is representable in the form

$$x = \sum_{n=0}^{\infty} \sigma_n^{-1}\ y_Y^{\wedge}(n) x_n\ =\ A^{\dagger} y. \tag{4.199}$$

Since the right-hand side y is error affected (due to the inaccuracy of the measuring instrument) in any practical application, the series (4.199) will not converge in general. For that purpose we are interested in regularized solutions.

The idea (cf. [77]) we follow is to represent the J-level regularization of the problem

$$Ax\ =\ y, \qquad x \in X, y \in Y \tag{4.200}$$

by means of a wavelet analysis. The overall advantage of such a method is that we obtain a $J+1$-level regularization by starting with the J-level regularization and adding so-called detail information. It becomes clear that any classical regularization method based on a filtered singular value decomposition can be reformulated in terms of our wavelet method. Thus, any known parameter choice strategy depending on the special method is also applicable and, moreover, any of the corresponding error estimates holds true, too. For that reason we omit these discussions here.

As is well known, a family $\{R_J\}_{J\in\mathbb{Z}}$ of linear operators $R_J : Y \to X$, $J \in \mathbb{Z}$, is a regularization of $A^{\dagger}$ if it satisfies the following properties:

(i) R_J is bounded on Y for all $J \in \mathbb{Z}$,
(ii) for any member $y \in \mathcal{R}(A)$, the limit relation $\lim_{J\to\infty} R_J y = A^{\dagger} y$ holds in the $\|\cdot\|_X$-sense.

The kernel $x_J = R_J y$ is called the J-level regularization of the problem (4.197).

Product kernels. A function $\Gamma^{P,Q}\langle\cdot,\cdot\rangle : \mathcal{G}_P \times \mathcal{G}_Q \to \mathbb{R}$, $P, Q \in \{X, Y\}$, of the form

$$\Gamma^{P,Q}(x,y) = \sum_{n=0}^{\infty} \Gamma^{\wedge}(n)\ p_n(x)\ q_n(y), \quad x \in \mathcal{G}_P, y \in \mathcal{G}_Q,\ \Gamma^{\wedge}(n) \in \mathbb{R}, n \in \mathbb{N}_0, \tag{4.201}$$

is called a (P,Q)-(*product*) *kernel*. Note that the indices P and Q in $\Gamma^{P,Q}$ are associated to the variables x and y, respectively, such that $\Gamma^{P,Q}(x,y) = \Gamma^{Q,P}(y,x)$. The sequence $\{\Gamma^{\wedge}(n)\}_{n=0,1,\ldots}$ is called the *symbol of the* (P,Q)*-kernel*. For brevity, a (P,P)-kernel is simply said to be a P-kernel.

An important question for our investigations is as follows: Fix one variable of a product kernel, what are the conditions for the product kernel (as a function of the remaining variable) to be of class $P \in \{X, Y\}$? The answer is provided by the concept of symbol admissibility.

A symbol $\{\Gamma^\wedge(n)\}_{n=0,1,\ldots}$ is called *P-admissible*, $P \in \{X, Y\}$, if it satisfies the following conditions:

(Ai)
$$\sum_{n=0}^{\infty} \left(\Gamma^\wedge(n)\right)^2 < \infty, \tag{4.202}$$

(Aii)
$$\sup_{x \in \mathcal{G}_P} \left(\sum_{n=0}^{\infty} (\Gamma^\wedge(n) p_n(x))^2 \right) < \infty. \tag{4.203}$$

A symbol $\{\Gamma^\wedge(n)\}_{n=0,1,\ldots}$ is called ((P,Q)-admissible, $P, Q \in \{X, Y\}$, or simply) *admissible*, if it is P-admissible as well as Q-admissible.

From the definition of admissibility we immediately obtain the result:

Let $\{\Gamma^\wedge(n)\}_{n=0,1,\ldots}$ be the symbol of an (P,Q)-kernel, $P, Q \in \{X, Y\}$.

(α) If $\{\Gamma^\wedge(n)\}_{n=0,1,\ldots}$ is P-admissible, then $\Gamma^{P,Q}(x,\cdot) \in Q$ for every (fixed) $x \in \mathcal{G}_P$,

(β) If $\{\Gamma^\wedge(n)\}_{n=0,1,\ldots}$ is Q-admissible, then $\Gamma^{P,Q}(\cdot,y) \in P$ for every (fixed) $y \in \mathcal{G}_Q$.

Convolutions. A fundamental tool for our wavelet theory is the concept of a *convolution* we introduce below: Let $\Gamma^{X,X}$ be an X-kernel with X-admissible symbol. Suppose that F is of class X. Then we understand the *convolution of* $\Gamma^{X,X}$ *and* F to be the function given by

$$(\Gamma^{X,X} * x)(t) = \left\langle \Gamma^{X,X}(t,\cdot), x \right\rangle_X = \sum_{n=0}^{\infty} \Gamma^\wedge(n) x_X^\wedge(n) x_n(x), \quad x \in \mathcal{G}_X. \tag{4.204}$$

We immediately see that $(\Gamma^{X,X} * x)_X^\wedge(n) = \Gamma^\wedge(n) x_X^\wedge(n)$ and $\Gamma^{X,X} * x \in X$. In analogous way we define the *convolution* of an (X,Y)-kernel $\Gamma^{X,Y}$ having an X-admissible symbol *with a function* $y \in Y$ to be the expression

$$(\Gamma^{X,Y} * y)(t) = \left\langle \Gamma^{X,Y}(t,\cdot), y \right\rangle_Y = \sum_{n=0}^{\infty} \Gamma^\wedge(n) y_Y^\wedge(n) x_n(t), \quad t \in \mathcal{G}_X, \tag{4.205}$$

and it follows that $\Gamma^{X,Y} * y \in X$. Lastly, the *convolution* of an (X,Y)-kernel $\Gamma^{X,Y}$ having an Y-admissible symbol *with a function* $F \in X$ is given by

$$(\Gamma^{X,Y} * x)(s) = \left\langle \Gamma^{X,Y}(\cdot,s), x \right\rangle_X = \sum_{n=0}^{\infty} \Gamma^\wedge(n) x_X^\wedge(n) y_n(s), \quad s \in \mathcal{G}_Y, \tag{4.206}$$

and we have $\Gamma^{X,Y} * x \in Y$. Next we proceed with the convolution of two product kernels leading to the following result: Let $\Gamma^{X,X}$ be an X-kernel with X-admissible

symbol and let $\Upsilon^{X,Y}$ be an (X,Y)-kernel with Y-admissible symbol. Then

$$\begin{aligned}(\Gamma * \Upsilon)^{X,Y}(t,s) &= (\Gamma^{X,X} * \Upsilon^{X,Y}(\cdot,s))(t)\\ &= \left\langle \Gamma^{X,X}(t,\cdot), \Upsilon^{X,Y}(\cdot,s)\right\rangle_X\\ &= \sum_{n=0}^{\infty} \Gamma^{\wedge}(n)\Upsilon^{\wedge}(n)x_n(t)y_n(s), \quad t \in \mathcal{G}_X, s \in \mathcal{G}_Y \end{aligned} \tag{4.207}$$

represents an (X,Y)-kernel with admissible symbol

$$((\Gamma * \Upsilon)^{X,Y})(n) = \Gamma^{\wedge}(n)\Upsilon^{\wedge}(n). \tag{4.208}$$

Dilation and shifting. In order to prepare the fundamentals of the forthcoming wavelet theory we are now interested in countable families $\{\Gamma_J^{P,Q}\}$, $J \in \mathbb{Z}$, of product kernels $\Gamma_J^{P,Q}$, $P,Q \in \{X,Y\}$. Observing our notations we are able to define a dilation operator acting on these families in the following way: let $\Gamma_J^{P,Q}$ be a member of the family of product kernels.

Then the *dilation operator* D_K, $K \in \mathbb{Z}$ is defined by $D_K\Gamma_J^{P,Q} = \Gamma_{J+K}^{P,Q}$. Especially, we obtain $\Gamma_J^{P,Q} = D_J\Gamma_0^{P,Q}$, $J \in \mathbb{Z}$. Thus we refer $\Gamma_0^{P,Q}$ to as a "mother kernel".

Moreover, we define a *shifting operator* S_t^P, $t \in \mathcal{G}_P$, $P \in \{X,Y\}$, by $S_t^P\Gamma_J^{P,Q} = \Gamma_J^{P,Q}(t,\cdot)$, $\quad t \in \mathcal{G}_P, J \in \mathbb{Z}$, resp. $S_s^Q\Gamma_J^{P,Q} = \Gamma_J^{P,Q}(\cdot,s)$, $s \in \mathcal{G}_Q, J \in \mathbb{Z}$. In doing so we consequently get $\Gamma_J^{P,Q}(t,\cdot) = S_t^P D_J\Gamma_0^{P,Q}$, $t \in \mathcal{G}_P, J \in \mathbb{Z}$, resp. $\Gamma_J^{P,Q}(\cdot,s) = S_s^Q D_J\Gamma_0^{P,Q}$, $s \in \mathcal{G}_Q, J \in \mathbb{Z}$.

Regularization scaling functions. Next we are concerned with a wavelet based regularization technique of problem (4.200): Let $\{(\Phi_J)^{\wedge}(n)\}_{n=0,1,\ldots}$, $J \in \mathbb{Z}$, be an (X,Y)-admissible symbol of a family of product kernels which additionally satisfies the following properties:

i) $\lim_{J\to\infty} \sigma_n((\Phi_J)^{\wedge}(n))^2 = 1$, $n \in \mathbb{N}$,
ii) $((\Phi_{J+1})^{\wedge}(n))^2 \geq ((\Phi_J)^{\wedge}(n))^2$, $J \in \mathbb{Z}$, $n \in \mathbb{N}$,
iii) $\lim_{J\to-\infty}((\Phi_J)^{\wedge}(n))^2 = 0$, $n \in \mathbb{N}$,
iv) $\sigma_0((\Phi_J)^{\wedge}(0))^2 = 1$, $J \in \mathbb{Z}$.

Then $\{(\Phi_J)^{\wedge}(n)\}_{n=0,1,\ldots}$ is said to be the *generating symbol of a regularization scaling function* (with respect to (4.197)). The (X,Y)-kernel

$${}^d\Phi_J^{X,Y} = \sum_{n=0}^{\infty} (\Phi_J)^{\wedge}(n)x_n y_n \tag{4.209}$$

is called a *decomposition regularization scaling function*, the (X,X)-kernel

$${}^r\Phi_J^{X,X} = \sum_{n=0}^{\infty} (\Phi_J)^{\wedge}(n)x_n x_n \tag{4.210}$$

is called a *reconstruction regularization scaling function.*

From the results of the previous subsection it is clear that ${}^d\Phi_J^{X,Y}(t,\cdot) \in Y, t \in \mathcal{G}_X, J \in \mathbb{Z}$, ${}^r\Phi_J^{X,X}(t,\cdot) \in X$, $t \in \mathcal{G}_X, J \in \mathbb{Z}$, and $({}^r\Phi_J * {}^d\Phi_J)^{X,Y}$ is an (X,Y)-kernel with (X,Y)-admissible symbol $\{((\Phi_J)^\wedge(n))^2\}_{n=0,1,\ldots}$. Observing these properties we are able to verify the following result which is central for our considerations:

Let $\{(\Phi_J)^\wedge(n)\}_{n=0,1,\ldots}$, $J \in \mathbb{Z}$, be the generating symbol of a regularization scaling function. Then, for any $y \in Y$,

$$x_J = {}^r\Phi_J^{X,X} * ({}^d\Phi_J^{X,Y} * y) \tag{4.211}$$

represents the J-level regularization of problem (4.200)*. If, in addition, $y \in \mathcal{R}(A)$, then*

$$\lim_{J\to\infty} \|x_J - A^\dagger y\| = 0. \tag{4.212}$$

For $R_J : Y \to X$ defined by

$$R_J = \left({}^r\Phi_J * {}^d\Phi_J\right)^{X,Y} * y,\ y \in Y$$

we have (see [77])

$$R_J y = \sum_{n=0}^{\infty} ((\Phi_J)^\wedge(n))^2\ y_Y^\wedge(n)\ x_n$$

and

$$\|R_J\|^2 \le \sum_{n=0}^{\infty} ((\Phi_J)^\wedge(n))^4, \quad J \in \mathbb{Z}. \tag{4.213}$$

As an immediate consequence of our results we obtain the result: Let y be a member of $\mathcal{R}(A)$. Suppose that $y^\varepsilon \in Y$ denotes the right-hand side of problem (4.200) with noise level $\|y - y^\varepsilon\| \le \varepsilon$. Moreover, assume that $\{(\Phi_J)^\wedge(n)\}_{n=0,1,\ldots}$, $J \in \mathbb{Z}$, is the generating symbol of a regularization scaling function, where the parameter $J = J(\varepsilon)$ is assumed to satisfy

(1) $\lim_{\varepsilon\to 0} J(\varepsilon) = \infty$,
(2) $\lim_{\varepsilon\to 0} \varepsilon((\Phi_J)^\wedge(n))^2 = 0$.

Then we have

$$\lim_{\varepsilon\to 0} \|({}^r\Phi_J * {}^d\Phi_J)^{X,Y} * y^\varepsilon - A^\dagger y\| = 0. \tag{4.214}$$

Condition iii) seems to be unnecessary for the proof of (4.214) and, in fact, it is. Nevertheless, in what follows we need this assumption for our multiresolution analysis and the (spectral) introduction of wavelets.

Multiresolution analysis. For any $y \in \mathcal{R}(A)$ each $({}^r\Phi_J * {}^d\Phi_J)^{X,Y} * y$ provides a regularization of the solution $A^\dagger y$ at scale J by "smoothing" the Fourier coefficients of $A^\dagger y$ with the symbol $\{((\Phi_J)^\wedge(n))^2\}_{n=0,1,\ldots}$. In terms of filtering, $({}^r\Phi_J * {}^d\Phi_J)^{X,Y}$ may be interpreted as a low-pass filter. Accordingly we understand the scale spaces $\mathcal{V}_J$ to be the image of $\mathcal{R}(A)$ under the operator R_J:

$$\mathcal{V}_J = R_J(\mathcal{R}(A)) = \{({}^r\Phi_J * {}^d\Phi_J)^{X,Y} * y : y \in \mathcal{R}(A)\}. \tag{4.215}$$

This leads us to the properties formulated in the following statement:

The scale spaces satisfy the following properties:

i) $\{X_0\} \subset \mathcal{V}_J \subset \mathcal{V}_{J'} \subset X$, $J \leq J'$, i.e., for any right-hand side $y \in \mathcal{R}(A)$ of problem (4.200), all J-level regularizations with fixed parameter J are sampled in a scale space $\mathcal{V}_J$ with the above property,
ii) $\bigcap_{J=-\infty}^{\infty} \mathcal{V}_J = \{x_0\}$,
iii) $\overline{\bigcup_{J=-\infty}^{\infty} \mathcal{V}_J}^{\|\cdot\|_X} = X$,
iv) if $x_J \in \mathcal{V}_J$, then $D_{-1}x_J \in \mathcal{V}_{J-1}$, $J \in \mathbb{Z}$.

If a collection of subspaces of X satisfies the above conditions we call them a *regularization multiresolution analysis* (*RMRA*).

Regularization wavelet functions. The definition of the regularization scaling function allows us to introduce regularization wavelets. An essential point is the definition of a decomposition and a reconstruction regularization wavelet associated to regularization mother wavelets. This definition, of course, has to be formulated in close relation to a prescribed regularization scaling function.

Let $\{(\Phi_j)^\wedge(n)\}_{n=0,1,\ldots}$, $j \in \mathbb{Z}$, be the generating symbol of a regularization scaling function. Then the (X,Y)-admissible generating symbol $\{(\Psi_j)^\wedge(n)\}_{n=0,1,\ldots}$, $j \in \mathbb{Z}$, and the (X,X)-admissible generating symbol $\{(\tilde{\Psi}_j)^\wedge(n)\}_{n=0,1,\ldots}$, $j \in \mathbb{Z}$, respectively, are defined by the "scaling equation"

$$(\tilde{\Psi}_j)^\wedge(n)(\Psi_j)^\wedge(n) = ((\Phi_{j+1})^\wedge(n))^2 - ((\Phi_j)^\wedge(n))^2. \tag{4.216}$$

Correspondingly, the (X,Y)-kernel

$${}^d\Psi_J^{X,Y} = \sum_{n=0}^{\infty} (\Psi_J)^\wedge(n)\ x_n\ y_n \tag{4.217}$$

is called the *decomposition regularization wavelet*, while the (X,X)-kernel

$${}^r\tilde{\Psi}_J^{X,X} = \sum_{n=0}^{\infty} (\tilde{\Psi}_J)^\wedge(n)\ x_n\ x_n \tag{4.218}$$

is called the *reconstruction regularization wavelet*. The corresponding regularization mother wavelets are denoted by ${}^d\Psi_0^{X,Y}$ and ${}^r\tilde{\Psi}_0^{X,X}$, respectively.

Using this notation, any decomposition regularization wavelet, respectively, any reconstruction regularization wavelet can be interpreted as a dilated and shifted copy of the corresponding mother wavelet.

$${}^d\Psi_J^{X,Y}(t,\cdot) = S_t^X D_J {}^d\Psi_0^{X,Y}, \tag{4.219}$$

$${}^r\Psi_J\tilde{\Psi}_J^{X,X}(t,\cdot) = S_t^X D_J {}^r\Psi_0^{X,X}. \tag{4.220}$$

At this stage it becomes obvious why we required Condition iv). A conclusion of (4.216) is a vanishing 0th moment of the regularization wavelets. Moreover, from

(4.216) it can be readily deduced that

$$((\Phi_0)^\wedge(n+1))^2 = \sum_{j=-\infty}^{J} (\tilde{\Psi}_j)^\wedge(n)(\Psi_j)^\wedge(n) = ((\Phi_0)^\wedge(n))^2 + \sum_{j=0}^{J} (\tilde{\Psi}_j)^\wedge(n)(\Psi_j)^\wedge(n). \tag{4.221}$$

Thus, we easily see in connection with 4.216 that

$$\left({}^r\tilde{\Phi}_{J+1} *^d \Phi_{J+1}\right)^{X,Y} = \sum_{j=-\infty}^{J} ({}^r\tilde{\Psi}_j *^d \Psi_j)^{X,Y} = ({}^r\Phi_0 *^d \Phi_0)^{X,Y} + \sum_{j=0}^{J} ({}^r\tilde{\Psi}_j *^r \Psi_j)^{X,Y}. \tag{4.222}$$

In analogy to the definition of the operator R_J we consider now convolution operators $S_J : Y \to X$, $J \in \mathbb{Z}$, defined by

$$S_J\, y = ({}^r\tilde{\Psi}_J *^d \Psi_J)^{X,Y} * y. \tag{4.223}$$

It describes the "detail information" of the right-hand side y at scale J. From Equation (4.222) it follows that the operator R_{J+1} can be decomposed in the following way:

$$R_{J+1} = R_0 + \sum_{j=0}^{J} S_j. \tag{4.224}$$

But this gives rise to introduce the *detail spaces* as follows:

$$\mathcal{W}_J = S_J(\mathcal{R}(A)) = \{({}^r\tilde{\Psi}_J *^d \Psi_J)^{X,Y} * y : \ y \in \mathcal{R}(A)\}. \tag{4.225}$$

The space $\mathcal{W}_J$ contains the "detail information" needed to go from a regularization at level J to a regularization at level $J+1$. Note that

$$\mathcal{V}_J = \mathcal{V}_{J-1} + \mathcal{W}_{J-1}, \tag{4.226}$$

$$\sum_{j=-\infty}^{J} \mathcal{W}_j = \mathcal{V}_0 + \sum_{j=0}^{J} \mathcal{W}_j = \mathcal{V}_{J+1}. \tag{4.227}$$

It is worth mentioning that, in general, the sum decomposition is neither direct nor orthogonal. Two examples leading to orthogonal and non-orthogonal multiresolution are introduced in the next subsection.

Any right-hand side $y \in Y$ can now be decomposed as follows. Starting with $R_0 y$ we find

$$R_{J+1} y = R_0 y + \sum_{j=0}^{J} S_j y \tag{4.228}$$

for any $J \in \mathbb{Z}$. In other words, the partial reconstruction $R_J y$ is nothing else than the "difference of two smoothings" at two consecutive scales,

$$S_J y = R_{J+1} y - R_J y. \tag{4.229}$$

For what follows we define the *regularization wavelet transform at scale $J \in \mathbb{Z}$ and position $t \in \mathcal{G}_X$* by letting

$$RWT(y)(J;t) = \langle {}^d\Psi_J^{X,Y}(t,\cdot), y\rangle_Y, \quad y \in Y. \tag{4.230}$$

From (4.222) it is not hard to verify the main result in this context:

Let $\{(\Phi_J)^\wedge(n)\}_{n=0,1,\ldots}$, $J \in \mathbb{Z}$, be the generating symbol of a regularization scaling function. Suppose that

$$\{(\Psi_j)^\wedge(n)\}_{n=0,1,\ldots}, \quad \{(\tilde{\Psi}_j)^\wedge(n)\}_{n=0,1,\ldots}, \quad j \in \mathbb{Z}$$

are the generating symbols of the corresponding regularization wavelets. Furthermore, let y be of class Y. Then,

$$x_J = ({}^r\Phi_0 * {}^d\Phi_0)^{X,Y} * y + \sum_{j=0}^{J} {}^r\tilde{\Psi}_j * RWT(y)(j;\cdot) \tag{4.231}$$

denotes the J-level regularization of the problem (4.200) satisfying

$$\lim_{J\to\infty} \|x_J - A^\dagger y\| = 0 \tag{4.232}$$

provided that $y \in \mathcal{R}(A)$.

Equation (4.231) shows the essential characteristic of regularization wavelets. By adding the so-called detail information of level J as the difference of two smoothings of two consecutive scales $J+1$ and J, we change the regularized solution from x_J to x_{J+1} thereby satisfying $\lim_{J\to\infty} x_J = A^\dagger y$ in the case that $y \in \mathcal{R}(A)$. Of course, this can be understood as a kind of a permanence principle.

Table: The multiscale method as presented here can be illustrated by the scheme shown in Figure 4.1.

$R_0(y)$		$R_1(y)$		$R_2(y)$	$\ldots$	$\underset{j\to\infty}{\longrightarrow}$	$A^\dagger y$
$\curvearrowright$		$\curvearrowright$		$\curvearrowright$			$\curvearrowright$
$\mathcal{V}_0$	$\subset$	$\mathcal{V}_1$	$\subset$	$\mathcal{V}_2$	$\ldots$	$=$	$A^\dagger X$
$\mathcal{V}_0+$	$\mathcal{W}_0$	$+$	$\mathcal{W}_1$	$+ \mathcal{W}_2$	$\ldots$	$=$	$A^\dagger X$
$\cup$	$\cup$		$\cup$	$\cup$			$\cup$
$R_0(y)+$	$S_0(y)$	$+$	$S_1(y)$	$+ S_2(y)+$	$\ldots$	$=$	$A^\dagger y$

FIGURE 4.1. Multiresolution regularization scheme.

Some generating symbols. The singular values $\{\sigma_n\}_{n=0,1,\ldots}$ of A satisfy $Ax_n = \sigma_n y_n$, $A^* y_n = \sigma_n x_n$, $n \in \mathbb{N}_0$. Keeping these facts in mind we are led to introduce the following examples of generating symbols of a regularization scaling function:

i) *Truncated singular value decomposition* (*bandlimited regularization*).

a) orthogonal:

$$(\Phi_J)^\wedge(n) = \begin{cases} \sigma_n^{-1/2} & \text{for} \quad n = 0,\ldots,N_J \\ 0 & \text{for} \quad n \geq N_J + 1 \end{cases}, \tag{4.233}$$

$$N_J = \begin{cases} 0 & \text{for} \quad J \in \mathbb{Z}, J < 0 \\ 2^J - 1 & \text{for} \quad J \in \mathbb{Z}, J \geq 0 \end{cases}, \tag{4.234}$$

b) non-orthogonal:

$$(\Phi_J)^\wedge(n) = \begin{cases} \sigma_n^{-1/2} & \text{for} \quad n = 0, \dots, M_J \\ \sigma_n^{-1/2}(\tau_J(n))^{1/2} & \text{for} \quad n = M_J + 1, \dots, N_J \\ 0 & \text{for} \quad n \geq N_J + 1 \end{cases}, \tag{4.235}$$

$$\begin{aligned} N_J &= \begin{cases} 0 & \text{for} \quad J \in \mathbb{Z}, J < 0 \\ 2^{J+1} - 1 & \text{for} \quad J \in \mathbb{Z}, J \geq 0 \end{cases}, \\ M_J &= \begin{cases} 0 & \text{for} \quad J \in \mathbb{Z}, J < 0 \\ 2^J - 1 & \text{for} \quad J \in \mathbb{Z}, J \geq 0 \end{cases} \end{aligned} \tag{4.236}$$

and

$$\tau_J(n) = 2 - 2^{-J}(n+1),\ n \in [2^J - 1, 2^{J+1} - 1],\ J \in \mathbb{N}_0. \tag{4.237}$$

It is easy to see that case a) leads to an orthogonal RMRA, i.e., the detail and the scale spaces satisfy the orthogonality conditions

$$\mathcal{V}_{J+1} = \mathcal{V}_J \oplus \mathcal{W}_J, \quad \mathcal{W}_J \perp \mathcal{W}_K,\ K \neq J,\ K, J \geq 0. \tag{4.238}$$

In case b) the scale and detail spaces are still finite dimensional, but the detail spaces are no longer orthogonal.

ii) *Tikhonov's regularization* (*non-bandlimited regularization*).

a) *classical*

$$(\Phi_J)^\wedge(n) = \left(\frac{\sigma_n}{\sigma_n^2 + \gamma_J^2} \right)^{\frac{1}{2}}, \quad n \in \mathbb{N}, J \in \mathbb{Z}, \tag{4.239}$$

b) *Tikhonov–Phillips*

$$(\Phi_J)^\wedge(n) = \left(\frac{\sigma_n}{\sigma_n^2 + \gamma_J^2 (n + \frac{1}{4})^4} \right)^{\frac{1}{2}}, \quad n \in \mathbb{N}, J \in \mathbb{Z} \tag{4.240}$$

with $\{\gamma_J\}$, $J \in \mathbb{Z}$, being a sequence of real numbers satisfying $\lim_{J \to \infty} \gamma_J = 0$ and $\lim_{J \to -\infty} \gamma_J = \infty$. Case a) leads to the minimization of the Tikhonov functional, where the penalty term is given by the norm of the regularized solution of problem (4.200). Case b) leads to the minimization of the Tikhonov–Phillips functional, where the penalty term is given by the linearized spherical bending energy of the regularized solution of problem (4.200). Both cases lead to infinite-dimensional scale and detail spaces and, furthermore, the RMRA is neither direct nor orthogonal.

4.4. Iterative methods as regularization schemes

From linear algebra we borrow the idea to use iterative solvers of the equation $Ax = y, y \in Y$, in the following way: Suppose that the matrix $A \in \mathbb{K}^{n \times n}$ is written in the form $A = B - C$ with an invertible B. Then we are able to write $Ax = y$, in equivalent form

$$Ax = y \Leftrightarrow (B - C)\, x = y \Leftrightarrow Bx = Cx + y \Leftrightarrow x = B^{-1}Cx + B^{-1}y. \tag{4.241}$$

Let us base the determination of a (uniquely determined) fixed point on the following iteration:

$$x^{(0)} \in \mathbb{K}^n, \quad \text{arbitrary},$$
$$Bx^{(n+1)} = Cx^{(n)} + y, \quad n \in \mathbb{N}_0.$$

Clearly, if $\left\{x^{(n)}\right\}$ converges to $x^* \in \mathbb{R}^n$, then it follows that $Ax^* = y$.

In numerical linear algebra the following procedures are convenient which should be recapitulated here: Let $A \in \mathbb{R}^{n\times n}$ be given. Let us decompose the matrix A in the form $A = L + D + U$ such that

$$L = \begin{pmatrix} 0 & & 0 \\ & \ddots & \\ * & & 0 \end{pmatrix} \quad D = \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix} \quad U = \begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix} \tag{4.242}$$

We distinguish the following concepts:

(1) (*Jakobi Method*) We let

$$B = D, \qquad C = -L - U. \tag{4.243}$$

Then we are led to

$$x^{(n+1)} = D^{-1}(L + U)\, x^{(n)} + D^{-1}y, \quad n \in \mathbb{N}. \tag{4.244}$$

(2) (*Gauss–Seidel Method*) We choose

$$B = D + L, \qquad C = -U. \tag{4.245}$$

Then we are led to

$$x^{(n+1)} = (D + L)^{-1} U x^{(n)} + (D + L)^{-1} y, \quad n \in \mathbb{N}. \tag{4.246}$$

(3) (*Richardson Method*) For $\omega > 0$, let

$$B = \frac{1}{\omega} I, \quad C = \frac{1}{\omega} I - A. \tag{4.247}$$

Then we are led to the recursion

$$\begin{aligned} x^{(n+1)} &= \left(\frac{1}{\omega} I\right)^{-1} \left(\frac{1}{\omega} I - A\right) x^{(n)} + \left(\frac{1}{\omega} I\right)^{-1} y \\ &= \omega \left(\frac{1}{\omega} I - A\right) x^{(n)} + \omega y \\ &= (I - \omega A)\, x^{(n)} + \omega y, \quad n \in \mathbb{N}. \end{aligned} \tag{4.248}$$

Since iteration methods known from linear algebra that use the upper or lower triangular part of a matrix cannot be adequately transferred to operator equations, we are not able to use the first two of the aforementioned iteration procedures. However, the third method (i.e., Richardson method or successive relaxation method) can be applied leading to a variant called Landweber iteration.

Landweber iteration. In order to solve the normal equation

$$A^* Ax = A^* y, \quad y \in Y,\ x \in X, \tag{4.249}$$

the *Landweber iteration* starting from the initial value $x^{(0)} \in X$ is defined by

$$\begin{aligned} x^{(m+1)} &= (I - \omega A^* A)\, x^{(m)} + \omega A^* y \\ &= x^{(m)} + \omega A^* \left(y - A x^{(m)} \right), \quad m = 0, 1, \ldots \end{aligned} \tag{4.250}$$

For simplicity, we introduce the family of operators $\{R_m\}_{m\in\mathbb{N}} \subset L(Y, X)$ by $R_m y = x^{(m)}$:

Let $0 < \omega < \frac{2}{\|A\|^2}$ and $A \in L(X, Y)$. Then, for $x^{(0)} \in X$, we have

$$\lim_{m\to\infty} R_m y = \begin{cases} A^\dagger y + P_{\mathcal{N}(A)} x^{(0)} & y \in D\left(A^\dagger\right) \\ \infty & y \notin D\left(A^\dagger\right) \end{cases} \tag{4.251}$$

The Landweber iteration is characterized by

$$x^{(m)} = R_m y = F_m(A^* A)\, A^* y + (I - \omega A^* A)^m\, x^{(0)}, \tag{4.252}$$

where

$$F_m(\lambda) = \omega \sum_{j=0}^{m-1} (1 - \omega\lambda)^j = \frac{1 - (1 - \omega\lambda)^m}{\lambda}. \tag{4.253}$$

Thus, the Landweber iteration represents a regularization with a filter that cannot be used as such if we do not know the singular values of our problem.

The polynomial $\lambda \mapsto p_m(\lambda) = 1 - \lambda F_m(\lambda) = (1 - \omega\lambda)^m$ is called *residual polynomial*.

- For $x^{(0)} = 0$ it follows that $A^\dagger y$ is the solution of the normal equation

$$A^* A x = A^* y \tag{4.254}$$

showing minimal norm (as usual). This result is also obtained for $x^{(0)} \in \mathcal{N}(A)^\perp$, since then $P_{\mathcal{N}(A)} x^{(0)} = 0$.

- For $x^{(0)} \neq 0$ we see that $A^\dagger y + P_{\mathcal{N}(A)} x^{(0)}$ is the solution of the normal equation

$$A^* A x = A^* y \tag{4.255}$$

with minimal distance to $x^{(0)}$, i.e., the $x^{(0)}$-minimum norm solution.

$$\left\| A^\dagger y + P_{\mathcal{N}(A)} x^{(0)} - x^{(0)} \right\| = \min\left\{ \left\| x - x^{(0)} \right\| : A^* A x = A^* y \right\} \tag{4.256}$$

Obviously, the choice of the initial value of the Landweber iteration for inverse problems is of great importance for the solution (see, e.g., [227]). We investigate the same choice for a noiseless and noisy right-hand side.

Let $A \in L(X, Y)$, $y, y^\varepsilon \in Y$ with $\|y - y^\varepsilon\|_Y < \varepsilon$. Let $\left\{x^{(m)}\right\}$ and $\left\{x^{(m),\varepsilon}\right\}$ be the Landweber iteration sequences with respect to y and y^ε corresponding to the same initial value $x^{(0)} = x^{(0),\varepsilon} \in X$ and the same parameter $\omega \in \left(0, \frac{2}{\|A\|^2}\right)$. Then

$$\left\| x^{(m)} - x^{(m),\varepsilon} \right\| \leq \varepsilon \begin{cases} \sqrt{2\omega}, & m = 1 \\ \sqrt{m\omega}, & m \geq 2. \end{cases} \tag{4.257}$$

The estimate (4.257) suggests that the parameter ω should be chosen as small as possible. However, we have seen that the convergence of the approximation error depends on the estimate $|1-\omega\lambda|<1$. As a consequence, for small ω the approximation error converges very slowly.

In more details, we separate the total error in the usual way

$$\left\|A^{\dagger}y-R_m y^{\varepsilon}\right\|=\underbrace{\left\|A^{\dagger}y-R_m y\right\|}_{\substack{\text{approximation}\\ \text{error}}}+\underbrace{\left\|R_m y-R_m y^{\varepsilon}\right\|}_{\substack{\text{data}\\ \text{error}}} \tag{4.258}$$

with $\left\|A^{\dagger}y-R_m y\right\| \overset{m\to\infty}{\longrightarrow} 0$ if $x^{(0)}=0$ and $\|R_m y-R_m y^{\varepsilon}\|\leq\sqrt{m\omega}\,\varepsilon$, i.e., the Landweber iteration is a regularization with a suitably chosen m. Our aim now is to find a stopping rule for m.

In fact, we obtain for the residual term

$$\begin{aligned}
Ax^{(m),\varepsilon}-y^{\varepsilon} &= A\left(\left(I-\omega A^*A\right)x^{(m-1),\varepsilon}+\omega A^*y^{\varepsilon}\right)-y^{\varepsilon}\\
&= Ax^{(m-1),\varepsilon}-\omega AA^*Ax^{(m-1),\varepsilon}+\omega AA^*y^{\varepsilon}-y^{\varepsilon}\\
&= \left(I-\omega AA^*\right)Ax^{(m-1),\varepsilon}-\left(I-\omega AA^*\right)y^{\varepsilon}\\
&= \left(I-\omega AA^*\right)\left(Ax^{(m-1),\varepsilon}-y^{\varepsilon}\right).
\end{aligned} \tag{4.259}$$

Under the choice $0<\omega<\frac{2}{\|A\|^2}$ we have

$$\begin{aligned}
\left\|Ax^{(m),\varepsilon}-y^{\varepsilon}\right\| &\leq \left\|I-\omega AA^*\right\|\cdot\left\|Ax^{(m-1),\varepsilon}-y^{\varepsilon}\right\|\\
&\leq \left\|Ax^{(m-1),\varepsilon}-y^{\varepsilon}\right\|,
\end{aligned} \tag{4.260}$$

i.e., the norms of the residuals are monotonously decreasing in m. Furthermore, if $y^{\varepsilon}\notin\mathcal{D}\left(A^{\dagger}\right)$, it follows that $Ax^{(m-1),\varepsilon}-y^{\varepsilon}\notin\mathcal{N}\left(A^*\right)$ (due to the fact that $y^{\varepsilon}\in\mathcal{D}(A^{\dagger})$).

If $0<\omega<\frac{2}{\|A\|^2}$ *and* $y\notin\mathcal{N}\left(A^*\right)$, *then*

$$\left\|\left(I-\omega AA^*\right)y\right\|<\|y\|. \tag{4.261}$$

If $y^{\varepsilon}\notin D\left(A^{\dagger}\right)$, then $\left\|Ax^{(m),\varepsilon}-y^{\varepsilon}\right\|<\left\|Ax^{(m-1),\varepsilon}-y^{\varepsilon}\right\|$, i.e., the residual is strictly monotonously decreasing. Hence, we are confronted with the typical dilemma of ill-posed problems that a small residual terms does not imply a small error. The monotonicity of the residual term suggests to use a discrepancy principle as a kind of "stopping rule". This observation goes back to [38]. In more detail, let $\tau>1$ be fixed. We are interested in determining $m^*\in\mathbb{N}_0$, such that

$$\left\|Ax^{(m^*),\varepsilon}-y^{\varepsilon}\right\|\leq\tau\varepsilon<\left\|Ax^{(m),\varepsilon}-y^{\varepsilon}\right\| \quad m=0,1,\ldots,m^*-1. \tag{4.262}$$

The discrepancy principle (4.262) seems to be particularly suitable for Landweber iteration, since a residual term larger than 2ε implies the monotonicity of the error:

Suppose that $A \in L(X,Y)$, $y \in \mathcal{R}(A)$ *and* $y^\varepsilon \in Y$ *with* $\|y - y^\varepsilon\| < \varepsilon$. *If* $\left\|Ax^{(m),\varepsilon} - y^\varepsilon\right\| > 2\varepsilon$ *and* $0 < \omega < \frac{1}{\|A\|^2}$, *then*

$$\left\|A^\dagger y - x^{(m+1),\varepsilon}\right\| < \left\|A^\dagger y - x^{(m),\varepsilon}\right\|, \quad m \in \mathbb{N}_0, \tag{4.263}$$

i.e., the error is also strictly monotonously decreasing.

In other words, as already announced the error decreases monotonously just like the residual as long as the residual stays larger than 2ε.

An upper bound for the number of iterations when using the discrepancy principle is as follows (cf. [227]):

Assume that $A \in L(X,Y)$, $y \in \mathcal{R}(A)$, *and* $y^\varepsilon \in Y$ *with* $\|y - y^\varepsilon\| < \varepsilon$. *Suppose that* $0 < \omega < \frac{1}{\|A\|^2}$. *The discrepancy principle* (4.262) *with* $\tau > 1$ *yields the stopping index* $m^* = m^*(\varepsilon, y^\varepsilon) \leq C_L \varepsilon^{-2}$ *with the constant* $C_L > 0$ *for the Landweber iteration.*

Note that this result does not require any assumptions on the smoothness of the solution. With such information the discrepancy principle enables us to stop the iteration much earlier.

Suppose that $A \in L(X,Y)$, $0 < \omega < 2/\|A\|^2$, $y \in \mathcal{R}(A)$, $y \in Y$ *with* $\|y - y^\varepsilon\| < \varepsilon$ *and* $x^{(0)} = 0$. *Then the Landweber iteration together with the discrepancy principle* (4.262) *is an order optimal regularization of* $A^\dagger$, *i.e., the Landweber iteration possesses infinite qualification. The stopping index can be estimated as follows*

$$m^* = m^*(\varepsilon, y^\varepsilon) \leq C_\mu \varepsilon^{\frac{-2}{\mu+1}} \tag{4.264}$$

with $C_\mu > 0$.

If we choose as an initial value $0 \neq x^{(0)}$ with $x^{(0)} \in \mathcal{N}(A)^\perp$, it suffices to study the convergence of the sequence $\{\tilde{x}^{(m),\varepsilon}\}$ with $\tilde{x}^{(0),\varepsilon} = 0$ which results from the Landweber method applied to the equation $Ax = y^\varepsilon - Ax^{(0)}$. The minimum norm solution of this equation is $A^\dagger y - x^{(0)}$ and if $A^\dagger y \in X_\mu$, we also need to have that $x^{(0)} \in X_\mu$ to obtain the optimal order of decay for the error, i.e., $O(\varepsilon^{\mu/(\mu+1)})$.

Unfortunately, for unknown μ we have to choose 0 as starting value.

The disadvantage of the Landweber iteration is that its convergence is rather slow, i.e., the stopping index m^* is often large. This is the reason why semi-iterative methods (see, e.g., [227] and the references therein) come into play to accelerate the convergence.

Semi-iterative methods. The characteristics of Landweber iteration are as follows;

$$x^{(0)} = 0 \tag{4.265}$$

and

$$x^{(m)} = F_m(A^*A)\, A^* y \tag{4.266}$$

with $F_m(\lambda)$ given by (4.253), i.e.,

$$F_m(\lambda) = \omega \sum_{j=0}^{m-1} (1-\omega\lambda)^j = \frac{1-(1-\omega\lambda)^m}{\lambda}. \tag{4.267}$$

For $\lambda \in [0, \|A\|^2]$ we have

$$F_m(\lambda) \to \frac{1}{\lambda}, \qquad m \to \infty. \tag{4.268}$$

Moreover, we are able to show that

$$\sup_{\lambda\in[0,\|A\|^2]} \lambda|F_m(\lambda)| = \sup_{\lambda\in[0,\|A\|^2]} |1-(1-\omega\lambda)^m| \le 2, \tag{4.269}$$

so that $\{F_m\}_{m\in\mathbb{N}}$ is a regularizing filter. The *residual polynomials* corresponding to the *filter polynomials* are of degree m:

$$p_m(\lambda) = 1 - \lambda F_m(\lambda) = (1-\omega\lambda)^m. \tag{4.270}$$

All in all, the Landweber iteration procedure is as follows:

$$y - Ax^{(m)} = p_m(AA^*)\, y. \tag{4.271}$$

In order to accelerate the Landweber iteration we are led to the idea (see, e.g., the monograph [227]), to replace the polynomial filter by another filter family, that shows a faster convergence to $1/\lambda$.

To this end we consider an alternative polynomial F_m of degree $m-1$, so that its residual polynomial $\lambda \mapsto p_m(\lambda) = 1 - \lambda F_m(\lambda)$ is of degree m. Letting $x^{(m)} := F_m(A^*A)\, A^*y$ we obtain as residuum $y - Ax^{(m)} = p_m(AA^*)\, y$. Now, if $\{F_m\}_{m\in\mathbb{N}}$ is a regularizing filter, the corresponding family $\{p_m\}_{m\in\mathbb{N}}$ has the following properties:

If $\{F_m\}$ is a regularizing filter, the residual polynomials $\{p_m\}$ satisfy the following properties:

(i) $\lim_{m\to\infty} p_m(\lambda) = 0$ *for* $\lambda \in [0, \|A\|^2]$.

(ii) $\{p_m\}$ *is uniformly bounded on the interval* $\left[0, \|A\|^2\right]$ *(by* $1 + C_F$*).*

(iii) $p_m(0) = 1$ *for all* m.

Conversely, if $\{p_m\}$ is a family of polynomials satisfying the properties (i), (ii) (iii)*, then*

$$F_m(\lambda) = \frac{1-p_m(\lambda)}{\lambda}, \quad m \in \mathbb{N}_0 \tag{4.272}$$

constitutes a regularizing filter.

From the theory on special functions of mathematical physics" (see, e.g., [73]) we borrow the following result:

Let the family $\{p_m\}$ fulfill the following conditions:

1. p_m is a polynomial of degree m on the interval $\left[0, \|A\|^2\right]$.
2. $\int_0^{\|A\|^2} p_m(\lambda)\, p_n(\lambda)\, w(\lambda)\, d\lambda = 0$ if $n \neq m$, where $w(\lambda) > 0$ for $\lambda \in \left[0, \|A\|^2\right]$ and w is piecewise continuous.
3. $p_m(0) = 1$, $m \in \mathbb{N}_0$.

Then there exist $A_m, B_m \in \mathbb{R}$ satisfying

$$p_m(\lambda) = p_{m-1}(\lambda) + A_m\left(p_{m-1}(\lambda) - p_{m-2}(\lambda)\right) - B_m \lambda p_{m-1}(\lambda). \tag{4.273}$$

The polynomials $\{p_m\}$ are *orthogonal polynomials* on the interval $\left[0, \|A\|^2\right]$ with respect to the weight function $w(\lambda)$. Note that $w(\lambda)$ induces a measure on $\left[0, \|A\|^2\right]$. In transition to iteration we are led to

$$x^{(0)} = 0, \tag{4.274}$$

$$x^{(1)} = F_1(A^*A)\, A^* y \quad \text{with } F_1(\lambda) = \frac{1 - p_1(\lambda)}{\lambda}, \tag{4.275}$$

$$x^{(m)} = x^{(m-1)} + A_m\left(x^{(m-1)} - x^{(m-2)}\right) + B_m A^*\left(y - Ax^{(m-1)}\right) \quad \text{for } m \geq 2. \tag{4.276}$$

This is the reason why methods of type (4.274), (4.275), (4.276) are called semi-iterative (note that, for each iteration step, two previous iteration values are required).

For simplicity, assume now that $\|A\| \leq 1$ which can be achieved by scaling the operator. As a consequence, we only need to consider polynomials on the interval $[0, 1]$. Then the following results are known for semi-iterative methods (see, e.g., [227]):

(1) *Let $\{p_m\}$ be a sequence of residual polynomials, so that it is uniformly bounded on $[0, 1]$, normalized by $p_m(0) = 1$, and it converges pointwise to 0 on $(0, 1]$. If $\{F_m\}$ is the corresponding sequence of filters, then for $A \in L(X, Y)$ and $x^{(0)} \in X$ it follows that*

$$\begin{aligned} \lim_{m\to\infty} x^{(m)} &= \lim_{m\to\infty} \left(x^{(0)} + F_m(A^*A)A^*(y - Ax^{(0)})\right) \\ &= \begin{cases} A^\dagger y + P_{\mathcal{N}(A)} x^{(0)} & : \quad y \in \mathcal{D}(A^\dagger), \\ \infty & : \quad y \notin \mathcal{D}(A^\dagger). \end{cases} \end{aligned}$$

(2) *Suppose that A is of class $L(X, Y)$. Let $\{x^{(m)}\}$, $\{x^{(m),\varepsilon}\}$ be the iterates of a semi-iterative method with respect to $y \in R(A)$ and $y^\varepsilon \in Y$ using the same initial value. Let the residual polynomials of the semi-iterative method be uniformly bounded by $C_p > 0$. Then, for the data error, we have*

$$\left\|x^{(m)} - x^{(m),\varepsilon}\right\| \leq 2C_p m\varepsilon. \tag{4.277}$$

Together with a stopping rule that fulfills

$$m^*(\varepsilon) \to \infty, \qquad \varepsilon m^*(\varepsilon) \to 0 \text{ for } \varepsilon \to 0 \tag{4.278}$$

the semi-iterative method is a regularization of $A^\dagger$.

In order to determine the speed of convergence we investigate

$$\omega_\mu(m) = \sup_{0\le\lambda\le 1} \lambda^{\mu/2}|p_m(\lambda)| \tag{4.279}$$

(3) *Each sequence of polynomials* $\{p_m\}_{m\in\mathbb{N}_0}$, $p_m(0) = 1$, *which satisfies the best asymptotic behavior*

$$\omega_\mu(m) = O(m^{-\mu}) \quad \text{for } m \to \infty \tag{4.280}$$

for some $\mu > 0$, *is uniformly bounded on* $[0,1]$ *and converges pointwise to* 0 *on* $(0,1]$. *In other words* $\{p_m\}_{m\in\mathbb{N}_0}$ *given in such a way is a sequence of residual polynomials, for which, in addition,* $\omega_\alpha(m) = O(m^{-\alpha})$ *for* $0 < \alpha \le \mu$.

The discrepancy principle (cf. (4.262)) can also be used as stopping rule.

(4) *Let* $A \in L(X,Y)$ *and* $y \in \mathcal{R}(A)$. *Let the normalized polynomials* $\{p_m\}_{m\in\mathbb{N}_0}$, $p_m(0) = 1$, *satisfy* (4.280) *for some* $\mu > 1$. *Then the corresponding semi-iterative method with starting value* $x^{(0)} = 0$ *is an order optimal regularization of* $A^\dagger$ *with respect to* X_α *for* $0 < \alpha \le \mu - 1$ *if it is combined with the discrepancy principle* (4.262) *as stopping rule with* $\tau > \sup\{\|p_m\|_{C[0,1]} \,|m \in \mathbb{N}_0\} \ge 1$.
The stopping index satisfies

$$m^* = m^*(\varepsilon, y^\varepsilon) = O\left(\varepsilon^{-1/(\alpha+1)}\right) \tag{4.281}$$

for $\varepsilon \to 0$.

Normalized polynomials that fulfill (4.280) automatically lead to semi-iterative order optimal regularization methods. The reason for this is that (4.280) implies the uniform boundedness of the polynomials on $[0,1]$ as well as

$$\sup\{|F_m(\lambda)| : \lambda \in [0,1]\} \le 2\tau m^2$$

for the corresponding filters.

In the general case that $p_t(\lambda) = 1 - \lambda F_t(\lambda)$ we are confronted with the situation (see, e.g., [227]) that

$$\omega_\mu(t) \le C_p t^{\,\mu/2} \quad \text{for } t \to 0 \tag{4.282}$$

which is the analogue of (4.280). However, this neither implies the uniform boundedness of $\{p_t\}_{t>0}$ nor an estimate like $|F_t(\lambda)| \le C_F t^{-\alpha}$.

Gradient method. The method of successive approximation gradient, and related iterative methods can be used for finding approximate solutions of ill-posed problems (see, e.g., [108, 133, 134, 172, 179] and the references therein for more details).

We let X and Y be two Hilbert spaces, both over $\mathbb{K}$, and let A be a bounded linear operator on X into Y. As already known, the linear equation

$$Ax = y, \quad y \in Y \tag{4.283}$$

may or may not have a solution depending on whether or not y is in $\mathcal{R}(A)$, the range of A, and even if $y \in \mathcal{R}(A)$ the solution of (4.283) need not be unique. For any bounded linear operator $A : X \to Y, \mathcal{R}(A)$ and $\mathcal{R}(A^*)$ are closed subspaces of X and Y, respectively, hence, $X = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$ and $Y = \mathcal{R}(A^*) \oplus \mathcal{R}(A^*)^\perp$. The

relations $\overline{\mathcal{R}(A)} = \mathcal{R}(A^*)^\perp$, $\overline{\mathcal{R}(A^*)} = \mathcal{R}(A)^\perp$, $\mathcal{R}(A)^\perp = \mathcal{R}(A^*), \overline{\mathcal{R}(A)} = \overline{\mathcal{R}(AA^*)}$ are also valid (see, for instance, [265]). Let P denote the orthogonal projection of X onto $\overline{\mathcal{R}(A^*)}$ and let Q denote the orthogonal projection of A on $\overline{\mathcal{R}(A)}$. Then $Ax = APx$ for all $x \in X, A^*y = A^*Qy$ for each $y \in Y$, and the restriction of A to $\mathcal{R}(A)^\perp$ has an inverse, which is not necessary.

In either case, i.e., $y \in \mathcal{R}(A)$ unique or non-unique solution of (4.283), one can seek a best approximate solution, i.e., a solution which minimizes the quadratic function $J(x) = \|Ax - y\|^2$. Such a solution always exist for all $y \in Y$ if $\mathcal{R}(A)$ is closed. If $\mathcal{R}(A)$ is arbitrary, a best approximation does not exist for all $y \in Y$, however, it does exist for all $y \in \mathcal{R}(A)^\perp$.

We consider the conjugate gradient (CG) method (cf. [133, 134]) that minimizes $J(x) = \|Ax - y\|^2$ at each step. That is, choose an initial vector $x_0 \in X$, then compute $r_0 = p_0 = A^*(Ax_0 - y)$, where A^* is the adjoint of A. If $p_0 \neq 0$, compute $x_1 = x_0 - \alpha_0 p_0$, where $\alpha_0 = \|r_0\|^2/\|Ap_0\|^2$. For $i = 1, 2, \ldots$, compute

$$r_i = A^*(Ax_i - y) = r_{i-1} - \alpha_{i-1} A^* A p_{i-1}, \tag{4.284}$$

where

$$\alpha_{i-1} = \frac{\langle r_{i-1}, p_{i-1}\rangle}{\|Ap_{i-1}\|^2}, \tag{4.285}$$

and if $r_i \neq 0$, then compute

$$p_i = r_i + \beta_{i-1} p_{i-1}, \tag{4.286}$$

where

$$\beta_{i-1} = -\frac{\langle r_i, A^* A p_{i-1}\rangle}{\|Ap_{i-1}\|^2}. \tag{4.287}$$

Set

$$x_{i+1} = x_i - \alpha_i p_i. \tag{4.288}$$

We examine some properties of the CG algorithm, for the case of an arbitrary bounded linear operator. To be more concrete, the domain of the generalized inverse of A is $\mathcal{D}(A^\dagger) = \mathcal{R}(A) + \mathcal{R}(A)^\dagger$. If $y \in \mathcal{D}(A^\dagger)$, then $Qy = \bar{y}$ is in the range of T and $v = A^\dagger y = A^\dagger \bar{y}$ and $\overline{y} = Qy = Av = AA^\dagger y$. Since Q is an orthogonal projection, the functional J can be written as $J(x) = \|Ax - y\|^2 = \|Ax - \bar{y}\|^2 + \|\bar{y} - y\|^2$. Thus, minimizing J is equivalent to minimizing the functional $\|Ax - \bar{y}\|^2$ which we denote by $g(x)$. Setting

$$u = v + (I - P)x_0 = A^\dagger y + (I - P)x_0 \tag{4.289}$$

one can define the error vector $e = x - u$ and the vector $r = A^*(Ax - y) = A^*(Ax - \bar{y})$. Then

$$(A^*A)e = r \tag{4.290}$$

and

$$\langle r, e\rangle = \|Ax - \bar{y}\|^2 = g(x). \tag{4.291}$$

The sequence of iterates $\{x_i\}$ generated by the CG method (4.284)–(4.288) is contained in the flat $x_0 + \mathcal{R}(A^*)$ with both r_i and p_i, for $i = 0, 1, 2, \ldots$, in $\mathcal{R}(A^*)$.

Moreover, $p_0, p_1, \ldots, p_{i-1}$ form an A^*A-orthogonal set of vectors and their span is an i-dimensional subspace of $\mathcal{R}(A^*)$. If at the ith step, $r_i = 0$, then both $Ax_i - y$ and $Ax_i - \bar{y}$ are vectors in $\mathcal{R}(A^*) = \overline{\mathcal{R}(A)^{\perp}}$. However, Ax_i and $\bar{y}$ are also in $\overline{\mathcal{R}(A)}$, and therefore $tx_i = \bar{y}$, implying that $g(x_i) = 0$. In this case, the iteration terminates at the ith step and we have

$$x_i = x_0 - \sum_{k=0}^{i-1} \alpha_k p_k = u, \tag{4.292}$$

as well as

$$A^{\dagger} y = P x_0 - \sum_{k=0}^{i-1} \alpha_i p_k. \tag{4.293}$$

Therefore, unless explicitly mentioned otherwise, we shall assume that the CG method does not terminate in a finite number of steps, that is $r_i \neq 0$ for $i = 0, 1, \ldots$. We list some known identities for the CG method.

For indices satisfying $k = 0, 1, 2, \ldots$, i and $i = 0, 1, 2, \ldots$, we have

$$\langle r_i, r_k \rangle = \langle p_i, r_k \rangle, \tag{4.294}$$

$$\|Ap_i\| \leqq \|Ar_i\|, \tag{4.295}$$

$$\beta_i = \frac{\|r_{i+1}\|^2}{\|r_i\|^2}, \tag{4.296}$$

$$\langle p_i, p_k \rangle = \frac{\|r_i\|^2 \|p_k\|^2}{\|r_k\|^2}, \tag{4.297}$$

$$p_i = \|r_i\|^2 \sum_{j=0}^{i} \frac{r_j}{\|r_j\|^2} \tag{4.298}$$

and

$$\|r_i\|^2 \leqq \|p_i\|. \tag{4.299}$$

Setting $g(x_i) = \langle r_i, e_i \rangle = \|Ax_i - \bar{y}\|^2$, where $e_i = x_i - u$, one finds that

$$g(x_i) - g(x_{i+1}) = \alpha_i \|r_i\|^2. \tag{4.300}$$

Finally, x_i minimizes the functionals $J(x)$ and $y(x)$ on the i-dimensional flat

$$x_0 + \operatorname{span}\{p_0, p_1, \ldots, p_{i-1}\}. \tag{4.301}$$

It is worth mentioning the following three properties (see [134]):

(i) For $k = 0, 1, 2, \ldots, i$

$$g(x_i) = \langle e_i, r_k \rangle = \langle r_i, e_k \rangle. \tag{4.302}$$

For $i = 0, 1, 2, \ldots$,

$$\langle p_i, e_i \rangle \|r_i\|^2 = g(x_i) \|p_i\|^2. \tag{4.303}$$

(ii) The inequality

$$\|e_{i+1}\|^2 \leqq \|e_i\|^2 - \alpha_i g(x_i) \tag{4.304}$$

holds for $i = 0, 1, 2, \ldots$.

(iii) For any non-negative integers i and j, both $\langle p_i, e_i\rangle$ and $\langle e_i, e_j\rangle$ are non-negative numbers.

The main result on the CG method for bounded linear operators with closed range depends heavily upon the following observation:

Let X and Y be two Hilbert spaces over the same field and let A be a bounded linear transformation mapping X into Y. If $\mathcal{R}(A)$ is closed, then $S = A|\mathcal{R}(A^*)$, the restriction of A to $\mathcal{R}(A^*)$, has a bounded inverse and $U = A^*A|\mathcal{R}(A^*)$ is a positive definite operator onto $\mathcal{R}(A^*)$.

Let $\mu(x) := \langle Ux, x\rangle/\langle x, x\rangle, x \neq 0$, be the *Rayleigh quotient of* U. Since U is a bounded symmetric positive definite linear operator on the Hilbert space $\mathcal{R}(A^*)$, the spectral bounds

$$m = \inf\{\mu(x) : x \in \mathcal{R}(A^*)\} \tag{4.305}$$

and

$$M = \sup\{\mu(x) : x \in \mathcal{R}(A^*)\} \tag{4.306}$$

are positive and finite.

Let X and Y be two Hilbert spaces over the real field and let A be a bounded linear transformation mapping X into Y. If the range of A is closed then the conjugate gradient method (4.284)–(4.288) *converges monotonously to the least squares solution $u = A^\dagger y + (I - P)x_0$ of $Ax = y$. Moreover, if m and M are the spectral bounds of $U = \{A^*A|\mathcal{R}(A^*)\}$, then*

$$\|x_i - u\|^2 \leqq \frac{g(x_0)}{m}\left(\frac{M-m}{M+m}\right)^{2i}, \quad i = 0, 1, 2, \ldots. \tag{4.307}$$

Altogether, let X and Y be two Hilbert space over the same field. If A is a linear transformation mapping X into Y of rank r, then the conjugate gradient method associated with the system $Ax = y$ converges in at most r steps to the least squares solution $u = A^\dagger y + (I - P)x_0$.

When $\mathcal{R}(A^*)$ is not closed, then the operator $U = A|\mathcal{R}(A^*)$ need not be positive definite, and therefore the generalization of an inequality cannot longer be utilized. However, it is still possible to establish convergence of the CG method under mild restrictions.

Let X and Y be two Hilbert spaces over the real field, and let A be a bounded linear operator mapping X to Y. If $Qy \in \mathcal{R}(AA^*A)$, then the conjugate gradient method (4.284)–(4.288)*, with initial value $x_0 \in \mathfrak{R}(A^*A)$, converges monotonously to the least squares solution of minimal norm $u = A^\dagger y$. In fact,*

$$\|x_i - u\|^2 \leqq \frac{\|A\|^2\|x_0 - A^\dagger y\|^2\|A^{*\dagger}x_0 - (AA^*)^\dagger\|^2}{\|A\|^2\|A^{*\dagger}x_0 - (AA^*)^\dagger y\|^2 + i\|x_0 - (AA^*)^\dagger y\|^2}, \quad i = 1, 2, \ldots. \tag{4.308}$$

4.5. Stochastic regularization methods

Let X be a Hilbert space. Our statistical approach starts from an equation of type (see, e.g., [101, 148])

$$Ax = y + \eta \tag{4.309}$$

where x, y, η are considered as values of jointly distributed random variables.

Random variables. On the probability space Ω equipped with the probability measure P we understand $\xi : \Omega \to X$ as Hilbert space-valued random variable. If X is a function space, then we denote by ξ a *stochastic process*. For $x \in X$ we obtain by $\xi_x = \langle x, \xi \rangle$ a real-valued random variable. Hence, for a complete orthogonal system $\{x_i\}$ in X, we are able to introduce by $\xi^{(i)} = \xi_{x_i} = \langle x_i, \xi \rangle$ an infinite number of jointly distributed random variables.

In what follows we suppose that ξ has a *vanishing expectation value*, i.e., we have

$$E[\langle x, \xi \rangle] = 0 \tag{4.310}$$

for all $x \in X$ (note that (4.310) does not mean any restriction, if (4.310) is violated we are allowed to go over to the random variable $\tilde{\xi} = \xi - E[\xi]$). Furthermore, we assume that the random variable has a finite second moment, so that

(i) $E[|\langle x, \xi \rangle|^2] < \infty$ for all $x \in X$,

(ii) the expectation value is continuous at x.

Then it follows that $E[\langle x, \xi \rangle \langle \xi, z \rangle]$ is a continuous, symmetric, non-negative bilinear form on X, hence, there exists a linear, continuous, selfadjoint, non-negative operator $R_{\xi\xi} : X \to X$ satisfying

$$\langle R_{\xi\xi} x, z \rangle = \mathrm{Cov}(x, z) = E[\langle x, \xi \rangle \langle \xi, z \rangle], \tag{4.311}$$

$R_{\xi\xi}$ is called *covariance operator*.

Example. Let $\mathcal{G}$ be a regular region. Assume that $X = L^2(\mathcal{G})$ and $\xi(\alpha)$ is defined for $\alpha \in \mathcal{G}$. Then we are able to identify the covariance operator with the covariance function resulting in the *autocovariance function* given by

$$R_{\xi\xi}(\alpha, \beta) = \mathrm{Cov}(\xi(\alpha), \xi(\beta)) = E[\xi(\alpha) \cdot \xi(\beta)]. \tag{4.312}$$

Application to an $L^2(\mathcal{G})$-function x yields the identity

$$R_{\xi,\xi} x(\alpha) = \int_{\mathcal{G}} R_{\xi\xi}(\alpha, \beta) x(\beta) \, d\beta. \tag{4.313}$$

As *white noise* we denote the Gaussian process characterized by

$$R_{\xi\xi}(\alpha, \beta) = \eta^2 \, \delta(\alpha - \beta), \tag{4.314}$$

i.e.,

$$R_{\xi\xi} = \eta^2 I. \tag{4.315}$$

Let ξ, η be jointly distributed random variables with $\xi : \Omega \to X, \zeta : \Omega \to Y$. Then we define the *cross-covariance operator* by $R_{\xi\zeta} : Y \to X$ as follows:

$$\langle R_{\xi\zeta} y, x \rangle = E\left[\langle y, \zeta \rangle_Y \langle x, \xi \rangle_X\right]. \tag{4.316}$$

Best linear estimator and Tikhonov–Phillips regularization. Next we have a look at the equation

$$A\xi = \beta + \zeta, \quad \xi : \Omega \to X, \quad \beta, \zeta : \Omega \to Y. \tag{4.317}$$

We assume that the inverse A^{-1} exists. The problem is to estimate ξ under the knowledge of β and ζ. To this end, we suppose that $E[\xi] = E[\beta] = 0$ such that $R_{\xi\zeta} = 0$ (i.e., ξ and ζ are uncorrelated) and $R_{\zeta\zeta}^{-1}$ exists. It is not difficult to show that (see, e.g., [101, 148])

$$R_{\beta\beta} = AR_{\xi\xi}A^* + R_{\zeta\zeta}, \tag{4.318}$$

$$R_{\xi\beta} = R_{\xi\xi}A^*. \tag{4.319}$$

In order to realize a least squares estimation we have to consider a *linear estimator* of ξ, i.e., a random variable $\xi_L = L\beta$, where $L : Y \to X$ is a linear and continuous operator such that $x_L = Ly$ is a solution of the operator equation. Central in our considerations is the following statement, that is standard in statistical geodesy (see, e.g., [101] and the references therein):

If $R_{\zeta\zeta}$ is assumed to have a continuous inverse, then

$$L = R_{\xi\beta}R_{\beta\beta}^{-1} = R_{\xi\xi}A^*(AR_{\xi\xi}A^* + R_{\zeta\zeta})^{-1} \tag{4.320}$$

minimizes the functional $E\left[|\langle x, \xi - L\beta\rangle|^2\right]$ for all x.

Under the simplifying assumptions

$$R_{\xi\xi} = I, \quad R_{\zeta\zeta} = \eta^2 I$$

we obtain for (4.320)

$$L = A^*(AA^* + \eta^2 I)^{-1} = (A^*A + \eta^2 I)^{-1}A^*. \tag{4.321}$$

In other words, in similarity to the Tikhonov–Philipps regularization, we are led to normal equations in order to determine the best linear estimator.

For arbitrary covariance operators we obtain $x = Ly$ by minimizing

$$\langle R_{\zeta\zeta}^{-1}(Ax - y), Ax - y\rangle + \langle R_{\xi\xi}^{-1}x, x\rangle = |||Ax - y|||^2 + ||x|| \tag{4.322}$$

with

$$|||y||| = \langle R_{\zeta\zeta}^{-1}y, y\rangle \tag{4.323}$$

and

$$\|x\| = \langle R_{\xi\xi}^{-1}x, x\rangle \ . \tag{4.324}$$

Observing these facts we are finally able to come to the following conclusion:

The best linear estimator is a special Tikhonov–Phillips regularization method, or, *the Tikhonov–Phillips regularization method is a special linear estimator.*

The stochastic approach also allows an interpretation of Bayes estimation as Tikhonov–Phillips regularization (for more details the reader is referred, e.g., to [101, 148] and the references therein).

4.6. Mollifier methods

The original idea of the mollifier method can be characterized as follows: We are interested in the solution $x^\dagger$ of $Ax = y$, but we realize that the problem is "too ill-posed" for being able to determine $x^\dagger$ accurately. Thus, we compromise by changing the problem into a more well-posed one, namely that of trying to determine a mollified version $E_\rho x^\dagger$ of the solution, where E_ρ is a suitable "mollification operator" depending on a parameter ρ. The heuristic motivation is that the trouble usually comes from high frequency components of the data and of the solution, which are damped out by mollification (which, in fact, defines mollification).

In abstract nomenclature, early mollifier methods for ill-posed problems were studied in [148]. Our approach is based on [63]. First we recapitulate the essential ingredients. Again we start from the operator equation

$$Ax = y,\ x \in X,\ y \in Y \tag{4.325}$$

with X, Y Hilbert spaces and $\mathcal{R}(A)$ non-closed. Our aim is to introduce operators $E_\rho : X \to X$ such that

$$E_\rho x \to x, \quad x \in X, \quad \rho \to 0. \tag{4.326}$$

IF X is a suitable function space, we are able to represent E_ρ by a *mollifier* e_ρ via the equation

$$E_\rho(x)(s) = \langle e_\rho(s, \cdot), x\rangle_X. \tag{4.327}$$

Instead of $x^\dagger$ we now look for $E_\rho x^\dagger$ for some $\rho > 0$, thereby assuming that e_ρ has a representation

$$A^* v_s^\rho = e_\rho(s, \cdot) \tag{4.328}$$

with $v_s^\rho \in Y$. Then, if $Ax^\dagger = y$, we can compute $E_\rho x^\dagger$ as follows:

$$(E_\rho x^\dagger)(x) = \langle e_\rho(s, \cdot), x^\dagger\rangle_X = \langle A^* v_s^\rho, x^\dagger\rangle_X = \langle v_s^\rho, Ax^\dagger\rangle_Y = \langle v_s^\rho, y\rangle_Y, \tag{4.329}$$

i.e.,

$$(E_\rho x^\dagger)(s) = \langle v_s^\rho, y\rangle_Y. \tag{4.330}$$

Hence, the problem of solving (4.325) reduces to that of solving (4.328), which is also ill posed as soon as (4.325) is. However, the right-hand side of (4.328) (which is actually a family of equations depending on the parameter s) is usually given analytically, since the mollifier e_ρ is chosen. Hence, there is no (or much less) error in the data of (4.328), and these equations can be solved (by regularization) much better than (4.325). As soon as an approximation for v_s^ρ has been computed, it can be used to solve (4.325) for any right-hand side y via (4.330). If we define the operator $S_\rho : Y \to X$ via the estimate

$$(S_\rho y)(s) = \langle v_s^\rho, y\rangle, \tag{4.331}$$

then, by (4.330), this operator maps the right-hand side of (4.325) to mollified solutions. This motivates the term *approximate inverse* of A used for S_ρ, also for the more general case that (4.328) is not solvable. In this case, (4.328) is replaced by

$$\|A^* v_s^\rho - e_\rho(s, \cdot)\|_X \to \min, \tag{4.332}$$

which, is equivalent to

$$AA^* v_s^\rho = Ae_\rho(s,\cdot). \tag{4.333}$$

Note that one needs the requirement that (4.333) is solvable. The function v_s^ρ is called *reconstruction kernel*; uniqueness can be enforced by solving (4.333) in the best-approximate sense, i.e., by selecting the solution of (4.333) with minimal norm: $v_s^\rho = (A^*)^\dagger e_\rho(s,\cdot)$. Assume, for simplicity, that $\mathcal{R}(A)$ is dense in Y, so that $(AA^*)^{-1}$ exist. Then we have with v_s^γ defined by (4.332):

$$\begin{aligned} S_\rho y &= \langle (AA^*)^{-1} Ae_\rho(s,\cdot), y\rangle_Y = \langle e_\rho(s,\cdot), A^*(AA^*)^{-1} y\rangle_X \\ &= \langle e_\rho(s,\cdot), (A^*A)^\dagger A^* y\rangle_X = (E_\rho x^\dagger)(s), \end{aligned} \tag{4.334}$$

i.e., $S_\rho y$ is the mollified version of the best-approximate solution of (4.330). This justifies (4.332).

Let A be compact with singular system $(\sigma_n; x_n, y_n)$. Let a regularization method realized in standard way, i.e.,

$$x_\alpha = \sum_{n=1}^{\infty} \sigma_n F_\alpha(\sigma_n^2) \langle y, y_n\rangle_Y y_n. \tag{4.335}$$

If we assume that X and Y are suitable function spaces, then (4.335) can be written as

$$x_\alpha(s) = \langle v_s^\rho, y\rangle_Y \tag{4.336}$$

with

$$v_s^\rho(t) = \sum_{n=1}^{\infty} \sigma_n F_\alpha(\sigma_n^2) y_n(t) y_n(s). \tag{4.337}$$

Now, v_s^ρ can be written in the form (4.333) with

$$e_\rho(s,t) = \sum_{n=1}^{\infty} \sigma_n^2 F_\alpha(\sigma_n^2) y_n(s) y_n(t). \tag{4.338}$$

Hence, x_α can be considered as a mollified solution $E_\rho x^\dagger$ with E_ρ given (in the sense of (4.326)) by the mollifier (4.338), so that linear regularization methods can also be viewed as mollifier methods.

The underlying "suitable function space" have to be such that point evaluation is continuous wherever used. The choice of the mollifier, of course, depends on what one wants to achieve; frequently used choices are singular integral kernels (such as Haar kernel, sinc kernel, etc.). Note again that each of these mollifiers can be applied only to equations, so that (4.333) admits a solution.

4.7. Backus–Gilbert method

The Backus–Gilbert method (cf. [18]) treats moment problems of the type

$$\langle x, k_i\rangle_X = \beta_i, \quad i \in \{1,\dots,n\}, \tag{4.339}$$

with given elements $k_i \in X$, for example $X = L^2(\mathcal{G})$, $\mathcal{G}$ regular region, so that (4.339) takes the form

$$\int_{\mathcal{G}} k_i(t)x(t)\ dt = \beta_i, \quad i \in \{1, \ldots, n\}, \tag{4.340}$$

which can be thought of as resulting from discretizing an integral equation of the first kind

$$\int_{\mathcal{G}} k(s,t)x(t)\ dt = \beta(s) \tag{4.341}$$

by collocating at points $s_1, \ldots, s_n$, so that $k_i(t) = k(s_i, t)$, $\beta_i = \beta(s_i)$.

With $Ax = (\langle x, k_1\rangle_X, \ldots, \langle x, k_n\rangle_X)^T$, $\beta = (\beta_1, \ldots, \beta_n)$, (4.339) can be written in the form $Ax = \beta$ with $X = L^2(\mathcal{G})$, $Y = \mathbb{R}^n$. In the Backus–Gilbert method, one looks for an approximate inverse $S : \mathbb{R}^n \to L^2(\mathcal{G})$ for A by defining

$$Sy = \sum_{i=1}^{n} y_i v_i, \tag{4.342}$$

with functions $v_i \in L^2(\mathcal{G})$ to be determined as follows: since

$$SAx = \sum_{i=1}^{n} \langle x, k_i\rangle_X v_i = \left\langle x, \sum_{i=1}^{n} k_i v_i \right\rangle, \tag{4.343}$$

i.e., for the concrete case (4.340)

$$(SAx)(s) = \int_{\mathcal{G}} x(t) \left[\sum_{i=1}^{n} k_i(t)v_i(s)\right]\ dt, \tag{4.344}$$

one should aim at determining the functions v_i such that

$$\sum_{i=1}^{n} k_i(t)v_i(s) \sim \delta(|s-t|). \tag{4.345}$$

The question is how to formalize the requirement (4.345). In the classical approach to the Backus–Gilbert method [18], this is done by minimizing, for any fixed $s \in \mathcal{G}$ and some chosen τ, the functional

$$(v_1, \ldots, v_n) \mapsto \int_{\mathcal{G}} |s-t|^{2\tau} \left|\sum_{k=1}^{n} k_i(t)v_i\right|^2\ dt. \tag{4.346}$$

Under the normalization constraint

$$\int_{\mathcal{G}} \sum_{i=1}^{n} k_i(t)v_i\ dt = 1, \tag{4.347}$$

we then take $v_i(s) := v_i$. The constraint (4.347) just says that for $x \equiv 1$, $(SAx)(s) = x(s)$ holds. The parameter τ (in [18] $\tau = 1$) determines the concrete method.

The common feature between mollification and the Backus–Gilbert method is the following: In both cases, an approximate inverse (determined by v_s^ϱ or by the $v_i(s)$) is determined independently from the right-hand side of the equation,

which can then be used to explicitly represent an approximate solution via (4.336) or via (4.342). By use of Lagrange multipliers, the Backus–Gilbert basis functions $v_1, \ldots, v_n$ can be determined pointwise from the linear system

$$\begin{pmatrix} G(s)\ w \\ w^T\ 0 \end{pmatrix} \begin{pmatrix} v \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad s \in \Omega, \tag{4.348}$$

with

$$G(s)_{ij} = \int_{\mathcal{G}} |s-t|^{2\tau} k_i(t) k_j(t)\ dt, \quad i,j \in \{1,\ldots,n\}, \tag{4.349}$$

$$w_i = \int_{\mathcal{G}} k_i(t)\ dt, \qquad i \in \{1,\ldots,n\}. \tag{4.350}$$

Note that the matrix of this system depends on s while in the corresponding system (4.333) for mollifier methods, s enters only in the right-hand side.

4.8. Numerical dilemmas and methodologies

The *numerical analysis of all ill-posed problem* ultimately involves solutions of finite-dimensional problems in order to obtain numerical approximations. This often entails a *two-stage regularization*. One first may "regularize" the problem in function spaces and then apply numerical methods to approximate the solution of a well-posed problem (or a family of such problems). On the other hand, one may "discretize" or approximate the IPP by finite-dimensional problems and then resolve numerical instability of these problems by methods of numerical linear algebra that are suitable for discrete ill-posed problems (see, e.g., [177, 184, 253]).

Two alternate routes are represented by the following diagram:

$$\begin{array}{ccccccc} & & P & & \overset{\mathbf{c}\ r}{\rightarrow} & P_t & \\ p & \swarrow & & \searrow p & & & \searrow p \\ F_n & \rightarrow & & P_n & \overset{\mathbf{d}\ \mathbf{r}}{\rightarrow} & P_{n,t} & P_{t,n} \end{array}$$

Here, P is a given ill-posed problem and P_t is a "regularized" version of P by use of a certain regularization scheme in a certain function space. F_n is a problem in a finite-dimensional function space and P_n is a finite-dimensional algebraic problem, both obtained by approximation of P; whereas $P_{n,t}$ denotes a "regularization" of the (numerically unstable) problem P_n and $P_{t,n}$ is the numerical approximation of the problem P_t, which is numerically stable for t not too small. In the diagram "**c r**" denotes **c**ontinuous **r**egularization, "**d r**" **d**iscrete **r**egularization and "p" denotes a generic "projection" (discretization, projection method, etc.).

At this stage, after having some knowledge of most of the regularization methods, we mention some procedures which provide concrete realizations of these schemes (cf. [184]):

$P \rightarrow P_t$ Tikhonov's regularization, (multiscale) regularizer operators in function spaces, quasi-reversibility methods, replacement of the IPP by

	a stable minimization problem depending on a parameter, iterative methods in function spaces;
$P \to F_n$	truncated singular value decompositions (TSVD), truncated series expansion, moment discretization, projection methods;
$P \to P_n$	finite difference method with collocation, reduction of F_n to algebraic equations;
$P_n \to P_{n,t}$	decomposition methods or regularization for linear algebraic equations, TSVD for matrices;
$P_\gamma \to P_{t,n}$	various numerical methods for solving well-posed problems, e.g., discretization, projection methods, multiscale procedures, etc.

Returning to the general scheme, we let $x, x_t, x_n, x_{t,n}, x_{n,t}$ denote, respectively, the "solution" (classical or least square of minimal norm) of the problems $P, P_t, P_n, P_{t,n}, P_{n,t}$ in the absence of contamination in the data, and let $x^\varepsilon, x_t^\varepsilon$, etc., denote the corresponding solution when the data are contaminated (y is replaced by y^ε, where $\|y^\varepsilon - y\| \le \varepsilon$ for some $\varepsilon > 0$). If we assume that the various *regularizer schemes* are convergent (e.g., $x_t \to x$ as $t \to 0$, $x_{n,t} \to x_n$ as $t \to 0$ for each fixed $n, x_n \to x$ as $n \to \infty$, etc.) in the absence of error, then estimates similar to (3.3) and (3.4) can be used to calculate an "approximate solution" x_ε as before. For example,

$$\|x_{t,n}^\varepsilon - x\| \le \|x_{t,n}^\varepsilon - x_t^\varepsilon\| + \|x_t^\varepsilon - x_t\| + \|x_t - x\|. \tag{4.351}$$

Here, $\|x_{t,n}^\varepsilon - x_t^\varepsilon\|_X$ is an *approximation error estimate* that provides a rate of convergence of the approximation scheme for the well-posed problem P_t for a fixed t. $\|x_t^\varepsilon - x_t\|$ is an *estimate for the contamination error* which can be estimated if the robustness of P_t is known and $\|x_t - x\|$ is a *regularization error*. Similarly,

$$\|x_{n,t}^\varepsilon - x\| \le \|x_{n,t}^\varepsilon - x_n^t\| + \|x_n^\varepsilon - x_n\| + \|x_n - x\|. \tag{4.352}$$

Note, however, that now $\|x_{n,t}^\varepsilon - x_n^t\|$ is an *error in the regularization of the problem* P_n, $\|x_n^\varepsilon - x_n\|$ is an *error due to the propagation of contamination into the discrete system*, etc.

The *dilemmas and methodologies of mathematical and numerical analysis of IPP* involve the following facts and observations:

(i) For most regularization and approximation schemes

$$\lim_{t \to 0} x_t = x \tag{4.353}$$

and

$$\lim_{n \to \infty} x_n = x \tag{4.354}$$

in the absence of contamination. For some approximation (projection) schemes, non-convergence can take place even without contamination.

(ii) Neither the double limit

$$\lim_{\substack{t \to 0, \\ n \to \infty}} x_{t,n}^\varepsilon, \tag{4.355}$$

nor the iterated limits [$\lim_{n\to\infty}\lim_{t\to 0}$ and $\lim_{t\to 0}\lim_{n\to\infty}$] of $x^{\varepsilon}_{t,n}$ and $x^{\varepsilon}_{n,t}$ exist. In fact, $\|x^{\varepsilon}_{t,n}\|_X$ and $\|x^{\varepsilon}_{n,t}\|_X$ blow up as $n\to\infty$ and $t\to 0$. It should be emphasized that this *blow-up is intrinsically inherent in all IPP regardless of any regularization-approximation scheme*. Thus, the best one can achieve for a numerical resolution of IPP is to minimize the error: $\|x-x^{\varepsilon}_{t,n}\|$ or $\|x-x^{\varepsilon}_{n,t}\|$, and to find "paths" along which $x_{t,n}$ and $x_{n,t}$ converge to x as $\varepsilon\to 0$.

(iii) The alternative routes diagram is *non-commutative*, in general. It is not always clear which path along the diagram is more effective. One has to compare the minimum errors of $\|x-x^{\varepsilon}_{t,n}\|_X$ and $\|x-x^{\varepsilon}_{n,t}\|_X$ and to take into consideration the computational complexity of the two paths. For some simple schemes, the diagram is commutative (e.g., Tikhonov regularization and TSVD for a compact operator commute).

(iv) Sharp resolutions of an ill-posed problem (i.e., an optimal compromise between accuracy and numerical stability) hung upon "optimal parameter choice criteria" (for t and n). Analytic criteria for this choice are often not available due to the lack of sharp rates of convergence in the preceding estimates, except for Tikhonov regularization or related methods based on simple variational principles. Often the parameter is chosen by an interactive computational scheme, based on rough analytic estimates.

References

[1] Acar, R., Vogel, C.R.: Analysis of Bounded Variation Penalty Methods for Ill-Posed Problems. *Inverse Problems*, **10**:1217–1229, 1994.

[2] Adams, R.A.: *Sobolev Spaces*. Academic Press, New York, San Francisco, 1975.

[3] Alber, Y.I.: The Solution of Nonlinear Equations with Monotone Operators in Banach Spaces, *Siberian Math. J.*, **16**:1–8, 1975.

[4] Alber, Y.I.: Iterative Regularization in Banach Spaces, *Soviet Math.* (*Iz. VUZ*), **30**:1–8, 1986

[5] Alber, Y.I.: The Regularization Method for Variational Inequalities with Nonsmooth Unbounded Operators in Banach Spaces. *Appl. Math. Lett.*, **6**:63–68, 1993.

[6] Alber, Y.I.: Generalized Projection Operators in Banach Spaces: Properties and Applications, *Functional Differential Equations, Proceedings of the Israel Seminar*, **1**:1–21, 1994.

[7] Alber, Y.I.: Metric and Generalized Projection Operators in Banach Spaces: Properties and Applications, In: *Theory and Applications of Nonlinear Operators of Monotone and Assertive Type* (A. Kartsatos, ed.). Marcel Dekker, New York, 15–50, 1996.

[8] Alber, Y.I., Notik, A.: Perturbed Unstable Variational Inequalities with Unbounded Operator on Approximately Given Sets. *Set-Valued Analysis*, Vol. 1, **4**:393–402, 1993.

[9] Alber, Y.I., Reich, S.: An Iterative Method for Solving a Class of Nonlinear Operator Equations in Banach Spaces, *Panamerican Math. J.*, **4**:39–54, 1994.

[10] Alber, Y.I., Rjazanceva, I.: Variational Inequalities with Discontinuous Monotone Mappings, *Soviet Math. Dokl.*, **25**: 206–210, 1982.

[11] Albert, A.: *Regressions and the Moore–Penrose Pseudoinverse*. Academic Press, 1972.

[12] Angell, T.S., Nashed, M.Z.: Operator-Theoretic and Computational Aspects of Ill-Posed Problems in Antenna Theory. In: *Proc. Int. Symp. Math. Theory of Networks and Syst.*, Delft Univ. of Technology, The Netherlands, 499–511, 1979.

[13] Anger, G.: A Characterization of Inverse Gravimetric Source Problem through Extremal Measures. *Rev. Geophys. Space Phys.*, **19**:299–306, 1981.

[14] Anger, G.: *Inverse Problems in Differential Equations*. Akademie-Verlag, Berlin, 1990.

[15] Aronszajn, N.: Theory of Reproducing Kernels. *Trans. Am. Math. Soc.*, **68**:337–404, 1950.

[16] Baart, M.L.: *Methods for Solution of Fredholm Integral Equations of the First Kind*. CSIR, Pretoria, Techn. Rep., **104**, 1979.

[17] Baer, R.: *Linear Algebra and Projective Geometry*. Academic Press, New York, 1952.

[18] Backus, G.E., Gilbert, F.: Numerical Applications of a Formalism for Geophysical Inverse Problems. *Geophys. J.R. Astron. Soc.*, **13**:247–276, 1967.

[19] Bakusinksii, A.B.: A general method for constructing regularizing algorithms for a linear incorrect equation in Hilbert space. U.S.S.R. Comp. Math. Meth. Phys. (7):279–284, 1967.

[20] Bakusinskii, A.B.: On the Principle of Iterative Regularization, *USSR Comput. Math. and Math. Phys.*, **19**:256–260, 1979.

[21] Barzaghi, R., Sansò, F.: Remarks on the Inverse Gravimetric Problem. *Boll. Geod. Scienze Affini*, **45**:203–216, 1986.

[22] Baumeister, J.: *Stable Solution of Inverse Problems*. Vieweg, Braunschweig, 1987.

[23] Ben-Israel, A., Greville, T.N.E., *Generalized Inverses: Theory and Applications*. New York: Wiley-Interscience, 1974.

[24] Bertero, M., de Mol, C., Viano, G.A.: Linear Inverse Problems With Discrete Data. *Inverse Problems*, **4**:573–594, 1985/88.

[25] Bertero, M., Brianzi, P., Pike, E.R., Rebolia, L.: Linear Regularizing Algorithms for Positive Solutions of Linear Inverse Problems. *Proc. R. Soc. Lond. A*, **415**:257–275, 1988.

[26] Bertero, M. DeMol, C., Viano, G.A.: The Stability of Inverse Problems. In: *Inverse Scattering Problems in Optics*, H.P. Baltes, ed., Berlin, Heidelberg, New York, Springer, 1979.

[27] Binder, A., Engl, H.W., Groetsch, C.W., Neubauer, A., Scherzer, O.: Weakly Closed Nonlinear Operators and parameter identification in Parabolic Equations by Tiknonov Regularization. *Appl. Anal.*, **55**:215–234, 1994.

[28] Bissantz, H., Hohage, T., Munk, A.: Ruymgaart, F.: Convergence Rates of General Regularization Methods for Statistical Inverse Problems and Applications. *SIAM J. Numer. Anal.*, **45**:2610–2626, 2007.

[29] Bjerhammar, A.: Rectangular Reciprocal Matrices, with Special Reference to Geodetic Calculations. Bulletin Géodésique, **25**, 188–220, 1951.

[30] Bjerhammar, A.: Theory of Errors and Generalized Matrix Inverses. Elsevier Scientific Publ. Co., Amsterdam, 1973.

[31] Boullion, T.L., Odell, P.L., *Generalized Inverse Matrices*. Wiley-Interscience, New York, (1971).

[32] Bruck, R.E.: A Strongly Convergent Iterative Solution of $0 \in Ux$ for a Maximal Monotone Operator U in Hilbert Space, *J. Math. Anal. Appl.*, **48**:114–126, 1974.

[33] Burger, M., Kaltenbacher, B.: Regularizing Newton–Kaczmarz Methods for Nonlinear Ill-Posed Problems. *SIAM J. Numer. Anal.*, **44**:1775–1797, 2006.

[34] Cavalier, L., Golubev, G.K.: Risk Hull Method and Regularization by Projections of Ill-posed Inverse Problems. *Ann. Stat.*, **34**:1653–1677, 2006.

[35] Cavalier, L., Golubev, G.K., Picard, D., Tsybakov, A.B.: Oracle Inequalities for Inverse Problems. *Ann. Stat.*, **30**:843–874, 2002.

[36] Craven, B.D., Nashed, M.Z.: Generalized Implicit Function Theorems when the Derivative has no Bounded Inverse: Theory, Methods, and Applications. *Nonlinear Analysis*, **6**:375–387, 1982.

[37] Davis, P.J.: *Interpolation and Approximation*. Blaisdell, New York, 1963.

[38] M. Defrise, C. de Mol.: A Note on Stopping Rules for Iterative Regulaization Methods and Filtered SVD. In: *Inverse Problems: An Interdisciplinary Study*. Academic Press, San Diego, 261–268, 1987.

[39] Desbat, L., Girard, D.: The "Minimum Reconstruction Error" Choice of Regularization Parameters: Some More Efficient Methods and Their Application of Deconvolution Problems. *SIAM, J. Sci. Compt.*, **16**:187–1403, 1995.

[40] Diaz, J.B. and Metcalf, F.T.: On Integration Procedures for Equations of the First Kind, $Ax = y$, and Picard's Criterion for the Existence of a Solution. *Math. Comp.* **24**:923–935, 1970.

[41] Dicken V. and Maass P.: Wavelet-Galerkin Methods for Ill-Posed Problems. *J. Inv. and Ill-posed Probl.*, **4**:203–222, 1996.

[42] Dimitiev,V.I., Il'inski, A.S., Svenshnikov, A.G.: The Developments of Mathematical Methods for the Study of Direct and Inverse Problems in Electrodynamics. *Russian Math. Surveys*, **31**:133–152, 1976.

[43] Dobson D.C., Scherzer, O.: Analysis of Regularized Total Variation Penalty Methods for Denoising. *Inverse Problems*, **12**:601–617, 1996.

[44] Donoho D.L.: Nonlinear Solution of Linear Inverse Problems by Wavelet-Vaguelette Decomposition. *Appl. Comp. Harm. Anal.*, **2**:101–126, 1995.

[45] Donoho D.L. and Johnstone, I.M.: Minimax Estimation Via Wavelet Shrinkage. *Ann. Statistics*, **26**:879–921, 1998.

[46] Dunford N., Schwarz, J.: *Linear Operators*, Vol. II, Wiley-Interscience, New York, 1963.

[47] Duris, C.S.: Optimal Quadrature Formulas Using Generalized Inverses. Part I. General Theory and Minimum Variance Formulas, *Math. Comp.* **25**:495–504, 1971.

[48] Eggermont, P.P.B.: Maximum Entropy Regularization for Fredholm Integral Equations of the First Kind. *SIAM J. Math. Anal.*, **24**:1557–1576, 1993.

[49] Eggermont P.P.B., LaRicca, V.N.: *Maximum Penalized Likelihood Estimation.* Volume II, Regression, Springer, New York, 2009. 10

[50] Eggermont P.P.B., LaRicca, V.N. Nashed, M.Z.: On Weakly Bounded Noise in Ill-Posed Problems. *Inverse Problems.* **25**:115018–115032, 2009.

[51] Eggermont, P.P.B.; LaRiccia, V.N.; Nashed, M.Z.: Noise Models for Ill-Posed Problems. In: *Handbook of Geomathematics, Vol.* 2, (Freeden, W., Nashed, M.Z., Sonar, T., Eds.), 741–762, Springer, Heidelberg, 2010.

[52] Eggermont P.P.B., LaRicca, V.N. Nashed, M.Z.: Moment Discretization of Ill-Posed Problems with Discrete Weakly Bounded Noise. *Int. J. Geomath.*, **3**:155–178, 2012.

[53] Eggermont, P.N., LaRiccia, V., Nashed, MZ: Noise Models for Ill-Posed Problems. In: *Handbook of Geomathematics, Vol.* 2*,* 2*nd edition*, (Freeden, W., Nashed, M.Z., Sonar, T., Eds.), 1633–1658, Springer, New York, 2015.

[54] Eicke, B.: Iteration Methods for Convexly Constrained Ill-posed Problems in Hilbert Space, *Numer. Funct. Anal. Optim.*, **13**:413–429, 1992.

[55] Eisenhart, C., Carl Friedrich Gauss. International Encyclopedia of Social Sciences, New York, Vol. VI, 74–81, 1986.

[56] Ekeland, I.; Temam, R. Convex Analysis and Variational Problems. North-Holland, Amsterdam, 1976.

[57] Elden, L.: Algorithms for the Regularization of Ill-Conditioned Least Squares Problems. *BIT*, 17:134–145, 1977.

[58] Elden, L.: An Algorithm for the Regularization of Ill-Conditioned Banded Least Squares Problems. *SIAM J. Sci. Stat. Comput.*,**5**:237–254,1984.

[59] Engl, H.: Discrepancy Principles for Tikhonov Regularization of Ill-Posed Problems, Leading to Optimal Convergence Rates, *J. Optim. Theory Appl.*, **52**:209–215, 1987.

[60] Engl, H.: *Integralgleichungen.* Springer Lehrbuch Mathematik, Wien, 1997.

[61] Engl, H.W., Nashed, M.Z.: Stochastic Projectional Schemes for Random Linear Operator Equations of the First and Second Kinds, *Numer. Funct. Anal. and Optimiz.*, **1**:451–473, 1979.

[62] Engl, H.W., Nashed, M.Z.: New Extremal Characterizations of Generalized Inverses of Linear Operators. *J. Math. Anal. Appl.*, **82**:566–586, 1981.

[63] Engl, H.W., Hanke, M., Neubauer, A.: *Regularization of Inverse Problems.* Kluwer, Dordrecht, 1996.

[64] Engl H.W., Kunisch K. and Neubauer, A.: Convergence Rates for Tikhonov Regularisation of Non-Linear Ill-Posed Problems. *Inverse Problems*, **5**:523–540, 1989.

[65] Engl, H., Louis, A.K., Rundell, W. (Eds.): *Inverse Problems in Geophysical Applications.* SIAM, Philadelphia, 1997.

[66] Flemming, J.; Hofmann, B.: A New Approach to Source Conditions in Regularization with General Residual Term. *Numer. Funct. Anal. Optimiz.*, **31**:254–284, 2010.

[67] Frankenberger H, Hanke M.: Kernel Polynomials for the Solution of Indefinite and Ill-Posed Problems. *Numer. Alg.*, **25**:197–212, 2002.

[68] Franklin J.N.: Well-posed Stochastic Extensions to Ill-Posed Linear Problems. *J. Math. Anal. Appl.*, **31**:682–716, 1970.

[69] Franklin, J.N.: On Thikhonov's Method for Ill-Posed Problems. *Math. Comp.*, **28**:889–907, 1974.

[70] Fredholm I.: Sur une classe d'équations functionelles. Acta Mathematica **27**:365–390, 1903.

[71] Freeden, W.: On Approximation by Harmonic Splines. *Manuscr. Geod.*, **6**:193–244, 1981.

[72] Freeden, W.: *Multiscale Modelling of Spaceborne Geodata*, B.G. Teubner, Stuttgart, Leipzig, 1999.

[73] Freeden, W., Gutting, M.: *Special Functions of Mathematical (Geo)Physics*. Birkhäuser, Basel, (2013).

[74] Freeden, W., Maier, T.: Spectral and Multiscale Signal-to-Noise Thresholding of Spherical Vector Fields. *Computational Geosciences*, **7**(3):215–250, 2003.

[75] Freeden, W., Michel, V.: *Multiscale Potential Theory (with Applications to Geoscience)*. Birkhäuser, Boston, 2004.

[76] Freeden, W., Nutz, H.: Satellite Gravity Gradiometry as Tensorial Inverse Problem., *Int. J. Geomath.*, **2**:177–218, 2012.

[77] Freeden, W., Schneider, F.: Regularization Wavelets and Multiresolution, *Inverse Problems*, **14**:493–515, 1998.

[78] Freeden, W., Schreiner, M.: Satellite Gravity Gradiometry (SGG): From Scalar to Tensorial Solution. In: *Handbook of Geomathematics, Vol.* 1 (Freeden, W., Nashed, M.Z., Sonar,T., Eds.), 269–302, Springer, Heidelberg, 2010.

[79] Freeden, W., Schreiner, M.: Satellite Gravity Gradiometry (SGG): From Scalar to Tensorial Solution. In: Handbook of Geomathematics, Freeden, W., Nashed, M.Z., and Sonar, T., Eds., Springer, 2nd edition, New York, Heidelberg, 2nd edition, 2015.

[80] Freeden, W., Witte, B.: A Combined (Spline-)Interpolation and Smoothing Method for the Determination of the Gravitational Potential from Heterogeneous Data. *Bull. Geod.*, **56**:53–62, 1982.

[81] Freeden, W., Michel, V., Nutz, H.: Satellite-to-Satellite Tracking and Satellite Gravity Gradiometry (Advanced Techniques for High-Resolution Geopotential Field Determination). *J. Eng. Math.*, **43**:19–56, 2002.

[82] Freeden, W., Nashed, Z., Sonar, T. (Eds.): *Handbook of Geomathematics*. Springer, New York, 2nd edition, Vol. 1, 2, and 3, 2015.

[83] Freeden, W., Schneider, F., Schreiner, M.: Gradiometry – An Inverse Problem in Modern Satellite Geodesy, in: Engl, H.W., Louis, A., Rundell, W. (Eds.), *GAMM-SIAM Symposium on Inverse Problems: Geophysical Applications*, 179–239, 1997.

[84] Friedrich, K.: Allgemeine für die Rechenpraxis geeignete Lösung für die Aufgaben der kleinsten Absolutsumme und der günstigsten Gewichtsverteilung, Zeitschrift für Vermessungswesen, 337–358, 1937.

[85] Gajewski, H., Gröger, K., Zacharias, K.: *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen*. Akademie-Verlag, Berlin, 1974.

[86] Galerkin, B.G.: Expansions in Stability Problems for Elastic Rods and Plates (in Russian). *Vestnik Inzkenorov*, **19**:897–908, 1915.

[87] Gauss, C.F.: Theoria motus corporum coelestium in sectionibus conicis solem ambientium, Hamburg, 1809, Werke 7. Translated into English by C.H. Davis, 1963.

[88] Gauss, C.F.: Theoria combinationis observationum erroribus minimis obnoxiae, 1. Teil, Göttingen, (1821).

[89] Gebbauer, B., Scherzer, O.: Impedance-Acoustic Tomography, *SIAM J. Appl. Math.*, **69**:565–576, 2008.

[90] Gelfand I.M., Shilov, G.E.: *Generalized Functions*. Vol. 1 and 2. Academic Press, New York, 1964.

[91] Gerhards, C.: *Spherical Multiscale Methods in Terms of Locally Supported Wavelets: Theory and Application to Geomagnetic Modeling*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern, 2011.

[92] Gfrerer, H.: An a Posteriori Parameter Choice for Ordinary and Iterated Tikhonov Regularization of Ill-Posed Problems Leading to Optimal Convergence Rates. *Math. Comp.*, **49**:523–542, 1987a.

[93] Gfrerer, H.: Supplement to: An a Posteriori Parameter Choice for Ordinary and Iterated Tikhonov Regularization of Ill-Posed Problems Leading to Optimal Convergence Rates. *Math. Comp.*, **49**:S5–S12, 1987b.

[94] Glockner, O.: *On Numerical Aspects of Gravitational Field Modelling from SST and SGG by Harmonic Splines and Wavelets* (*with Application to CHAMP Data*), Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern, 2001.

[95] Gholami, A.; Siahkoohi, H.R.: Regularization of Linear and Nonlinear Geophysical Ill-Posed Problems with Joint Sparsity Constraints. *Geophys. J. Int.*, **180**:871–882, 2010.

[96] Gohberg, I., Goldberg, S., Kaashoek, M.A.: Classes of Linear Operators, Vol. II. Birkhäuser OT 63, 1993.

[97] Goldstine, H.H.: A History of Numerical Analysis from the 16th Through the 19th Century. Springer-Verlag, New York, 1977.

[98] Golomb M. and Weinberger, H.F.: Optimal Approximation and Error Bounds, In One Numerical Approximation,*Proc. Sympos. Madison, Wis.*, 1958, (R.E. Langer, ed.), University of Wisconsin Press, 117–190, 1959.

[99] Golub, G.H., Van Loan, C.F.: *Matrix Computations*. 3rd ed., The John Hopkins Univ. Press, Baltimore, M.D., 1996.

[100] Grafarend, E.W.: Six Lectures on Geodesy and Global Geodynamics. In: *Proceedings of the Third International Summer School in the Mountains* (Moritz, H., Sünkel, H. (Eds.)), 531–685, 1982.

[101] Grafarend, E.W., Awange, J.L.: *Applications of Linear and Nonlinear Models*. Springer, Berlin, Heidelberg, 2012.

[102] Graves, J., Prenter, P.M.: On Generalized Iterative Filters for Ill-Posed Problems. *Numer. Math.*, **30**:281–299, 1978.

[103] Grenander U.: *Abstract Inference*. Wiley, New York, 1981.

[104] Groetsch, C.W.: *Generalized Inverses of Linear Operators*. Marcel Dekker. Inc., New York, 1977.

[105] Groetsch, C.W.: *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman, London-Boston, 1984.

[106] Groetsch, C.W.: *Inverse Problems in the Mathematical Science*. Vieweg, Braunschweig, 1993.

[107] Groetsch, C.W.: Comments on Morozov's Discrepancy Principle. In Hämmerlin, G.; Hoffmann, K.H., Eds., *Improperly Posed Problems and Their Numerical Treatment*, 97–104, Birkhäuser, Basel, 1983.

[108] Groetsch C.W., Scherzer, O.: Iterative Stabilization and Edge Detection. In: *Contemp. Math.*, **313**:129–141. Providence, RI: Amer. Math. Soc., 2002.

[109] Hadamard, J.: Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton Univ. Bull.*, **13**:49–52, 1902.

[110] Hadamard, J.: *Lectures on the Cauchy Problem in Linear Partial Differential Equations*, Yale University Press, New Haven, 1923.

[111] Haddad R.A. Parsons, T.W.: *Digital Signal Processing: Theory, Applications and Hardware*. Computer Science Press, 1991.

[112] Hanke, M.: *Conjugate Gradient Type Methods for Ill-Posed Problems*. Pitman Research Notes in Mathematics, Longman House, Harlow, Essex, 1995.

[113] Hanke, M., Hansen, P.C.: Regularization Methods for Large Scale Problems. *Surveys Math. Indust.*, **3**:253–315, 1993.

[114] Hanke, M., Scherzer, O.: Inverse Problems Light: Numerical Differentiation. *Amer. Math. Monthly*, **108**:512–521, 2001.

[115] Hanke, M.; Vogel, C.R.: Two-Level Preconditioners for Regularized Inverse Problems. *Numer. Math.*, **83**:385–402, 1999.

[116] Hansen, P.C.: Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. *SIAM Rev.*, **34**:561–580, 1992.

[117] Hanson, R.J.: A Numerical Methods for Solving Fredholm Integral Equations of the First Kind. *SIAM J. Numer. Anal.* **8**:616–662, 1971.

[118] Hegland, M.: Variable Hilbert Scales and their Interpolation Inequalities with Applications to Tikhonov Regularization. *Appl. Anal.*, **59**:207–223, 1995

[119] Helmert, F.: *Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate*. Teubner, Berlin, 1907.

[120] Heuser, H.: *Funktionalanalysis*. 4. Auflage, Teubner, 1975.

[121] Hilbert, D.: *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*. Teubner, Leipzig, 1912.

[122] Hille, E.: Introduction to the General Theory of Reproducing Kernels, *Rocky Mountain, J. Math.*, **2**:321–368, 1972.

[123] Hirzebruch, Scharlau: *Funktional Analysis I*, BI 296a, 1971.

[124] Hofmann, B.: *Inverse Probleme*. Teubner 1999.

[125] Hofmann, B.; Mathé, P.; Weiszäcker, H. von: Regularisation in Hilbert Space under Unbounded Operators and General Source Conditions. *Inverse Problems*, **25**:115–130, 2009.

[126] Hohage, T.; Pricop, M.: Nonlinear Tikhonov Regularization in Hilbert Scales for Inverse Boundary Value Problems with Random Noise. *Inverse Problems and Imaging*, **2**:271–290, 2008.

[127] Hurwitz, W.A.: On the Pseudo-Resolvent to the Kernel of an Integral Equation. *Transactions of the American Mathematical Society*, **13**:405–418, 1912.

[128] Ismail, M., Nashed, M.Z., Zayed, A., Ghaleb, A.: *Mathematical Analysis, Wavelets and Signal Processing*. Contemporary Mathematics, Vol. 190, American Mathematical Society. Providence, RI, 1995.

[129] Ivanov V.K., Kudrinskii, V.Yu.: Approximate Solution of Linear Operator Equations in Hilbert Space by the Method of Least Squares. *I. Z. Vycisl. Mat. i Mat. Fiz.*, **6**:831–944, 1966.

[130] Jacobsen, M., Hansen, P.C.; Saunders, M.A.: Subspace preconditioned LSQR for Discrete Ill-Posed Problems. *BIT Numer. Math.*, **43**:975–989, 2003.

[131] Jorgensen P., Tian, F.: Graph Laplacians and Discrete Reproducing Kernel Hilbert Spaces from Restrictions. *Stochastic Analysis and Applications*, **34**:722–747, 2016.

[132] Kaipio J., Somersalo, E.: *Statistical and Computational Inverse Problems*. Springer, Berlin, 2005.

[133] Kammerer, W.J., Nashed, M.Z.: The Convergence of the Conjugate Gradient Method for Singular Linear Operator Equations. *SIAM J. Numer. Anal.*, **9**:165–181, 1972.

[134] W.J. Kammerer, M.Z. Nashed: Iterative Methods for Best Approximate Solutions of Linear Integral Equation of the First and Second Kind. *J. Math. Anal. Appl.*, **40**:547–573, 1972.

[135] Kantorowitsch, L.W., Akilow, G.P.: *Funktionalanalysis in normierten Räumen*. Akademie-Verlag, Berlin, 1964.

[136] Kato, T.: Perturbation Theory for Nullity Deficiency and Other Quantities of Linear Operators. *J. Analyse Math.*, **6**:271–322, 1958.

[137] Kirsch, A.: *An Introduction to the Mathematical Theory of Inverse Problems*. 2nd ed., Springer, Heidelberg, 1996.

[138] Kowar, R., Scherzer, O.: Convergence analysis of a Landweber–Kaczmarz method for solving nonlinear ill-posed problems, in: S. Romanov, S.I. Kabanikhin, Y.E. Anikonov, A.L. Bukhgeim, (Eds.), Ill-Posed and Inverse Problems, VSP Publishers, Zeist, 2002.

[139] Kress, R.: *Linear Integral Equations*. 2nd ed., Springer, Berlin, 1989.

[140] Kreyszig, E.: *Introductory Functional Analysis with Applications*. John Wiley and Sons, New York, 1978.

[141] Larson, D., Massopust, P., Nashed, M.Z., Nguyen, M.C., Papadakis, M., Zayed, A., (Eds.): *Frames and Operator Theory in Analysis and Signal Processing*. Contemporary Mathematics. Vol. 451, American Mathematical Society. Providence, RI, 2008.

[142] Lavrentiev, M.M.: *Some Improperly Posed Problems of Mathematical Physics*, Izdat. Sibirsk. Otdel, Akad. Nauk. SSSR, Novosibirsk, 1962, *English Transl., Springer Tracts in Natural Philosophy*, Vol. 11, Springer-Verlag, Berlin, 1967.

[143] Lieusternik, L.A., Sobolev, V.J.: *Elements of Functional Analysis*. New York: Ungar, 1961.

[144] Lin, Y., Brown, L.D.: Statistical Properties of the Method of Regularization with Periodic Gaussian Reproducing Kernel. *Ann. Statist.*, **32**(4):1723–1743, 2004.

[145] Liskovets, O.A.: Regularization of Variational Inequalities with Pseudo-Monotone Operators on Approximately Given Sets, *Differ. Equations*, **11**:1970–1977, 1989.

[146] Liu, F.; Nashed, M.Z.: Tikhonov Regularization of Nonlinear Ill-Posed Problems with Closed Operators in Hilbert Scales. *J. Inverse Ill-Posed Problems*, **5**:363–376, 1997.

[147] Locker J., Prenter, P.M.: Regularization with Differential Operators. *J. Math. Anal. Appl.*, **74**:504–529, 1980.

[148] Louis, A.K.: *Inverse und schlecht gestellte Probleme*. Teubner, Stuttgart, 1989.

[149] Louis, A.K., Maass, P.: A Mollifier Method for Linear Equations of the First Kind. *Inverse Problems*, **6**:427–440, 1989.

[150] Louis, A.K., Maass, P., Rieder, A.: *Wavelets: Theorie und Anwendungen*. B.G. Teubner Studienbücher, 1998.

[151] Mair, B.A.; Ruymgaart, F.H.: Statistical Inverse Estimation in Hilbert Scales. *SIAM J. Appl. Math.*, **56**:1424–1444, 1996

[152] Mathé, P.: The Lepskii Principle Revisited, *Inverse Problems*, **2**2:111–115, 2006.

[153] Mathé, P.; Hofmann, B.: How General are General Source Conditions? *Inverse Problems*, **24**: 2008.

[154] Mathé, P., Pereverzev, S.V: The Discretized Discrepancy Principle Under General Source Conditions, *J. Complexity*, **22**:371–381, 2006a.

[155] Mathé, P., Pereverzev, S.V: Regularization of Some Linear Ill-Posed Problems with Discretized Random Noisy Data. *Math. Comp.*, **75**:1913–1929, 2006b.

[156] Mathé, P.; Pereverzev, S.V.: Geometry of Linear Ill-Posed Problems in Variable Hilbert Scales. *Inverse Problems*, **19**:789–803, 2003.

[157] Marti, J.T.: An Algorithm For Computing Minimum Norm Solutions of Fredholm Integral Equations of the First Kind. *SIAM J. Numer. Anal.*, **15**:1071–1076, 1978.

[158] Marti, J.T.: On the Convergence of an Algorithm Computing Minimum-Norm Solutions of Ill-Posed Problems. *Math. Comp.*, **34**:521–527, 1980.

[159] Meissl, P.A.: *A Study of Covariance Functions Related to the Earth's Disturbing Potential*. Department of Geodetic Science, No. 151, The Ohio State University, Columbus, OH, (1971).

[160] Meissl, P.A.: Hilbert Spaces and Their Applications to Geodetic Least Squares Problems. *Boll. Geod. Sci. Aff.*, N. 1, 181–210, 1976.

[161] Michel, V.: *A Multiscale Method for the Gravimetry Problem: Theoretical and Numerical Aspects of Harmonic and Anharmonic Modelling*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen, 1999.

[162] Michel, V.: Scale Continuous, Scale Discretized and Scale Discrete Harmonic Wavelets for the Outer and the Inner Space of a Sphere and Their Application to an Inverse Problem in Geomathematics. *Appl. Comp. Harm. Anal.* (*ACHA*), **12**:77–99, 2002.

[163] Michel, V.: *A Multiscale Approximation for Operator Equations in Separable Hilbert Spaces – Case Study: Reconstruction and Description of the Earth's Interior*. Habilitation Thesis, University of Kaiserslautern, Geomathematics Group, Shaker, Aachen, 2002a.

[164] Miller, K.: Least Squares Methods for Ill-Posed Problems with a Prescribed Bounded. *SIAM J. Math. Anal.*, **1**:52–74, 1970.

[165] Moore, E.H.: On the Reciprocal of the General Algebraic Matrix. *Bulletin of the American Mathematical Society*, **26**:394–395, 1920.

[166] Moore, E.H.: General Analysis. *Memoirs of the American Mathematical Society*, **1**:197–209, 1935.

[167] Moritz, H.: *Advanced Physical Geodesy*. Herbert Wichmann Verlag, Karlsruhe, Abacus Press, Tunbridge, 1980.

[168] Morozov, V.A.: On the Solution of Functional Equations by the Method of Regularization. *Soviet Math. Doklady*, **7**:414–41, 1966.

[169] Morozov, V.A.: *Methods for Solving Incorrectly Posed Problems*. (In Russian), Moscow Univ., 1975 (English transl., editor M.Z. Nashed), Springer, New York, 1984.

[170] Morozov, V.A.: *Methods for Solving Incorrectly Posed Problems*. Springer, New York, 1984.

[171] Nagel, H.-H., On the Estimation of Optical Flow: Relations Between New Approaches and Some New Results. Artif. Intell. **33**,299–324 1987.

[172] Nashed, M.Z.: Steepest Descent for Singular Linear Operator Equations. *SIAM J. Numer. Anal.*, **7**:358–362, 1970.

[173] Nashed, M.Z.: Generalized Inverses, Normal Solvability and Iteration for Singular Operator Equations. In: *Nonlinear Functional Analysis and Applications*, L.B. Rall, ed., Academic, New York, 311–359, 1971.

[174] Nashed, M.Z.: Differentiability and Related Properties of Nonlinear Operators: Some Aspects of the Role of Differentials in Nonlinear Functional Analysis, in: *Nonlinear Functional Analysis and Applications*, L.B. Rall, ed., 103–309, Academic Press, New York, 1971.

[175] Nashed, M.Z.: Some Aspects of Regularization and Approximations of Solutions of Ill-Posed Operator Equations. In: *Proc.* 1972 *Army Numerical Analysis Conf., Edgewood, MD*, 163–181, 1972.

[176] Nashed, M.Z.: Approximate Regularized Solutions to Improperly Posed Linear Integral and Operator Equations. In: *Constructive and Computational Methods for Differential and Integral Equations*. D.L. Colton and R.P. Gilbert, Eds., Springer, New York, LNM 430, 289–322, 1974.

[177] Nashed, M.Z. (ed.): *Generalized Inverses and Applications*. Academic Press, New York, 1976.

[178] Nashed, M.Z.: Aspects of Generalized Inverses in Analysis and Regularization. In: *Generalized Inverses and Applications*. Academic Press, New York, 193–244, 1976.

[179] Nashed, M.Z.: Perturbations and Approximation for Generalized Inverses and Linear Operators. In: *Generalized Inverses and Applications*. Academic Press, New York, 325–396, 1976.

[180] Nashed, M.Z.: On Moment-Discretization and Least-Squares Solutions of Linear Integration Equations of the First Kind. *J. Math. Anal. Appl.*, **53**:359–366, 1976.

[181] Nashed, M.Z.: Regularization and Approximation of Ill-Posed Problems in System Theory. In: *Proc.* 1979 *Conf. on Information Sciences and Systems*, G.G.L. Meyer and C.R. Westgate, Eds., The Johns Hopkins Univ., New York, 568–575, 1979.

[182] Nashed, M.Z.: New Applications of Generalized Inverses in System and Control Theory. In: *Proc.* 1980 *Conf. on Information Sciences and Systems, J.B. Thomas. ed. Princeton.* NJ, Princeton, 353–358, 1980

[183] Nashed, M.Z.: Continuous and Semicontinuous Analogous of Iterative Method of Cimmino and Kaczmarz with Applications to the Inverse Radon Transform. In: *Mathematical Aspects of Computerized Tomography.* G.T. Herman and F. Natterer, Eds., Springer, New York, 160–178, 1981.

[184] Nashed, M.Z.: Operator-Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory. *IEEE Trans. Antennas Propagation*, **29**: 220–231, 1981.

[185] Nashed, M.Z.: A New Approach to Classification and Regularization of Ill-Posed Operator Equations. In: H. Engl and C.W. Groetsch, Eds., Inverse and Ill-Posed Problems, Band 4, *Notes and Reports in Mathematics and Science and Engineering*, Academic Press, Boston, 1987.

[186] Nashed, M.Z.: Inner, Outer, and Generalized Inverses in Banach and Hilbert Spaces, *Numer. Funct. Anal. Optim.*, **9**:261–326, 1987

[187] Nashed, M.Z.: *Inverse Problems, Moment Problems and Signal Processing: Un Menage a Trois, Mathematics in Science and Technology.* Siddiqi, A.H., Singh, R.C. Manchanda, P. (Eds.), World Scientific, New Jersey, 1–19, 2010.

[188] Nashed, M.Z., Engl, H.W.: Random Generalized Inverses and Approximate Solution of Random Operator Equations. In: *Approximate Solution of Random Equations*, A.T. Bharucha-Reid, ed., North-Holland, New York, 149–210, 1979.

[189] Nashed, M.Z., Lin, F.: On Nonlinear Ill-Posed Problems II: Monotone Operator Equations and Monotone variational Inequalities. In: *Theory and Applications of Nonlinear Operators of Monotone and Assertive Type*, A. Kartsatos, ed., Marcel Dekker, New York, 223–240, 1996.

[190] Nashed, M.Z., Scherzer, O.: Stable Approximation of Nondifferentiable Optimization Problems with Variational Inequalities, *Contemp. Math.*, **204**:155–170, 1997.

[191] Nashed, M.Z., Scherzer, O.: Stable Approximation of a Minimal Surface Problem with Variational Inequalities. *Abst. and Appl. Anal.*, **2**:137–161, 1997.

[192] Nashed, M.Z., Scherzer, O. (Eds.): *Inverse Problems, Image Analysis and Medical Imaging.* Contemporary Mathematics, Vol. 313, American Mathematical Society, Providence, R.I., 2002.

[193] Nashed, Z.M., Sun, Q.: Function Spaces for Sampling Expansions, in: X. Shen, A.I. Zayed (Eds.), *Multiscale Signal Analysis and Modeling*, Springer Science + Business Media, 81–104, 2013.

[194] Nashed, M.Z., Votruba, F.G.: A Unified Operator Theory of Generalized Inverses. In: *Generalized Inverses and Applications.* M.Z. Nashed ed., Academic Press, New York, 1–109, 1976.

[195] Nashed, M.Z., Wahba, G.: Generalized Inverses in Reproducing Kernel Spaces: An Approach to Regularization of Linear Operator Equations. *SIAM J. Math. Anal.*, **5**:974–987, 1974.

[196] Nashed, M.Z., Wahba, G.: Approximate Regularized Pseudosolution of Liner Operator Equations when the Data-Vector is not in the Range of the Operator. *Bull. Amer. Math. Soc.*, **80**:1213–1218, 1974a.

[197] Nashed, M.Z., Wahba, G.: Convergence Rates of Approximate Least Squares Solutions of Linear Integral and Operator Equations of the First Kind. *Math. Comput.*, **28**:69–80, 1974b.

[198] Natanson, I.P.: *Constructive Function Theory.* Frederick Ungar Publ. Co., New York, 1965.

[199] Natterer F.: The Finite Element Method for Ill-Posed Problems. *RAIRO Analyse Numérique*, **11**:271–278, 1977.

[200] Natterer F.: Regularisierung schlecht gestellter Probleme durch Projektionsverfahren. *Numer. Math.*, **28**:329–341, 1977.

[201] Natterer F.: Error Bounds for Tikhonov Regularization in Hilbert Scales. *Appl Anal.* **18**:29–37, 1984.

[202] Neubauer, A.: On Converse and Saturation Results for Tikhonov Regularization of Linear Ill-Posed Problems. *SIAM J. Numer. Anal.*, **34**:517–527, 1997.

[203] Novikoff, P.: Sur le problème inverse du potentiel. *Comptes Rendus de l'Académie des Sciences de l'URSS* **18**:165–168, 1938.

[204] Ortega, J.M. and Rheinboldt, W.C.: On Discretization and Differentiation of Operators with Applications to Newton's Method, *SIAM J. Numer. Anal.*, **3**, 1966.

[205] Parker, R.L.: The Theory of Ideal Bodies for Gravity Interpretation. *Geophys. J. R. Astr. Soc.*, **42**:315–334, 1975.

[206] Payne, L.E.: *Improperly Posed Problems in Partial Differential Equations.* SIAM Publications, Philadelphia, 1975.

[207] Pedlosky, J.: *Geophysical Fluid Dynamics.* Springer, New York, Heidelberg, Berlin, 1979.

[208] Penrose, R.: A Generalized Inverse for Matrices. *Proceedings of the Cambridge Philosophical Society*, **51**:406–413, 1955.

[209] Penrose, R.: On Best Approximate Solutions of Linear Matrix Equations. *Proceedings of the Cambridge Philosophical Society*, **25**:17–19, 1956.

[210] Pereverzev, S.V., Schock E.: On the Adaptive Selection of the Parameter in Regularization of Ill-Posed Problems, *SIAM Journal on Numerical Analysis,* **43**:2060–2076, 2005.

[211] Perry, W.L.: On the Bojarski–Lewis Inverse Scattering Method. *IEEE Trans. Antennas Propagat.*, **6**:826–829, 1974.

[212] Perry, W.L.: Approximate Solution of Inverse Problems with Piecewise Continuous Solution. *Radio Science*, **12**:634–642, 1977.

[213] Petrini, H.: Sur l'existence des dérivées secondes du potentiel. *C.R. Acad. Sci. Paris*, **130**:233–235, 1900.

[214] Petryshyn, W.V.: On generalized inverses and uniform convergence of $(I - \beta K)^n$ with applications to iterative methods. J. Math. Anal. Appl., (18):417–439, 1967, MR 34 , 8191.

[215] Petrov, G.I.: Application of Galerkin's Method to a Problem of the Stability of the Flow of a Viscous Fluid (in Russian). *Priklad. Mate. Mekh.*, **4**:3–12, 1940.

[216] Phillips, B.L.: A Technique for the Numerical Solution of Certain Integral Equations of the First Kind. *J. Assoc. Comput. Math.* **9**:84–97, 1962.

[217] Plackett, R.L.: A Historical Note on the Method of Least Squares. *Biometrika*, **36**:458–460, 1949.

[218] Plato, R.: Optimal Algorithms for Linear Ill-Posed Problems Yielding Regularization Methods. *Numer. Funct. Anal. Optimization*, **11**:111–118, 1990.

[219] Rado, R.: Note on Generalized Inverses of Matrices. *Proceedings of the Cambridge Philosophical Society*, **52**:600–601, 1956.

[220] Rao, C.R., Mitra S.K.: *Generalized Inverse of Matrices and Its Applications*. Wiley, New York, 1971.

[221] Rayleigh, L.: *The Theory of Sound*. Oxford Univ. Press, 1896.

[222] Reed, M., Simon, B.: *Functional Analysis*. I, Academic Press, New York, 1972.

[223] Reid, W.T.: Generalized Inverses of Differential and Integral Operators. Theory and Applications of Generalized Inverses of Matrices (T.L. Boullion and P.L. Odell, Eds.). Symposium Proceedings, Texas Tech University Mathematics Series, Vol. 4., Lubbock; Texas.

[224] Ribière G.: Régularisation d'opérateurs. *Rev. Informat. Recherche Opérationnelle*, **1**:57–79, 1967.

[225] Richter G.R.: Numerical Solution of Integral Equations of the First Kind with Non-Smooth Kernels. *SIAM J. Numer. Anal.*, **15**:511–522, 1978.

[226] Robinson, D.W.: Gauss and Generalized Inverses. *Historia Mathematics*, **7**:118–125, 1980.

[227] Rieder, A.: *Keine Probleme mit Inversen Problemen*. Vieweg 2003.

[228] Ritz, W.: Über lineare Funktionalgleichungen. *Acta Math.*, **41**:71–98, 1918.

[229] Rudin, L.I.: *Functional Analysis*. Mc Graw-Hill, New York, 1973.

[230] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, **60**: 259, 1992.

[231] Rummel R., Balmino, G., Johannessen, J., Visser, P., Woodworth P.: Dedicated Gravity Field Missions – Principles and Aims. *J. Geodyn.*, **33**:3–20, 2002.

[232] Saitoh, S.: *Theory of Reproducing Kernels and its Applications*. Longman, New York, 1988.

[233] Scherzer, O. (Ed.): *Handbook of Mathematical Methods in Imaging*. Springer, New York, 2015.

[234] Seidmann, T.I.: Non-Convergence Results for the Application of Least-Squares Estimations onto Ill-Posed Problems. *J. Optim. Theory Appl.*, **30**:535–547, 1980.

[235] Shinozaki, Sibuya M., and Ranabe, K.: Numerical algorithms for the Moore–Penrose Inverse of a Matrix: Direct Methods. *Ann. Inst. Statist. Math.* **24**:193–203, 1972.

[236] Showalter D.W., Ben-Israel, B.: Representation and Computation of the Generalized Inverse of a Bounded Linear Operator between two Hilbert Spaces. *Atti Accad. Naz. Kincei Rend. Cl. Sci, Fis. Mat. Natur.*, (8), v. 48:184–194, MR 42, 8310, 1970.

[237] Shure, L., Parker, R.L., Backus, G.E.: Harmonic Splines for Geomagnetic Modelling. *Phys. Earth Planet. Inter.*, **28**:215–229, 1982.

[238] Siegel, C.L.: Über die analytische Theorie der quadratischen Formen, III. Annals of Mathematics **38**:212–291, 1937.

[239] Shin C.E., Sun, Q.; Stability of localized operators. *J. Funct. Anal.*, **256**:2417–2439, 2009.

[240] Sjöberg, L.: Station Adjustment of Directions Using Generalized Inverses. In: Borre, K., Welsch, W. (Eds.), International Federation of Surveyors FIG – Proceedings Survey Control Networks Meeting of Study Group 5B, 7th–9th July, 1982, Aalborg University Centre, Denmark. Schriftenreihe des Wissenschaftlichen Studiengangs Vermessungswesen der Hochschule der Bundeswehr München, Heft 7, 381–399, 1982.

[241] Song, M.: *Regularization-Projection Methods and Finite Element Approximations for Ill-Posed Linear Operator Equations*. Ph.-D. Thesis, Univ. Michigan, 1978.

[242] Strand, O.N.: Theory and Methods Related to the Singular Function Expansion and Landweber's Iteration for Integral Equations of the First Kind. *SIAM J. Numer. Anal.*, **11**:798–825, 1974.

[243] Sudakov V.N., Khalfin, L.A.: A Statistical Approach to the Correctness of the Problems of Mathematical Physics. *Dokl. Akad. Nauk SSSR.*, 157-1058-1060, 1964.

[244] Tadmor, E., Nezzar, S., and Vese, L.: A Multiscale Image Representation Using Hierarchical (BV, L^2) Decompositions. *Multiscale Model. Simul.*, **2**:554–579, 2004.

[245] Taylor, A.E., Lay, D.: *Functional Analysis*. Wiley, New York, 1979.

[246] Tikhonov, A.N.: On the Stability of Inverse Problems. *Dokl. Akad. Nauk SSSR*, **39**:195–198, 1943.

[247] Tikhonov, A.N.: On the Solution of Incorrectly Formulated Problems and the Regularization Method, *Dokl. Akad. Nauk SSSR*, **151**:501–504, 1963.

[248] Tikhonov, A.N.: On Methods of Solving Incorrect Problems, *Amer. Math. Soc. Transl.*, **2**:222–224, 1968.

[249] Tikhonov, A.N., Arsenin V.Y.: *Solutions of Ill-Posed Problems*. Wiley, New York, 1977.

[250] Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V., Yagola, A.G.: *Numerical Methods for the Solution of Ill-Posed Problems*. Kluwer, Dordrecht. 1995.

[251] Twomey, S.: On the Numerical Solution of Fredholm Integral Equations of the First Kind by the Inversion of the Linear System Produced by Quadrature. *J. Assoc. Comp. Mach.*, **10**:97–101, 1963.

[252] Vainberg, M.M.: *Variational Method and Method of Monotone Operators*, Wiley, New York, 1973.

[253] Varah, J.: On the Numerical Solution of Ill-Conditioned Linear Systems with Applications to Ill-Posed Problems. *SIAM J. Numer. Anal.*, **10**:257–267, 1973.

[254] Vogel, C.R.: *Computational Methods for Inverse Problems*. SIAM, Philadelphia, 2002.

[255] Vogel, C.R. Oman, M.E.: Iterative Methods for Total Variation Denoising. *SIAM J.Sci. Comput.*, **17**:227–238, 1996.

[256] Wahba, G.: Convergence Rates of Certain Approximate Solutions of Fredholm Integral Equations of the First Kind. *J. Approx. Theo.*, **7**:167–185. 1973.

[257] Weck, N.: Zwei inverse Probleme in der Potentialtheorie. In: *Mitt. Inst. Theor. Geodäsie, Universität Bonn*, **4**:27–36, 1972.

[258] Weiner, H.W. (Eds.): *Reproducing Kernel Hilbert Spaces*. Applications in Statistical Signal Processing. Hutchinson Ross, Stroudsburg, PA, 1982.

[259] Werner, D.: *Funktionalanalysis*. 4. Auflage; Springer, Berlin, 2002.

[260] Werner, J., Numerische Mathematik 1. Vieweg Studium, Braunschweig, 1991.

[261] Wolf, H.: *Ausgleichungsrechnung. Formeln zur praktischen Anwendung*. Dümmler Verlag, Bonn, 197).

[262] Xia X.G., Nashed M.Z.: The Backus–Gilbert Method for Signals in Reproducing Hilbert Spaces and Wavelet Subspaces. *Inverse Problems*, **10**:785–804, 1994.

[263] Xia X.G., Nashed M.Z.: A Modified Minimum Norm Solution Method for Bandlimited Signal Extrapolation with Inaccurate Data. *Inverse Problems*, **13**:1641–1661, 1997.

[264] Yao, K.: Applications of Reproducing Kernel Hilbert Spaces – Bandlimited Signal Models. *Inform. and Control*, **11**:429–444, 1967.

[265] Yosida, K.: *Functional Analysis*. 5th ed., Springer, Berlin, 1965.

[266] Zhou, L., Li, X, Pan, F.: Gradient-Based Iterative Identification for Wiener Nonlinear Systems with Non-Uniform Sampling. *Nonlinear Dynam.*, **76**:627–634, 2014.

[267] Zwaan, M.: Approximation of the Solution to the Moment Problem in a Hilbert Space, *Numer. Functional Anal. Optimization*, **11**:601–612, 1990.

[268] Zwaan, M.: MRI Reconstruction as a Moment Problem, *Math. Methods Appl. Sci*, **15**:661–675, 1992.

Willi Freeden
Geomathematics Group
University of Kaiserslautern
MPI-Gebäude, Paul-Ehrlich-Str. 26
D-67663 Kaiserslautern, Germany
e-mail: freeden@rhrk.uni-kl.de

M. Zuhair Nashed
Mathematics Department
University of Central Florida
Orlando, USA

Handbook of Mathematical Geodesy
Geosystems Mathematics, 315–458

Geodetic Observables and Their Mathematical Treatment in Multiscale Framework

Willi Freeden and Helga Nutz

Abstract. For the determination of the Earth's gravitational field various types of observations are available nowadays, e.g., from terrestrial gravimetry, airborne gravimetry, satellite-to-satellite tracking, satellite gravity gradiometry, etc. The mathematical relation between these observables on the one hand and the gravitational field and the shape of the Earth on the other hand is called the *integrated concept of physical geodesy*. In this paper, an integrated concept of physical geodesy in terms of harmonic wavelets is presented. Essential tools for approximation are Runge–Walsh type integration formulas relating an integral over an internal sphere to suitable linear combinations of observational functionals, i.e., linear functionals representing the geodetic observables in terms of gravitational quantities on and outside the Earth. A scale discrete version of multiresolution is described for approximating the gravitational potential on and outside the Earth's surface. Furthermore, an exact fully discrete wavelet approximation is developed for the case of bandlimited wavelets. A method for combined global outer harmonic and local harmonic wavelet modeling is proposed corresponding to realistic Earth's models.

Keywords. Integrated wavelet concept, scaling function, Runge–Walsh approximation, geodetic observables, Meissl schemata.

1. Introduction

Gravity as observed on the Earth's surface is the combined effect of the gravitational mass attraction and the centrifugal force due to the Earth's rotation. Under the assumption that the centrifugal force is explicitly known, the determination of the gravity mainly reduces to getting knowledge of the gravitation. According to the classical Newton Law of Gravitation (1687), knowing the density distribution of a body such as the Earth, the gravitational potential can be computed everywhere in the Euclidean space $\mathbb{R}^3$.

Although Earth's gravitational field modeling is always governed by the same physical laws, it changes its nature when it is seen from different spatial and time scales. To be more concrete, if one looks at gravitational field determination on the basis of an increasing spatial magnification and accuracy, we have to go from something that is suitably characterized by a simple mass point, on astronomical scale, to what is described by a global truncated multipole (i.e., outer harmonic) model, at scales corresponding to satellite altimetry, down to wavelengths of about 100 km. By further zooming in we can reach a spatial resolution of about 1 km showing a very complicated pattern, strongly related to the shape of the Earth and to irregular masses inside the Earth's crust. Simultaneously, the error in the knowledge of the gravitational field models goes from 5 Gal, the flattening effect, down to 10 mGal in a today's global model, down to about 10^{-1} mGal at the regional 1 km resolution or even better. There is also a change of the gravitational field in the time scale depending on the time interval under consideration, for instance, gravitational changes due to geotectonic displacements of masses inside the Earth on very long time scales. It changes because of motions of the rotational axis inside the Earth's body and it shows a periodic change because of the continent and ocean reactions to the torques generated by the moon and the sun. Finally, gravitation shows a change because of human activities, for instance, because of the presence of artificial lakes, height's variations in the water-bearing stratum under cities, etc. It is also worth mentioning that there are certain relations between different scales in the time-like behaviour and in the space-like behaviour of the gravitational field. In any way, it may be assumed for global up to regional modeling purposes that the time-like variations of the field are either well predictable (like tides etc.) or so slow as to be neglected, e.g., on the scale of a decade, or so small and local as to be beyond the scope of interest. Thus, global gravitational field modeling as scientific issue is by definition based on the assumption of a stationary gravitational field with a spatial resolution ranging from a worldwide scale down to about 1 km and from about 1000 Gal of the full field down to, at least, 10^{-1} mGal, or even better in some regional areas.

What we would like to present in this contribution are mathematical structures in straightforward continuation to the monograph [19] by which the gravitational part of the gravity field can be approximated progressively better and better, reflecting an increasing flow of observations of terrestrial, airborne and/or satellite type, e.g., terrestrial gravimetry, airborne gravimetry, satellite altimetry, satellite-to-satellite tracking (SST), satellite gravity gradiometry (SGG), etc. More precisely, we shall try to outline the canonical bridge of gravitational field determination from the well-established global outer harmonic approximation corresponding to a spherical Earth to modern multiscale methods involving the actual geometry of the Earth's surface (thereby neglecting, e.g., the small effect of the atmosphere in the outer space).

The so-called disturbing potential is probably the most crucial quantity in gravity field modeling. The disturbing potential is a scalar quantity which is obtained as the difference between the gravity potential of the Earth and the normal

gravity potential of a reference surface, usually an ellipsoid. The deviations of the gravity potential from the normal potential are relatively small. Note that both the gravity potential and the normal gravity potential contain the same centrifugal potential. Thus, the disturbing potential is harmonic in the outer space.

At this stage some remarks should be made in order to clarify our approach in more detail:

1. The mathematical connection between the observables, the gravity field and the shape of the Earth is called the *integrated concept of physical geodesy*.
2. The foundation of the integrated geodesy approach is the fact that every geodetic measurement is a functional which may assumed to be suitably linearizable by introducing, e.g., normal potentials associated to a reference surface such as an ellipsoid. In other words, the relation between the object function, i.e., the geopotential and the data, may be supposed to be linear.
3. More and more measurements refer to satellites and cannot be modeled as functionals of the gravitational potential on the boundary. Although these observations show a denser observational distribution, they are much more difficult to handle, since they show an exponentially spectral smoothing while moving to the outer space. As a consequence, essential knowledge of the gravitational potential should be based on ground observations, but gravitational field modeling cannot be treated only within a boundary-value formulation because of spaceborne observations. This fact is the reason why we do not speak of the "geodetic boundary-value problem (GBVP)" but of the "integrated concept".
4. Concerning the layout of this contribution a particular interest is focussed on the satellite methods SST und SGG, which are introduced within the framework of pseudodifferential operators assuming non-spherical (orbital) geometry.
5. An important feature of our contribution are the so-called Meissl schemata which are graphical illustrations for the conversion of data both on different heights (terrestrial level, satellite orbit) and of different degrees of derivative of the gravitational potential. The comparison between data on the (spherical) Earth's surface and the orbital sphere was primarily carried out by Meissl (1971) and has been transformed by Rummel [60, 61, 63] and by Rummel and van Gelderen [64, 65] into a more general framework concerning relations between different gravity quantities in the framework of outer harmonics. One of our objectives is the extension of the Meissl schemata to the concept of multiscale decomposition of scalar functions, vector, and tensor fields. In principle, we follow the ideas of mathematical classification first presented in [19, 29, 32–34] for the scalar case and extended in the Ph.D.-thesis [58] to the vector and tensor approach.

2. Current state of gravity field determination

Positioning systems are ideally located as far as possible from the Earth, whereas gravity field sensors are ideally located as close as possible to the Earth. Following these basic principles, various positioning and gravity field determination techniques have been designed. Sensors may be sensitive to local or global features of the gravity field. Considering the spatial location of the data, we may distinguish between terrestrial (surface), airborne, and spaceborne methods. Regarding the data type we have various measurement principles of the gravity field (see, for example, [9–11, 51] and the references therein for more details) leading to different types of data.

2.1. Important geodetic observables

(a) *Gravity Measurements:* The force of gravity provides a directional structure to the space above the Earth's surface. It is tangential to the vertical plumb lines and perpendicular to all (level) equipotential surfaces. Any water surface at rest is part of a level surface. Level (equipotential) surfaces are ideal reference surfaces, for example, for heights. The geoid is defined as that level surface of the gravity field which best fits the mean sea level. Gravity vectors can be measured by absolute or relative gravimeters. The highest available accuracy relative gravity measurements are conducted at the Earth's surface. Measurements on ships and in aircrafts deliver reasonably good data only after the removal of inertial noise. Gravity data are converted into gravity anomalies by subtracting a corresponding reference potential derived from a simple gravity field model associated to an, e.g., ellipsoidal surface (see also Appendix A). Gravity anomalies are furthermore converted into mean gravity anomalies by a proper averaging process over well defined areas. It should be pointed out that the distribution of Earth's gravity data on a global scale is far from being homogeneous with large gaps, in particular over oceans but also over land. In addition, the quality of the data is very distinct. Thus, terrestrial gravity data coverage now and in the foreseeable future is far from being satisfactory for the global purpose of geoidal determination (at an accuracy of essentially less than one centimeter).

(b) *Vertical Deflections.* The direction of the gravity vector can be obtained by astronomical positioning. Measurements are only possible on the Earth's surface. Observations of the gravity vector are converted into so-called vertical deflections by subtracting a corresponding reference direction derived from a simple gravity field model associated to an ellipsoidal surface. Vertical deflections are tangential fields of the anomalous potential in a spherical Earth's model. Due to the high measurement effort required to acquire these types of data compared to a gravity measurement, the data density of vertical deflections is much less than that of gravity anomalies. Gravitational field determination based on the observation of vertical deflections and combined with gravity is feasible in smaller areas with good data coverage.

(c) *Satellite Radar Altimetry.* Satellite radar altimetry has demonstrated an impressive capability of mapping the surface of the oceans. The ocean surface is a good approximation of an equipotential surface and, as such, its offset from the geoid at mean sea level (mean in terms of time) is called sea surface topography. This offset, which can be as large as two meters, reflects many effects including the variables salinity, ocean temperature, ocean currents, variable atmospheric conditions such as wind and air pressure perturbations, tides, etc. Since the sea surface topography refers to the geoid, the precise and sufficiently detailed knowledge of the geoid is mandatory.

(d) *Global Gravitational Field Models.* On the basis of all satellite data, collected over the last decades in orbits at different altitudes and inclinations, only long wavelength components of the global gravity field can be recovered. There are two reasons for this fact: First, an orbit as such is rather insensitive to local features of the gravitational field, and this insensitivity increases with increasing orbit altitude. Second, the satellites which can and are being used are flying at altitudes which are too high for a better purpose such as local gravimetry. Therefore, satellite-only global gravity field models are reliable to a moderate maximum degree expressed in a potential representation in terms of spherical harmonics. Considering the shortcomings of satellite-only gravity field models and of the information content of surface data, several institutions have been working for many years on the combination of both data sets. This work in geodesy has resulted in various gravitational field models in terms of spherical harmonics. All gravity field data available worldwide have entered into the production of this model. Therefore, such models represent the latest state of the art in global gravitational field knowledge.

2.2. Satellite concepts and airborne data

The three satellite concepts which are of importance for gravity field determination are satellite-to-satellite tracking in the high-low mode (SST hi-lo), satellite-to-satellite tracking in the low-low mode (SST lo-lo), and satellite gravity gradiometry (SGG). Common to all three concepts is that the determination of the Earth's gravitational field is based on the measurement of the relative motion (in the Earth's gravity field) of test masses.

1. *Satellite-to-Satellite Tracking.* In the case of SST hi-lo the low flying test mass is a low earth orbiter (LEO) and the high flying test masses are the satellites of the GNSS-system (i.e., GPS, GLONASS, Galileo, and Beidou). As, for example, the GNSS-receiver mounted on the LEO always "contacts" four or even more of the GNSS satellites the relative motion of the LEO can be monitored three-dimensionally, i.e., in all three coordinate directions. The lower the orbit of the LEO the higher is its sensitivity with respect to the spatial variations of the gravitational forces but to skin forces as well (atmospheric drag, solar radiation, albedo, etc.). The latter have either to be compensated for by a drag-free mechanism or be measured by a three axis accelerometer.

Also the high orbiters, the GNSS satellites, are affected by non-gravitational forces. However the latter can be modeled quite well. They affect mainly the very long spatial scales, and to a large extent their effect averages out. In addition, the ephemerides of the GNSS satellites are determined very accurately by the large network of ground stations. In the case of SST lo-lo the relative motion between two LEOs, chasing each other, is measured with highest precision. The quantity of interest is the relative motion of the centre of mass of the two satellites. Again, the effect of non-gravitational forces on the two spacecraft either has to be compensated actively or be measured.

2. *Satellite Gravity Gradiometry.* The satellite gravity gradiometry technique is the measurement of the relative acceleration, not between free falling test masses like satellites, but of test masses at different locations inside one satellite. Each test mass is enclosed in a housing and kept levitated (floating, without ever touching the walls) by a capacitive or inductive feedback mechanism. The difference in feedback signals between two test masses is proportional to their relative acceleration and exerted purely by the differential gravitational field. Non-gravitational acceleration of the spacecraft affects all accelerometers inside the satellite in the same manner and so ideally drops out during differencing. The rotational motion of the satellite affects the measured differences. However, the rotational signal (angular velocities and accelerations) can be separated from the gravitational signal, if acceleration differences are taken in all possible (spatial) combinations (= full tensor gradiometer). In order to achieve a higher sensitity, an orbit as low as possible is of great importance.

In a unified view on spaceborne missions (see, e.g., [9–11, 51]), one can argue that the basic observable in all three cases is gravitational acceleration. In the case of SST hi-lo, with the motion of the high orbiting GNSS satellites assumed to be perfectly known, this corresponds to an in situ 3-D acceleration measurement in the LEO. For SST lo-lo it is the measurement of acceleration difference over the intersatellite distance and in the line-of-sight (LOS) of the LEOs. Finally, in the case of gradiometry, it is the measurement of acceleration differences in 3-D over the tiny baseline of the gradiometer. In short we are confronted with the following situation:

SST hi-lo:	3-D acceleration	= gravitational gradient,
SST lo-lo:	acceleration difference	= difference in gradient,
SGG:	differential	= gradient of gradient (“tensor”).

As explained in more detail by W. Freeden [19], in mathematical sense, it is a transition from the first derivative of the gravitational potential via a difference in the first derivative to the second derivative. The guiding parameter that determines sensitivity with respect to the spatial scales of the Earth’s gravitational potential is the distance between the test masses, being almost infinite for SST hi-lo and almost zero for gradiometry.

3. *Airborne Gravimetry.* Airborne gravimetry is a highly sensitive detection method of the gravitational potential of the Earth by a gravity accelerometer mostly for regional and/or local purposes. Proposals to implement airborne gravimetry go back to the late fifties of the last century, and first flight experiments were already done in the early sixties. A major obstacle of such techniques at that time was the inaccuracy of navigational information (e.g., velocity and acceleration of the space vehicle) which is needed to obtain the desired precision. Although at an appropriate level of accuracy airborne gravimetry is vastly superior in economy and efficiency to pointwise terrestrial methods, there were serious doubts in the seventies and eighties of ever achieving useful results. In the early nineties, however, great advances in GNSS technology opened new ways to resolve the navigational problems. More explicitly, altitude, position, and velocity of the airborne gravity system become sufficiently computable from the inertial measurements updated by GNSS carrier phase and Doppler observations. Vehicle accelerations are derivable from GNSS data only, so that in a third step the airborne gravity disturbance is determinable from the difference between the force vector and the GNSS-derived acceleration vector. Nowadays, some industrial companies are perfecting their system concepts by paying careful attention to the operational conditions under which an airborne gravimeter works best, also for progress in gravimetric exploration.

All in all, over the last decades, geoscientists have realized the great complexity of the Earth and its environment. In particular, the knowledge of the gravity potential and its level (equipotential) surfaces have become an important issue. It was realized that dedicated highly accurate gravity field sensors, when operating in an isolated manner, have their shortcomings, and combining data from different sensors is therefore the way forward. At this stage of development, the global determination of the Earth's gravitational field is a mathematical challenge which should include the numerical progress obtainable by modern multiscale approximation.

2.3. Gravity field applications

The knowledge of the gravitational field of the Earth is of great importance for many applications from which we only mention some significant examples (cf. [19, 61]):

(i) *Geodesy and Civil Engineering.* Accurate heights are needed for civil constructions, mapping, etc. They are obtained by leveling, a very time consuming and expensive procedure. Nowadays, geometric heights can be obtained fast and efficiently from space positioning (GNSS). The geometric heights are convertible to leveled heights by subtracting the precise geoid, which is achieved by a high resolution gravitational potential. To be more specific, in those areas where good gravity information is available already, the future data information will eliminate all medium and long wavelength distortions in

unsurveyed areas. For example, GNSS (GPS, GLONASS, Galileo, or Beidou) together with today's satellite missions provide high quality height information at global scale.

(ii) *Satellite Orbits.* For any positioning from space, the uncertainty in the orbit of the spacecraft is the limiting factor. The spaceborne techniques eliminate basically all gravitational uncertainties in satellite orbits.

(iii) *Solid Earth Physics.* The gravity anomaly field derivable from future satellite observations has its origin mainly in mass inhomogeneities of the continental and oceanic lithosphere. Together with height information and regional tomography, a much deeper understanding of tectonic processes is obtainable.

(iv) *Physical Oceanography.* Altimeter satellites in combination with a precise geoid deliver global dynamic ocean topography. From ocean topography, global surface circulation and its variations in time can be computed resulting in efficient ocean modeling. Circulation allows the determination of transport processes of, e.g., polluted material. Moreover, ocean modeling is an important indicator of climate change.

(v) *Earth System.* There is a growing awareness of global environmental problems (for example, the CO_2-question, the rapid decrease of rain forests, global sea level changes, etc.). What is the role of the airborne methods and satellite missions in this context? They do not tell us the reasons for physical processes, but it is essential to bring the phenomena into one system (e.g., to make sea level records comparable in different parts of the world). In other words, equipotential surfaces such as the geoid may be viewed as an almost static reference for many rapidly changing processes and at the same time as a "frozen picture" of tectonic processes that evolve in geological time spans.

(vi) *Exploration Geophysics and Prospecting.* Knowledge of local geologic structures can easily be gained by means of terrestrial and airborne data so gravity prospecting can be done over land or sea areas using different techniques and equipment. Terrestrial gravimetry was first applied to prospect for salt domes (e.g., in the Gulf of Mexico), and later for looking for anticlines in continental areas. In future, embedded in (regional) airborne and (global) spaceborne gravity information such as satellite-to-satellite tracking (SST) and/or satellite gravity gradiometry (SGG) (see, e.g., [19, 27, 32] and the references therein), new promising components in gravimetrically oriented modeling can be expected, for example, based on multiscale modeling providing reconstruction and decomposition of geological signatures.

2.4. Principles of multiscale approximation

Spaceborne observation combined with terrestrial and airborne activities provide huge datasets of the order of millions of data (see [9–11, 51, 63]). Standard mathematical theory and numerical methods are not at all adequate for the solution of data systems with such a structure, because these methods are not adapted to the specific properties of the data set. They quickly reach their capacity limit even on very powerful computers. An adequate reconstruction of the gravitational

field from the huge and heterogeneous data material requires a careful multiscale analysis of the gravitational potential, fast solution techniques, and a proper stabilization of the inverse character of satellite problems by regularization. In order to achieve these objectives various strategies and structures must be introduced reflecting the different aspects of geopotential determination. While global long-wavelength modeling can be adequately done by use of spherical harmonic expansions, it becomes more and more obvious that harmonic splines and/or wavelets are most likely the candidates for medium and short-wavelength approximation. The concept of harmonic wavelets, however, demands its own nature which only on exploration areas of small size may be developed to some extend from the theory in Euclidean spaces. Fundamental results known from the Euclidean wavelet approach have to be recovered. Nevertheless, the stage is set for working out and improving essential ideas and results involving harmonic wavelets. Why are harmonic wavelets important in future gravitational potential determination? Following [19], the answer is summarized in the following sentence:

Harmonic wavelets are "building blocks" that enable fast decorrelation of gravitational data. Thus three features are incorporated in this way of thinking about georelevant harmonic wavelets, namely basis property, decorrelation, and efficient algorithms. These aspects should be discussed in more detail:

(i) *Basis property*
Wavelets are building blocks for the approximation of arbitrary functions (signals). In mathematical understanding this formulation expresses that the set of wavelets forms a "frame" (see, e.g., [6] for details in classical one-dimensional theory).

(ii) *Decorrelation*
Wavelets possess the ability to decorrelate the signal. This means, that the representation of the signal via wavelet coefficients occurs in a "more constituting" form as in the original form reflecting a certain amount of space and frequency (more accurately, momentum) information. The decorrelation enables the extraction of specific information contained in a signal through a particular number of coefficients. Signals usually show a correlation in the frequency (momentum) domain as well as in the space domain. Obviously, since data points in a local neighborhood are stronger correlated as those data points far-off from each other, signal characteristics often appear in certain frequency bands. In order to analyze and reconstruct such signals, we need "auxiliary functions" providing localized information in the space as well as in the frequency domain. In applications, different approaches have been realized in the field of signal analysis before the occurrence of wavelets: on the one hand, the Fourier theory allows a trendsetting bandlimited decomposition, on the other hand, the Haar theory offers short-wavelets spacelimited decomposition. The (Heisenberg) uncertainty principle (see, e.g., [21]) tells us that a simultaneous sharp localization in frequency as well as space domain is exclusive. Even more within a "zooming-in process", the amount of

frequency as well as space contribution can be specified in quantitative way. A so-called scaling function forms a compromise in which a certain balanced amount of frequency and space localization in the sense of the uncertainty principle is realized. In consequence, each scaling function depends on two variables, namely a "shifting" and a scaling parameter, which control the amount of the space localization to be available at the price of the frequency localization, and vice versa. Associated to each scaling function is a wavelet function, which here is simply understood to be the difference of two successive scaling functions. All in all, filtering (convolution) with a scaling function takes the part of a lowpass filter, while convolution with the corresponding wavelet function provides a bandpass filtering. A multiscale approximation of a signal is the successive execution of an efficient evaluation process by use of scaling and wavelet functions which show more and more space localization at the cost of frequency localization. The wavelet transform within a multiscale approximation lays the foundation for the decorrelation of a signal.

(iii) *Efficient algorithms*
Wavelet transformation provides efficient algorithms because of the space-localizing character. The successive decomposition of the signal by use of wavelets at different scales offers the advantage for efficient and economic numerical calculation (e.g., tree algorithm). The detail information stored in the wavelet coefficients leads to a reconstruction from a rough to a fine resolution and to a decomposition from fine to rough resolution in form of tree algorithms. In particular, the decomposition algorithm is an excellent tool for the post-processing of a signal into "constituting blocks" by decorrelation, e.g., the specification of signature bands corresponding to certain geological formations.

3. Geodetically relevant Sobolev spaces

We start our mathematical foundation of Meissl schemata by introducing some basic information related to the theory of geodetic observables within the framework of Sobolev spaces. We adopt the following general scheme of notation which is non-standard in geodesy, but extremely helpful in establishing Meissl schemata especially for the vectorial and tensorial framework. Capital letters (F, G, ...) are used for scalar functions, small letters (f, g, ...) represent vector fields and small boldface letters ($\mathbf{f}$, $\mathbf{g}$, ...) represent tensor fields of second rank. As usual, a scalar function having k continuous derivatives is said to be of class $C^{(k)}$ whereas L^2 denotes the Hilbert space of square integrable functions. A vector field having k continuous derivatives is said to be of class $c^{(k)}$ and l^2 denotes the Hilbert space of square-integrable vector fields. Finally, the space of all tensor fields having k continuous derivatives is denoted by $\mathbf{c}^{(k)}$ and $\mathbf{l}^2$ denotes the Hilbert space of all square-integrable tensor fields.

$\Sigma \subset \mathbb{R}^3$ is called a regular surface if Σ is the boundary of a regular region $\Sigma^{\text{int}} \subset \mathbb{R}^3$, i.e., $\Sigma = \partial\Sigma^{\text{int}}$, with the following properties (cf. [20]):

(i) Σ constitutes an orientable piecewise smooth Lipschitzian manifold of dimension 2.
(ii) The origin is contained in Σ^{int}.
(iii) Σ divides $\mathbb{R}^3$ into the "inner space" Σ^{int} and the "outer space" $\Sigma^{\text{ext}} = \mathbb{R}^3 \backslash \overline{\Sigma^{\text{int}}}$, $\overline{\Sigma^{\text{int}}} = \Sigma^{\text{int}} \cup \Sigma$.

Georelevant regular surfaces Σ are, for example, the sphere, the ellipsoid, the telluroid, the geoid, and the regular Earth's surface.

The geometric concept to be discussed in our approach is as follows (see Figure 3.1): Σ denotes the Earth's surface which we assume to be known and

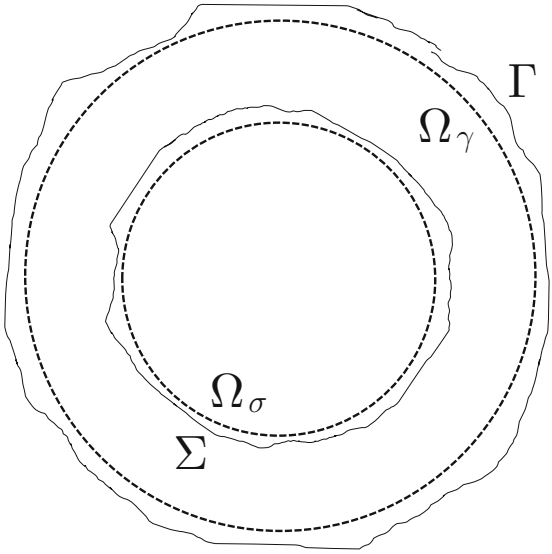


FIGURE 3.1. Geometric concept characterizing the surface of the Earth Σ and the orbit of a satellite Γ.

regular. Γ is the orbit of a satellite which is not necessarily a closed surface. σ is the radius of a so-called Runge (in the jargon of geodesy, Bjerhammar) sphere inside the Earth, that is $\sigma < \alpha = \inf_{x\in\Sigma} |x|$. The value γ is a lower bound of the lowest possible altitude of the satellite, i.e., $\gamma < \inf_{x\in\Gamma} |x|$. $\Omega_\sigma^{\text{ext}} = \{x \in \mathbb{R}^3 \ : \ |x| > \sigma\}$ denotes the outer space of the sphere Ω_σ with radius σ around the origin 0, whereas Σ^{ext} denotes the outer space of the (actual) Earth.

Let $V : \Omega_\sigma^{\text{ext}} \to \mathbb{R}$, $v : \Omega_\sigma^{\text{ext}} \to \mathbb{R}^3$, and $\mathbf{v} : \Omega_\sigma^{\text{ext}} \to \mathbb{R}^3 \otimes \mathbb{R}^3$, respectively, be a scalar, vector, and tensor field on the set $\Omega_\sigma^{\text{ext}}$. We say that V, v, $\mathbf{v}$, respectively, are *harmonic* on $\Omega_\sigma^{\text{ext}}$ if V, v, $\mathbf{v}$ are twice continuously differentiable on $\Omega_\sigma^{\text{ext}}$ and $\Delta V = 0$, $\Delta v = 0$, $\Delta \mathbf{v} = 0$ on $\Omega_\sigma^{\text{ext}}$.

Without proof we mention some well-known theorems concerning harmonic fields on $\Omega_\sigma^{\text{ext}}$ (for the proofs see, for example, [20, 38, 47]):

(1) Every harmonic field in $\Omega_\sigma^{\text{ext}}$ is analytic in $\Omega_\sigma^{\text{ext}}$, i.e., every harmonic field is determined by its local properties.

(2) *Harnack's convergence theorem:* Let $V_\delta : \Omega_\sigma^{\text{ext}} \to \mathbb{R}$, $v_\delta : \Omega_\sigma^{\text{ext}} \to \mathbb{R}^3$, and $\mathbf{v}_\delta : \Omega_\sigma^{\text{ext}} \to \mathbb{R}^3 \otimes \mathbb{R}^3$, respectively, be harmonic on $\Omega_\sigma^{\text{ext}}$ for each value δ ($0 < \delta < \delta_0$), and regular at infinity. Moreover, let

$$V_\delta \to V, \quad \delta \to 0,\ \delta > 0, \tag{3.1}$$

$$v_\delta \to v, \quad \delta \to 0,\ \delta > 0, \tag{3.2}$$

$$\mathbf{v}_\delta \to \mathbf{v}, \quad \delta \to 0,\ \delta > 0, \tag{3.3}$$

uniformly on each subset K of $\Omega_\sigma^{\text{ext}}$ with $\text{dist}(\overline{K}, \partial\Omega_\sigma^{\text{ext}}) > 0$. Then $V : \Omega_\sigma^{\text{ext}} \to \mathbb{R}$, $v : \Omega_\sigma^{\text{ext}} \to \mathbb{R}^3$, and $\mathbf{v} : \Omega_\sigma^{\text{ext}} \to \mathbb{R}^3 \otimes \mathbb{R}^3$, respectively, is harmonic on $\Omega_\sigma^{\text{ext}}$ and regular at infinity.

(3) Let $V : \overline{\Omega_\sigma^{\text{ext}}} \to \mathbb{R}$ be twice continuously differentiable on $\Omega_\sigma^{\text{ext}}$ and continuous on $\overline{\Omega_\sigma^{\text{ext}}}$, i.e., $V \in C^{(0)}(\overline{\Omega_\sigma^{\text{ext}}}) \cap C^{(2)}(\Omega_\sigma^{\text{ext}})$, harmonic on $\Omega_\sigma^{\text{ext}}$, and regular at infinity. Then the *maximum/minimum principle* tells us that

$$\sup_{x \in \overline{\Omega_\sigma^{\text{ext}}}} |V(x)| \leq \sup_{x \in \Omega_\sigma} |V(x)|\,. \tag{3.4}$$

(4) There is a so-called fundamental solution (singularity function) $S : x \mapsto |x-y|^{-1}$, $x \neq y$, with respect to the Laplace operator Δ such that *the fundamental theorem of potential theory*

$$\int_{\partial\Omega_\sigma^{\text{ext}}} \left(\frac{1}{|x-y|} \frac{\partial V}{\partial \nu}(y) - V(y) \frac{\partial}{\partial \nu_y} \frac{1}{|x-y|} \right) d\omega(y)$$
$$= \begin{cases} -4\pi V(x), & x \in \Omega_\sigma^{\text{ext}}, \\ -2\pi V(x), & x \in \partial\Omega_\sigma^{\text{ext}}, \\ 0, & x \notin \overline{\Omega_\sigma^{\text{ext}}}, \end{cases}$$

holds true.

3.1. Scalar outer harmonic and Sobolev theory

As already explained, we let $\Omega_\sigma \subset \mathbb{R}^3$ be the sphere around the origin with radius $\sigma > 0$, $\Omega_\sigma^{\text{int}}$ is the inner space of Ω_σ, and $\Omega_\sigma^{\text{ext}}$ is the outer space. We let $\Omega = \Omega_1$. By virtue of the isomorphism $\Omega \ni \xi \mapsto \sigma\xi \in \Omega_\sigma$ we assume functions $F : \Omega_\sigma \to \mathbb{R}$ to be defined on Ω. It is clear that the function spaces defined on Ω admit their natural generalizations as spaces of functions defined on Ω_σ. We have, for example, $C^{(\infty)}(\Omega_\sigma)$, $L^p(\Omega_\sigma)$, etc.

Let $\{Y_{n,m}\}_{n \in \mathbb{N}_0; m=1,\ldots,2n+1}$ be an L^2-orthonormal system of (surface) spherical harmonics. Obviously, such an $L^2(\Omega)$-orthonormal system of spherical harmonics forms an orthogonal system on Ω_σ (with respect to $(\cdot,\cdot)_{L^2(\Omega_\sigma)}$). More explicitly,

we have

$$(Y_{n,k}, Y_{p,q})_{L^2(\Omega_\sigma)} = \int_{\Omega_\sigma} Y_{n,k}\left(\frac{x}{|x|}\right) Y_{p,q}\left(\frac{x}{|x|}\right) d\omega(x) = \sigma^2 \delta_{n,p}\delta_{k,q}, \tag{3.5}$$

where $\delta_{n,p}$ is the Kronecker symbol and $d\omega$ is the surface element. With the relationship $\xi \leftrightarrow \sigma\xi$, the *surface gradient* $\nabla^{*;\sigma}$ and the *Beltrami operator* $\Delta^{*;\sigma}$ on Ω_σ, respectively, have the representation $\nabla^{*;\sigma} = (1/\sigma)\nabla^{*;1} = (1/\sigma)\nabla^*$, $\Delta^{*;\sigma} = (1/\sigma^2)\Delta^{*;1} = (1/\sigma^2)\Delta^*$, where ∇^*, Δ^* are the surface gradient and the Beltrami operator of the unit sphere Ω.

We now introduce the system $\{Y^\sigma_{n,k}\}_{n=0,1,\ldots;\, k=1,\ldots,2n+1}$ by letting

$$Y^\sigma_{n,k}(x) = \frac{1}{\sigma} Y_{n,k}\left(\frac{x}{|x|}\right), \quad x \in \Omega_\sigma. \tag{3.6}$$

Due to (3.5) the system $\{Y^\sigma_{n,k}\}_{n=0,1,\ldots;k=1,\ldots,2n+1}$ is an orthonormal basis in $L^2(\Omega_\sigma)$:

$$L^2(\Omega_\sigma) = \overline{\operatorname{span}_{\substack{n=0,1,\ldots;\\ k=1,\ldots,2n+1}} (Y^\sigma_{n,k})}^{\|\cdot\|_{L^2(\Omega_\sigma)}}. \tag{3.7}$$

The system $\{H_{n,m}(\sigma;\cdot)\}_{n\in\mathbb{N}_0;m=1,\ldots,2n+1}$, of *scalar outer harmonics* defined by

$$H_{n,m}(\sigma;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n+1} Y_{n,m}\left(\frac{x}{|x|}\right), \quad x \in \overline{\Omega^{\text{ext}}_\sigma},$$

satisfies the following properties:

- $H_{n,m}(\sigma;\cdot)$ is of class $C^{(\infty)}(\Omega^{\text{ext}}_\sigma)$,
- $H_{n,m}(\sigma;\cdot)$ is harmonic in $\Omega^{\text{ext}}_\sigma$, i.e., $\Delta_x H_{n,m}(\sigma;x) = 0$ for $x \in \Omega^{\text{ext}}_\sigma$,
- $H_{n,m}$ is regular at infinity, i.e., $|H_{n,m}(\sigma;x)| = \mathcal{O}(|x|^{-1})$, $x| \to \infty$,
- $H_{n,m}(\sigma;\cdot)|_{\Omega_\sigma} = \frac{1}{\sigma} Y_{n,m}$,
- $\int_{\Omega_\sigma} H_{n,m}(\sigma;x) H_{k,l}(\sigma;x) d\omega(x) = \delta_{n,k}\delta_{m,l}$.

As it is well known (cf., e.g., [32, 57]), the *addition theorem of outer harmonics* reads as follows:

$$\sum_{m=1}^{2n+1} H_{n,m}(\sigma;x) H_{n,m}(\sigma;y) = \frac{2n+1}{4\pi\sigma^2}\left(\frac{\sigma^2}{|x|\ |y|}\right)^{n+1} P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right), \tag{3.8}$$

for all $(x,y) \in \overline{\Omega^{\text{ext}}_\sigma} \times \overline{\Omega^{\text{ext}}_\sigma}$ and $n \in \mathbb{N}_0$, where P_n denotes the Legendre polynomial of degree n. $Harm_n(\overline{\Omega^{\text{ext}}_\sigma})$ denotes the space of all outer harmonics of order n, $n \in \mathbb{N}_0$:

$$Harm_n(\overline{\Omega^{\text{ext}}_\sigma}) = \operatorname{span}_{m=1,\ldots,2n+1}(H_{n,m}(\sigma;\cdot)).$$

It is well known that $\dim(Harm_n(\overline{\Omega^{\text{ext}}_\sigma})) = 2n+1$. We let $Harm_{p,\ldots,q}(\overline{\Omega^{\text{ext}}_\sigma})$ be the space of all linear combinations of the functions $H_{n,m}(\sigma;\cdot)$ on $\overline{\Omega^{\text{ext}}_\sigma}$, $n = p,\ldots,q$, $m = 1,\ldots,2n+1$, i.e.,

$$Harm_{p,\ldots,q}(\overline{\Omega^{\text{ext}}_\sigma}) = \bigoplus_{n=p}^{q} Harm_n(\overline{\Omega^{\text{ext}}_\sigma}).$$

The space $Harm_{p,\dots,q}(\overline{\Omega_\sigma^{\rm ext}})$ has the reproducing kernel $K_{Harm_{p,\dots,q}(\overline{\Omega_\sigma^{\rm ext}})}(\cdot,\cdot)$ given by

$$K_{Harm_{p,\dots,q}(\overline{\Omega_\sigma^{\rm ext}})}(x,y) = \sum_{n=p}^{q}\sum_{m=1}^{2n+1} H_{n,m}(\sigma;x)H_{n,m}(\sigma;y)$$
$$= \sum_{n=p}^{q}\frac{2n+1}{4\pi\sigma^2}\left(\frac{\sigma^2}{|x|\,|y|}\right)^2 P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right). \tag{3.9}$$

$Pot(\Sigma^{\rm ext})$ denotes the space of all functions (potentials) $U:\Sigma^{\rm ext}\to\mathbb{R}$ with

- $U\in C^{(2)}(\Sigma^{\rm ext})$,
- U satisfies the Laplace equation in the outer space, i.e., $\Delta_x U(x)=0, x\in\Sigma^{\rm ext}$,
- U is regular at infinity, i.e., $|U(x)|=O(|x|^{-1})$, $|x|\to\infty$.

As usual, for $k=0,1,\dots$, we let $Pot^{(k)}(\overline{\Sigma^{\rm ext}})$ be the space of functions $F:\overline{\Sigma^{\rm ext}}\to\mathbb{R}$ such that $F|_{\Sigma^{\rm ext}}\in Pot(\Sigma^{\rm ext})$ and $F\in C^{(k)}(\overline{\Sigma^{\rm ext}})$, in brief,

$$Pot^{(k)}(\overline{\Sigma^{\rm ext}}) = Pot(\Sigma^{\rm ext})\cap C^{(k)}(\overline{\Sigma^{\rm ext}}). \tag{3.10}$$

It is known from [13] and [17] that

$$L^2(\Sigma) = \overline{\operatorname{span}_{\substack{n=0,1,\dots;\\ m=1,\dots,2n+1}}(H_{n,m}(\sigma;\cdot))|_\Sigma}^{\|\cdot\|_{L^2(\Sigma)}}, \tag{3.11}$$

$$C^{(0)}(\Sigma) = \overline{\operatorname{span}_{\substack{n=0,1,\dots;\\ m=1,\dots,2n+1}}(H_{n,m}(\sigma;\cdot))|_\Sigma}^{\|\cdot\|_{C^{(0)}(\Sigma)}}. \tag{3.12}$$

Furthermore (cf. [13]),

$$Pot^{(0)}(\overline{\Sigma^{\rm ext}}) = \overline{\operatorname{span}_{\substack{n=0,1,\dots;\\ m=1,\dots,2n+1}}(H_{n,m}(\sigma;\cdot))|_{\overline{\Sigma^{\rm ext}}}}^{\|\cdot\|_{C^{(0)}(\overline{\Sigma^{\rm ext}})}}. \tag{3.13}$$

Next we introduce *Sobolev spaces* $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ (cf. [14]). We start with a general definition based on the concept of summable sequences, give some examples for spaces with a reproducing kernel structure, and, finally, introduce the well-known $\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$-spaces.

The introduction of the Sobolev spaces may be based on a linear space $\mathcal{A}$ consisting of all sequences $\{A_n\}$ of real numbers $A_n, n=0,1,\dots$, i.e.,

$$\mathcal{A} = \{\{A_n\}\ :\ A_n\in\mathbb{R},\ n=0,1,\dots\}.$$

For given sequences $\{A_n\},\{B_n\}\in\mathcal{A}$ we denote by $\mathcal{N}(B_n^{-1}A_n)$ the set of all non-negative integers n for which $B_nA_n^{-1}$ exists and is different from 0. Let $\mathcal{N}_0(B_n^{-1}A_n)$ denote the complement of $\mathcal{N}(B_n^{-1}A_n)$ in $\mathbb{N}_0$. Consequently, it follows that $\mathbb{N}_0=\mathcal{N}(B_n^{-1}A_n)\cup\mathcal{N}_0(B_n^{-1}A_n)$ and $\mathcal{N}(B_n^{-1}A_n)\cap\mathcal{N}_0(B_n^{-1}A_n)=\emptyset$. In particular, if $\{B_n\}$ is chosen such that $B_n=1$ for all $n\in\mathbb{N}_0$, $\mathcal{N}(A_n)$ is the set of all integers $n\in\mathbb{N}_0$ for which $A_n\neq 0$, and $\mathcal{N}_0(A_n)$ is the set of all integers $n\in\mathbb{N}_0$ with $A_n=0$. Further on $\mathcal{N}(A_n)$ is always assumed to be non-void. Moreover, we write $\mathcal{N}$ instead of $\mathcal{N}(A_n)$ if no confusion is likely to arise.

Consider the set $\mathcal{E}(\overline{\Omega_\sigma^{\rm ext}})$ $\left(=\mathcal{E}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}})\right)$ of all functions $F\in Pot^{(\infty)}(\overline{\Omega_\sigma^{\rm ext}})$ of the form

$$F=\sum_{n\in\mathcal{N}}\sum_{m=1}^{2n+1}F^\wedge(n,m)H_{n,m}(\sigma;\cdot) \tag{3.14}$$

with

$$F^\wedge(n,m)=F^{\wedge_{L^2(\Omega_\sigma)}}(n,m)=\int_{\Omega_\sigma}F(y)H_{n,m}(\sigma;y)\,d\omega(y)$$

satisfying

$$\sum_{n\in\mathcal{N}}\sum_{m=1}^{2n+1}A_n^2\,(F^\wedge(n,m))^2<\infty \tag{3.15}$$

(note that $\Sigma_{n\in\mathcal{N}}$ means that the sum is extended over all non-negative integers n with $n\in\mathcal{N}$). From the Cauchy–Schwarz inequality it follows that

$$\begin{aligned}&\left|\sum_{n\in\mathcal{N}}\sum_{m=1}^{2n+1}A_n^2\,F^\wedge(n,m)G^\wedge(n,m)\right| \\ &\quad\le\left(\sum_{n\in\mathcal{N}}\sum_{m=1}^{2n+1}A_n^2\,(F^\wedge(n,m))^2\right)^{1/2}\left(\sum_{n\in\mathcal{N}}\sum_{m=1}^{2n+1}A_n^2\,(G^\wedge(n,m))^2\right)^{1/2}\end{aligned} \tag{3.16}$$

for all $F,G\in\mathcal{E}(\overline{\Omega_\sigma^{\rm ext}})$, hence, the left-hand side of (3.16) is finite whenever each member of the right-hand side is finite. This is the reason why we are able to impose on $\mathcal{E}(\overline{\Omega_\sigma^{\rm ext}})$ an inner product $(\cdot,\cdot)_{\mathcal{H}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}})}$ by letting

$$(F,G)_{\mathcal{H}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}})}=\sum_{n\in\mathcal{N}}\sum_{m=1}^{2n+1}A_n^2\,F^\wedge(n,m)G^\wedge(n,m). \tag{3.17}$$

The associated norm is given by

$$\|F\|_{\mathcal{H}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}})}=\left(\sum_{n\in\mathcal{N}}\sum_{m=1}^{2n+1}A_n^2\,(F^\wedge(n,m))^2\right)^{1/2}. \tag{3.18}$$

Summarizing our results we therefore obtain the following definition.

Definition 3.1. The *Sobolev space* $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ (more accurately: $\mathcal{H}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}})$) is the completion of $\mathcal{E}(\overline{\Omega_\sigma^{\rm ext}})(=\mathcal{E}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}}))$ under the norm $\|\cdot\|_{\mathcal{H}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}})}$:

$$\mathcal{H}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}})=\overline{\mathcal{E}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}})}^{\|\cdot\|_{\mathcal{H}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}})}}.$$

$\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ equipped with the inner product corresponding to the norm (3.18) is a Hilbert space. The system $\{H_{n,m}^{*\{A_n\}}(\sigma;\cdot)\}$ given by

$$H_{n,m}^{*\{A_n\}}(\sigma;x)=A_n^{-1}H_{n,m}(\sigma;x),\quad x\in\overline{\Omega_\sigma^{\rm ext}}, \tag{3.19}$$

is a Hilbert basis. We simply write $H_{n,m}^*(\sigma;\cdot)$ instead of $H_{n,m}^{*\{A_n\}}(\sigma;\cdot)$ if no confusion is likely to arise.

Consider the Beltrami operator $\Delta^{*;\sigma}$ on the sphere Ω_σ. We know that

$$\Delta^{*;\sigma} Y_{n,m} = \frac{1}{\sigma^2}\Delta^* Y_{n,m} = -\frac{1}{\sigma^2} n(n+1) Y_{n,m}$$

for $n \in \mathbb{N}_0$; $m = 1, \ldots, 2n+1$ (note that $\Delta^{*;1} = \Delta^*$). Thus we formally have

$$\left(-\Delta^{*;\sigma} + \frac{1}{4\sigma^2}\right)^{s/2} Y_{n,k} = \left(\frac{n+\frac{1}{2}}{\sigma}\right)^s Y_{n,m}$$

and

$$\left(\left(-\Delta^{*;\sigma} + \frac{1}{4\sigma^2}\right)^{s/2} F\right)^\wedge (n,m) = \left(\frac{n+\frac{1}{2}}{\sigma}\right)^s F^\wedge(n,m)$$

for all $n \in \mathbb{N}_0$; $m = 1, \ldots, 2n+1$.

Definition 3.2. For any given value $s \in \mathbb{R}$, the *Sobolev space* $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ is the completion of $\mathcal{E}(\overline{\Omega_\sigma^{\text{ext}}})$ under the norm $\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}$:

$$\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) = \overline{\mathcal{E}(\overline{\Omega_\sigma^{\text{ext}}})}^{\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}}.$$

$\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ equipped with the inner product $(\cdot,\cdot)_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}$ is a Hilbert space. The system $\{H_{n,m}^s(\sigma;\cdot)\}$ given by

$$H_{n,m}^s(\sigma;x) = \left(\frac{\sigma}{n+\frac{1}{2}}\right)^s H_{n,m}(\sigma;x), \quad x \in \overline{\Omega_\sigma^{\text{ext}}}, \tag{3.20}$$

is a Hilbert basis.

Hence, the norm in $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ reads as follows:

$$\|F\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} = \left(\int_{\Omega_\sigma} \left(\left(-\Delta_x^{*;\sigma} + \frac{1}{4\sigma^2}\right)^{s/2} F(x)\right)^2 d\omega(x)\right)^{1/2}. \tag{3.21}$$

$\mathcal{H}_0(\overline{\Omega_\sigma^{\text{ext}}})$ may be understood as the space of all harmonic functions in $\Omega_\sigma^{\text{ext}}$, regular at infinity, corresponding to L^2-restrictions (note that the potentials in $\mathcal{H}_0(\overline{\Omega_\sigma^{\text{ext}}})$ are uniquely determined by their L^2-(Dirichlet) boundary conditions on Ω_σ). According to our construction, $Pot^{(\infty)}(\overline{\Omega_\sigma^{\text{ext}}})$ is a dense subspace of $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ for each s. If $t < s$, then $\|F\|_{\mathcal{H}_t(\overline{\Omega_\sigma^{\text{ext}}})} \le \|F\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}$ and $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \subset \mathcal{H}_t(\overline{\Omega_\sigma^{\text{ext}}})$.

If we associate to U the outer harmonic expansion (3.14) it is of fundamental importance to know when the series (3.14) converges uniformly on the whole set $\overline{\Omega_\sigma^{\text{ext}}}$. To this end we need the concept of summable sequences.

Definition 3.3. A sequence $\{A_n\}_{n\in\mathbb{N}_0} \in \mathcal{A}$ is called summable if

$$\sum_{n=0}^{\infty} \frac{2n+1}{A_n^2} < \infty. \tag{3.22}$$

Lemma 3.4 (Sobolev Lemma). *Assume that the sequences* $\{A_n\}_{n\in\mathbb{N}_0}, \{B_n\}_{n\in\mathbb{N}_0} \in \mathcal{A}$ *are such that* $\{B_n^{-1}A_n\}_{n\in\mathbb{N}_0}$ *is summable. Then each* $F \in \mathcal{H}\left(\{B_n^{-1}A_n\}; \overline{\Omega_\sigma^{\text{ext}}}\right)$ *corresponds to a potential of class* $Pot^{(0)}(\overline{\Omega_\sigma^{\text{ext}}})$.

The Sobolev Lemma which is proved in [19] states that in the case of summability of the sequence $\{B_n^{-1}A_n\}_{n\in\mathbb{N}_0}$, the Fourier series in terms of the basis functions $H_{n,m} \in \mathcal{H}\left(\{B_n^{-1}A_n\}; \overline{\Omega_\sigma^{\text{ext}}}\right)$ is continuous on the boundary Ω_σ. In particular, we have the following statement (cf. [19]).

Lemma 3.5. *If* $U \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $s > k+1$, *then* U *corresponds to a potential of class* $Pot^{(k)}(\overline{\Omega_\sigma^{\text{ext}}})$.

3.2. Vectorial outer harmonic and Sobolev theory

We now extend the theory of scalar outer harmonics and scalar Sobolev spaces to the vectorial case. We use a system of vector spherical harmonics (cf. [21]) in order to generate the set of vector outer harmonics in such a way, that the Laplace equation is fulfilled componentwise.

Let $\{\tilde{y}_{n,m}^{(i)}\}_{i=1,2,3;n\in\mathbb{N}_{0_i};m=1,\ldots,2n+1}$ be a set of vector spherical harmonics satisfying the condition of being a set of eigenfunctions of the Beltrami operator, with

$$0_i = \begin{cases} 0, & i = 1, \\ 1, & i = 2, 3. \end{cases} \tag{3.23}$$

(see, e.g., [21, 32, 58], for a detailed introduction and profound discussion of these vector spherical harmonics). In the nomenclature of [32], the vector outer harmonics $h_{n,m}^{(i)}(\sigma;\cdot)$ of degree n and kind i are defined by

$$h_{n,m}^{(1)}(\sigma;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n+2} \tilde{y}_{n,m}^{(1)}\left(\frac{x}{|x|}\right), \quad n = 0,1,\ldots; m = 1,\ldots,2n+1, \tag{3.24}$$

$$h_{n,m}^{(2)}(\sigma;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n} \tilde{y}_{n,m}^{(2)}\left(\frac{x}{|x|}\right), \quad n = 1,2,\ldots; m = 1,\ldots,2n+1, \tag{3.25}$$

$$h_{n,m}^{(3)}(\sigma;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n+1} \tilde{y}_{n,m}^{(3)}\left(\frac{x}{|x|}\right), \quad n = 1,2,\ldots; m = 1,\ldots,2n+1, \tag{3.26}$$

for $x \in \overline{\Omega_\sigma^{\text{ext}}}$. The following properties are satisfied:

- $h_{n,m}^{(i)}(\sigma;\cdot)$ is of class $c^{(\infty)}(\Omega_\sigma^{\text{ext}})$,
- $\Delta_x h_{n,m}^{(i)}(\sigma;x) = 0$ for $x \in \Omega_\sigma^{\text{ext}}$, i.e., every component function $h_{n,m}^{(i)} \cdot \varepsilon^k$ satisfies the Laplace equation,
- $h_{n,m}^{(i)}$ is regular at infinity, i.e., $|h_{n,m}^{(i)}(\sigma;x)| = \mathcal{O}(|x|^{-1})$, $|h_{n,m}^{(2)}(\sigma \cdot x)| = \mathcal{O}(|x|^{-2})$, $|x| \to \infty$
- $h_{n,m}^{(i)}(\sigma;\cdot)|_{\Omega_\sigma} = (1/\sigma)\tilde{y}_{n,m}^{(i)}$,
- $(h_{n,m}^{(i)}(\sigma;\cdot), h_{l,s}^{(j)}(\sigma;\cdot))_{l^2(\Omega_\sigma)} = \int_{\Omega_\sigma} h_{n,m}^{(i)}(\sigma;x) h_{l,s}^{(j)}(\sigma;x)\, d\omega(x) = \delta_{i,j}\delta_{n,l}\delta_{m,s}$.

We introduce

$$harm^{(i)}(\overline{\Omega_\sigma^{\text{ext}}}) = \overline{\text{span}_{\substack{n=0_i,\dots;\\ m=1,\dots,2n+1}} h_{n,m}^{(i)}(\sigma;\cdot)}^{\|\cdot\|_{c^{(0)}(\overline{\Omega_\sigma^{\text{ext}}})}}, \tag{3.27}$$

$$harm(\overline{\Omega_\sigma^{\text{ext}}}) = \overline{\text{span}_{\substack{i=1,2,3;n=0_i,\dots;\\ m=1,\dots,2n+1}} h_{n,m}^{(i)}(\sigma;\cdot)}^{\|\cdot\|_{c^{(0)}(\overline{\Omega_\sigma^{\text{ext}}})}}. \tag{3.28}$$

Some results concerning addition theorems for outer harmonics using Legendre tensors and Legendre vectors can be found in the Ph.D.-thesis [58] and are not discussed here.

Lemma 3.6. *Let $\{H_{n,m}(\sigma;\cdot)\}_{n\in\mathbb{N}_0;m=1,\dots,2n+1}$ be a system of scalar outer harmonics. Then*

$$\overline{\text{span}\{H_{n,m}(\sigma;\cdot)\varepsilon^i|_\Sigma\}_{i=1,2,3}}^{\|\cdot\|_{l^2(\Sigma)}} = l^2(\Sigma),$$

$$\overline{\text{span}\{H_{n,m}(\sigma;\cdot)\varepsilon^i|_\Sigma\}_{i=1,2,3}}^{\|\cdot\|_{c^{(0)}(\Sigma)}} = c^{(0)}(\Sigma).$$

Theorem 3.7. *Let $\{h_{n,m}^{(i)}(\sigma;\cdot)\}_{\substack{i=1,2,3;n=0_i,\dots;\\ m=1,\dots,2n+1}}$ be a system of vector outer harmonics as defined in* (3.24)–(3.26)*. Then the following statements hold true:*

$$l^2(\Sigma) = \overline{\text{span}_{\substack{i=1,2,3;n=0_i,\dots;\\ m=1,\dots,2n+1}} (h_{n,m}^{(i)}(\sigma;\cdot))|_\Sigma}^{\|\cdot\|_{l^2(\Sigma)}},$$

$$c^{(0)}(\Sigma) = \overline{\text{span}_{\substack{i=1,2,3;n=0_i,\dots;\\ m=1,\dots,2n+1}} (h_{n,m}^{(i)}(\sigma;\cdot))|_\Sigma}^{\|\cdot\|_{c^{(0)}(\Sigma)}}.$$

In order to define the vectorial potential space $pot(\Sigma^{\text{ext}})$ we need the divergence and curl operator, which are defined by

$$\text{div}\, f(x) = \sum_{i=1}^{3} \frac{\partial F_i}{\partial x_i}(x), \quad f = \sum_{i=1}^{3} F_i\varepsilon^i, \tag{3.29}$$

and

$$(\text{curl}\, f(x))_i = \sum_{j,k=1}^{3} \varepsilon_{ijk}\frac{\partial F_k}{\partial x_j}(x), \tag{3.30}$$

where ε_{ijk} is the alternator defined by

$$\varepsilon_{ijk} = \begin{cases} +1, & (i,j,k) \text{ is an even permutation of } (1,2,3),\\ -1, & (i,j,k) \text{ is an odd permutation of } (1,2,3),\\ 0, & (i,j,k) \text{ is not a permutation of } (1,2,3). \end{cases} \tag{3.31}$$

By $pot(\Sigma^{\text{ext}})$ we denote the space of all vector fields $f : \Sigma^{\text{ext}} \to \mathbb{R}^3$ satisfying the following properties:

(i) $f \in c^{(1)}(\Sigma^{\text{ext}})$,
(ii) f is a harmonic vector field: $\text{div} f = 0$, $\text{curl} f = 0$ in Σ^{ext},
(iii) f is regular at infinity: $|f(x)| = \mathcal{O}(|x|^{-2})$, $|x| \to \infty$.

Furthermore, we let

$$pot^{(k)}(\overline{\Sigma^{\text{ext}}}) = pot(\Sigma^{\text{ext}}) \cap c^{(k)}(\overline{\Sigma^{\text{ext}}}), \tag{3.32}$$

which is meant in the same sense as we explained in the scalar case. It is well known (see, e.g., [38]), that every function $f \in c^{(k)}(\Sigma^{\text{ext}})$ satisfying $\text{curl} f = 0$ is the gradient of a function $V \in C^{(k+1)}(\Sigma^{\text{ext}})$: $f = \nabla V$. As a consequence, we get that every $f \in pot(\Sigma^{\text{ext}})$ can be represented as a gradient field $f = \nabla V$, where $V \in Pot(\Sigma^{\text{ext}})$, and vice versa. Furthermore, it is obvious, that a function $f \in pot(\Sigma^{\text{ext}})$ of the form $f = \sum_{i=1}^{3} F_i \varepsilon^i$ fulfills $F_i \in Pot(\Sigma^{\text{ext}})$.

For arbitrary $\varepsilon > 0$, we have an integer $N = N(\varepsilon)$ and coefficients $a_{n,m}$, $n = 0, \ldots, N$; $m = 1, \ldots, 2n+1$, such that

$$\sup_{x \in \Sigma} \left| F(x) - \sum_{n=0}^{N} \sum_{m=1}^{2n+1} a_{n,m} H_{n,m}(\sigma; x) \right| < \varepsilon. \tag{3.33}$$

For the gradient of $H_{n,m}(\sigma; \cdot)$ we obtain

$$\nabla_x H_{n,m}(\sigma; x) = C\, h^{(1)}_{n,m}(\sigma; x), \tag{3.34}$$

with a constant factor C, which leads us to (cf. [25])

$$pot^{(0)}(\overline{\Sigma^{\text{ext}}}) = \overline{\text{span}_{\substack{n \in \mathbb{N}_0; \\ m=1,\ldots,2n+1}} (h^{(1)}_{n,m}(\sigma; \cdot))|_{\overline{\Sigma^{\text{ext}}}}}^{\|\cdot\|_{c^{(0)}(\overline{\Sigma^{\text{ext}}})}} \tag{3.35}$$

(Runge–Walsh approximation property).

In analogy to the scalar case, we define Sobolev spaces for vector fields. We do not restrict our considerations to $pot^{(\infty)}(\overline{\Omega^{\text{ext}}_\sigma})$ as a reference space for the definition of vectorial Sobolev spaces, because in this case only the $h^{(1)}_{n,m}$-part would be taken into account.

Consider the space a defined by

$$a = \{\{a_n\} \,|\, a_n = \left(A^{(1)}_n, A^{(2)}_n, A^{(3)}_n\right)^T \in \mathbb{R}^3,\ A^{(i)}_n \neq 0,\ n \in \mathbb{N}_0\}. \tag{3.36}$$

Obviously, we have $\{A^{(i)}_n\}_{n \in \mathbb{N}_0} \in \mathcal{A}$ for $i \in \{1, 2, 3\}$.

For $\{a_n\}_{n \in \mathbb{N}_0} \in a$ we define

$$e^{(i)}(\overline{\Omega^{\text{ext}}_\sigma}) = \left\{ f \in harm^{(i)}(\overline{\Omega^{\text{ext}}_\sigma}) \ :\ \sum_{n=0_i}^{\infty} \sum_{m=1}^{2n+1} |A^{(i)}_n|^2 (f, h^{(i)}_{n,m})^2_{l^2(\Omega_\sigma)} < \infty \right\}, \tag{3.37}$$

$i \in \{1, 2, 3\}$. Equipped with the inner product

$$(f, g)_{h(\overline{\Omega^{\text{ext}}_\sigma})} = \sum_{i=1}^{3} \sum_{n=0_i}^{\infty} \sum_{m=1}^{2n+1} |A^{(i)}_n|^2 (f, h^{(i)}_{n,m})_{l^2(\Omega_\sigma)} (g, h^{(i)}_{n,m})_{l^2(\Omega_\sigma)}, \tag{3.38}$$

$f, g \in e^{(i)}(\overline{\Omega^{\text{ext}}_\sigma})$, the space $e^{(i)}(\overline{\Omega^{\text{ext}}_\sigma})$ becomes a pre-Hilbert space. We define the Sobolev space $h^{(i)}(\overline{\Omega^{\text{ext}}_\sigma}) = h^{(i)}(\{A^{(i)}_n\}; \overline{\Omega^{\text{ext}}_\sigma})$ to be the completion of $e^{(i)}(\overline{\Omega^{\text{ext}}_\sigma})$

under the norm $\|\cdot\|_{h(\overline{\Omega_\sigma^{\text{ext}}})}$, which denotes the norm associated to $(\cdot,\cdot)_{h(\overline{\Omega_\sigma^{\text{ext}}})}$:

$$h^{(i)}(\{A_n^{(i)}\};\overline{\Omega_\sigma^{\text{ext}}}) = \overline{e^{(i)}(\overline{\Omega_\sigma^{\text{ext}}})}^{\|\cdot\|_{h(\overline{\Omega_\sigma^{\text{ext}}})}}. \tag{3.39}$$

We use the following notation

$$h(\overline{\Omega_\sigma^{\text{ext}}}) = h(\{a_n\};\overline{\Omega_\sigma^{\text{ext}}}) = \bigoplus_{i=1}^{3} h^{(i)}(\overline{\Omega_\sigma^{\text{ext}}}) = \bigoplus_{i=1}^{3} h^{(i)}(\{A_n^{(i)}\};\overline{\Omega_\sigma^{\text{ext}}}). \tag{3.40}$$

The space $h(\overline{\Omega_\sigma^{\text{ext}}})$ equipped with the inner product $(\cdot,\cdot)_{h(\overline{\Omega_\sigma^{\text{ext}}})}$ is a Hilbert space with Hilbert basis $\{h_{n,m}^{(i)*\{A_n^{(i)}\}}(\sigma;\cdot)\}_{i=1,2,3;\,n=0_i,\ldots;\,m=1,\ldots,2n+1}$ given by

$$h_{n,m}^{(i)*\{A_n^{(i)}\}}(\sigma;x) = (A_n^{(i)})^{-1} h_{n,m}^{(i)}(\sigma;x), \quad x \in \overline{\Omega_\sigma^{\text{ext}}}. \tag{3.41}$$

We can, therefore, expand a function $f \in h(\overline{\Omega_\sigma^{\text{ext}}})$ as a Fourier series in terms of the basis $h_{n,m}^{(i)*\{A_n^{(i)}\}}$:

$$f = \sum_{i=1}^{3} \sum_{n=0_i}^{\infty} \sum_{m=1}^{2n+1} f^{(i)\wedge_{h(\{a_n\};\overline{\Omega_\sigma^{\text{ext}}})}}(n,m) h_{n,m}^{(i)*\{A_n^{(i)}\}}, \tag{3.42}$$

where

$$f^{(i)\wedge_{h(\{a_n\};\overline{\Omega_\sigma^{\text{ext}}})}}(n,m) = f^{(i)\wedge}(n,m) = (f, h_{n,m}^{(i)*\{A_n^{(i)}\}})_{h(\overline{\Omega_\sigma^{\text{ext}}})}. \tag{3.43}$$

In analogy to the scalar spaces $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, we define the vectorial spaces $h_s(\overline{\Omega_\sigma^{\text{ext}}})$ by

$$h_s^{(i)}(\overline{\Omega_\sigma^{\text{ext}}}) = h^{(i)}\left(\left\{\left(\frac{n+\frac{1}{2}}{\sigma}\right)^s\right\};\overline{\Omega_\sigma^{\text{ext}}}\right), \tag{3.44}$$

$$h_s(\overline{\Omega_\sigma^{\text{ext}}}) = \bigoplus_{i=1}^{3} h_s^{(i)}(\overline{\Omega_\sigma^{\text{ext}}}). \tag{3.45}$$

The space $h_s(\overline{\Omega_\sigma^{\text{ext}}})$ equipped with the inner product $(\cdot,\cdot)_{h_s(\overline{\Omega_\sigma^{\text{ext}}})}$ is a Hilbert space with Hilbert basis $\{h_{n,m}^{(i)s}(\sigma;\cdot)\}_{i=1,2,3;\,n=0_i,\ldots;\,m=1,\ldots,2n+1}$ given by

$$h_{n,m}^{(i)s}(\sigma;x) = \left(\frac{\sigma}{n+\frac{1}{2}}\right)^s h_{n,m}^{(i)}(\sigma;x), \quad x \in \overline{\Omega_\sigma^{\text{ext}}}. \tag{3.46}$$

In the case of the space $h_0(\overline{\Omega_\sigma^{\text{ext}}})$ we understand the norm $\|\cdot\|_{h_0(\overline{\Omega_\sigma^{\text{ext}}})}$ to be the $\|\cdot\|_{l^2(\Omega_\sigma)}$- norm.

Next, the scalar Sobolev Lemma 3.4 will be extended to vector fields.

Definition 3.8. A sequence $\{a_n\}_{n\in\mathbb{N}_0} \in a$ is called summable if

$$\sum_{n=0_i}^{\infty} \frac{2n+1}{\left(A_n^{(i)}\right)^2} < \infty, \tag{3.47}$$

for $i = 1, 2, 3$.

In the sequel, $\{b_n^{-1}\}_{n\in\mathbb{N}_0} \in a$ means the sequence given by

$$b_n^{-1} = \left(\left(B_n^{(1)}\right)^{-1}, \left(B_n^{(2)}\right)^{-1}, \left(B_n^{(3)}\right)^{-1}\right)^T, \tag{3.48}$$

and

$$b_n^{-1}a_n = \left(A_n^{(1)}\left(B_n^{(1)}\right)^{-1}, A_n^{(2)}\left(B_n^{(2)}\right)^{-1}, A_n^{(3)}\left(B_n^{(3)}\right)^{-1}\right)^T. \tag{3.49}$$

Lemma 3.9 (Vectorial Sobolev Lemma). *Assume, that* $\{a_n\}_{n\in\mathbb{N}_0}, \{b_n\}_{n\in\mathbb{N}_0} \in a$ *are sequences such that* $\{b_n^{-1}a_n\}_{n\in\mathbb{N}_0} \in a$ *is summable. Then each* $f \in h(\{b_n^{-1}a_n\}; \overline{\Omega_\sigma^{\mathrm{ext}}})$ *corresponds to a function of class harm*$(\overline{\Omega_\sigma^{\mathrm{ext}}})$.

3.3. Tensorial outer harmonic and Sobolev theory

The extension of vectorial to tensorial theory is straightforward (see [21, 32, 58]). With the help of a system $\{\tilde{\mathbf{y}}_{n,m}^{(i,k)}\}$ of tensor spherical harmonics we can derive a set of tensor outer harmonics $\{\mathbf{h}_{n,m}^{(i,k)}(\sigma;\cdot)\}$ satisfying the Laplace equation componentwise.

Let $\{\tilde{\mathbf{y}}_{n,m}^{(i,k)}\}_{i,k=1,2,3;n\in\mathbb{N}_0;m=1,\ldots,2n+1}$ with

$$0_{ik} = \begin{cases} 0, & (i,k) \in \{(1,1),(2,1),(3,1)\}, \\ 1, & (i,k) \in \{(1,2),(1,3),(2,3),(3,3)\}, \\ 2, & (i,k) \in \{(2,2),(3,2)\}, \end{cases} \tag{3.50}$$

be a set of tensorial spherical harmonics satisfying the condition of being eigenfunctions of the Beltrami operator (see, e.g., the Ph.D.-thesis [58] for a detailed introduction and profound discussion of these tensor spherical harmonics). The tensor outer harmonics $\mathbf{h}_{n,m}^{(i,k)}(\sigma;\cdot)$ of degree n and kind (i,k) are then defined by

$$\mathbf{h}_{n,m}^{(1,1)}(\sigma;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n+3} \tilde{\mathbf{y}}_{n,m}^{(1,1)}\left(\frac{x}{|x|}\right), \tag{3.51}$$

$$\mathbf{h}_{n,m}^{(1,2)}(\sigma;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n+1} \tilde{\mathbf{y}}_{n,m}^{(1,2)}\left(\frac{x}{|x|}\right), \tag{3.52}$$

$$\mathbf{h}_{n,m}^{(2,1)}(\sigma;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n+1} \tilde{\mathbf{y}}_{n,m}^{(2,1)}\left(\frac{x}{|x|}\right), \tag{3.53}$$

$$\mathbf{h}_{n,m}^{(2,2)}(R;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n-1} \tilde{\mathbf{y}}_{n,m}^{(2,2)}\left(\frac{x}{|x|}\right), \tag{3.54}$$

$$\mathbf{h}_{n,m}^{(3,3)}(\sigma;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n+1} \tilde{\mathbf{y}}_{n,m}^{(3,3)}\left(\frac{x}{|x|}\right), \tag{3.55}$$

$$\mathbf{h}_{n,m}^{(1,3)}(\sigma;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n+2} \tilde{\mathbf{y}}_{n,m}^{(1,3)}\left(\frac{x}{|x|}\right), \tag{3.56}$$

$$\mathbf{h}_{n,m}^{(2,3)}(\sigma;x) = \frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n} \tilde{\mathbf{y}}_{un,m}^{(2,3)}\left(\frac{x}{|x|}\right), \tag{3.57}$$

$$\mathbf{h}_{n,m}^{(3,1)}(\sigma;x)=\frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n+2}\tilde{\mathbf{y}}_{n,m}^{(3,1)}\left(\frac{x}{|x|}\right), \tag{3.58}$$

$$\mathbf{h}_{n,m}^{(3,2)}(\sigma;x)=\frac{1}{\sigma}\left(\frac{\sigma}{|x|}\right)^{n}\tilde{\mathbf{y}}_{n,m}^{(3,2)}\left(\frac{x}{|x|}\right), \tag{3.59}$$

where $x\in\overline{\Omega_\sigma^{\mathrm{ext}}}$, $n=0_{ik},\ldots;\ m=1,\ldots,2n+1$. The following properties are satisfied:

- $\mathbf{h}_{n,m}^{(i,k)}(\sigma;\cdot)$ is of class $\mathbf{c}^{(\infty)}(\Omega_\sigma^{\mathrm{ext}})$,
- $\Delta_x\mathbf{h}_{n,m}^{(i,k)}(\sigma;x)=0$ for $x\in\Omega_\sigma^{\mathrm{ext}}$, i.e., the component functions of $\mathbf{h}_{n,m}^{(i,k)}(\sigma;\cdot)$ fulfill the Laplace equation,
- $\mathbf{h}_{n,m}^{(i,k)}$ is regular at infinity, i.e., $|\mathbf{h}_{n,m}^{(i,k)}(\sigma;x)|=\mathcal{O}(|x|^{-3}),\quad |x|\to\infty$.
- $\mathbf{h}_{n,m}^{(i,k)}(\sigma;\cdot)|_{\Omega_\sigma}=(1/\sigma)\tilde{\mathbf{y}}_{n,m}^{(i,k)}$,
- $(\mathbf{h}_{n,m}^{(i,k)}(\sigma;\cdot),\mathbf{h}_{l,s}^{(p,q)}(\sigma;\cdot))_{\mathbf{l}^2(\Omega_\sigma)}=\int_{\Omega_\sigma}\mathbf{h}_{n,m}^{(i,k)}(\sigma;x)\mathbf{h}_{l,s}^{p,q}(\sigma;x)d\omega(x)$ $=\delta_{i,p}\delta_{k,q}\delta_{n,l}\delta_{m,s}$.

Moreover, we define

$$\mathbf{harm}^{(i,k)}(\overline{\Omega_\sigma^{\mathrm{ext}}})=\overline{\mathrm{span}\underset{m=1,\ldots,2n+1}{n=0_{ik}\ldots;}\mathbf{h}_{n,m}^{(i,k)}(\sigma;\cdot)}^{\|\cdot\|_{\mathbf{c}^{(0)}(\overline{\Omega_\sigma^{\mathrm{ext}}})}}, \tag{3.60}$$

$$\mathbf{harm}(\overline{\Omega_\sigma^{\mathrm{ext}}})=\overline{\mathrm{span}_{i,k\in\{1,2,3\};n=0_{ik}\ldots;}\underset{m=1,\ldots,2n+1}{\mathbf{h}_{n,m}^{(i,k)}(\sigma;\cdot)}}^{\|\cdot\|_{\mathbf{c}^{(0)}(\overline{\Omega_\sigma^{\mathrm{ext}}})}}. \tag{3.61}$$

Some results concerning addition theorems for outer harmonics can be formulated both for the tensor product of two tensor outer harmonics and for the product of a scalar and a tensor outer harmonic. They can be found in the Ph.D.-thesis [58] and are not discussed in this contribution.

Lemma 3.10. *Let $\{H_{n,m}(\sigma;\cdot)\}_{n\in\mathbb{N}_{0_{ik}};m=1,\ldots,2n+1}$ be a system of scalar outer harmonics. Then*

$$\overline{\mathrm{span}\{H_{n,m}(\sigma;\cdot)\varepsilon^i\otimes\varepsilon^k|_\Sigma\}}^{\|\cdot\|_{\mathbf{l}^2(\Sigma)}}=\mathbf{l}^2(\Sigma), \tag{3.62}$$

$$\overline{\mathrm{span}\{H_{n,m}(\sigma;\cdot)\varepsilon^i\otimes\varepsilon^k|_\Sigma)\}}^{\|\cdot\|_{\mathbf{c}^{(0)}(\Sigma)}}=\mathbf{c}^{(0)}(\Sigma). \tag{3.63}$$

Theorem 3.11. *Let $\{\mathbf{h}_{n,m}^{(i,k)}\}_{\substack{i,k=1,2,3;n=0_{ik},\ldots;\\ m=1,\ldots,2n+1}}$ be a system of tensor outer harmonics. Then the following statements hold true:*

$$\mathbf{l}^2(\Sigma)=\overline{\mathrm{span}\underset{\substack{i,k=1,2,3;n=0_{ik},\ldots;\\ m=1,\ldots,2n+1}}{(\mathbf{h}_{n,m}^{(i,k)}(\sigma;\cdot))|_\Sigma}}^{\|\cdot\|_{\mathbf{l}^2(\Sigma)}}, \tag{3.64}$$

$$\mathbf{c}(\Sigma)=\overline{\mathrm{span}\underset{\substack{i,k=1,2,3;n=0_{ik},\ldots;\\ m=1,\ldots,2n+1}}{(\mathbf{h}_{n,m}^{(i,k)}(\sigma;\cdot))|_\Sigma}}^{\|\cdot\|_{\mathbf{c}(\Sigma)}}. \tag{3.65}$$

In order to define a tensorial counterpart $\mathbf{pot}(\overline{\Sigma^{\mathrm{ext}}})$ of the space $pot(\overline{\Sigma^{\mathrm{ext}}})$, we need the divergence and the curl operator of tensor fields. Having (3.29) in

mind, we define $\operatorname{div}\mathbf{f}$ by

$$(\operatorname{div}\mathbf{f}(x))_i = \sum_{j=1}^{3} \frac{\partial F_{i,j}}{\partial x_j}(x), \quad \mathbf{f} = \sum_{i,j=1}^{3} F_{i,j}\varepsilon^i \otimes \varepsilon^j. \tag{3.66}$$

Furthermore, based on (3.30) we have the following definition of $\operatorname{curl}\mathbf{f}$:

$$(\operatorname{curl}\mathbf{f}(x))_{i,j} = \sum_{p,k=1}^{3} \varepsilon_{ipk} \frac{\partial F_{j,k}}{\partial x_p}(x). \tag{3.67}$$

The space $\mathbf{pot}(\Sigma^{\text{ext}})$ denotes the space of all tensor fields $\mathbf{f} : \Sigma^{\text{ext}} \to \mathbb{R}^3 \otimes \mathbb{R}^3$ satisfying the following properties:

(i) $\mathbf{f} \in \mathbf{c}^{(1)}(\Sigma^{\text{ext}})$,
(ii) $\mathbf{f}$ is a harmonic tensor field: $\operatorname{div}\mathbf{f} = 0,\ \operatorname{curl}\mathbf{f} = 0$ in Σ^{ext},
(iii) $\mathbf{f}$ is regular at infinity: $|\mathbf{f}(x)| = \mathcal{O}(|x|^{-3}),\ |x| \to \infty$.

Furthermore, we let

$$\mathbf{pot}^{(k)}(\overline{\Sigma^{\text{ext}}}) = \mathbf{pot}(\Sigma^{\text{ext}}) \cap \mathbf{c}^{(k)}(\overline{\Sigma^{\text{ext}}}), \tag{3.68}$$

which we understand in the same sense as in the scalar and vectorial case. As shown, e.g., in [38], every tensor function $\mathbf{f} \in \mathbf{c}^{(k)}(\Sigma^{\text{ext}})$ with $\operatorname{curl}\mathbf{f} = 0$ is the gradient of a vector field $v \in c^{(k+1)}(\Sigma^{\text{ext}})$:

$$\mathbf{f} = \nabla v, \tag{3.69}$$

where ∇v is the tensor of second rank defined by

$$(\nabla_x v)_{ij}(x) = \frac{\partial v_i}{\partial x_j}(x). \tag{3.70}$$

Therefore, every member $\mathbf{v} \in \mathbf{pot}(\Sigma^{\text{ext}})$ can be represented as a gradient field $\mathbf{v} = \nabla v$, where v is of class $pot(\Sigma^{\text{ext}})$, and vice versa. As a consequence of this, in connection with the fact that every $v \in pot(\Sigma^{\text{ext}})$ can be represented as a gradient field $v = \nabla V$ with $V \in Pot(\Sigma^{\text{ext}})$, we finally get that a tensor field $\mathbf{v} \in \mathbf{pot}(\Sigma^{\text{ext}})$ can be represented as the Hesse tensor of a scalar field $V \in Pot(\Sigma^{\text{ext}})$:

$$\mathbf{v} = \nabla \otimes \nabla V, \tag{3.71}$$

and vice versa.

It is obvious, that $\mathbf{f} \in \mathbf{pot}(\Sigma^{\text{ext}})$ of the form $\mathbf{f} = \sum_{i,k=1}^{3} F_{i,k}\varepsilon^i \otimes \varepsilon^k$ fulfills $F_{i,k} \in Pot(\Sigma^{\text{ext}})$. In addition, we are able to show that

$$\mathbf{pot}^{(0)}(\overline{\Sigma^{\text{ext}}}) = \overline{\operatorname{span}_{\substack{n\in\mathbb{N}_0;\\ m=1,\dots,2n+1}} (\mathbf{h}_{n,m}^{(1,1)}(\sigma;\cdot))|_{\overline{\Sigma^{\text{ext}}}}}^{\|\cdot\|_{\mathbf{c}^{(0)}(\overline{\Sigma^{\text{ext}}})}} \tag{3.72}$$

(Runge–Walsh approximation property).

Our purpose is now to define Sobolev spaces for tensor fields in analogy to the vectorial Sobolev spaces. We introduce the linear space $\mathbf{a}$ in the following way:

$$\mathbf{a} = \{\{\mathbf{a}_n\} \,|\, \mathbf{a}_n \in \mathbb{R}^3 \otimes \mathbb{R}^3,\ A_n^{(i,k)} \neq 0,\ n \in \mathbb{N}_0; m = 1, \dots, 2n+1; i,k \in \{1,2,3\}\}, \tag{3.73}$$

where

$$\mathbf{a}_n = \begin{pmatrix} A_n^{(1,1)} & A_n^{(1,2)} & A_n^{(1,3)} \\ A_n^{(2,1)} & A_n^{(2,2)} & A_n^{(2,3)} \\ A_n^{(3,1)} & A_n^{(3,2)} & A_n^{(3,3)} \end{pmatrix}, \tag{3.74}$$

with $\{A_n^{(i,k)}\}_{n\in\mathbb{N}_0} \in \mathcal{A}$ for $i,k \in \{1,2,3\}$.

Let us now consider a sequence $\{\mathbf{a}_n\}_{n\in\mathbb{N}_0} \in \mathbf{a}$. Then we define

$$\mathbf{e}^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}) = \left\{ \mathbf{f} \in \mathbf{harm}^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}) \ : \ \sum_{n=0_{ik}}^{\infty} \sum_{m=1}^{2n+1} |A_n^{(i,k)}|^2 (\mathbf{f}, \mathbf{h}_{n,m}^{(i,k)})_{\mathbf{l}^2(\Omega_\sigma)}^2 < \infty \right\}, \tag{3.75}$$

$i,k \in \{1,2,3\}$. Equipped with the inner product

$$(\mathbf{f},\mathbf{g})_{\mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}})} = \sum_{i,k=1}^{3} \sum_{n=0_{ik}}^{\infty} \sum_{m=1}^{2n+1} |A_n^{(i,k)}|^2 (\mathbf{f}, \mathbf{h}_{n,m}^{(i,k)})_{\mathbf{l}^2(\Omega_\sigma)} (\mathbf{g}, \mathbf{h}_{n,m}^{(i,k)})_{\mathbf{l}^2(\Omega_\sigma)}, \tag{3.76}$$

$\mathbf{f},\mathbf{g} \in \mathbf{e}^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$, the space $\mathbf{e}^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$ becomes a pre-Hilbert space. We define the Sobolev space $\mathbf{h}^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}) = \mathbf{h}^{(i,k)}(\{A_n^{(i,k)}\}; \overline{\Omega_\sigma^{\text{ext}}})$ to be the completion of $\mathbf{e}^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$ under the norm $\|\cdot\|_{\mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}})}$, which denotes the norm associated to $(\cdot,\cdot)_{\mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}})}$:

$$\mathbf{h}^{(i,k)}(\{A_n^{(i,k)}\}; \overline{\Omega_\sigma^{\text{ext}}}) = \overline{\mathbf{e}^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})}^{\|\cdot\|_{\mathbf{h}(\{\overline{\Omega_\sigma^{\text{ext}}})}}}. \tag{3.77}$$

We use the following notation

$$\mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}}) = \bigoplus_{i,k=1}^{3} \mathbf{h}^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}). \tag{3.78}$$

The space $\mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}})$ equipped with the inner product $(\cdot,\cdot)_{\mathbf{h}(\{\overline{\Omega_\sigma^{\text{ext}}})}$ is a Hilbert space. The system $\{\mathbf{h}_{n,m}^{(i,k)*\{A_n^{(i,k)}\}}(\sigma;\cdot)\}_{\substack{i,k\in\{1,2,3\};n\in\mathbb{N}_{0_{ik}};\\ m=1,\dots,2n+1}}$, given by

$$\mathbf{h}_{n,m}^{(i,k)*\{A_n^{(i,k)}\}}(\sigma;x) = (A_n^{(i,k)})^{-1} \mathbf{h}_{n,m}^{(i,k)}(\sigma;x), \quad x \in \overline{\Omega_\sigma^{\text{ext}}}, \tag{3.79}$$

represents an $\mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}})$-orthonormal Hilbert basis in $\mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}})$.

As a consequence, we can expand a function $\mathbf{f} \in \mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}})$ as a Fourier series in terms of the basis $\mathbf{h}_{n,m}^{(i,k)*\{A_n^{(i,k)}\}}$:

$$\mathbf{f} = \sum_{i,k=1}^{3} \sum_{n=0_{ik}}^{\infty} \sum_{m=1}^{2n+1} \mathbf{f}^{(i,k)\wedge_{\mathbf{h}(\{\mathbf{a}_n\};\overline{\Omega_\sigma^{\text{ext}}})}}(n,m) \mathbf{h}_{n,m}^{(i,k)*\{A_n^{(i,k)}\}}, \tag{3.80}$$

where

$$\mathbf{f}^{(i,k)\wedge_{\mathbf{h}(\{\mathbf{a}_n\};\overline{\Omega_\sigma^{\text{ext}}})}}(n,m) = \mathbf{f}^{(i,k)\wedge}(n,m) = (\mathbf{f}, \mathbf{h}_{n,m}^{(i,k)*\{A_n^{(i,k)}\}})_{\mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}})}. \tag{3.81}$$

Finally, in analogy to the vectorial spaces $h_s^{(i)}(\overline{\Omega_\sigma^{\text{ext}}})$, we define

$$\mathbf{h}_s^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}) = \mathbf{h}^{(i,k)}\left(\left\{\left(\frac{n+\frac{1}{2}}{\sigma}\right)^s\right\}; \overline{\Omega_\sigma^{\text{ext}}}\right), \tag{3.82}$$

$$\mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}}) = \bigoplus_{i,k=1}^{3} \mathbf{h}_s^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}). \tag{3.83}$$

The space $\mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})$ equipped with the inner product $(\cdot,\cdot)_{\mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})}$ is a Hilbert space. The system $\{\mathbf{h}_{n,m}^{(i,k)s}(\sigma;\cdot)\}_{i,k\in\{1,2,3\};\, n\in\mathbb{N}_{0_{ik}};\, m=1,\ldots,2n+1}$, given by

$$\mathbf{h}_{n,m}^{(i,k)s}(\sigma;x) = \left(\frac{\sigma}{n+\frac{1}{2}}\right)^s \mathbf{h}_{n,m}^{(i,k)}(\sigma;x), \quad x \in \overline{\Omega_\sigma^{\text{ext}}}, \tag{3.84}$$

represents an $\mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}})$-orthonormal Hilbert basis in $\mathbf{h}(\overline{\Omega_\sigma^{\text{ext}}})$.

Our next goal is to extend the Sobolev Lemma 3.4 to tensor fields.

Definition 3.12. A sequence $\{\mathbf{a}_n\}_{n\in\mathbb{N}_0} \in \mathbf{a}$ is called summable if

$$\sum_{n=0_{ik}}^{\infty} \frac{2n+1}{\left(A_n^{(i,k)}\right)^2} < \infty \tag{3.85}$$

for $i,k \in \{1,2,3\}$.

In the sequel, $\{\mathbf{b}_n^{-1}\}_{n\in\mathbb{N}_0} \in \mathbf{a}$ represents the sequence given by

$$\mathbf{b}_n^{-1} = \begin{pmatrix} \left(B_n^{(1,1)}\right)^{-1} & \left(B_n^{(1,2)}\right)^{-1} & \left(B_n^{(1,3)}\right)^{-1} \\ \left(B_n^{(2,1)}\right)^{-1} & \left(B_n^{(2,2)}\right)^{-1} & \left(B_n^{(2,3)}\right)^{-1} \\ \left(B_n^{(3,1)}\right)^{-1} & \left(B_n^{(3,2)}\right)^{-1} & \left(B_n^{(3,3)}\right)^{-1} \end{pmatrix}, \tag{3.86}$$

and $\{\mathbf{a}_n^{-1}\mathbf{b}_n\}_{n\in\mathbb{N}_0} \in \mathbf{a}$ is given by

$$\mathbf{b}_n^{-1}\mathbf{a}_n = \begin{pmatrix} A_n^{(1,1)}\left(B_n^{(1,1)}\right)^{-1} & A_n^{(1,2)}\left(B_n^{(1,2)}\right)^{-1} & A_n^{(1,3)}\left(B_n^{(1,3)}\right)^{-1} \\ A_n^{(2,1)}\left(B_n^{(2,1)}\right)^{-1} & A_n^{(2,2)}\left(B_n^{(2,2)}\right)^{-1} & A_n^{(2,3)}\left(B_n^{(2,3)}\right)^{-1} \\ A_n^{(3,1)}\left(B_n^{(3,1)}\right)^{-1} & A_n^{(3,2)}\left(B_n^{(3,2)}\right)^{-1} & A_n^{(3,3)}\left(B_n^{(3,3)}\right)^{-1} \end{pmatrix}. \tag{3.87}$$

Lemma 3.13 (Tensorial Sobolev Lemma). *Assume, that the sequences* $\{\mathbf{a}_n\}_{n\in\mathbb{N}_0}$, $\{\mathbf{b}_n\}_{n\in\mathbb{N}_0} \in \mathbf{a}$ *are such that* $\{\mathbf{b}_n^{-1}\mathbf{a}_n\}_{n\in\mathbb{N}_0} \in \mathbf{a}$ *is summable. Then each* $\mathbf{f} \in \mathbf{h}\left(\{\mathbf{b}_n^{-1}\mathbf{a}_n\}; \overline{\Omega_\sigma^{\text{ext}}}\right)$ *corresponds to a function of class* $\mathbf{harm}(\overline{\Omega_\sigma^{\text{ext}}})$.

4. Pseudodifferential operators and geodetic nomenclature

All gravitational information under discussion in physical geodesy leads to operator equations relating the (disturbing) potential to geodetically relevant observables. In physical geodesy, one can think of observables as operating on an "input signal" F (e.g., the (disturbing) potential) to produce an (scalar, vectorial or tensorial) output signal of the form

$$\Lambda F = G \tag{4.1}$$

(for example, geoidal undulation, gravity anomaly, radial or tangential derivatives), where Λ is a certain (scalar, vectorial or tensorial) operator. Note, that later on we will differentiate in our notation weather we deal with scalar, vectorial or tensorial observables, but in this introductory part of the text for reason of readability we do not distinguish the geodetic quantities. Fortunately, it is the case in geodetic applications involving the (disturbing) potential that large portions of interest can be well approximated by operators that represent linear, rotation-invariant pseudodifferential operators.

The standard pseudodifferential operators Λ occurring in physical geodesy (cf. [69]) have to reflect the Pizzetti concept (cf. [36, 59]):

1. The mass within the reference ellipsoid for establishing the disturbing potential F is equal to the mass of the Earth.
2. The center of the reference ellipsoid coincides with the center of the Earth.
3. The value of the potential on the geoidal surface and the value of the normal potential on the reference ellipsoidal surface are the same.
4. There are no masses outside the geoid (remove-restore-principle from masses outside the geoid).
5. The constructive approximation is simplified for reasons of computational economy from an ellipsoidal to a spherical framework by Runge–Walsh justification (see the contribution [4] in this volume).

The presentation of the classical quantities in gravitational potential determination can be formulated within the framework of pseudodifferential operators. To be more concrete, in our approach we deal with radial, tangential and mixed (first- and second-order) derivatives of the Earth gravitational potential. Two important properties have to be taken into account specifying the operators which we study in the sequel. On the one hand, the mathematical modeling should lead to a consistent setup. It turns out that this requirement is, in fact, assured by the operators. On the other hand, we demand the assigned operators to be isotropic for structural reasons. In consequence (see also [63]), the (scalar) tangential derivatives $\frac{\partial}{\partial \varphi}$ and $\frac{\partial}{\partial t}$ are of no interest for us because they do not lead to isotropic operators in a scalar framework. Instead of using scalar tangential operators we decide to go over to the vectorial (and tensorial) tangential derivative using the surface gradient ∇^*. Indeed, we want to point out that we have the choice between two viable variants namely either to develop a scalar anisotropic theory for component modeling, or to turn over to vectorial/tensorial isotropic theory. In this contribution, we prefer the

second variant, expecting that the development of a vector/tensor theory provides us with a versatile tool for modeling geodetically relevant vector and tensor fields and solving the SST and SGG problem in a simply structured isotropic framework. The observables we discuss are presented in Tables 1, 2 and 3.

Quantity	Operator	Symbol	Order
gravity anomaly	Λ_A	$\frac{n-1}{\sigma}$	1
geoid undulations	Λ_U	σ^2	0
Stokes operator	Λ_{St}	$\frac{\sigma}{n-1}$	-1
first radial derivative	$\Lambda_{\frac{\partial}{\partial r}}$	$-\frac{n+1}{\sigma}$	1
second radial derivative	$\Lambda_{\frac{\partial^2}{\partial r^2}}$	$\frac{(n+1)(n+2)}{\sigma^2}$	2
upward continuation	Λ_{UPC}	$\left(\frac{\sigma}{\gamma}\right)^n$	$-\infty$
scalar SST	Λ_{SST}	$\left(\frac{\sigma}{\gamma}\right)^n \frac{n+1}{\gamma}$	$-\infty$
scalar SGG	Λ_{SGG}	$\left(\frac{\sigma}{\gamma}\right)^n \frac{(n+1)(n+2)}{\gamma^2}$	$-\infty$

TABLE 1. Scalar geodetic observables leading to isotropic pseudodifferential operators (note that the symbol is given with respect to $H_{n,m}$).

4.1. Scalar theory

We start with the scalar definition and give some examples.

Definition 4.1. Let $\mathcal{H}_s(\overline{\Omega_\tau^{\text{ext}}})$ and $\mathcal{H}_s(\overline{\Omega_\rho^{\text{ext}}})$ be Sobolev spaces, $\tau, \rho > 0$. Furthermore, let $\{\Lambda^\wedge(n)\}_{n\in\mathbb{N}_0}$ be a sequence of real numbers. The operator $\Lambda : \mathcal{H}_s(\overline{\Omega_\tau^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\rho^{\text{ext}}})$ defined by

$$\Lambda F = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \Lambda^\wedge(n) F^\wedge(n,m) H_{n,m}^s(\rho;\cdot) \tag{4.2}$$

is called a *scalar pseudodifferential operator* of order t, if

$$\lim_{n\to\infty} \frac{|\Lambda^\wedge(n)|}{\left(n+\frac{1}{2}\right)^t} = \text{const} \neq 0 \tag{4.3}$$

for some $t \in \mathbb{R}$. The sequence $\{\Lambda^\wedge(n)\}_{n\in\mathbb{N}_0}$ is called the *symbol* of Λ. Moreover, if the limit relation

$$\lim_{n\to\infty} \frac{|\Lambda^\wedge(n)|}{\left(n+\frac{1}{2}\right)^t} = 0 \tag{4.4}$$

holds for all $t \in \mathbb{R}$, then the operator is called a *pseudodifferential operator of order* $-\infty$.

Quantity	Operator	Symbol	Order
first tangential derivative	∇^*	$\frac{n}{\sigma}\sqrt{\frac{n+1}{2n+1}}$, $\frac{n+1}{\sigma}\sqrt{\frac{n}{2n+1}}$, 0	1
second mixed derivative	$\nabla^*\frac{\partial V}{\partial r}$,	$\frac{n(n+1)}{\sigma^2}\sqrt{\frac{n+1}{2n+1}}$, $\frac{(n+1)^2}{\sigma^2}\sqrt{\frac{n}{2n+1}}$, 0	2
vectorial SST	λ_{SST}	$\left(\frac{\sigma}{\gamma}\right)^n \frac{n}{\gamma}\sqrt{\frac{n+1}{2n+1}}$, $\left(\frac{\sigma}{\gamma}\right)^n \frac{n+1}{\gamma}\sqrt{\frac{n}{2n+1}}$, 0	$-\infty$,
vectorial SGG	λ_{SGG}	$\left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+1)}{\gamma^2}\sqrt{\frac{n+1}{2n+1}}$ $\left(\frac{\sigma}{\gamma}\right)^n \frac{(n+1)^2}{\gamma}\sqrt{\frac{n}{2n+1}}$, 0	$-\infty$

TABLE 2. Vectorial geodetic observables leading to isotropic pseudodifferential operators (note that the symbol is given with respect to $h^{(i)}_{n,m}$, $i = 1, 2, 3$ from top to down for each operator).

Note that the convergence of the series in (4.2) is understood in $\mathcal{H}_s(\overline{\Omega^{\text{ext}}_\rho})$-topology. As an immediate consequence (cf. [69]), we have the important relation

$$\Lambda H^s_{n,m}(\tau;\cdot) = \Lambda^\wedge(n) H^s_{n,m}(\rho;\cdot). \tag{4.5}$$

In other words, we have the requirement that the outer harmonics are the eigenfunctions of the operator Λ, and the invertibility has to be controlled by the invertibility of the values $\Lambda^\wedge(n)$, $n \in \mathbb{N}_0$. The symbol has many appealing properties (cf. [69]): It is easily seen that

$$(\Lambda' + \Lambda'')^\wedge(n) = (\Lambda')^\wedge(n) + (\Lambda'')^\wedge(n), \tag{4.6}$$

$$(\Lambda'\Lambda'')^\wedge(n) = (\Lambda')^\wedge(n)(\Lambda'')^\wedge(n), \tag{4.7}$$

for all $n \in \mathbb{N}_0$.

As any "output function" (output signal) can be expanded into an orthogonal series of outer harmonics

$$G = \Lambda F = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1} \Lambda^\wedge(n) F^\wedge(n,m) H^s_{n,m}(\rho;\cdot) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1} G^\wedge(n,m) H^s_{n,m}(\rho;\cdot) \tag{4.8}$$

Quantity	Operator	Symbol	Order
second tangential derivative	$\nabla^* \otimes \nabla^*$	$\frac{n(n+1)}{\sigma^2(2n+1)(2n+3)}\sqrt{(n+2)(n+1)(2n+1)(2n+3)}$, $\frac{-(n+1)(n-1)}{\sigma^2(2n-1)(2n+1)}\sqrt{3}n^2$, 0, $\frac{-n(n+2)}{\sigma^2(2n+3)(2n+1)}(n+1)\sqrt{(2n+1)(2n+3)}$, $\frac{n(n+1)(n+2)}{\sigma^2(2n-1)(2n+1)}\sqrt{n(n-1)(2n-1)(2n+1)}$, 0, for $(i,k) \in \{(2,3),(3,1),(3,2),(3,3)\}$	2
tensorial SGG	$\boldsymbol{\lambda}_{SGG}$	$\left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+1)}{\gamma^2(2n+1)(2n+3)}\sqrt{(n+2)(n+1)(2n+1)(2n+3)}$, $-\left(\frac{\sigma}{\gamma}\right)^n \frac{-(n+1)(n-1)}{\gamma^2(2n-1)(2n+1)}\sqrt{3}n^2$, 0, $-\left(\frac{\sigma}{\gamma}\right)^n \frac{-n(n+2)}{\gamma^2(2n+3)(2n+1)}(n+1)\sqrt{(2n+1)(2n+3)}$, $\left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+1)(n+2)}{\sigma^2(2n-1)(2n+1)}\sqrt{n(n-1)(2n-1)(2n+1)}$, 0, for $(i,k) \in \{(2,3),(3,1),(3,2),(3,3)\}$	$-\infty$

TABLE 3. Tensorial geodetic observables leading to isotropic pseudodifferential operators (note that the symbol is given with respect to $\mathbf{h}_{n,m}^{(i,k)}$, $i,k = 1,2,3$, from top to down $((1,1),(1,2),\ldots,(3,2),(3,3))$ for each operator).

in the sense of $\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\rho^{\text{ext}}})}$, we are confronted with a spectral representation of the form

$$G^\wedge(n,m) = (\Lambda F)^\wedge(n,m) = \Lambda^\wedge(n)\ F^\wedge(n,m), \quad n \in \mathbb{N}_0,\ k = 1,\ldots,2n+1. \tag{4.9}$$

This means that the "amplitude spectrum" $\{G^\wedge(n,m)\}$ of the response of Λ is described in terms of the amplitude spectrum of functions (signals) F by a simple multiplication by the "transfer" $\Lambda^\wedge(n)$.

The following list contains (scalar) pseudodifferential operators which are of importance for geodetic applications.

Consider a potential F of the class $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, that is

$$F = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} F^\wedge(n,m) H_{n,m}^s(\sigma;\cdot), \tag{4.10}$$

where we use the geometric concept as explained in Section 3 and shown in Figure 3.1.

(i) *Gravity Anomalies*. The problem of determining the disturbing potential U with $\Lambda(U) = F$ from prescribed gravity anomalies F is the "fundamental

problem of classical physical geodesy" (see, e.g., [37, 43, 53, 69]). The operator related to gravity anomalies $\Lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ has the symbol

$$\Lambda^\wedge(n) = \frac{n-1}{\sigma}. \tag{4.11}$$

(ii) *Geoid Undulations*. The operator related to geoid undulations $\Lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ has the symbol

$$\Lambda^\wedge(n) = \sigma^2. \tag{4.12}$$

(iii) *Stokes Operator*. This operator is defined by

$$\Lambda(F)(x) = \frac{\sigma}{4\pi} \int_{\Omega_\sigma} St(x,y)F(y), d\omega(y), \quad x \in \Omega_\sigma \tag{4.13}$$

where $St(\cdot,\cdot)$ is the Stokes kernel (cf. [32, 68, 69]).
The Stokes operator $\Lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ has the symbol

$$(\Lambda)^\wedge(n) = \begin{cases} 0\,, & \text{for} \quad n = 1 \\ \frac{\sigma}{n-1}\,, & \text{for} \quad n = 0, 2, 3, 4, \dots\ . \end{cases} \tag{4.14}$$

(iv) *Upward Continuation Operator*. The upward continuation operator associates to $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ the solution ΛF of the Dirichlet problem $\Lambda F \in Pot^{(0)}(\overline{\Omega_\gamma^{\text{ext}}})$ corresponding to the boundary values $(\Lambda F)|_{\Omega_\gamma} = F|_{\Omega_\gamma}$. The upward continuation operator $\Lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$ has the symbol

$$\Lambda^\wedge(n) = \left(\frac{\sigma}{\gamma}\right)^n, \quad n \in \mathbb{N}_0. \tag{4.15}$$

The upward continuation operator indeed plays an important role in the mathematical treatment of spaceborne problems, since it relates potential values at height σ to potential values at height $\gamma (> \sigma)$.

(v) *Operator of the (Negative) First-order Radial Derivative on* Ω_σ. This operator associates to $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ the solution ΛF of the Dirichlet problem $\Lambda F \in Pot^{(0)}(\overline{\Omega_\sigma^{\text{ext}}})$ corresponding to the boundary values $(\Lambda F)|_{\Omega_\sigma} = -\frac{\partial}{\partial r} F|_{\Omega_\sigma}$. Λ is a pseudodifferential operator of order 1 with symbol $\{\Lambda^\wedge(n)\}_{n\in\mathbb{N}_0}$ given by

$$\Lambda^\wedge(n) = \frac{n+1}{\sigma}, \quad n \in \mathbb{N}_0. \tag{4.16}$$

In fact, Λ is the "harmonic continuation" of the radial derivative on Ω_σ into the outer space $\Omega_\sigma^{\text{ext}}$ and is important in case of the SST problem.

(vi) *Operator of the Second-order Radial Derivative on* Ω_σ. This operator associates to $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ the solution ΛF of the Dirichlet problem $\Lambda F \in Pot^{(0)}(\overline{\Omega_\sigma^{\text{ext}}})$ corresponding to the boundary values $(\Lambda F)|_{\Omega_\sigma} = \frac{\partial^2}{\partial r^2} F|_{\Omega_\sigma}$. Λ is a pseudodifferential operator of order 2 with symbol $\{\Lambda^\wedge(n)\}_{n\in\mathbb{N}_0}$ given by

$$\Lambda^\wedge(n) = \frac{(n+1)(n+2)}{\sigma^2}, \quad n \in \mathbb{N}_0. \tag{4.17}$$

Λ is the "harmonic continuation" of the second radial derivative on Ω_σ into the outer space $\Omega_\sigma^{\text{ext}}$ and is important in case of the SGG problem.

4.2. Vectorial theory

We now introduce vectorial pseudodifferential operators and give two examples.

Definition 4.2. Let $\mathcal{H}_s(\overline{\Omega_\tau^{\text{ext}}})$ be a scalar Sobolev space and $h_s^{(i)}(\overline{\Omega_\rho^{\text{ext}}})$ a vectorial Sobolev space, $\tau, \rho > 0$, $i \in \{1,2,3\}$. Furthermore, let $\{\lambda^{(i)\wedge}(n)\}_{n\in\mathbb{N}_{0_i}}$ be a sequence of real numbers for $i = 1,2,3$. The operator $\lambda^{(i)} : \mathcal{H}_s(\overline{\Omega_\tau^{\text{ext}}}) \to h_s^{(i)}(\overline{\Omega_\rho^{\text{ext}}})$ defined by

$$\lambda^{(i)}F = \sum_{n=0_i}^{\infty} \sum_{m=1}^{2n+1} \lambda^{(i)\wedge}(n) F^\wedge(n,m) h_{n,m}^{(i)s}(\rho;\cdot) \tag{4.18}$$

is called a *vectorial pseudodifferential operator* of kind i and order t, if

$$\lim_{n\to\infty} \frac{|\lambda^{(i)\wedge}(n)|}{(n+\frac{1}{2})^t} = \text{ const } \neq 0 \tag{4.19}$$

for some $t \in \mathbb{R}$. Moreover, if the limit relation

$$\lim_{n\to\infty} \frac{|\lambda^{(i)\wedge}(n)|}{(n+\frac{1}{2})^t} = 0 \tag{4.20}$$

holds for all $t \in \mathbb{R}$, then the operator $\lambda^{(i)}$ is called a *vectorial pseudodifferential operator of kind i and order $-\infty$*. The sequence $\{\lambda^{(i)\wedge}(n)\}$ is called the *symbol* of $\lambda^{(i)}$. Further on, the operator $\lambda : \mathcal{H}_s(\overline{\Omega_\tau^{\text{ext}}}) \to h_s(\overline{\Omega_\rho^{\text{ext}}})$ defined by

$$\lambda = \sum_{i=1}^{3} \lambda^{(i)}, \tag{4.21}$$

is called a *vectorial pseudodifferential operator of order t*, where $t = \max_{i=1}^3$ (*order of* $\lambda^{(i)}$). Moreover, if the limit relation

$$\lim_{n\to\infty} \frac{|\lambda^{(i)\wedge}(n)|}{(n+\frac{1}{2})^t} = 0 \tag{4.22}$$

holds for all $t \in \mathbb{R}$, and all $i \in \{1,2,3\}$, then the operator λ is called a *vectorial pseudodifferential operator of order $-\infty$*.

We now give two examples of vectorial pseudodifferential operators which are important for geodetic applications. We use the surface gradient on the sphere Ω_σ defined by

$$\nabla^{*;\sigma} = \frac{1}{\sigma}\nabla^*. \tag{4.23}$$

(iv) *The Operator of the First-order Tangential Derivatives on Ω_σ*. This operator associates to $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ the solution λF of the Dirichlet problem $\lambda F \in h_s(\overline{\Omega_\sigma^{\text{ext}}})$ corresponding to the boundary value $(\lambda F)|_{\Omega_\sigma} = \nabla^{*,\sigma} F|_{\Omega_\sigma}$. λ is a

pseudodifferential operator of order 1 with symbol $\{\lambda^{(i)\wedge}(n)\}_{n\in\mathbb{N}_{0_i}}$ given by

$$\lambda^{(i)\wedge}(n)=\begin{cases}\frac{n}{\sigma}\sqrt{\frac{n+1}{2n+1}}, & i=1,\\ \frac{n+1}{\sigma}\sqrt{\frac{n}{2n+1}}, & i=2,\\ 0, & i=3.\end{cases}\tag{4.24}$$

In fact, Λ is the "harmonic continuation" of the tangential derivative on Ω_σ into the outer space $\Omega_\sigma^{\text{ext}}$ and is important in case of the SST problem.

(v) *The Operator of the (Negative) Second-order Mixed Derivatives on* Ω_σ. This operator associates to $F\in\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ the solution λF of the Dirichlet problem $\lambda F\in h_s(\overline{\Omega_\sigma^{\text{ext}}})$ corresponding to the boundary values $(\lambda F)|_{\Omega_\sigma}=-\frac{\partial}{\partial r}\nabla_\xi^{*,\sigma}F|_{\Omega_\sigma}$. λ is a pseudodifferential operator of second order with symbol $\{\lambda^{(i)\wedge}(n)\}_{n\in\mathbb{N}_{0_i}}$ given by

$$\lambda^{(i)\wedge}(n)=\begin{cases}\frac{n(n+1)}{\sigma^2}\sqrt{\frac{n+1}{2n+1}}, & i=1,\\ \frac{(n+1)^2}{\sigma^2}\sqrt{\frac{n}{2n+1}}, & i=2,\\ 0, & i=3.\end{cases}\tag{4.25}$$

Λ is the "harmonic continuation" of the second-order mixed derivatives on Ω_σ into the outer space $\Omega_\sigma^{\text{ext}}$ and is important in case of the SGG problem.

4.3. Tensorial theory

The introduction of tensorial pseudodifferential operators is straightforward.

Definition 4.3. Let $\mathcal{H}_s(\overline{\Omega_\tau^{\text{ext}}})$ be a scalar Sobolev space and $\mathbf{h}_s^{(i,k)}(\overline{\Omega_\rho^{\text{ext}}})$ a tensorial Sobolev space, $\tau,\rho>0$, $i,k\in\{1,2,3\}$. Furthermore, for $i,k\in\{1,2,3\}$, let $\boldsymbol{\lambda}^{(i,k)\wedge}(n)_{n\in\mathbb{N}_{0_{ik}}}$ be a sequence of real numbers. The operator $\boldsymbol{\lambda}^{(i,k)\wedge}(n):\mathcal{H}_s(\overline{\Omega_\tau^{\text{ext}}})\to\mathbf{h}_s^{(i,k)}(\overline{\Omega_\rho^{\text{ext}}})$ defined by

$$\boldsymbol{\lambda}^{(i,k)}F=\sum_{n=0_{ik}}^{\infty}\sum_{m=1}^{2n+1}\boldsymbol{\lambda}^{(i,k)\wedge}(n)\}F^\wedge(n,m)\mathbf{h}_{n,m}^{(i,k)s}(\rho;\cdot)\tag{4.26}$$

is called a *tensorial pseudodifferential operator* of kind (i,k) and order t, if the limit relation

$$\lim_{n\to\infty}\frac{|\,\boldsymbol{\lambda}^{(i,k)\wedge}(n)|}{(n+\frac{1}{2})^t}=\text{ const }\neq 0\tag{4.27}$$

is satisfied for some $t\in\mathbb{R}$. Moreover, if the limit relation

$$\lim_{n\to\infty}\frac{|\,\boldsymbol{\lambda}^{(i,k)\wedge}(n)|}{(n+\frac{1}{2})^t}=0\tag{4.28}$$

holds for all $t\in\mathbb{R}$, then the operator $\boldsymbol{\lambda}$ is called a *pseudodifferential operator of kind* (i,k) *and order* $-\infty$. The sequence $\{\,\boldsymbol{\lambda}^{(i,k)\wedge}(n)\}$ is called the (*spherical*)

symbol of $\boldsymbol{\lambda}^{(i,k)}$. Further on, the operator $\boldsymbol{\lambda} : \mathcal{H}_s(\overline{\Omega_\tau^{\text{ext}}}) \to \mathbf{h}_s(\overline{\Omega_\rho^{\text{ext}}})$ defined by

$$\boldsymbol{\lambda} = \sum_{i=1}^{3} \sum_{k=1}^{3} \boldsymbol{\lambda}^{(i,k)}, \tag{4.29}$$

is called a *tensorial pseudodifferential operator of order* t, where $t = \max_{i,k=1}^{3}$ (*order of* $\boldsymbol{\lambda}^{(i,k)}$). Moreover, if the limit relation

$$\lim_{n\to\infty} \frac{|\,\boldsymbol{\lambda}^{(i,k)\wedge}(n)|}{(n+\frac{1}{2})^t} = 0 \tag{4.30}$$

holds for all $t \in \mathbb{R}$, and all $i, k \in \{1, 2, 3\}$, then the operator $\boldsymbol{\lambda}$ is called a *pseudodifferential operator of order* $-\infty$.

Finally, we mention one important example.

(iv) *The Operator of the Second-order Tangential Derivatives on* Ω_σ. This operator associates to $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ the solution $\boldsymbol{\lambda}F$ of the Dirichlet problem $\boldsymbol{\lambda}F \in \mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})$ corresponding to the boundary values $(\boldsymbol{\lambda}F)|_{\Omega_\sigma} = \nabla^{*,\sigma} \otimes \nabla^{*,\sigma} F|_{\Omega_\sigma}$. It is a pseudodifferential operator of order 2 with the symbol $\{\boldsymbol{\lambda}^{(i,k)\wedge}(n)\}_{n\in\mathbb{N}_{0_{ik}}}$ given by

$$\boldsymbol{\lambda}^{(i,k)\wedge}(n) \tag{4.31}$$
$$= \begin{cases} \frac{n(n+1)}{\sigma^2(2n+1)(2n+3)}\sqrt{(n+2)(n+1)(2n+1)(2n+3)}, & (i,k) = (1,1), \\ \frac{-(n+1)(n-1)}{\sigma^2(2n-1)(2n+1)}\sqrt{3}n^2, & (i,k) = (1,2), \\ \frac{-n(n+2)}{\sigma^2(2n+3)(2n+1)}(n+1)\sqrt{(2n+1)(2n+3)}, & (i,k) = (2,1), \\ \frac{n(n+1)(n+2)}{\sigma^2(2n-1)(2n+1)}\sqrt{n(n-1)(2n-1)(2n+1)}, & (i,k) = (2,2), \\ 0, & \text{else.} \end{cases}$$

Λ is the "harmonic continuation" of the second-order tangential derivatives on Ω_σ into the outer space $\Omega_\sigma^{\text{ext}}$ and is important in case of the SGG problem.

5. Reproducing kernel structure and observational functionals

Of great importance for our considerations are Sobolev spaces equipped with a reproducing kernel structure. The importance of the reproducing kernel lies in the fact that it determines the norm of the dual space. Furthermore, no computational work must be done to evaluate inner products involving reproducing kernel expressions. Within this section, we focus on scalar theory and essentially follow [19]. The extension to vectorial and tensorial reproducing kernel Sobolev spaces is not hard to perform.

5.1. Reproducing Hilbert spaces

Theorem 5.1. *Let the sequence $\{A_n\}$ be summable in the sense of Definition* 3.3. *Then $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ (more explicitly, $\mathcal{H}(\{A_n\};\overline{\Omega_\sigma^{\rm ext}})$) is a Hilbert subspace of the space $Pot^{(0)}(\overline{\Omega_\sigma^{\rm ext}})$. The space $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ has the reproducing kernel function*

$$K_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}(\cdot,\cdot) : \overline{\Omega_\sigma^{\rm ext}} \times \overline{\Omega_\sigma^{\rm ext}} \to \mathbb{R}$$

given by

$$K_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}(x,y) = \sum_{n\in\mathcal{N}(A_n)} \sum_{m=1}^{2n+1} H_{n,m}^{*\{A_n\}}(\sigma;x) H_{n,m}^{*\{A_n\}}(\sigma;y),$$

$x, y \in \overline{\Omega_\sigma^{\rm ext}}$.

If $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ has a reproducing kernel, then the Fourier (orthogonal) expansion of a potential in terms of the Hilbert basis $\{H_{n,k}^*(\sigma;\cdot)\}$ in $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ converges uniformly on the domain $\overline{\Omega_\sigma^{\rm ext}}$ (cf. [3, 7]). To be more specific, the relation

$$\lim_{N\to\infty} \left\| F - \sum_{\substack{n\in\mathcal{N}\\ n\le N}} \sum_{m=1}^{2n+1} F^\wedge(n,m) H_{n,m}^{*\{A_n\}}(\sigma;\cdot) \right\|_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})} = 0$$

implies

$$\lim_{N\to\infty} \sup_{x\in\overline{\Omega_\sigma^{\rm ext}}} \left| F(x) - \sum_{\substack{n\in\mathcal{N}\\ n\le N}} \sum_{m=1}^{2n+1} F^\wedge(n,m) H_{n,m}^*(\sigma;x) \right| = 0.$$

The representer of a bounded linear functional $\mathcal{L}$ on $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ has a simple expression. More explicitly, $L(x) = \mathcal{L}K_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}(\cdot,x), x \in \overline{\Omega_\sigma^{\rm ext}}$, is in $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$, and for all $F \in \mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ we have $\mathcal{L}F = (F,L)_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}$ (note that x is held fixed and $\mathcal{L}$ is applied to $K_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}(\cdot,x)$ as a function of the first variable). Obviously, $(L,L)_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})} = \mathcal{L}\mathcal{L}K_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}(\cdot,\cdot) = (\mathcal{L},\mathcal{L})_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})^*}$. The *dual space* $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})^*$ of $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ (i.e., the space of all linear bounded functionals on $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$) is a Hilbert space with respect to $\|\cdot\|_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})^*} = (\cdot,\cdot)^{\frac{1}{2}}_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})^*}$; the spaces $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ and $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})^*$ are known as isomorphic and isometric (see, e.g., [7]).

Reproducing kernel representations may be used to act as basis system in reproducing Sobolev spaces.

Theorem 5.2. *Let $\{A_n\}$ be summable in the sense of Definition* 3.3. *Assume that X is a countable dense set of points on a regular surface $\Xi \subset \overline{\Omega_\sigma^{\rm ext}}$ (for example, Runge sphere Ω_σ, real Earth's surface Σ). Then*

$$\overline{\mathrm{span}_{x\in X} K_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}(x,\cdot)}^{\|\cdot\|_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}} = \mathcal{H}(\overline{\Omega_\sigma^{\rm ext}}).$$

Theorem 5.2 allows an obvious generalization by means of bounded linear functionals on $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$.

Theorem 5.3. *Let $\{A_n\}$ be summable. Assume that X is a countable dense set of linear functionals in $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})^*$. Then*

$$\overline{\text{span}_{\mathcal{L}\in X}\mathcal{L}K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot)}^{\|\cdot\|_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}} = \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}}).$$

The set of all finite linear combinations of outer harmonics is dense in the space $Pot^{(0)}(\overline{\Omega_\sigma^{\text{ext}}})$ in the sense of $\|\cdot\|_{C^{(0)}(\overline{\Omega_\sigma^{\text{ext}}})}$. Hence, $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ is a dense subset of $Pot^{(0)}(\overline{\Omega_\sigma^{\text{ext}}})$, too. This leads us to the following corollary.

Corollary 5.4. *Under the assumption of Theorem* 5.3

$$\overline{\text{span}_{\mathcal{L}\in X}\mathcal{L}K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot)}^{\|\cdot\|_{C^{(0)}(\overline{\Omega_\sigma^{\text{ext}}})}} = Pot^{(0)}(\overline{\Omega_\sigma^{\text{ext}}}).$$

Next we come to the problem of specifying certain types of sequences $\{A_n\}$ such that $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})(=\mathcal{H}(\{A_n\};\overline{\Omega_\sigma^{\text{ext}}}))$ is a reproducing kernel Hilbert space. We restrict ourselves to those kernel functions which are usable later on in multiscale approximation. Other types of kernel functions which are known from spline interpolation or smoothing procedures (see, for example, [14–16, 18, 20, 49, 55, 56, 72]) are not discussed here.

Our list of (reproducing) kernel functions is divided into two parts, namely *bandlimited kernel functions* such as Shannon's kernel, smoothed Shannon kernels, etc., and *non-bandlimited kernel functions* such as rational kernel functions, exponential kernel functions, (smoothed) Haar kernel functions, etc.

5.2. Bandlimited kernel functions

These kernel functions are characterized by the property that only a finite number of coefficients A_n does not vanish. Consequently, the reproducing kernel Hilbert space is of finite dimension.

At this stage two important cases of bandlimited kernels should be mentioned:

(a) *The Shannon Kernel* (see Figure 5.1). For a non-negative integer N we let

$$A_n = \begin{cases} 1, & n \in [0, N+1), \\ 0, & n \in [N+1, \infty), \end{cases}$$

i.e., $\mathcal{N}(A_n) = \{0, \ldots, N\}$. Obviously, the reproducing kernel Hilbert space $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ is equal to the space $Harm_{0,\ldots,N}(\overline{\Omega_\sigma^{\text{ext}}})$ of outer harmonics of degree $\leq N$. The reproducing kernel function $K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) : \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\sigma^{\text{ext}}} \to \mathbb{R}$, i.e., the Shannon kernel, reads as follows:

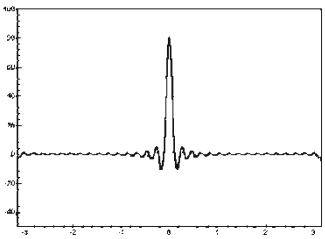
$$\begin{aligned} K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(x,y) &= \sum_{0\leq n\leq N} \sum_{m=1}^{2n+1} H_{n,m}^*(\sigma;x) H_{n,m}^*(\sigma;y) \\ &= \sum_{0\leq n\leq N} \frac{2n+1}{4\pi\sigma^2} \left(\frac{\sigma^2}{|x|\,|y|}\right)^{n+1} P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right). \end{aligned} \tag{5.1}$$

Observing the well-known recursion relation for Legendre polynomials

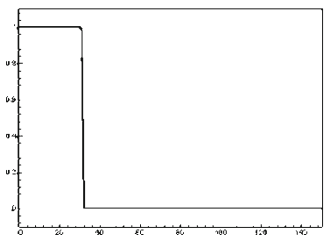
$$(n+1)(P_{n+1}(t)-P_n(t))-n(P_n(t)-P_{n-1}(t)) = (2n+1)(t-1)P_n(t), \quad n \geq 1, \tag{5.2}$$

we obtain for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$

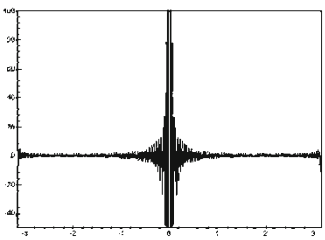
$$\left(\frac{x}{|x|}\cdot\frac{y}{|y|}-1\right) K_{\mathcal{H}(\overline{\Omega_\sigma^{\mathrm{ext}}})}(x,y) = \frac{N+1}{4\pi\sigma^2}\left(P_{N+1}\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right)-P_N\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right)\right). \tag{5.3}$$



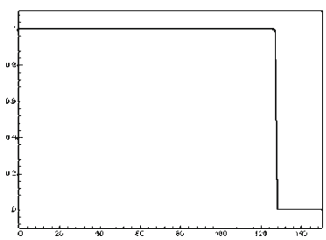
(a) $K(x,y)$ for $N = 2^5 - 1$



(b) $K^\wedge(n)$ for $N = 2^5 - 1$



(c) $K(x,y)$ for $N = 2^7 - 1$



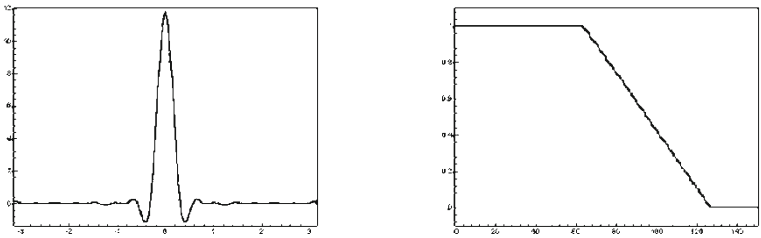
(d) $K^\wedge(n)$ for $N = 2^7 - 1$

FIGURE 5.1. Shannon kernel with $N = 2^5 - 1$ (above) and $N = 2^7 - 1$ (below): space domain, i.e., $K(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $K^\wedge(n) = A_n$ (right).

(b) *Smoothed Shannon Kernels* (see Figure 5.2). For (fixed) non-negative integers N, M with $N > M+1$ we let

$$A_n = \begin{cases} 1, & n \in [0, M+1), \\ \frac{N-m}{N-M}, & n \in [M+1, N+1), \\ 0, & n \in [N+1, \infty). \end{cases}$$

Of course, many other suitable choices can be found for practical purposes.



(a) $K(x,y)$ (b) $K^\wedge(n)$

FIGURE 5.2. Smoothed Shannon kernel with $M = 2^6$ and $N = 2^7 - 1$: space domain, i.e., $K(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $K^\wedge(n) = A_n$ (right).

5.3. Non-bandlimited kernel functions

All non-bandlimited kernels share the property that an infinite number of coefficients A_n is different from zero. The corresponding reproducing Hilbert kernel spaces are infinite-dimensional. We mention rational kernels, exponential kernels, and "locally supported" kernels, i.e., (smoothed) Haar kernels.

(a) *Rational Kernels* (see Figure 5.3). Let $\{A_n\}$ be a sequence of real numbers A_n satisfying the following conditions:
   (i) $n \mapsto A_n^2$, $n \in \mathbb{N}_0$, is a (real) rational function (in the integer variable n).
   (ii) There exist two positive constants C, C' with

$$C\left(\frac{n+(\frac{1}{2})}{\sigma}\right)^{2+\varepsilon} \le A_n^2 \le C'\left(\frac{n+(\frac{1}{2})}{\sigma}\right)^{\alpha} \tag{5.4}$$

   for some $\varepsilon > 0$, $\alpha \ge 2+\varepsilon$.

   Then the norm reads

$$\|F\|^2_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})} = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1} A_n^2 \left(F^\wedge(n,m)\right)^2.$$

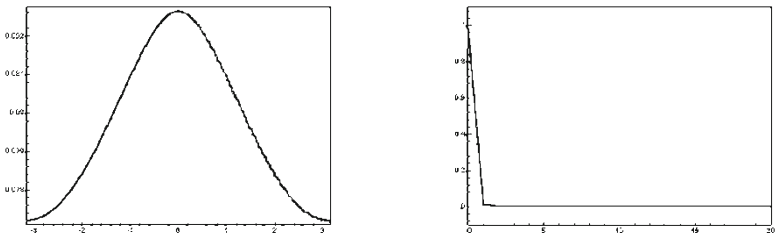
   For the reproducing kernel in $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ we find the representation

$$K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(x,y) = \sum_{n=0}^{\infty} \frac{1}{A_n^2}\frac{2n+1}{4\pi\sigma^2}\left(\frac{\sigma^2}{|x|\,|y|}\right)^{n+1} P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right).$$

   $\left\{A_n\left(\frac{n+\frac{1}{2}}{\sigma}\right)^{-\beta}\right\}$ is summable for all $\beta < \varepsilon/2$.

(b) *Exponential Kernels*. An alternative to come to candidates of reproducing kernel sum representations with an exponential rate of convergence is to use a sequence $\{A_n\}$ of the form

$$A_n = \left(\frac{\sigma}{\sigma'}\right)^n B_n, \quad n \in \mathcal{N}, \tag{5.5}$$



(a) $K(x,y)$ (b) $K^\wedge(n)$

FIGURE 5.3. Rational kernel with $A_n^2 = (1+n)^{-s}$, $s = 6.5$: space domain, i.e., $K(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $K^\wedge(n) = A_n$ (right).

with $\sigma' < \sigma$ and B_n satisfying

$$0 < B_n^2 \leq C' \left(\frac{n + (\frac{1}{2})}{\sigma} \right)^\alpha \tag{5.6}$$

for all $n \in \mathcal{N}$, some value α and a positive constant C'. The radius $\sigma' (< \sigma)$ should be taken close to the value σ (i.e., σ' is assumed to be the radius of a Runge sphere so that σ/σ' is close to 1). It is evident that an "inner radius" σ' gives additional flexibility in choosing the norm of the Hilbert space and also results in more general sequences $\{A_n\}$ being possible. On the other hand, the radius σ' appears as an artificial value in the infinite sum of the kernel to force an exponential rate of sum convergence. In conclusion, the sequence $\left\{ A_n \left(\frac{n+\frac{1}{2}}{\sigma} \right)^{-\beta} \right\}$ is summable for every β.

Kernel representations of type (5.5) for $(x,y) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\sigma^{\text{ext}}}$

$$K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(x,y) = \sum_{n \in \mathcal{N}} \frac{1}{B_n^2} \frac{2n+1}{4\pi\sigma'^2} \left(\frac{\sigma'^2}{|x|\,|y|} \right)^{n+1} P_n \left(\frac{x}{|x|} \cdot \frac{y}{|y|} \right)$$

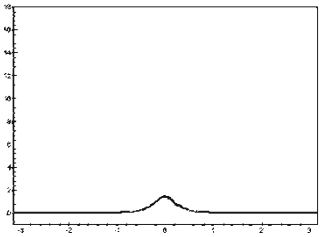
are well known from geophysical applications (see, for example, [14, 32, 55, 72]).

Far- and Near-Field Methods as well as Multipole Methods are explained in the Ph.D.-thesis [39] and can also be found in [24, 40, 41] and in the contribution [42] in this volume.

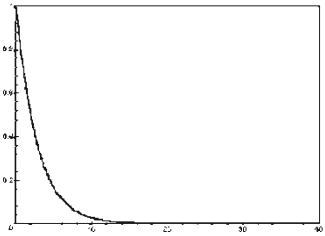
Of particular importance for purposes of minimum norm (spline) interpolation and smoothing (cf., e.g., [14–16, 18, 72]) are kernels, which are available in terms of elementary functions. We only mention here (cf. [52]):

(i) *Abel–Poisson kernel* (see Figure 5.4):

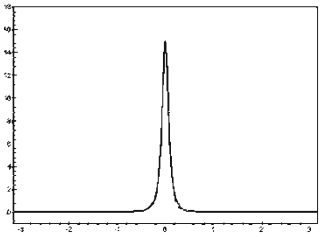
$$B_n^2 = 1, \quad n \in \mathbb{N}_0. \tag{5.7}$$



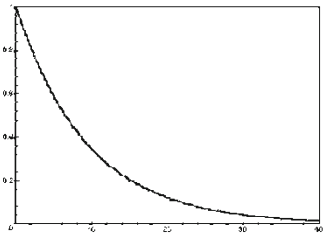
(a) $K(x,y)$ for $\frac{\sigma'}{\sigma} = 0.7$



(b) $K^\wedge(n)$ for $\frac{\sigma'}{\sigma} = 0.7$



(c) $K(x,y)$ for $\frac{\sigma'}{\sigma} = 0.9$



(d) $K^\wedge(n)$ for $\frac{\sigma'}{\sigma} = 0.9$

FIGURE 5.4. Abel–Poisson kernel with $\frac{\sigma'}{\sigma} = 0.7$ (above) and $\frac{\sigma'}{\sigma} = 0.9$ (below): space domain, i.e., $K(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $K^\wedge(n) = A_n$ (right).

The kernel reads as follows:

$$K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(x,y) = \frac{|x|\,|y|}{4\pi\sigma'^2}\,\frac{|x|^2|y|^2 - \sigma'^4}{(L(x,y))^{3/2}}, \quad x,y \in \overline{\Omega_\sigma^{\text{ext}}},$$

where we have used the abbreviation

$$L(x,y) = |x|^2|y|^2 - 2\sigma'^2 x \cdot y + \sigma'^4.$$

(ii) *"Singularity kernel"* (see Figure 5.5)

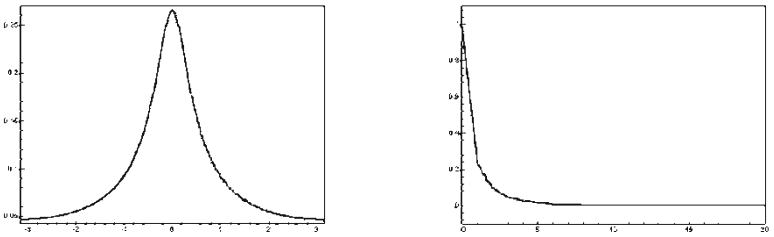
$$B_n^2 = (2n+1)/2, \quad n \in \mathbb{N}_0. \tag{5.8}$$

The kernel is given by

$$K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(x,y) = \frac{1}{4\pi}\,\frac{1}{(L(x,y))^{\frac{1}{2}}}, \quad x,y \in \overline{\Omega_\sigma^{\text{ext}}}.$$

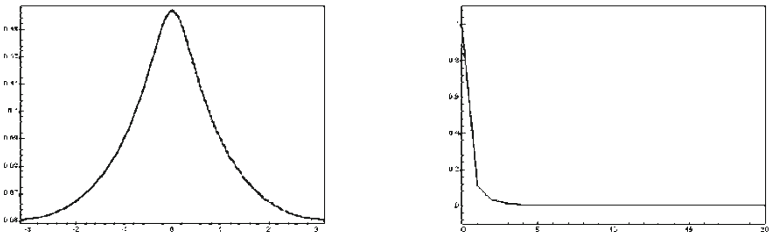
(iii) *"Logarithmic kernel"* (see Figure 5.6)

$$B_n^2 = (2n+1)(n+1), \quad n \in \mathbb{N}_0. \tag{5.9}$$



(a) $K(x,y)$ (b) $K^\wedge(n)$

FIGURE 5.5. Singularity kernel with $\frac{\sigma'}{\sigma} = 0.7$: space domain, i.e., $K(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $K^\wedge(n) = A_n$ (right).



(a) $K(x,y)$ (b) $K^\wedge(n)$

FIGURE 5.6. Logarithmic kernel with $\frac{\sigma'}{\sigma} = 0.7$: space domain, i.e., $K(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $K^\wedge(n) = A_n$ (right).

Now we have

$$K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(x,y) = \frac{1}{4\pi\sigma'^2} \ln\left(1 + \frac{2\sigma'^2}{M(x,y)}\right), \quad x,y \in \overline{\Omega_\sigma^{\text{ext}}},$$

with

$$M(x,y) = (L(x,y))^{\frac{1}{2}} + |x|\,|y| - \sigma'^2.$$

(c) *"Locally Supported" Kernels* (*Smoothed Haar Kernels,* see Figure 5.7): Consider the piecewise polynomial function $B_h^{(k)} : [-1,+1] \to \mathbb{R}$, $k = 0,1,\ldots$ and $h \in (0,1)$ given by

$$B_h^{(k)}(t) = \begin{cases} 0, & t \in [-1,h), \\ \frac{(t-h)^k}{(1-h)^k}, & t \in [h,1], \end{cases} \tag{5.10}$$

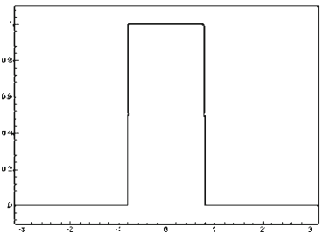
(cf. [5, 20, 21, 26, 35, 67]). Let $\xi \in \Omega = \Omega_1$ be fixed. Then the ξ-zonal function $B_h^{(k)}(\xi\ \cdot) : \Omega \to \mathbb{R}$ has a local support. More explicitly, the support of $B_h^{(k)}(\xi\ \cdot)$ is the cap with centre ξ characterized by

$$\mathrm{supp} B_h^{(k)}(\xi\ \cdot) = \{\eta \in \Omega\ :\ h \leq \xi \cdot \eta \leq 1\}.$$

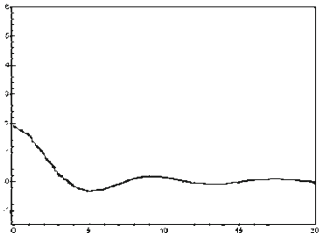
The ξ-zonal function $B_h^{(0)}(\xi\ \cdot) : \Omega \to \mathbb{R}$ given by

$$B_h^{(0)}(\xi \cdot \eta) = \begin{cases} 0 & \text{for} \quad \xi \cdot \eta \in [-1, h), \\ 1 & \text{for} \quad \xi \cdot \eta \in [h, 1]. \end{cases}$$

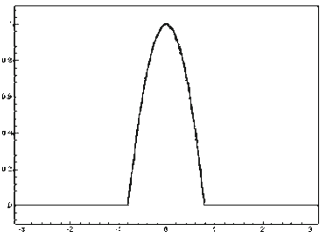
is called the *Haar kernel at position* $\xi \in \Omega$, while $B_h^{(k)}(\xi\ \cdot)$, $k > 0$, are called *"smoothed" Haar kernels at position* $\xi \in \Omega$.



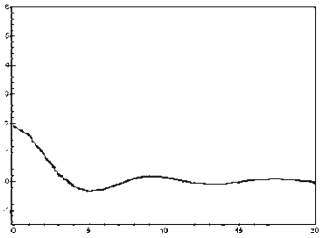
(a) Haar kernel $K(x, y)$



(b) Symbol $K^\wedge(n)$ of the Haar kernel



(c) Smoothed Haar kernel $K(x, y)$



(d) Symbol $K^\wedge(n)$ of the smoothed Haar kernel

FIGURE 5.7. Haar kernel (above) and smoothed Haar kernel (below) with $h = 0.7$: space domain, i.e., $K(x, y)$ for $(x, y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $K^\wedge(n) = A_n$ (right).

An easy calculation shows that the *iterated "Haar kernel"*

$$(B_h^{(k)})^{(2)}(\xi\ \cdot) = (B_h^{(k)} *_{L^2(\Omega)} B_h^{(k)})(\xi\ \cdot)$$

also has a cap with centre ξ as a local support:

$$\mathrm{supp}(B_h^{(k)})^{(2)}(\xi\ \cdot) = \{\eta \in \Omega\ :\ 2h^2 - 1 \leq \xi \cdot \eta \leq 1\}.$$

Expanding $B_h^{(k)}$ in terms of Legendre polynomials we obtain

$$B_h^{(k)} = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} (B_h^{(k)})^\wedge(n) P_n, \tag{5.11}$$

where

$$(B_h^{(k)})^\wedge(n) = 2\pi \int_{-1}^{+1} \left(\frac{t-h}{1-h} \right)^k P_n(t)\, dt, \quad n = 0, 1, \ldots.$$

The recurrence formulae for Legendre polynomials give us

$$(k+1)(B_h^{(k)})^\wedge(0) = 2\pi(1-h), \tag{5.12}$$

$$(k+2)(B_h^{(k)})^\wedge(1) = (k+1+h)(B_h^{(k)})^\wedge(0), \tag{5.13}$$

$$\begin{aligned}(n+k+2)(B_h^{(k)})^\wedge(n+1) &= (2n+1)h(B_h^{(k)})^\wedge(n) \\ &\quad + (k+1-n)(B_h^{(k)})^\wedge(n-1)\end{aligned} \tag{5.14}$$

(for more details the reader is referred to [26]).

For $k = 0$ it is easy to see that $\left|(B_h^{(0)})^\wedge(n)\right| = \mathcal{O}(n^{-3/2})$, $n \to \infty$. Moreover, from the recurrence relations Eqs. (5.12)–(5.14) it follows that

$$\left|(B_h^{(k)})^\wedge(n)\right| = \mathcal{O}(n^{-(3/2)-k}), \quad n \to \infty.$$

Furthermore, [67] has shown the following statements:

(i) $(B_h^{(k)})^\wedge(n) \neq 0$ for $n = 0, 1, \ldots, k+2$.

(ii) For $n \geq k+2$, $(B_h^{(k)})^\wedge(n) = 0$ if and only if $C_{n-k-1}^{k+\frac{3}{2}}(h) = 0$ (where $C_m^{k+\frac{3}{2}}$ is the Gegenbauer polynomial of order m with respect to $k + \frac{3}{2}$).

This leads us to the following result: For $k \geq 0$, $h \in (0,1)$, the sequence

$$A_n = \begin{cases} ((B_h^{(k)})^\wedge(n))^{-1}, & n \in \mathcal{N}, \\ 0, & n \in \mathcal{N}_0 \end{cases} \tag{5.15}$$

is summable.

In case of locally supported kernels we have the following lemma:

Lemma 5.5. *$\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}}) \subset Pot^{(0)}(\overline{\Omega_\sigma^{\text{ext}}})$, as defined by (5.15), is a reproducing kernel Hilbert space with the reproducing kernel*

$$K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(x,y) = \sum_{n \in \mathcal{N}} \left(\left(B_h^{(k)} \right)^{(2)} \right)^\wedge (n) \frac{2n+1}{4\pi\sigma^2} \left(\frac{\sigma^2}{|x|\,|y|} \right)^{n+1} P_n \left(\frac{x}{|x|} \cdot \frac{y}{|y|} \right). \tag{5.16}$$

Moreover, for $x = \sigma\xi$, $y = \sigma\eta$, we have

$$\sigma^2 K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(x,y)\Big|_{\substack{|x|=\sigma, \\ |y|=\sigma}} = \left(B_h^{(k)} \right)^{(2)} \left(\frac{x}{|x|} \cdot \frac{y}{|y|} \right) = \left(B_h^{(k)} \right)^{(2)} (\xi \cdot \eta),$$

where

$$\operatorname{supp}\left(B_h^{(k)}\right)^{(2)}\left(\cdot\,\frac{x}{|x|}\right)=\left\{y\in\Omega_\sigma\ :\ 2h^2-1\le\frac{x}{|x|}\cdot\frac{y}{|y|}\le 1\right\}.$$

In other words, reproducing kernel Hilbert spaces of potentials defined on and outside the sphere Ω_σ are found such that the "restriction" $(x,y)\mapsto K_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}(x,y),\quad (x,y)\in\Omega_\sigma\times\Omega_\sigma$, is a locally supported (zonal) function on Ω_σ (note that $(B_h^{(k)})^{(2)}(\xi\cdot\eta)$ is a zonal function, i.e., depends only on the scalar product of the unit vectors ξ and η).

6. Ill-posedness of the satellite problems

The question of subsets $X\subset\overline{\Omega_\gamma^{\rm ext}}$ on which observations are required in order to uniquely determine the potential $F|_{\overline{\Sigma^{\rm ext}}}$, is answered in this section. In order to handle existence and stability of the solution we give a reformulation of the pseudodifferential operators as convolution operators.

6.1. Scalar SST and SGG problem

Throughout the remaining part of this contribution, the sequence $\{A_n\}\in\mathcal{A}$ generating the reference space $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ for gravitational field determination is assumed to satisfy the so-called '*consistency conditions*':

Definition 6.1. A sequence $\{A_n\}\in A$ is said to satisfy the *consistency conditions* (CC1) and (CC2) relative to $[\sigma,\sigma^{\rm inf})$, if the following conditions are satisfied:

(CC1) A_n is different from 0 for all $n\in\mathbb{N}_0$, i.e.,

$$A_n\neq 0,\quad n=0,1,\dots, \tag{6.1}$$

and

(CC2) there exists a value τ with $\sigma\le\tau<\sigma^{\rm inf}$ such that

$$\sum_{n=0}^{\infty}(2n+1)\left(\frac{\sigma}{\tau}\right)^n\frac{1}{A_n^2}<\infty. \tag{6.2}$$

The "downward continuation problem" of determining the potential $F\in\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$ from "satellite data" $G\in\mathcal{H}_s(\overline{\Omega_\gamma^{\rm ext}})$ reads as follows.

(i) (*Scalar*) *SST Problem* (*Corresponding to the First-order Radial Derivative*). Let the values $G(x)$, $x\in X$, for some subset $X\subset\overline{\Omega_\gamma^{\rm ext}}$ be known from a function G of the class $\mathcal{H}_s(\overline{\Omega_\gamma^{\rm ext}})$. We search for a potential $F|_{\overline{\Sigma^{\rm ext}}}$ with F being from $\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$ that fulfills the (scalar) SST operator equation with the SST operator $\Lambda:\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})\to\mathcal{H}_s(\overline{\Omega_\gamma^{\rm ext}})$ given by

$$\Lambda F(x)=G(x),\quad x\in X, \tag{6.3}$$

where

$$(\Lambda F)(x)=\left(-\frac{x}{|x|}\cdot\nabla_x\right)F(x)|_{|x|=\gamma}=G(x),\quad x\in X. \tag{6.4}$$

Equation (6.4) means that the SST operator is the composition of the radial derivative and the upward continuation operator. Having in mind that the symbol of a pseudodifferential operator $\Lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}}) \to \mathcal{H}_s(\overline{\Omega_\gamma^{\rm ext}})$ satisfies $\Lambda H_{n,m}^s(\sigma;\cdot) = \Lambda^\wedge(n) H_{n,m}^s(\gamma;\cdot)$, we have

$$\Lambda^\wedge(n) = \frac{n+1}{\gamma}\left(\frac{\sigma}{\gamma}\right)^n, \quad n = 0, 1, \ldots, \tag{6.5}$$

and the SST operator is given by

$$\Lambda F(x) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \Lambda^\wedge(n) F^\wedge(n,m) H_{n,m}^s(\gamma; x). \tag{6.6}$$

(ii) (*Scalar*) *SGG problem* (*Corresponding to the Second-order Radial Derivative*). Let the values $G(x)$, $x \in X$, for some subset $X \subset \overline{\Omega_\gamma^{\rm ext}}$ be known from a function G of the class $\mathcal{H}_s(\overline{\Omega_\gamma^{\rm ext}})$. We search for a potential $F|_{\overline{\Sigma^{\rm ext}}}$ with F being from $\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$ such that

$$\Lambda F(x) = G(x), \quad x \in X, \tag{6.7}$$

where the SGG operator $\Lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}}) \to \mathcal{H}_s(\overline{\Omega_\gamma^{\rm ext}})$ with the symbol

$$\Lambda^\wedge(n) = \frac{(n+1)(n+2)}{\gamma^2}\left(\frac{\sigma}{\gamma}\right)^n, \quad n = 0, 1, \ldots, \tag{6.8}$$

is given by

$$\begin{aligned} \Lambda F(x) &= \left(-\frac{x}{|x|}\cdot\nabla_x\right)\left(-\frac{x}{|x|}\cdot\nabla_x\right) F(x)|_{|x|=\gamma} \\ &= \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1} \Lambda^\wedge(n) F^\wedge(n,m) H_{n,m}^s(\gamma; x). \end{aligned} \tag{6.9}$$

In the case of combined SST/SGG data we have the following formulation in terms of pseudodifferential operators.

(iii) *Combined* (*scalar*) *SST/SGG problem.* Let the values $G_1(x)$, $x \in X_1 \subset \overline{\Omega_\gamma^{\rm ext}}$ and $G_2(x)$, $x \in X_2 \subset \overline{\Omega_\gamma^{\rm ext}}$ be known from a function of class $\mathcal{H}_s(\overline{\Omega_\gamma^{\rm ext}})$. Let the symbols of the two corresponding pseudodifferential operators Λ_1 and Λ_2 be given by

$$\Lambda_1^\wedge(n) = \left(\frac{\sigma}{\gamma}\right)^n \frac{n+1}{\gamma}, \qquad n = 0, 1, \ldots \quad \text{for SST}, \tag{6.10}$$

$$\Lambda_2^\wedge(n) = \left(\frac{\sigma}{\gamma}\right)^n \frac{(n+1)(n+2)}{\gamma^2}, \quad n = 0, 1, \ldots \quad \text{for SGG}. \tag{6.11}$$

Find a potential $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})|_{\overline{\Sigma^{\rm ext}}}$ such that

$$(\Lambda_1 F)(x) = G_1(x), \quad x \in X_1, \tag{6.12}$$

$$(\Lambda_2 F)(x) = G_2(x), \quad x \in X_2. \tag{6.13}$$

In order to give an answer to the question of subsets $X \subset \overline{\Omega_\gamma^{\text{ext}}}$ on which data are necessary to assure uniqueness of the solution F, we define $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-*fundamental systems.*

Definition 6.2. A system $X = \{x_n\}_{n=0,1,\ldots}$ of points $x_n \in \overline{\Omega_\gamma^{\text{ext}}}$ is called an $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system in $\overline{\Omega_\gamma^{\text{ext}}}$, if the conditions $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ and $F(x_n) = 0$ for $n = 0, 1, \ldots$ imply $F = 0$.

For fundamental systems we get the following uniqueness theorems which are proved in the Ph.D.-thesis [58].

Theorem 6.3. *Let $X = \{x_n\}_{n=0,1,\ldots}$ be an $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system in $\overline{\Omega_\gamma^{\text{ext}}}$. Then the potential $F|_{\overline{\Sigma^{\text{ext}}}}$ solving the (scalar) SST or SGG problem is uniquely defined.*

Theorem 6.4. *Let $X_1 \subset \overline{\Omega_\sigma^{\text{ext}}}$, $X_2 \subset \overline{\Omega_\sigma^{\text{ext}}}$ such that $X = X_1 \cup X_2 = \{x_n\}_{n=0,1,\ldots}$ is an $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system in $\overline{\Omega_\gamma^{\text{ext}}}$. Then the potential $F|_{\overline{\Sigma^{\text{ext}}}}$ solving the combined (scalar) SST/SGG problem is uniquely defined.*

In order to present the results concerning the ill-posedness of the satellite problems, we essentially follow [19]. We reformulate the SST and SGG problem as a convolution equation using kernel functions.

Definition 6.5. Let $\alpha,\ \beta \in \mathbb{R}$, $\alpha \geq \sigma$ and $\beta \geq \sigma$. Then any kernel $K^{\alpha,\beta}(\cdot,\cdot) : \overline{\Omega_\alpha^{\text{ext}}} \times \overline{\Omega_\beta^{\text{ext}}} \to \mathbb{R}$ of the form

$$K^{\alpha,\beta}(x,y) = \sum_{n=0}^{\infty} K^{\wedge}(n) \sum_{m=1}^{2n+1} H_{n,m}^{s}(\alpha;x) H_{n,m}^{s}(\beta;y) \tag{6.14}$$

$(x,y) \in \overline{\Omega_\alpha^{\text{ext}}} \times \overline{\Omega_\beta^{\text{ext}}}$, is called an $\mathcal{H}_{\alpha,\beta}$-kernel.

The sequence $\{(K^{\alpha,\beta})^{\wedge}(n)\}_{n\in\mathbb{N}_0}$ with $(K^{\alpha,\beta})^{\wedge}(n) = \left(\frac{\alpha\beta}{\sigma^2}\right)^n K^{\wedge}(n)$, $n = 0, 1, \ldots$, is called the (α,β)-*symbol of the $\mathcal{H}_{\alpha,\beta}$-kernel* $K^{\alpha,\beta}(\cdot,\cdot)$. The (σ,σ)-symbol of the $\mathcal{H}_{\alpha,\beta}$-kernel $K^{\alpha,\beta}(\cdot,\cdot)$ is simply called the *symbol of the $\mathcal{H}_{\alpha,\beta}$-kernel.*

Definition 6.6. An $\mathcal{H}_{\alpha,\beta}$-kernel $K^{\alpha,\beta}(\cdot,\cdot)$ with symbol $\{K^{\wedge}(n)\}_{n=0,1,\ldots}$ is called admissible, if the following conditions are satisfied:

(i) $\sum_{n=0}^{\infty} (K^{\wedge}(n))^2 < \infty$,

(ii) $\sum_{n=0}^{\infty} (2n+1)\,(K^{\wedge}(n))^2 \left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s} < \infty$.

The first property in Definition 6.6 ensures that $K^{\wedge}(n) \to 0$ as $n \to \infty$, whereas the second condition implies the following lemma.

Lemma 6.7. *Let $\alpha, \beta \in \mathbb{R}$, $\alpha \geq \sigma$, $\beta \geq \sigma$.*

(i) *If $K^{\alpha,\beta}(\cdot,\cdot)$ is an admissible $\mathcal{H}_{\alpha,\beta}$-kernel with the symbol $\{K^{\wedge}(n)\}_{n=0,1,\ldots}$, then $K^{\alpha,\beta}(x,\cdot)$ is an element of $\mathcal{H}_s(\overline{\Omega_\beta^{\text{ext}}})$ for every (fixed) $x \in \overline{\Omega_\alpha^{\text{ext}}}$.*

(ii) *If* $K^{\alpha,\beta}(\cdot,\cdot)$ *is an admissible* $\mathcal{H}_{\alpha,\beta}$*-kernel with the symbol* $\{K^\wedge(n)\}_{n=0,1,\ldots}$, *then* $K^{\alpha,\beta}(\cdot,y)$ *is an element of* $\mathcal{H}_s(\overline{\Omega_\alpha^{\rm ext}})$ *for every* (*fixed*) $x \in \overline{\Omega_\beta^{\rm ext}}$.

Suppose now that F, G are elements of class $\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$. Then we understand the $\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$-*convolution* of F and G simply to be the inner product in $\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$, i.e.:

$$F * G = (F, G)_{\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})}. \tag{6.15}$$

(More precisely, we had to write $F * G = F *_{\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})} G$.) By definition, we let

$$F^\wedge(n,k) = F * H_{n,k}^s(\sigma;\cdot) \tag{6.16}$$

for $n \in \mathcal{N}(A_n)$; $k = 1, \ldots, 2n+1$. It follows from (6.15) via the Parseval identity that

$$F * G = \sum_{n\in\mathcal{N}} \sum_{k=1}^{2n+1} F^\wedge(n,k) G^\wedge(n,k),$$

for $F, G \in \mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$.

We now define the convolution of an admissible $\mathcal{H}_{\alpha,\beta}$-kernel against a function $F \in \mathcal{H}_s(\overline{\Omega_\beta^{\rm ext}})$ as follows:

$$\begin{aligned}(K^{\alpha,\beta} * F)(x) &= K^{\alpha,\beta}(x,\cdot) * F \\ &= \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} K^\wedge(n) F^\wedge(n,m) H_{n,m}^s(\alpha; x), \quad x \in \overline{\Omega_\alpha^{\rm ext}}. \end{aligned} \tag{6.17}$$

It directly follows that $(K^{\alpha,\beta} * F)^\wedge(n,m) = K^\wedge(n) F^\wedge(n,m)$ and $K^{\alpha,\beta} * F \in \mathcal{H}_s(\overline{\Omega_\alpha^{\rm ext}})$. In analogous way we define the convolution of an $\mathcal{H}_{\alpha,\beta}$-kernel $K^{\alpha,\beta}(\cdot,\cdot)$ against a function $F \in \mathcal{H}_s(\overline{\Omega_\alpha^{\rm ext}})$ by

$$\begin{aligned}(K^{\alpha,\beta} * F)(y) &= K^{\alpha,\beta}(\cdot,y) * F \\ &= \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} K^\wedge(n) F^\wedge(n,m) H_{n,m}^s(\beta; y), \quad y \in \overline{\Omega_\beta^{\rm ext}}, \end{aligned} \tag{6.18}$$

and $K^{\alpha,\beta} * F$ is an element of $\mathcal{H}_s(\overline{\Omega_\beta^{\rm ext}})$.

If L, K are admissible $\mathcal{H}_{\sigma,\sigma}$-kernels, then the $\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$-*convolution* $L * K$ is defined by

$$(L * K)(x,y) = (L(x,\cdot), K(\cdot,y))_{\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})}, \quad (x,y) \in \overline{\Omega_\sigma^{\rm ext}} \times \overline{\Omega_\sigma^{\rm ext}}.$$

Obviously, $(L * K)(\cdot,\cdot)$ is an admissible $\mathcal{H}_{\sigma,\sigma}$-kernel, and it is not difficult to see that

$$(L * K)^\wedge(n) = L^\wedge(n) K^\wedge(n), \quad n \in \mathcal{N}\left((K^\wedge(n) L^\wedge(n))^{-1} A_n\right).$$

We usually write $K^{(2)}(\cdot,\cdot) = (K * K)(\cdot,\cdot)$ to indicate the convolution of an $\mathcal{H}_{\sigma,\sigma}$-kernel with itself. $K^{(2)}(\cdot,\cdot) = (K * K)(\cdot,\cdot)$ is said to be the *iterated kernel* of $K(\cdot,\cdot)$. More generally, $K^{(p)}(\cdot,\cdot) = (K^{(p-1)} * K)(\cdot,\cdot)$ for $p = 2, 3, \ldots$, and

$K^{(1)}(\cdot,\cdot) = K(\cdot,\cdot)$ for $p = 1$. Obviously, we have

$$(K^{(2)})^{\wedge}(n) = (K^{\wedge}(n))^2.$$

In order to give an answer to the question of ill-posedness of the (scalar) SST or SGG problem, the continuity of the inverse additionally has to be investigated. The answer to this question requires the reformulation of the problem as convolution equation. Starting from a pseudodifferential operator $\Lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$ given by

$$\Lambda F = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \Lambda^{\wedge}(n) F^{\wedge}(n,m) H_{n,m}^s(\gamma;\cdot), \tag{6.19}$$

we can interpret the symbol of the pseudodifferential operator as the symbol of an $\mathcal{H}_{\sigma,\gamma}$-kernel $(K^{\Lambda})^{\sigma,\gamma}$ presuming that the symbol satisfies the admissibility conditions. The pseudodifferential operator is then given by the convolution identity

$$\Lambda F(x) = (K^{\Lambda})^{\sigma,\gamma}(\cdot,x) * F, \quad x \in \overline{\Omega_\gamma^{\text{ext}}}, \tag{6.20}$$

for $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, where $(K^{\Lambda})^{\wedge}(n) = \Lambda^{\wedge}(n)$, $n = 0, 1, \ldots$. Obviously, we have

$$\begin{aligned}(K^{\Lambda})^{\sigma,\gamma}(\cdot,x) * H_{n,m}^s(\sigma;\cdot) &= (K^{\Lambda})^{\gamma,\sigma}(x,\cdot) * H_{n,m}^s(\sigma;\cdot) \\ &= \Lambda^{\wedge}(n) H_{n,m}^s(\gamma;x),\end{aligned} \tag{6.21}$$

for all $n \in \mathbb{N}$; $m = 1, \ldots, 2n+1$, or, equivalently,

$$\Lambda H_{n,m}^s(\sigma;\cdot) = \Lambda^{\wedge}(n) H_{n,m}^s(\gamma;\cdot). \tag{6.22}$$

Having a look at the (scalar) SST and SGG operator, we get the following result.

Theorem 6.8. *The $\mathcal{H}_{\sigma,\gamma}$-kernel $(K^{\Lambda})^{\sigma,\gamma}$ defined by the symbol*

$$\Lambda^{\wedge}(n) = \begin{cases} \left(\frac{\sigma}{\gamma}\right)^n \frac{n+1}{\gamma}, & n = 0, 1, \ldots \quad \textit{for SST}, \\ \left(\frac{\sigma}{\gamma}\right)^n \frac{(n+1)(n+2)}{\gamma^2}, & n = 0, 1, \ldots \quad \textit{for SGG}, \end{cases} \tag{6.23}$$

is admissible, if $\{\left(\frac{n+\frac{1}{2}}{\sigma}\right)^s\}$ is summable in the sense of Eq. (3.3).

Theorem 6.9. *Let $\Lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$ be a pseudodifferential operator with $(K^{\Lambda})^{\sigma,\gamma}$ satisfying the admissibility conditions. Then the pseudodifferential operator Λ is bounded and $\|\Lambda\| = \max_{n\in\mathbb{N}_0} |\Lambda^{\wedge}(n)|$. Further on, Λ is an injective operator.*

From functional analysis (see, e.g., [70, 77]), we know that the SST and SGG operators are compact as being so-called Hilbert–Schmidt operators. Summing up the preceding considerations we finally get the following result.

Theorem 6.10. *Let*

$$\Lambda F = G, \quad F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}), \quad G \in \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}}), \tag{6.24}$$

be the (scalar) SST or SGG problem. Then Λ is a compact operator with infinite-dimensional range. Furthermore, Λ^{-1} is not bounded on $\mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$. The (scalar) SST or SGG problem is solvable if and only if

$$\sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\left(\frac{G^\wedge(n,m)}{\Lambda^\wedge(n)}\right)^2<\infty. \tag{6.25}$$

Remembering Hadamard's definition of a well-posed problem (existence, uniqueness and continuity of the inverse), we consequently see that the (scalar) SST or SGG problem is ill posed, as it violates the first and third condition.

6.2. Vectorial SST and SGG problem

Following [58], we additionally formulate uniqueness results for the (vectorial) SST and SGG problems. Let $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ be a (scalar) Sobolev space with $\left(\frac{n+\frac{1}{2}}{\sigma}\right)^s$ satisfying the consistency condition (CC2) relative to $[\sigma,\tau)$ (see Eq. (6.2)). Further on, let $h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}})$, $i=1,2$, be (vectorial) Sobolev spaces. Then the "downward continuation problem" of determining the potential $F\in\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ from "satellite data" $g\in h_s^{(1)}(\overline{\Omega_\gamma^{\text{ext}}})\oplus h_s^{(2)}(\overline{\Omega_\gamma^{\text{ext}}})$ reads as follows.

(i) (*Vectorial*) *SST problem* (*Corresponding to the First-order Tangential Derivative*). Let the values $g(x)$, $x\in X$, for some subset $X\subset\overline{\Omega_\gamma^{\text{ext}}}$ be known from a function g of the class $h_s^{(1)}(\overline{\Omega_\gamma^{\text{ext}}})\oplus h_s^{(2)}(\overline{\Omega_\gamma^{\text{ext}}})$. We search for a potential $F|_{\overline{\Sigma^{\text{ext}}}}$ with F being of the class $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ such that

$$\lambda F(x)=g(x),\quad x\in X, \tag{6.26}$$

where the SST Operator $\lambda:\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})\to h_s^{(1)}(\overline{\Omega_\gamma^{\text{ext}}})\oplus h_s^{(2)}(\overline{\Omega_\gamma^{\text{ext}}})$ is given by

$$(\lambda F)(x)=\nabla_\xi^{*,\sigma}F(x)|_{|x|=\gamma}, \tag{6.27}$$

with $x=|x|\xi$. Observing the symbol

$$\lambda^{(i)\wedge}(n)=\begin{cases}\left(\frac{\sigma}{\gamma}\right)^n\frac{n}{\gamma}\sqrt{\frac{n+1}{2n+1}}, & i=1;\quad n=0,1,\dots,\\ \left(\frac{\sigma}{\gamma}\right)^n\frac{n+1}{\gamma}\sqrt{\frac{n}{2n+1}}, & i=2;\quad n=1,2,\dots,\end{cases} \tag{6.28}$$

the (vectorial) SST operator can be written as

$$\lambda F(x)=\sum_{i=1}^{2}\sum_{n=1}^{\infty}\sum_{m=1}^{2n+1}\lambda^{(i)\wedge}(n)F^\wedge(n,m)h_{n,m}^{(i)s}(\gamma;x). \tag{6.29}$$

In the case of SGG-data the mixed derivatives can be handled within vectorial framework.

(ii) (*Vectorial*) *SGG problem* (*Corresponding to the Second-order Mixed Derivatives*). Let the values $g(x)$, $x\in X$, for some subset $X\subset\overline{\Omega_\gamma^{\text{ext}}}$ be known from

a function g of the class $h_s^{(1)}(\overline{\Omega_\gamma^{\rm ext}}) \oplus h_s^{(2)}(\overline{\Omega_\gamma^{\rm ext}})$. We search for a potential $F|_{\overline{\Sigma^{\rm ext}}}$ with F being of the class $\mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$ such that

$$\lambda F(x) = g(x), \quad x \in X, \tag{6.30}$$

where the SGG operator $\lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}}) \to h_s^{(1)}(\overline{\Omega_\gamma^{\rm ext}}) \oplus h_s^{(2)}(\overline{\Omega_\gamma^{\rm ext}})$ with symbol

$$\lambda^{(i)\wedge}(n) = \begin{cases} \left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+1)}{\gamma^2}\sqrt{\frac{n+1}{2n+1}}, & i=1; \quad n=0,1,\ldots, \\ \left(\frac{\sigma}{\gamma}\right)^n \frac{(n+1)^2}{\gamma^2}\sqrt{\frac{n}{2n+1}}, & i=2; \quad n=1,2,\ldots, \end{cases} \tag{6.31}$$

is given by

$$\sum_{i=1}^{2}\sum_{n=1}^{\infty}\sum_{m=1}^{2n+1} \lambda^{(i)\wedge}(n) F^\wedge(n,m) h_{n,m}^{(i)s}(\gamma;x). \tag{6.32}$$

In order to give an answer to the question of subsets $X \subset \overline{\Omega_\gamma^{\rm ext}}$ on which data are necessary to get uniqueness of the solution F, we define $h_s^{(i)}(\overline{\Omega_\sigma^{\rm ext}})$-*fundamental systems.*

Definition 6.11. A system $X = \{x_n\}_{n=0,1,\ldots}$ of points $x_n \in \overline{\Omega_\sigma^{\rm ext}}$ is called an $h_s^{(i)}(\overline{\Omega_\sigma^{\rm ext}})$-fundamental system in $\overline{\Omega_\sigma^{\rm ext}}$, if the conditions $g \in h_s^{(i)}(\overline{\Omega_\sigma^{\rm ext}})$ and $g(x_n) = 0$ for $n \in \mathbb{N}_0$ imply $g = 0$, $i \in \{1,2,3\}$. Further on, X is called an $h_s^{(i)}(\overline{\Omega_\sigma^{\rm ext}}) \oplus h_s^{(j)}(\overline{\Omega_\sigma^{\rm ext}})$-fundamental system, if $g \in h_s^{(i)}(\overline{\Omega_\sigma^{\rm ext}}) \oplus h_s^{(j)}(\overline{\Omega_\sigma^{\rm ext}})$ and $g(x_n) = 0$ for $n \in \mathbb{N}_0$ imply $g = 0$ for $i, j \in \{1,2,3\}$ with $i \neq j$.

We now obtain the following uniqueness theorem.

Theorem 6.12. *Let $X = \{x_n\}_{n=0,1,\ldots}$ be an $h_s^{(1)}(\overline{\Omega_\gamma^{\rm ext}}) \oplus h_s^{(2)}(\overline{\Omega_\gamma^{\rm ext}})$-fundamental system in $\overline{\Omega_\gamma^{\rm ext}}$. Then the potential $F|_{\overline{\Sigma^{\rm ext}}}$ solving the (vectorial) SST or SGG problem is uniquely defined up to an additive constant C.*

Definition 6.13. Let $\alpha, \beta \in \mathbb{R}$, $\alpha \geq \sigma$ and $\beta \geq \sigma$. Then any kernel $k^{(i),\alpha,\beta}(\cdot,\cdot) : \overline{\Omega_\alpha^{\rm ext}} \times \overline{\Omega_\beta^{\rm ext}} \to \mathbb{R}^3$ of the form

$$k^{(i),\alpha,\beta}(x,y) = \sum_{n=0_i}^{\infty} k^{(i)\wedge}(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\alpha;x) h_{n,m}^{(i)s}(\beta;y), \tag{6.33}$$

$(x,y) \in \overline{\Omega_\alpha^{\rm ext}} \times \overline{\Omega_\beta^{\rm ext}}$, is called an $h_{\alpha,\beta}^{(i)}$-kernel. Furthermore,

$$k^{\alpha,\beta}(x,y) = \sum_{i=1}^{3} k^{(i),\alpha,\beta}(x,y), \tag{6.34}$$

$(x,y) \in \overline{\Omega_\alpha^{\rm ext}} \times \overline{\Omega_\beta^{\rm ext}}$, is called an $h_{\alpha,\beta}$-kernel.

The sequence $\{\left(k^{(i),\alpha,\beta}\right)^\wedge(n)\}_{n\in\mathbb{N}_{0_i}}$ with

$$\left(k^{(i),\alpha,\beta}\right)^\wedge(n) = \left(\frac{\alpha\beta}{\sigma^2}\right)^n k^{(i)\wedge}(n), \quad n = 0_i, \ldots, \tag{6.35}$$

is called the (α,β)-*symbol of the* $h^{(i)}_{\alpha,\beta}$-*kernel* $k^{(i),\alpha,\beta}(\cdot,\cdot)$. The (σ,σ)-symbol of the $h^{(i)}_{\alpha,\beta}$-kernel $k^{(i),\alpha,\beta}(\cdot,\cdot)$ is simply called the *symbol of the* $h^{(i)}_{\alpha,\beta}$-*kernel*.

Definition 6.14. An $h^{(i)}_{\alpha,\beta}$-kernel $k^{(i),\alpha,\beta}(\cdot,\cdot)$ with symbol $\{k^{(i)\wedge}(n)\}_{n=0_i,\dots}$ is called admissible, if the following conditions are satisfied:

(i) $\sum_{n=0_i}^{\infty}(k^{(i)\wedge}(n))^2<\infty$,

(ii) $\sum_{n=0_i}^{\infty}(2n+1)\left(k^{(i)\wedge}(n)\right)^2\left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s}<\infty$,

(iii) (a) $\sum_{n=0}^{\infty}(2n+1)(2n+3)\left(k^{(1)\wedge}(n)\right)^2\left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s}<\infty$,

(b) $\sum_{n=1}^{\infty}(2n+1)(2n-1)\left(k^{(2)\wedge}(n)\right)^2\left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s}<\infty$,

(c) $\sum_{n=1}^{\infty}(2n+1)(2n+1)\left(k^{(3)\wedge}(n)\right)^2\left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s}<\infty$.

Furthermore, the $h_{\alpha,\beta}$-kernel is called admissible, if the $h^{(i)}_{\alpha,\beta}$-kernels, $i\in\{1,2,3\}$, are admissible.

The second and the third condition imply the following lemma.

Lemma 6.15. *Let* $\alpha,\beta\in\mathbb{R}$, $\alpha\geq\sigma$, $\beta\geq\sigma$.

(i) *If* $k^{(i),\alpha,\beta}(\cdot,\cdot)$ *is an admissible* $h^{(i)}_{\alpha,\beta}$*-kernel with the symbol* $\{k^{(i)\wedge}(n)\}_{n=0_i,\dots}$, *then* $k^{(i),\alpha,\beta}(x,\cdot)$ *is an element of* $h^{(i)}_s(\overline{\Omega^{\rm ext}_\beta})$ *for every (fixed)* $x\in\overline{\Omega^{\rm ext}_\alpha}$.

(ii) *If* $k^{(i),\alpha,\beta}(\cdot,\cdot)$ *is an admissible* $h^{(i)}_{\alpha,\beta}$*-kernel with the symbol* $\{k^{(i)\wedge}(n)\}_{n=0_i,\dots}$, *then the component functions* $k^{(i),\alpha,\beta}(\cdot,y)^T\varepsilon^l$ *are elements of* $\mathcal{H}_s(\overline{\Omega^{\rm ext}_\alpha})$ *for every (fixed)* $x\in\overline{\Omega^{\rm ext}_\beta}$, $l\in\{1,2,3\}$.

Our next step is the definition of the convolution of an admissible $h^{(i)}_{\alpha,\beta}$-kernel against a function $f\in h_s(\overline{\Omega^{\rm ext}_\beta})$ as follows:

$$\begin{aligned}(k^{(i),\alpha,\beta}*f)(x)&=k^{(i),\alpha,\beta}(x,\cdot)*f\\&=\sum_{n=0_i}^{\infty}\sum_{m=1}^{2n+1}k^{(i)\wedge}(n)f^{(i)\wedge}(n,m)H^s_{n,m}(\alpha;x),\quad x\in\overline{\Omega^{\rm ext}_\alpha}.\end{aligned}\tag{6.36}$$

It directly follows that $(k^{(i),\alpha,\beta}*f)^\wedge(n,m)=k^{(i)\wedge}(n)f^{(i)\wedge}(n,m)$, $n=0_i$, $i\in\{1,2,3\}$, and $k^{(i),\alpha,\beta}*f\in\mathcal{H}_s(\overline{\Omega^{\rm ext}_\alpha})$. In an analogous way we define the convolution of an $h_{\alpha,\beta}$-kernel $k^{\alpha,\beta}(\cdot,\cdot)$ against a function $F\in\mathcal{H}_s(\overline{\Omega^{\rm ext}_\alpha})$ by

$$\begin{aligned}(k^{\alpha,\beta}\star F)(y)&=k^{\alpha,\beta}(\cdot,y)\star F\\&=\sum_{i=1}^{3}\sum_{n=0_i}^{\infty}\sum_{m=1}^{2n+1}k^{(i)\wedge}(n)F^\wedge(n,m)h^{(i)s}_{n,m}(\beta;y),\quad y\in\overline{\Omega^{\rm ext}_\beta},\end{aligned}\tag{6.37}$$

and $k^{\alpha,\beta}\star F$ is an element of $h_s(\overline{\Omega^{\rm ext}_\beta})$.

Our next purpose is to present the formulation of the vectorial SST respectively SGG operators with the help of convolutions. This enables us to give an

answer to the question of continuity of the inverse. We start from a pseudodifferential operator $\lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to h_s(\overline{\Omega_\gamma^{\text{ext}}})$ given by

$$\lambda F = \sum_{i=1}^{3} \sum_{n=0_i}^{\infty} \sum_{m=1}^{2n+1} \lambda^{(i)\wedge}(n) F^\wedge(n,m) h_{n,m}^{(i)s}(\gamma;\cdot), \tag{6.38}$$

and interpret the symbol of the pseudodifferential operator as the symbol of an $h_{\sigma,\gamma}$-kernel $(k^\lambda)^{\sigma,\gamma}$ presuming that the symbol satisfies the admissibility conditions. The pseudodifferential operator is then given by the convolution identity

$$\lambda F(x) = (k^\lambda)^{\sigma,\gamma}(\cdot, x) \star F, \quad x \in \overline{\Omega_\gamma^{\text{ext}}}, \tag{6.39}$$

for $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, where $(k^\lambda)^{(i)\wedge}(n) = \lambda^{(i)\wedge}(n)$, $i = 1,2,3$; $n = 0_i, \ldots$. Obviously, we have

$$(k^{\lambda^{(i)}})^{\sigma,\gamma}(\cdot, x) \star H_{n,m}^s(\sigma;\cdot) = \lambda^{(i)\wedge}(n) h_{n,m}^{(i)s}(\gamma; x), \tag{6.40}$$

for all $i = 1,2,3$; $n = 0_i, \ldots$; $m = 1, \ldots, 2n+1$, or, equivalently

$$\lambda^{(i)} H_{n,m}^s(\sigma;\cdot) = \lambda^{(i)\wedge}(n) h_{n,m}^{(i)s}(\gamma;\cdot). \tag{6.41}$$

Having a look at the (vectorial) SST and SGG operator, we get the following result.

Theorem 6.16. *The $h_{\sigma,\gamma}$-kernel $(k^\lambda)^{\sigma,\gamma}$ defined by the symbol*

$$(k^\lambda)^{(1)\wedge}(n) = \lambda^{(1)\wedge}(n) = \begin{cases} \left(\frac{\sigma}{\gamma}\right)^n \frac{n}{\gamma}\sqrt{\frac{n+1}{2n+1}}, & n = 0,1,\ldots \quad \textit{for SST}, \\ \left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+1)}{\gamma^2}\sqrt{\frac{n+1}{2n+1}}, & n = 0,1,\ldots \quad \textit{for SGG}, \end{cases} \tag{6.42}$$

and

$$(k^\lambda)^{(2)\wedge}(n) = \lambda^{(2)\wedge}(n) = \begin{cases} \left(\frac{\sigma}{\gamma}\right)^n \frac{n+1}{\gamma}\sqrt{\frac{n}{2n+1}}, & n = 1,2,\ldots \quad \textit{for SST}, \\ \left(\frac{\sigma}{\gamma}\right)^n \frac{(n+1)^2}{\gamma^2}\sqrt{\frac{n}{2n+1}}, & n = 1,2,\ldots \quad \textit{for SGG}, \end{cases} \tag{6.43}$$

is admissible, if $\{\left(\frac{n+\frac{1}{2}}{\sigma}\right)^s\}$ is summable and satisfies, in addition, condition (iii) *in Definition* 6.14.

Theorem 6.17. *Let $\lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to h_s(\overline{\Omega_\gamma^{\text{ext}}})$ be a pseudodifferential operator with $(k^\lambda)^{\sigma,\gamma}$ satisfying the admissibility conditions, and $\lambda^{(i)\wedge}(n) \neq 0$, $i \in \{1,2,3\}$, $n = 0_i, \ldots$. Then the pseudodifferential operator λ is bounded and*

$$\|\lambda\| = \max_{n \in \mathbb{N}_0} \left| \sum_{i=1}^{3} \lambda^{(i)\wedge}(n) \right|, \tag{6.44}$$

where we let $\lambda^{(2)\wedge}(0) = \lambda^{(3)\wedge}(0) = 0$. Further on, λ is an injective operator.

Finally, we get the following result.

Theorem 6.18. *Let*

$$\lambda F = g, \quad F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}),\ g \in h_s^{(1)}(\overline{\Omega_\gamma^{\text{ext}}}) \oplus h_s^{(2)}(\overline{\Omega_\gamma^{\text{ext}}}) \tag{6.45}$$

be the (vectorial) SST or SGG problem. Then λ is a compact operator with infinite-dimensional range. Furthermore, λ^{-1} is not bounded on $h_s^{(1)}(\overline{\Omega_\gamma^{\text{ext}}}) \oplus h_s^{(2)}(\overline{\Omega_\gamma^{\text{ext}}})$. The SST/SGG problem is solvable if and only if

$$\sum_{i=1}^{2} \sum_{n=0_i}^{\infty} \sum_{m=1}^{2n+1} \left(\frac{g^{(i)\wedge}(n,m)}{\lambda^{(i)\wedge}(n)} \right)^2 < \infty. \tag{6.46}$$

We consequently get that the (vectorial) SST/SGG problem is ill posed because existence and continuity of the inverse are violated.

6.3. Tensorial SGG problem

The formulation of the definitions and theorems for the tensorial case is straightforward. Let $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ be a (scalar) Sobolev space satisfying the consistency condition (CC2) relative to $[\sigma, \tau)$ (see Eq. (6.2)). Further on, let $\mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}})$, $(i,k) \in \{(1,1),(1,2),(2,1),(2,2)\}$, be (tensorial) Sobolev spaces. Then the "downward continuation problem" of determining the potential $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ from "satellite data" $\mathbf{g} \in \mathbf{h}_s^{SGG}(\overline{\Omega_\gamma^{\text{ext}}})$, where we use the abbreviation

$$\mathbf{h}_s^{SGG}(\overline{\Omega_\gamma^{\text{ext}}}) = \mathbf{h}^{(1,1)}(\overline{\Omega_\gamma^{\text{ext}}}) \oplus \mathbf{h}_s^{(1,2)}(\overline{\Omega_\gamma^{\text{ext}}}) \oplus \mathbf{h}_s^{(2,1)}(\overline{\Omega_\gamma^{\text{ext}}}) \oplus \mathbf{h}_s^{(2,2)}(\overline{\Omega_\gamma^{\text{ext}}}), \tag{6.47}$$

reads as follows.

(i) (*Tensorial*) *SGG problem* (*Corresponding to the Second-order Tangential Derivative*). Let the values $\mathbf{g}(x)$, $x \in X$, for some subset $X \subset \overline{\Omega_\gamma^{\text{ext}}}$ be known from a function $\mathbf{g}$ of the class $\mathbf{h}_s^{SGG}(\overline{\Omega_\gamma^{\text{ext}}})$. We search for a potential $F|_{\overline{\Sigma^{\text{ext}}}}$ with F being from $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ such that

$$\boldsymbol{\lambda} F(x) = \mathbf{g}(x), \quad x \in X, \tag{6.48}$$

where the SGG operator $\boldsymbol{\lambda} : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathbf{h}_s^{SGG}(\overline{\Omega_\gamma^{\text{ext}}})$ is given by

$$(\boldsymbol{\lambda} F)(x) = \nabla^{*,\sigma} \otimes \nabla^{*,\sigma} F(x)|_{|x|=\gamma}, \tag{6.49}$$

with $x = |x|\xi$. With the help of the symbol

$$\boldsymbol{\lambda}^{(i,k)\wedge}(n) = \begin{cases} \left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+1)}{\gamma^2(2n+1)(2n+3)} \sqrt{\nu_n^{(1,1)}}, & (i,k) = (1,1), \quad n = 0,1,\ldots, \\ -\left(\frac{\sigma}{\gamma}\right)^n \frac{(n+1)(n-1)}{\gamma^2((2n-1)(2n+1)} \sqrt{\nu_n^{(1,2)}}, & (i,k) = (1,2), \quad n = 1,2,\ldots, \\ -\left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+2)}{\gamma^2(2n+3)(2n+1)} \sqrt{\nu_n^{(2,1)}}, & (i,k) = (2,1), \quad n = 0,1,\ldots, \\ \left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+1)(n+2)}{\gamma^2(2n-1)(2n+1)} \sqrt{\nu_n^{(2,2)}}, & (i,k) = (2,2), \quad n = 2,3,\ldots, \end{cases} \tag{6.50}$$

with

$$\nu_n^{(1,1)} = (n+1)(n+2)(2n+1)(2n+3), \tag{6.51}$$

$$\nu_n^{(1,2)} = 3n^4, \tag{6.52}$$

$$\nu_n^{(2,1)} = (n+1)^2(2n+1)(2n+3), \tag{6.53}$$

$$\nu_n^{(2,2)} = n(n-1)(2n-1)(2n+1), \tag{6.54}$$

the SGG operator can be written as

$$\boldsymbol{\lambda} F(x) = \sum_{(i,k)\in\mathcal{I}^{SGG}} \sum_{n=0_{ik}}^{\infty} \sum_{m=1}^{2n+1} \boldsymbol{\lambda}^{(i,k)\wedge}(n) F^{\wedge}(n,m) \mathbf{h}_{n,m}^{(i,k)s}(\gamma; x), \tag{6.55}$$

where $\mathcal{I}^{SGG} = \{(1,1),(1,2),(2,1),(2,2)\}$ is the index set for the tensorial SGG problem.

In order to give an answer to the question of subsets $X \subset \overline{\Omega_\gamma^{\text{ext}}}$ on which data are necessary to get uniqueness of the solution F, we define $\mathbf{h}_s^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$-*fundamental systems.*

Definition 6.19. A system $X = \{x_n\}_{n=0,1,\ldots}$ of points $x_n \in \overline{\Omega_\sigma^{\text{ext}}}$ is called an $\mathbf{h}_s^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system in $\overline{\Omega_\sigma^{\text{ext}}}$, if the conditions $\mathbf{g} \in \mathbf{h}_s^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$ and $\mathbf{g}(x_n) = 0$ for $n \in \mathbb{N}_0$ imply $\mathbf{g} = 0$, $i,k \in \{1,2,3\}$. In analogy the fundamental systems are defined for spaces which are direct sums of the spaces $\mathbf{h}_s^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$, $i,k \in \{1,2,3\}$.

As in the scalar and vectorial case we have the following theorem.

Theorem 6.20. *Let $X = \{x_n\}_{n=0,1,\ldots}$ be an $\mathbf{h}_s^{SGG}(\overline{\Omega_\gamma^{\text{ext}}})$-fundamental system in $\overline{\Omega_\gamma^{\text{ext}}}$. Then the potential $F|_{\overline{\Sigma^{\text{ext}}}}$ solving the (tensorial) SGG problem is uniquely defined up to a term of the form*

$$V(x) = \sum_{n=0}^{1} \sum_{m=1}^{2n+1} c_{nm} \left(\frac{\sigma}{|x|}\right)^{n+1} \frac{1}{\sigma} Y_{n,m}\left(\frac{x}{|x|}\right), \quad x \in \overline{\Omega_\sigma^{\text{ext}}}, \tag{6.56}$$

for constants $c_{01}, c_{11}, c_{12}, c_{13} \in \mathbb{R}$.

We finally shortly present the results using the reformulation as convolution equation.

Definition 6.21. Let $\alpha, \beta \in \mathbb{R}$, $\alpha \geq \sigma$ and $\beta \geq \sigma$. Then any kernel $\mathbf{k}^{(i,k),\alpha,\beta}(\cdot,\cdot) : \overline{\Omega_\alpha^{\text{ext}}} \times \overline{\Omega_\beta^{\text{ext}}} \to \mathbb{R}^3 \otimes \mathbb{R}^3$ of the form

$$\mathbf{k}^{(i,k),\alpha,\beta}(x,y) = \sum_{n=0_{ik}}^{\infty} \mathbf{k}^{(i,k)\wedge}(n) \sum_{m=1}^{2n+1} H_{n,m}^{s}(\alpha; x) \mathbf{h}_{n,m}^{(i,k)s}(\beta; y), \tag{6.57}$$

$(x,y) \in \overline{\Omega_\alpha^{\text{ext}}} \times \overline{\Omega_\beta^{\text{ext}}}$, is called an $\mathbf{h}_{\alpha,\beta}^{(i,k)}$-kernel. Furthermore,

$$\mathbf{k}^{\alpha,\beta}(x,y) = \sum_{i,k=1}^{3} \mathbf{k}^{(i,k),\alpha,\beta}(x,y), \tag{6.58}$$

$(x,y) \in \overline{\Omega_\alpha^{\text{ext}}} \times \overline{\Omega_\beta^{\text{ext}}}$, is called an $\mathbf{h}_{\alpha,\beta}$-kernel.

The sequence $\{\left(\mathbf{k}^{(i,k),\alpha,\beta}\right)^\wedge(n)\}_{n\in\mathbb{N}_{0_{ik}}}$ with

$$\left(\mathbf{k}^{(i,k),\alpha,\beta}\right)^\wedge(n) = \left(\frac{\alpha\beta}{\sigma^2}\right)^n \mathbf{k}^{(i,k)\wedge}(n), \quad n = 0_{ik}, \ldots, \tag{6.59}$$

is called the (α,β)-*symbol of the* $\mathbf{h}_{\alpha,\beta}^{(i,k)}$-*kernel* $\mathbf{k}^{(i,k),\alpha,\beta}(\cdot,\cdot)$. The (σ,σ)-symbol of the $\mathbf{h}_{\alpha,\beta}^{(i,k)}$-kernel $\mathbf{k}^{(i,k),\alpha,\beta}(\cdot,\cdot)$ is simply called the *symbol of the* $\mathbf{h}_{\alpha,\beta}^{(i,k)}$-*kernel*.

Definition 6.22. An $\mathbf{h}_{\alpha,\beta}^{(i,k)}$-kernel $\mathbf{k}^{(i,k),\alpha,\beta}(\cdot,\cdot)$ with the symbol $\{\mathbf{k}^{(i,k)\wedge}(n)\}_{n=0_{ik},\ldots}$ is called admissible, if the following conditions are satisfied:

(i) $\sum_{n=0_{ik}}^{\infty}(\mathbf{k}^{(i,k)\wedge}(n))^2 < \infty$,

(ii) $\sum_{n=0_{ik}}^{\infty}(2n+1)(\mathbf{k}^{(i,k)\wedge}(n))^2\left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s} < \infty$,

(iii) (a) $\sum_{n=0}^{\infty}(2n+1)(2n+5)(\mathbf{k}^{(1,1)\wedge}(n))^2\left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s} < \infty$,

(b) $\sum_{n=0_{ik}}^{\infty}(2n+1)(2n+3)(\mathbf{k}^{(i,k)\wedge}(n))^2\left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s} < \infty$, $(i,k) \in \{(1,3),(3,1)\}$,

(c) $\sum_{n=0_{ik}}^{\infty}(2n+1)(2n+1)(\mathbf{k}^{(i,k)\wedge}(n))^2\left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s} < \infty$, $(i,k) \in \{(1,2),(2,1),(3,3)\}$,

(d) $\sum_{n=0_{ik}}^{\infty}(2n+1)(2n-1)(\mathbf{k}^{(i,k)\wedge}(n))^2\left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s} < \infty$, $(i,k) \in \{(2,3),(3,2)\}$,

(e) $\sum_{n=2}^{\infty}(2n+1)(2n-3)(\mathbf{k}^{(2,2)\wedge}(n))^2\left(\frac{\sigma}{n+\frac{1}{2}}\right)^{2s} < \infty$.

Furthermore, the $\mathbf{h}_{\alpha,\beta}$-kernel is called admissible, if all $\mathbf{h}_{\alpha,\beta}^{(i,k)}$-kernels, $i,k \in \{1,2,3\}$, are admissible.

The second and the third condition imply the following lemma.

Lemma 6.23. *Let* $\alpha,\beta \in \mathbb{R}$, $\alpha \geq \sigma$, $\beta \geq \sigma$.

1. *If the kernel* $\mathbf{k}^{(i,k),\alpha,\beta}(\cdot,\cdot)$ *is an admissible* $\mathbf{h}_{\alpha,\beta}^{(i,k)}$*-kernel with the symbol given by* $\{\mathbf{k}^{(i,k)\wedge}(n)\}_{n=0_{ik},\ldots}$, *then* $\mathbf{k}^{(i,k),\alpha,\beta}(x,\cdot)$ *is an element of* $\mathbf{h}_s^{(i,k)}(\overline{\Omega_\beta^{\text{ext}}})$ *for every (fixed)* $x \in \overline{\Omega_\alpha^{\text{ext}}}$.
2. *If the kernel* $\mathbf{k}^{(i,k),\alpha,\beta}(\cdot,\cdot)$ *is an admissible* $\mathbf{h}_{\alpha,\beta}^{(i,k)}$*-kernel with the symbol* $\{\mathbf{k}^{(i,k)\wedge}(n)\}_{n=0_{ik},\ldots}$, *then the component functions* $\mathbf{k}^{(i,k),\alpha,\beta}(\cdot,y)\cdot\varepsilon^j\otimes\varepsilon^l$ *are elements of* $\mathcal{H}_s(\overline{\Omega_\alpha^{\text{ext}}})$ *for every (fixed)* $x \in \overline{\Omega_\beta^{\text{ext}}}$, $j,l \in \{1,2,3\}$.

We now define the convolution of an admissible $\mathbf{h}^{(i,k)}_{\alpha,\beta}$-kernel against a function $\mathbf{f} \in \mathbf{h}_s(\overline{\Omega^{\text{ext}}_\beta})$ as follows:

$$(\mathbf{k}^{(i,k),\alpha,\beta} * \mathbf{f})(x) = \mathbf{k}^{(i,k),\alpha,\beta}(x,\cdot) * \mathbf{f} \tag{6.60}$$
$$= \sum_{n=0_{ik}}^{\infty} \sum_{m=1}^{2n+1} \mathbf{k}^{(i,k)\wedge}(n)\mathbf{f}^{(i,k)\wedge}(n,m) H^s_{n,m}(\alpha;x), \quad x \in \overline{\Omega^{\text{ext}}_\alpha}.$$

It follows directly that $(\mathbf{k}^{(i,k),\alpha,\beta} * \mathbf{f})^\wedge(n,m) = \mathbf{k}^{(i,k)\wedge}(n)\mathbf{f}^{(i,k)\wedge}(n,m)$, $n = 0_{ik}$, $i,k \in \{1,2,3\}$, and $\mathbf{k}^{(i,k),\alpha,\beta} * \mathbf{f} \in \mathcal{H}_s(\overline{\Omega^{\text{ext}}_\alpha})$. In an analogous way we define the convolution of an $\mathbf{h}_{\alpha,\beta}$-kernel $\mathbf{k}^{\alpha,\beta}(\cdot,\cdot)$ against a function $F \in \mathcal{H}_s(\overline{\Omega^{\text{ext}}_\alpha})$ by

$$(\mathbf{k}^{\alpha,\beta} \star F)(y) = \mathbf{k}^{\alpha,\beta}(\cdot,y) \star F \tag{6.61}$$
$$= \sum_{i,k=1}^{3} \sum_{n=0_{ik}}^{\infty} \sum_{m=1}^{2n+1} \mathbf{k}^{(i,k)\wedge}(n) F^\wedge(n,m) \mathbf{h}^{(i,k)s}_{n,m}(\beta;y), \quad y \in \overline{\Omega^{\text{ext}}_\beta},$$

and $\mathbf{k}^{\alpha,\beta} \star F$ is an element of $\mathbf{h}_s(\overline{\Omega^{\text{ext}}_\beta})$. Our next purpose is to present the formulation of the tensorial SGG operator with the help of convolutions. This enables us to give an answer to the question of continuity of the inverse. We start from a pseudodifferential operator $\boldsymbol{\lambda} : \mathcal{H}_s(\overline{\Omega^{\text{ext}}_\sigma}) \to \mathbf{h}_s(\overline{\Omega^{\text{ext}}_\gamma})$ given by

$$\boldsymbol{\lambda} F = \sum_{i,k=1}^{3} \sum_{n=0_{ik}}^{\infty} \sum_{m=1}^{2n+1} \boldsymbol{\lambda}^{(i,k)\wedge}(n) F^\wedge(n,m) \mathbf{h}^{(i,k)s}_{n,m}(\gamma;\cdot), \tag{6.62}$$

and interpret the symbol of the pseudodifferential operator as the symbol of an $\mathbf{h}_{\sigma,\gamma}$-kernel $(\mathbf{k}\,\boldsymbol{\lambda})^{\sigma,\gamma}$ presuming that the symbol satisfies the admissibility conditions. The pseudodifferential operator is then given by the convolution identity

$$\boldsymbol{\lambda} F(x) = (\mathbf{k}\,\boldsymbol{\lambda})^{\sigma,\gamma}(\cdot,x) \star F, \quad x \in \overline{\Omega^{\text{ext}}_\gamma}, \tag{6.63}$$

for $F \in \mathcal{H}_s(\overline{\Omega^{\text{ext}}_\sigma})$, where $(\mathbf{k}\,\boldsymbol{\lambda})^{(i,k)\wedge}(n) = \boldsymbol{\lambda}^{(i,k)\wedge}(n)$, $i,k = 1,2,3; n = 0_{ik}, \ldots$. Obviously, we have

$$(\mathbf{k}\,\boldsymbol{\lambda})^{\sigma,\gamma}(\cdot,x) \star H^s_{n,m}(\sigma;\cdot) = \boldsymbol{\lambda}^{(i,k)\wedge}(n) \mathbf{h}^{(i,k)s}_{n,m}(\gamma;x), \tag{6.64}$$

for all $i,k = 1,2,3$; $n = 0_{ik}, \ldots$; $m = 1, \ldots, 2n+1$, or, equivalently,

$$\boldsymbol{\lambda}^{(i,k)} H^s_{n,m}(\sigma;\cdot) = \boldsymbol{\lambda}^{(i,k)\wedge}(n) \mathbf{h}^{(i,k)s}_{n,m}(\gamma;\cdot). \tag{6.65}$$

Having a look at the (tensorial) SGG operator, we get the following result.

Theorem 6.24. *The* $\mathbf{h}_{\sigma,\gamma}^{SGG}$*-kernel* $(\mathbf{k}\ \boldsymbol{\lambda})^{\sigma,\gamma}$ *defined by the symbol*

$$\boldsymbol{\lambda}^{(i,k)\wedge}(n) = \begin{cases} \left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+1)}{\gamma^2(2n+1)(2n+3)}\sqrt{\nu_n^{(1,1)}}, & (i,k)=(1,1), \\ -\left(\frac{\sigma}{\gamma}\right)^n \frac{(n+1)(n-1)}{\gamma^2((2n-1)(2n+1)}\sqrt{\nu_n^{(1,2)}}, & (i,k)=(1,2), \\ -\left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+2)}{\gamma^2(2n+3)(2n+1)}\sqrt{\nu_n^{(2,1)}}, & (i,k)=(2,1), \\ \left(\frac{\sigma}{\gamma}\right)^n \frac{n(n+1)(n+2)}{\gamma^2(2n-1)(2n+1)}\sqrt{\nu_n^{(2,2)}}, & (i,k)=(2,2), \end{cases} \tag{6.66}$$

is admissible, if $\left(\frac{n+\frac{1}{2}}{\sigma}\right)^s$ *is summable and satisfies, in addition, condition* (iii) *in Definition* 6.22.

We finally get the following results.

Theorem 6.25. *Let* $\boldsymbol{\lambda} : \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathbf{h}(\overline{\Omega_\gamma^{\text{ext}}})$ *be a pseudodifferential operator with* $(\mathbf{k}\ \boldsymbol{\lambda})^{\sigma,\gamma}$ *satisfying the admissibility conditions, and* $\boldsymbol{\lambda}^{(i,k)\wedge}(n) \neq 0$*,* $i \in \{1,2,3\}$*,* $n = 0_{ik}, \ldots$*. Then the pseudodifferential operator* $\boldsymbol{\lambda}$ *is bounded and*

$$\|\boldsymbol{\lambda}\| = \max_{n\in\mathbb{N}_0} \left| \sum_{i,k=1}^{3} \boldsymbol{\lambda}^{(i,k)\wedge}(n) \right|, \tag{6.67}$$

where the sum has to be understood in the same sense as in the vectorial case. Further on, $\boldsymbol{\lambda}$ *is an injective operator.*

Theorem 6.26. *Let*

$$\boldsymbol{\lambda} F = \mathbf{g}, \quad F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}), \quad \mathbf{g} \in \mathbf{h}_s^{SGG}(\overline{\Omega_\gamma^{\text{ext}}}), \tag{6.68}$$

be the (tensorial) SGG problem. Then $\boldsymbol{\lambda}$ *is a compact operator with infinite-dimensional range. Furthermore,* $\boldsymbol{\lambda}^{-1}$ *is not bounded on* $\mathbf{h}_s^{SGG}(\overline{\Omega_\gamma^{\text{ext}}})$*. The SGG problem is solvable if and only if*

$$\sum_{(i,k)\in\mathcal{I}^{SGG}} \sum_{n=0_{ik}}^{\infty} \sum_{m=1}^{2n+1} \left(\frac{\mathbf{g}^{(i,k)\wedge}(n,m)}{\boldsymbol{\lambda}^{(i,k)\wedge}(n)} \right)^2 < \infty. \tag{6.69}$$

We consequently have that the (tensorial) SGG problem is ill posed because existence and continuity of the inverse are violated.

7. Geodetically oriented wavelet approximation

In this section we present a multiscale approach based on wavelet approximation. Note that all modern multiscale approaches have a conception of wavelets as constituting multiscale building blocks in common, which provide a fast and efficient way to decorrelate a given signal data set. As already mentioned in Section 2.4, this characterization contains three basic attributes (basis property, decorrelation and efficient algorithms), which are common features of all classical wavelets and

form the key for a variety of applications, particularly for signal reconstruction and decomposition, thresholding, data compression, denoising, etc.

7.1. Scalar wavelet theory

We start with the presentation of the scalar theory, where we follow the approach given in [19]. First, we define an $\mathcal{H}_{\sigma,\sigma}$-multiresolution analysis. We use the abbreviation $\Phi^{(2)}(\cdot,\cdot) = (\Phi * \Phi)(\cdot,\cdot)$, where Φ is an $\mathcal{H}_{\sigma,\sigma}$-kernel.

Definition 7.1. Let $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be a family of admissible $\mathcal{H}_{\sigma,\sigma}$-kernels as defined in Definition 6.6. Then the family $\{\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})\}_{j\in\mathbb{N}_0}$ of scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ defined by

$$\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) = \{\Phi_j^{(2)} * F \ : \ F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})\}, \tag{7.1}$$

is called an $\mathcal{H}_{\sigma,\sigma}$*-multiresolution analysis*, if the following properties are satisfied:

(i) $\mathcal{V}_0(\overline{\Omega_\sigma^{\text{ext}}}) \subset \cdots \subset \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) \subset \mathcal{V}_{j+1}(\overline{\Omega_\sigma^{\text{ext}}}) \subset \cdots \subset \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$,

(ii) $\overline{\bigcup_{j\in\mathbb{N}_0} \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})}^{\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}} = \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$.

Wavelet analysis is based on the idea of splitting the function into a lowpass part and several bandpass parts. The so-called scaling function corresponds to the lowpass filter, whereas the bandbass filters are the shifted and dilated versions of the wavelet, which are defined as differences between successive scaling functions with the help of a so-called *refinement equation*.

Definition 7.2. A family $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ of sequences $\{\varphi_j(n)\}_{n\in\mathbb{N}_0}$ is called a *generator of a scaling function*, if it satisfies the following requirements:

(i) $(\varphi_j(0))^2 = 1$, for all $j \in \mathbb{N}_0$,

(ii) $(\varphi_j(n))^2 \le (\varphi_{j'}(n))^2$, for all $j, j' \in \mathbb{N}_0$ with $j \le j'$ and all $n \in \mathbb{N}$,

(iii) $\lim_{j\to\infty} (\varphi_j(n))^2 = 1$, for all $n \in \mathbb{N}$.

Based on the definition of a generator of a scaling function, we now introduce $\mathcal{H}_{\sigma,\sigma}$-scaling functions.

Definition 7.3. A family $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of $\mathcal{H}_{\sigma,\sigma}$-kernels $\Phi_j(\cdot,\cdot)$ defined by $\Phi_j^\wedge(n) = \varphi_j(n)$, $n, j \in \mathbb{N}_0$, i.e.,

$$\Phi_j(x,y) = \sum_{n=0}^{\infty} \varphi_j(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\sigma;x) H_{n,m}^s(\sigma;y), \quad x,y \in \overline{\Omega_\sigma^{\text{ext}}}, \tag{7.2}$$

is called an $\mathcal{H}_{\sigma,\sigma}$*-scaling function*, if it satisfies the following properties:

(i) $\Phi_j(\cdot,\cdot)$ is an admissible $\mathcal{H}_{\sigma,\sigma}$-kernel for every $j \in \mathbb{N}_0$ (in the sense of Definition 6.6),

(ii) $\{\Phi_j^\wedge(n)_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ constitutes a generator of a scaling function (in the sense of Definition 7.2).

The following theorem shows the approximation property of an $\mathcal{H}_{\sigma,\sigma}$-scaling function.

Theorem 7.4. *Let $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathcal{H}_{\sigma,\sigma}$-scaling function. Then*

$$\lim_{j\to\infty} \|F - \Phi_j^{(2)} * F\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} = 0 \tag{7.3}$$

holds for all $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$.

We now introduce the dilation and the shifting operator in order to define an $\mathcal{H}_{\sigma,\sigma}$-approximate identity. Let $J, J_1, J_2 \in \mathbb{N}_0$ and $x \in \overline{\Omega_\sigma^{\text{ext}}}$. Then we define the *dilation operator* D_{J_1} and the *shifting operator* S_x by

$$D_{J_1} : \Phi_{J_2}(\cdot,\cdot) \mapsto D_{J_1}\Phi_{J_2}(\cdot,\cdot) = \Phi_{J_1+J_2}(\cdot,\cdot), \tag{7.4}$$

$$S_x : \Phi_J(\cdot,\cdot) \mapsto S_x\Phi_J(\cdot,\cdot) = \Phi_J(x,\cdot). \tag{7.5}$$

The shifting operator S_y acting on the second variable is defined in an analogous way. Note that by definition $\Phi_J(\cdot,\cdot) = D_J\Phi_0(\cdot,\cdot)$ for any $J \in \mathbb{N}_0$.

Definition 7.5. Let $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathcal{H}_{\sigma,\sigma}$-scaling function. Then $\{P_j\}_{j\in\mathbb{N}_0}$ with $P_j : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ defined by

$$\begin{aligned} P_j(F)(x) &= \left(S_x D_j \Phi_0^{(2)}(\cdot,\cdot), F\right)_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} \\ &= \left(\Phi_j^{(2)}(x,\cdot), F\right)_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} \\ &= (\Phi_j^{(2)} * F)(x), \end{aligned} \tag{7.6}$$

for $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $x \in \overline{\Omega_\sigma^{\text{ext}}}$, is called an *$\mathcal{H}_{\sigma,\sigma}$-approximate identity.*

The kernel Φ_0 is called mother kernel of the $\mathcal{H}_{\sigma,\sigma}$-scaling function. Theorem 7.4 leads to

$$\lim_{j\to\infty} \|F - P_j(F)\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} = 0. \tag{7.7}$$

The following theorem clarifies the connection between the concept of multiresolution analysis and the scaling functions.

Theorem 7.6. *Let $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathcal{H}_{\sigma,\sigma}$-scaling function. Then $\{\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})\}_{j\in\mathbb{N}_0}$ forms an $\mathcal{H}_{\sigma,\sigma}$-multiresolution analysis.*

We now turn to the definition of the primal and dual wavelet.

Definition 7.7. Let $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathcal{H}_{\sigma,\sigma}$-scaling function. Then the families of $\mathcal{H}_{\sigma,\sigma}$-kernels $\{\Psi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$, $\{\tilde{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by

$$\Psi_j^\wedge(n) = \psi_j(n), \quad n, j \in \mathbb{N}_0, \tag{7.8}$$

$$\tilde{\Psi}_j^\wedge(n) = \tilde{\psi}_j(n), \quad n, j \in \mathbb{N}_0, \tag{7.9}$$

are called (*primal*) *$\mathcal{H}_{\sigma,\sigma}$-wavelet* and *dual $\mathcal{H}_{\sigma,\sigma}$-wavelet*, respectively, if all $\mathcal{H}_{\sigma,\sigma}$-kernels $\Psi_j(\cdot,\cdot)$, $\tilde{\Psi}_j(\cdot,\cdot)$, $j \in \mathbb{N}_0$, are admissible and the symbols $\{\psi_j(n)\}$, $\{\tilde{\psi}_j(n)\}$, in addition, satisfy the (scalar) refinement equation

$$\tilde{\psi}_j(n)\psi_j(n) = (\varphi_{j+1}(n))^2 - (\varphi_j(n))^2 \tag{7.10}$$

for all $j, n \in \mathbb{N}_0$.

The following equation is a direct consequence of the refinement equation:

$$(\varphi_{J+1}(n))^2 = (\varphi_0(n))^2 + \sum_{j=0}^{J} \tilde{\psi}_j(n)\psi_j(n), \quad J \in \mathbb{N}_0, \tag{7.11}$$

for all $n \in \mathbb{N}_0$. This property finally leads to the reconstruction formula which states how the original function $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ can be derived from a lowpass part and the corresponding bandpass parts (see Theorem 7.9).

We now turn to the definition of the wavelet transform. To this end we define $\mathbb{N}_{-1} = \mathbb{N}_0 \cup \{-1\}$ and let $\psi_{-1}(n) = \tilde{\psi}_{-1}(n) = \varphi_0(n)$, for $n \in \mathbb{N}_0$, $\Psi_{-1}(\cdot,\cdot) = \tilde{\Psi}_{-1}(\cdot,\cdot) = \Phi_0(\cdot,\cdot)$. This abbreviation simplifies our notation. Then we define the space

$$\mathcal{H}_s(\mathbb{N}_{-1} \times \overline{\Omega_\sigma^{\text{ext}}}) = \{H : \mathbb{N}_{-1} \times \overline{\Omega_\sigma^{\text{ext}}} \to \mathbb{R} \ : \ \sum_{j=-1}^{\infty} (H(j;\cdot), H(j;\cdot))_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} < \infty\} \tag{7.12}$$

with inner product

$$(H_1, H_2)_{\mathcal{H}_s(\mathbb{N}_{-1}\times\overline{\Omega_\sigma^{\text{ext}}})} = \sum_{j=-1}^{\infty} (H_1(j;\cdot), H_2(j;\cdot))_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} \tag{7.13}$$

and corresponding norm

$$\|H\|_{\mathcal{H}_s(\mathbb{N}_{-1}\times\overline{\Omega_\sigma^{\text{ext}}})} = \left(\sum_{j=-1}^{\infty} \|H(j;\cdot)\|^2_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} \right)^{1/2}. \tag{7.14}$$

With the help of the dilation operator D_j and the shifting operator S_y we introduce the following abbreviation:

$$\Psi_{j;y}(\cdot) = \Psi_j(\cdot, y) = S_y\Psi_j(\cdot,\cdot) = S_y D_j \Psi_0(\cdot,\cdot), \tag{7.15}$$

$$\tilde{\Psi}_{j;y}(\cdot) = \tilde{\Psi}_j(\cdot, y) = S_y\tilde{\Psi}_j(\cdot,\cdot) = S_y D_j \tilde{\Psi}_0(\cdot,\cdot). \tag{7.16}$$

Definition 7.8. Let $\{\Psi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_{-1}}$ be a (primal) $\mathcal{H}_{\sigma,\sigma}$-wavelet. Then

$$WT : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\mathbb{N}_{-1} \times \overline{\Omega_\sigma^{\text{ext}}}),$$

defined by

$$(WT)(F)(j;y) = (\Psi_{j;y}, F)_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})} = (\Psi_j * F)(y), \tag{7.17}$$

is called $\mathcal{H}_{\sigma,\sigma}$-wavelet transform of F at position $y \in \overline{\Omega_\sigma^{\text{ext}}}$ and scale $j \in \mathbb{N}_{-1}$.

Having the definition of the scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ in mind, we now define the detail spaces $\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$ at scale j by

$$\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}}) = \left\{ \tilde{\Psi}_j * \Psi_j * F \ : \ F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \right\}, \quad j \in \mathbb{N}_0. \tag{7.18}$$

Theorem 7.9 (Scalar Reconstruction Formula for the Outer Space). *Let the families* $\{\Psi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ *and* $\{\tilde{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$, *respectively, be a (primal)* $\mathcal{H}_{\sigma,\sigma}$*-wavelet and its dual corresponding to an* $\mathcal{H}_{\sigma,\sigma}$*-scaling function* $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$. *Then*

$$F=\sum_{j=-1}^{\infty}\tilde{\Psi}_j*\Psi_j*F \tag{7.19}$$

holds for all $F\in\mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$ *(in* $\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})}$*-sense).*

We now solve the (scalar) SST or SGG problem using bandlimited harmonic wavelets. First, we define $\mathcal{H}_{\alpha,\alpha}$-scaling functions with the help of a generator of a scaling function $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$. Since the generator does not depend on σ, we can directly extend the theory to the case of $\mathcal{H}_{\alpha,\alpha}$-scaling functions $\Phi_j^{\alpha,\alpha}$ with $\alpha\geq\sigma$:

$$\Phi_j^{\alpha,\alpha}(x,y)=\sum_{n=0}^{\infty}\varphi_j(n)\sum_{m=1}^{2n+1}H_{n,m}^s(\alpha;x)H_{n,m}^s(\alpha;y), \tag{7.20}$$

where

$$(\Phi_j^{\alpha,\alpha})^\wedge(n)=\varphi_j(n). \tag{7.21}$$

As a consequence, Theorem 7.4 is valid substituting σ by α. Furthermore, the definition of the scale spaces can be directly transferred in the following way:

$$\mathcal{V}_j(\overline{\Omega_\alpha^{\mathrm{ext}}})=\{(\Phi_j^{(2)})^{\alpha,\alpha}*F\ :\ F\in\mathcal{H}_s(\overline{\Omega_\alpha^{\mathrm{ext}}})\}, \tag{7.22}$$

where

$$(\Phi_j^{(2)})^{\alpha,\alpha}=\Phi_j^{\alpha,\alpha}*\Phi_j^{\alpha,\alpha}. \tag{7.23}$$

The system $\{\mathcal{V}_j(\overline{\Omega_\alpha^{\mathrm{ext}}})\}$ of scale spaces forms a multiresolution analysis due to Theorem 7.6. We now investigate the solution of the restriction of an operator $\Lambda:\mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})\to\mathcal{H}_s(\overline{\Omega_\gamma^{\mathrm{ext}}})$ to a scale space $\mathcal{V}_j$:

$$\Lambda:\mathcal{V}_j(\overline{\Omega_\sigma^{\mathrm{ext}}})\to\mathcal{V}_j(\overline{\Omega_\gamma^{\mathrm{ext}}}). \tag{7.24}$$

Note that $\Lambda(\mathcal{V}_j(\overline{\Omega_\sigma^{\mathrm{ext}}}))\subset\mathcal{V}_j(\overline{\Omega_\gamma^{\mathrm{ext}}})$ is automatically fulfilled, because every $F\in\mathcal{V}_j(\overline{\Omega_\sigma^{\mathrm{ext}}})$ of the form

$$F=\Phi_j^{(2)}*Q,\quad Q\in\mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}}) \tag{7.25}$$

with Fourier coefficients $F^\wedge(n,m)=(\varphi_j^\wedge(n))^2Q^\wedge(n,m)$ leads to

$$\begin{aligned}\Lambda F(x)&=\sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\Lambda^\wedge(n)F^\wedge(n,m)H_{n,m}^s(\gamma;x)\\&=\sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\Lambda^\wedge(n)(\varphi_j^\wedge(n))^2Q^\wedge(n,m)H_{n,m}^s(\gamma;x)\\&=(\Phi_j^{(2)})^{\gamma,\gamma}*(\Lambda Q)=(\Phi_j^{(2)})^{\gamma,\gamma}*G,\end{aligned} \tag{7.26}$$

where we let $G=\Lambda Q\in\mathcal{H}_s(\overline{\Omega_\gamma^{\mathrm{ext}}})$. Thus, we get the following theorem.

Theorem 7.10. *The restriction of the operator* $\Lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$ *to a scale space* $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, *i.e.,*

$$\Lambda|_{\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})} : \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{V}_j(\overline{\Omega_\gamma^{\text{ext}}}) \tag{7.27}$$

is injective. Moreover, we have the following results:

(i) *If the families* $\{\{\psi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ *and* $\{\{\tilde{\psi}_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ *are bandlimited (for example,* $\psi_j(n) = \tilde{\psi}_j(n) = 0$ *for all* $n \geq 2^j$*), then the restricted operator is even bijective. To be more specific, for* $G \in \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$ *the unique solution* $F_j \in \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, *of the equation*

$$\Lambda F_j = (\Phi_j^{(2)})^{\gamma,\gamma} * G \tag{7.28}$$

is given by

$$F_j = (\Phi_j^{(2)})^{\sigma,\sigma} * Q, \tag{7.29}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ *is given by*

$$Q^\wedge(n,m) = \begin{cases} \frac{G^\wedge(n,m)}{\Lambda^\wedge(n)}, & n \in [0, 2^j), \\ 0, & n \in [2^j, \infty), \end{cases} \tag{7.30}$$

$n = 0, 1, \ldots; m = 1, \ldots, 2n+1.$

(ii) *If the families* $\{\{\psi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ *and* $\{\{\tilde{\psi}_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ *are not bandlimited, the equation*

$$\Lambda F_j = (\Phi_j^{(2)})^{\gamma,\gamma} * G \tag{7.31}$$

has a solution $F_j \in \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ *provided that* $G \in \mathcal{H}_s^\Lambda(\overline{\Omega_\sigma^{\text{ext}}})$*, where* $\mathcal{H}_s^\Lambda(\overline{\Omega_\sigma^{\text{ext}}})$ *is a suitable Sobolev space (see the Ph.D.-thesis* [58] *for a detailed introduction). In this case, the unique solution of the equation is given by*

$$F_j = (\Phi_j^{(2)})^{\sigma,\sigma} * Q, \tag{7.32}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ *is obtainable by*

$$Q^\wedge(n,m) = \frac{G^\wedge(n,m)}{\Lambda^\wedge(n)}, \tag{7.33}$$

$n = 0, 1, \ldots; m = 1, \ldots, 2n+1.$

We now define the primal wavelets $\{\Psi_j^{\alpha,\alpha}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ and the dual wavelets $\{\tilde{\Psi}_j^{\alpha,\alpha}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ for $\alpha \geq \sigma$ in the way as we did in the case of the scaling functions and get

$$\Psi_j^{\alpha,\alpha}(x,y) = \sum_{n=0}^{\infty} \psi_j(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\alpha;x) H_{n,m}^s(\alpha;y), \tag{7.34}$$

$$\tilde{\Psi}_j^{\alpha,\alpha}(x,y) = \sum_{n=0}^{\infty} \tilde{\psi}_j(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\alpha;x) H_{n,m}^s(\alpha;y), \tag{7.35}$$

where

$$(\Psi_j^{\alpha,\alpha})^\wedge(n) = \psi_j(n), \quad (\tilde{\Psi}_j^{\alpha,\alpha})^\wedge(n) = \tilde{\psi}_j(n). \tag{7.36}$$

The detail spaces are defined in canonical manner:

$$\mathcal{W}_j(\overline{\Omega_\alpha^{\text{ext}}}) = \{(\Psi_j * \tilde{\Psi}_j)^{\alpha,\alpha} * F \ : \ F \in \mathcal{H}_s(\overline{\Omega_\alpha^{\text{ext}}})\}, \tag{7.37}$$

where

$$(\Psi_j * \tilde{\Psi}_j)^{\alpha,\alpha} = \Psi_j^{\alpha,\alpha} * \tilde{\Psi}_j^{\alpha,\alpha}. \tag{7.38}$$

The reconstruction formula given in Theorem 7.9 is valid substituting $\tilde{\Psi}_j * \Psi * F$ by $(\tilde{\Psi}_j * \Psi)^{\alpha,\alpha} * F$. Theorem 7.10 can now be transferred to the restriction on detail spaces and we get the following theorem.

Theorem 7.11. *The restriction of the operator $\Lambda : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$ to a detail space $\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, i.e.,*

$$\Lambda|_{\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})} : \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{W}_j(\overline{\Omega_\gamma^{\text{ext}}}) \tag{7.39}$$

is injective. Moreover, we have the following results:

(i) *If the family $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ is bandlimited (for example, $\varphi_j(n) = 0$ for all $n \geq 2^j$), then the restricted operator is even bijective. To be more specific, for $G \in \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$ the unique solution $H_j \in \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, of the equation*

$$\Lambda H_j = (\tilde{\Psi}_j * \Psi_j)^{\gamma,\gamma} * G \tag{7.40}$$

is given by

$$H_j = (\tilde{\Psi}_j * \Psi_j)^{\sigma,\sigma} * Q, \tag{7.41}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ is obtainable by

$$Q^\wedge(n,m) = \begin{cases} \frac{G^\wedge(n,m)}{\Lambda^\wedge(n)}, & n \in [0, 2^{j+1}), \\ 0, & n \in [2^{j+1}, \infty), \end{cases} \tag{7.42}$$

$n \in \mathbb{N}_0$; $m = 1, \ldots, 2n+1$.

(ii) *If the family $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ is non-bandlimited, the equation*

$$\Lambda H_j = (\tilde{\Psi}_j * \Psi_j)^{\gamma,\gamma} * G \tag{7.43}$$

has a solution $H_j \in \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$ provided that $G \in \mathcal{H}_s^\Lambda(\overline{\Omega_\sigma^{\text{ext}}})$, where $\mathcal{H}_s^\Lambda(\overline{\Omega_\sigma^{\text{ext}}})$ is a suitable Sobolev space (cf. the Ph.D.-thesis [58] for a detailed definition). In this case, the unique solution of the equation is given by

$$H_j = (\tilde{\Psi}_j * \Psi_j)^{\sigma,\sigma} * Q, \tag{7.44}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ is obtainable by

$$Q^\wedge(n,m) = \frac{G^\wedge(n,m)}{\Lambda^\wedge(n)}, \tag{7.45}$$

$n \in \mathbb{N}_0$; $m = 1, \ldots, 2n+1$.

Up to now, we have summarized some results about the filtered solution, i.e., the solution when we restrict the operator to scale or detail spaces. In the case of the unfiltered solution, we have the following theorem.

Theorem 7.12. *Let $G \in \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$ satisfy the condition $G \in \text{im}(\Lambda)$. Then the unique solution $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ of the equation $\Lambda F = G$ is given by*

$$F^\wedge(n,m) = \frac{G^\wedge(n,m)}{\Lambda^\wedge(n)}, \tag{7.46}$$

$n \in \mathbb{N}_0$; $m = 1, \ldots, 2n+1$.

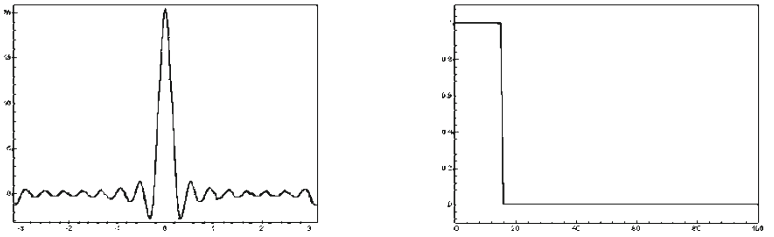
Examples for scaling functions

To make the preceding considerations more concrete, we would like to show that all reproducing kernel functions introduced in Section 5 may be used as $\mathcal{H}_{\sigma,\sigma}$-scaling functions. We essentially follow [19] and distinguish in accordance with Definition 7.2 two cases, viz. (1) *bandlimited $\mathcal{H}_{\sigma,\sigma}$-scaling functions* and (2) *non-bandlimited $\mathcal{H}_{\sigma,\sigma}$-scaling functions.*

(1) *Bandlimited $\mathcal{H}_{\sigma,\sigma}$-scaling Functions.* Suppose that $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ is a Sobolev space (satisfying the consistency conditions (CC1) and (CC2) relative to $[\sigma, \sigma^{\text{inf}})$). Consider sequences $\{\varphi_j(n)\}_{n\in\mathbb{N}_0}$ with "local support" (for example, $\varphi_j(n) = 0$ for all $n \geq 2^j$, $j \in \mathbb{N}_0$). Thus all members $\Phi_j(\cdot,\cdot)$ of an associated $\mathcal{H}_{\sigma,\sigma}$-scaling function $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ with $(\Phi_j)^\wedge(n) = \varphi_j(n)$, $n \in \mathbb{N}_0$, are bandlimited. This allows to deal with finite-dimensional scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$. Consequently, all spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ have finite-dimensional basis systems.

(1a) *Shannon $\mathcal{H}_{\sigma,\sigma}$-scaling function* (see Figure 7.1). Consider the family

$$\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$$



(a) $\Phi_4(x,y)$ (b) $\varphi_4(n)$

FIGURE 7.1. Shannon $\mathcal{H}_{\sigma,\sigma}$-scaling function for $j = 4$ and $A_n = 1$: space domain, i.e., $\Phi_j(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $\varphi_j(n)$ (right).

given by

$$\varphi_j(n) = \begin{cases} 1, & n \in [0, 2^j), \\ 0, & n \in [2^j, \infty). \end{cases}$$

The family $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ forms a generator of a scaling function in the sense of Definition 7.2. The $\mathcal{H}_{\sigma,\sigma}$-scaling function $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ reads as follows:

$$\Phi_j(x,y) = \sum_{n\le 2^j-1} \frac{1}{A_n^2} \frac{2n+1}{4\pi\sigma^2} \left(\frac{\sigma^2}{|x|\,|y|}\right)^{n+1} P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right),$$

$(x,y) \in \overline{\Omega_\sigma^{\rm ext}} \times \overline{\Omega_\sigma^{\rm ext}}$. A remarkable property is that $\Phi_j(\cdot,\cdot)$ coincides with its iterations:

$$\Phi_j^{(k)}(\cdot,\cdot) = (\Phi_j *_{\mathcal{H}} \Phi_j^{(k-1)})(\cdot,\cdot), \quad k = 2,3,\dots\ .$$

The scale spaces

$$\mathcal{V}_j(\overline{\Omega_\sigma^{\rm ext}}) = P_j(\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})) = \bigoplus_{n\le 2^j-1} Harm_n(\overline{\Omega_\sigma^{\rm ext}}), \quad j \in \mathbb{N}_0,$$

satisfy the properties:

(i) $\mathcal{V}_0(\overline{\Omega_\sigma^{\rm ext}}) \subset \cdots \subset \mathcal{V}_j(\overline{\Omega_\sigma^{\rm ext}}) \subset \mathcal{V}_{j+1}(\overline{\Omega_\sigma^{\rm ext}}) \subset \cdots \subset \mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$,

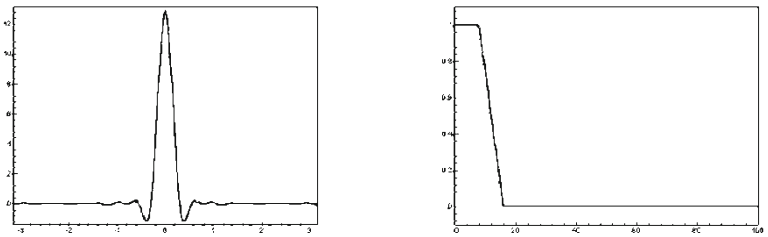
(ii) $\overline{\bigcup_{j\in\mathbb{N}_0} \mathcal{V}_j(\overline{\Omega_\sigma^{\rm ext}})}^{\|\cdot\|_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}} = \mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$,

(iii) $\bigcap_{j\in\mathbb{N}_0} V_j(\overline{\Omega_\sigma^{\rm ext}}) = Harm_0(\overline{\Omega_\sigma^{\rm ext}})$.

The multiresolution analysis is orthogonal. As a matter of fact, the Shannon "detail spaces" $\mathcal{W}_j(\overline{\Omega_\sigma^{\rm ext}}) = \mathcal{V}_{j+1}(\overline{\Omega_\sigma^{\rm ext}}) \ominus \mathcal{V}_j(\overline{\Omega_\sigma^{\rm ext}})$ of different scales j do not have any common frequencies. Consequently, the orthogonality of the outer harmonics immediately implies the orthogonality of the Shannon detail spaces. The scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\rm ext}})$, $j \in \mathbb{N}_0$, form an $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$-multiresolution analysis. Apart from this, it can be even verified that the decomposition of the scale space $\mathcal{V}_{j+1}(\overline{\Omega_\sigma^{\rm ext}})$ into the scale space $\mathcal{V}_j(\overline{\Omega_\sigma^{\rm ext}})$ and the detail space $\mathcal{W}_j(\overline{\Omega_\sigma^{\rm ext}})$ is orthogonal. This orthogonality of the decomposition easily follows from the already known fact that

$$\begin{aligned} \mathcal{V}_{j+1}(\overline{\Omega_\sigma^{\rm ext}}) &= \bigoplus_{0\le n\le 2^{j+1}-1} Harm_n(\overline{\Omega_\sigma^{\rm ext}}) \\ &= \bigoplus_{0\le n\le 2^j-1} Harm_n(\overline{\Omega_\sigma^{\rm ext}}) \oplus \bigoplus_{2^j\le n\le 2^{j+1}-1} Harm_n(\overline{\Omega_\sigma^{\rm ext}}) \\ &= \mathcal{V}_j(\overline{\Omega_\sigma^{\rm ext}}) \oplus \mathcal{W}_j(\overline{\Omega_\sigma^{\rm ext}}). \end{aligned} \tag{7.47}$$

On the one hand, the orthogonal structure of the Shannon multiresolution analysis seems to be very profitable. On the other hand, it is not surprising that the Shannon $\mathcal{H}_{\sigma,\sigma}$-scaling function shows strong oscillations. This is the price to be paid for the sharp separation "in momentum space". For numerical purposes it is often advisable to discuss "smoothed versions" of the Shannon



(a) $\Phi_4(x,y)$ (b) $\varphi_4(n)$

FIGURE 7.2. Smoothed Shannon $\mathcal{H}_{\sigma,\sigma}$-scaling function for $j = 4$ and $A_n = 1$, $h = \frac{1}{2}$: space domain, i.e., $\Phi_j(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $\varphi_j(n)$ (right).

kernels. But this automatically implies the loss of the orthogonality in the multiresolution analysis.

(1b) *Smoothed Shannon $\mathcal{H}_{\sigma,\sigma}$-scaling Function* (see Figure 7.2). For fixed $h \in [0,1)$ we now consider the family $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ given by

$$\varphi_j(n) = \begin{cases} 1, & n \in [0, 2^j h), \\ \frac{1-2^{-j}n}{1-h}, & n \in [2^j h, 2^j), \\ 0, & n \in [2^j, \infty). \end{cases}$$

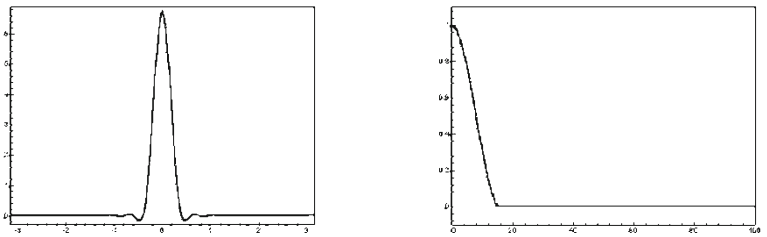
The family $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ defines a generator of an $\mathcal{H}_{\sigma,\sigma}$-scaling function. Obviously, $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ with $(\Phi_j)^\wedge(n) = \varphi_j(n)$ for $n,j \in \mathbb{N}_0$ is an $\mathcal{H}_{\sigma,\sigma}$-scaling function. Clearly, for each $n \in \mathbb{N}_0$, $\{\varphi_j(n)\}_{j\in\mathbb{N}_0}$ is monotonously increasing. The kernels $\Phi_j(\cdot,\cdot)$: $\overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\sigma^{\text{ext}}} \to \mathbb{R}$ read as follows:

$$\Phi_j(x,y) = \sum_{n\le 2^j-1} \frac{2n+1}{4\pi\sigma^2} \frac{\varphi_j(n)}{A_n^2} \left(\frac{\sigma^2}{|x|\,|y|}\right)^{n+1} P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right).$$

The value $h \in [0,1)$ represents a "control parameter" of the smoothing effect of the $\mathcal{H}_{\sigma,\sigma}$-scaling function $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$. The scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, form an $\mathcal{H}_{\sigma,\sigma}$-multiresolution analysis. This multiresolution analysis, however, is *not* orthogonal, since $\mathcal{V}_{j+1}(\overline{\Omega_\sigma^{\text{ext}}}) = \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) + \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, cannot be understood as orthogonal sum decomposition.

(1c) *Cubic Polynomial (CP) $\mathcal{H}_{\sigma,\sigma}$-scaling Function* (see Figure 7.3). In order to gain a higher intensity of the smoothing effect than in the case of the $\mathcal{H}_{\sigma,\sigma}$-scaling function (1b), we introduce a function $\varphi_0 : [0,\infty) \to \mathbb{R}$ in such a way that $\varphi_0|_{[0,1]}$ coincides with the uniquely determined cubic polynomial $p : [0,1] \to [0,1]$ with the properties:

$$p(0) = 1,\ p(1) = 0, \quad p'(0) = 0,\ p'(1) = 0.$$



(a) $\Phi_4(x,y)$ (b) $\varphi_4(n)$

FIGURE 7.3. CP $\mathcal{H}_{\sigma,\sigma}$-scaling function for $j = 4$ and $A_n = 1$: space domain, i.e., $\Phi_j(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $\varphi_j(n)$ (right).

It is not difficult to see that these properties are fulfilled by

$$p(t) = (1-t)^2(1+2t), \quad t \in [0,1].$$

This leads us to a function $\varphi_0 : [0,\infty) \to \mathbb{R}$ given by

$$\varphi_0(t) = \begin{cases} (1-t)^2(1+2t), & t \in [0,1), \\ 0, & t \in [1,\infty). \end{cases}$$

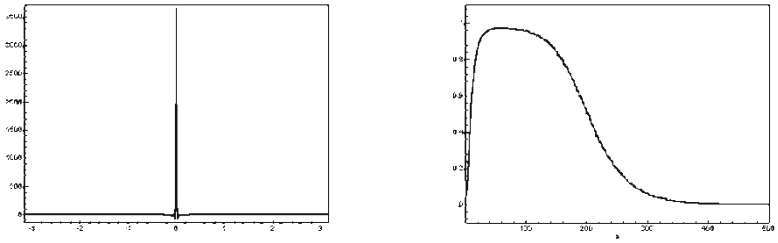
It is obvious that φ_0 is a monotonously decreasing function. In [31] a construction principle of deriving scaling functions from a "mother function" $\varphi_0 : [0,\infty) \to \mathbb{R}$ by letting $\varphi_j(t) = \varphi_0(2^{-j}t), t \in [0,\infty)$, is described and we thus define the family $\{\{\varphi_j\}_{j\in\mathbb{N}_0}\}_{n\in\mathbb{N}_0}$ with $\varphi_j(t) = \varphi_0(2^{-j}t)$, $t \in [0,\infty)$, by

$$\varphi_j(t) = \varphi_0(2^{-j}t) = \begin{cases} (1-2^{-j}t)^2(1+2^{-j+1}t), & t \in [0,2^j), \\ 0, & t \in [2^j,\infty). \end{cases}$$

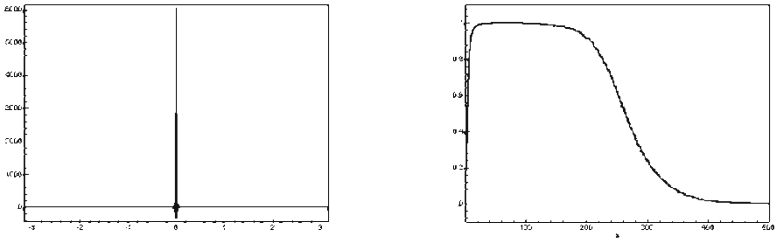
$\{\varphi_j(n)\}_{j\in\mathbb{N}_0}$ is a monotonously increasing sequence for each $n \in \mathbb{N}_0$, hence, $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ defines an $\mathcal{H}_{\sigma,\sigma}$-scaling function. The finite-dimensional scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, represent a non-orthogonal $\mathcal{H}_{\sigma,\sigma}$-multiresolution analysis.

Finally, it should be remarked that one can think of other ways to "smooth" the Shannon generator but these are not discussed.

(2) *Non-bandlimited $\mathcal{H}_{\sigma,\sigma}$-scaling functions.* Next we take a look at non-bandlimited generators of scaling functions. In other words, all $\mathcal{H}_{\sigma,\sigma}$-scaling functions $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ discussed in the following share the property that their "generators" $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ have a "global support". Since there are only a few conditions for a family $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ to generate an $\mathcal{H}_{\sigma,\sigma}$-scaling function, there are various possibilities for its concrete realization. In our approach we concentrate on three types: Tikhonov, rational, and exponential $\mathcal{H}_{\sigma,\sigma}$-scaling functions.



(a) $\Phi_5(x,y)$ (b) $\varphi_5(n)$



(c) $\Phi_7(x,y)$ (d) $\varphi_7(n)$

FIGURE 7.4. Tikhonov $\mathcal{H}_{\sigma,\sigma}$-scaling function for $j = 5$ (above) and $j = 7$ (below) and $A_n = 1$: space domain, i.e., $\Phi_j(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $\varphi_j(n)$ (right).

(2a) *Tikhonov $\mathcal{H}_{\sigma,\sigma}$-scaling Function* (see Figure 7.4). Consider the family

$$\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$$

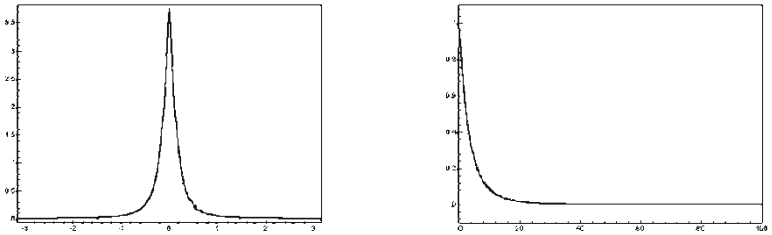
given by

$$\varphi_j(n) = \begin{cases} 1, & n = 0, \\ \left(\frac{\tau_n^2}{\tau_n^2+(2^{-j})^2}\right)^{1/2}, & n = 1,2,\ldots, \end{cases} \tag{7.48}$$

where the sequence $\{\tau_n\}_{n\in\mathbb{N}_0}$ with $\tau_n \neq 0$ for all $n \in \mathbb{N}_0$ is given in such a way that

$$\text{(i)} \qquad \sum_{n=0}^{\infty} \tau_n^2 < \infty\,, \qquad \text{(ii)} \qquad \sum_{n=0}^{\infty} (2n+1)\left(\frac{\tau_n}{A_n}\right)^2 < \infty.$$

It is not hard to see that the family $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ constitutes an $\mathcal{H}_{\sigma,\sigma}$-scaling function. The Tikhonov $\mathcal{H}_{\sigma,\sigma}$-scaling function plays an important role in the theory of regularization wavelets.



(a) $\Phi_4(x,y)$ (b) $\varphi_4(n)$

FIGURE 7.5. Rational $\mathcal{H}_{\sigma,\sigma}$-scaling function for $j = 4$ and $A_n = 1$, $\tau = 5$: space domain, i.e., $\Phi_j(x,y)$ for $(x,y) \in \Omega_\sigma \times \Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $\varphi_j(n)$ (right).

(2b) *Rational $\mathcal{H}_{\sigma,\sigma}$-scaling Functions* (see Figure 7.5). Consider $\varphi_j : [0,\infty) \to \mathbb{R}$ given by

$$\varphi_j(t) = (1 + 2^{-j}t)^{-\tau}, \quad t \in [0,\infty), \quad \tau > 1. \tag{7.49}$$

Clearly, for all values $\tau > 1$, the family $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ forms a generator of a scaling function. All functions φ_j, $j \in \mathbb{N}_0$, define admissible $\mathcal{H}_{\sigma,\sigma}$-kernels $\Phi_j(\cdot,\cdot)$, $j \in \mathbb{N}_0$, if, in addition, $\tau > 1$ is chosen in such a way that

$$\sum_{n=0}^{\infty}(2n+1)\frac{(1+2^{-j}n)^{-2\tau}}{A_n^2} < \infty \tag{7.50}$$

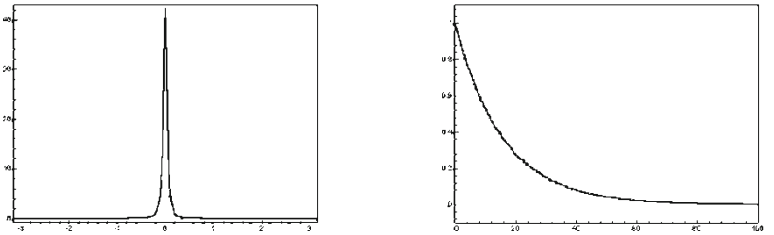
for $j \in \mathbb{N}_0$. For example, in the case of $\mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$, i.e., $A_n = \left(\frac{n+\frac{1}{2}}{\sigma}\right)^s$ for $n = 0,1,\ldots$, we find $s + \tau > 1$ to satisfy the estimate (7.50). More generally, $(1+n)^{-2\tau}A_n^{-2} = \mathcal{O}(n^{-2-\varepsilon})$ for $n \to \infty$ with $\varepsilon > 0$ together with $\tau > 1$ is a sufficient condition to define an admissible $\mathcal{H}_{\sigma,\sigma}$-kernel $\Phi_j(\cdot,\cdot)$, $j \in \mathbb{N}_0$. The $\mathcal{H}_{\sigma,\sigma}$-scaling function $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ consists of the kernels

$$\Phi_j(x,y) = \sum_{n=0}^{\infty}\frac{(1+2^{-j}n)^{-\tau}}{A_n^2}\frac{2n+1}{4\pi\sigma^2}\left(\frac{\sigma^2}{|x|\,|y|}\right)^{n+1}P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right),$$

$(x,y) \in \overline{\Omega_\sigma^{\mathrm{ext}}} \times \overline{\Omega_\sigma^{\mathrm{ext}}}$. The functions φ_j, $j \in \mathbb{N}_0$, are monotonously decreasing on the interval $[0,\infty)$ for all values $\tau > 1$ and all $j \in \mathbb{N}_0$. Therefore, the scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\mathrm{ext}}})$ form an $\mathcal{H}_{\sigma,\sigma}$-multiresolution analysis provided that both $\tau > 1$ and the summability condition (7.50) is valid.

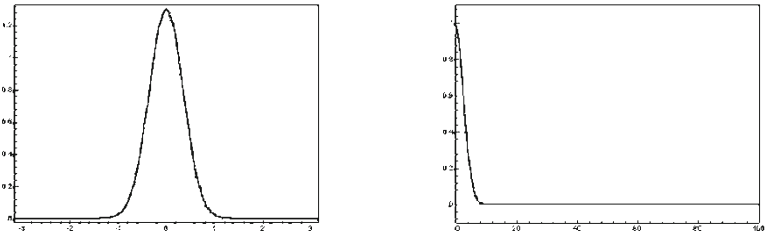
(2c) *Exponential $\mathcal{H}_{\sigma,\sigma}$-scaling Functions* (see Figures 7.6 and 7.7). Choose $\varphi_j : [0,\infty) \to \mathbb{R}$, $j \in \mathbb{N}_0$, to be defined by

$$\varphi_j(t) = e^{-2^{-j}H(t)}, \qquad t \in [0,\infty), \tag{7.51}$$



(a) $\Phi_4(x,y)$ (b) $\varphi_4(n)$

FIGURE 7.6. Abel–Poisson $\mathcal{H}_{\sigma,\sigma}$-scaling function for $j=4$ and $A_n=1$, $\tau=1$: space domain, i.e., $\Phi_j(x,y)$ for $(x,y)\in\Omega_\sigma\times\Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $\varphi_j(n)$ (right).



(a) $\Phi_4(x,y)$ (b) $\varphi_4(n)$

FIGURE 7.7. Gauss–Weierstraß $\mathcal{H}_{\sigma,\sigma}$-scaling function for $j=4$ and $A_n=1$, $\tau=1$: space domain, i.e., $\Phi_j(x,y)$ for $(x,y)\in\Omega_\sigma\times\Omega_\sigma$ in sectional representation (left) and frequency domain, i.e., $\varphi_j(n)$ (right).

where $H:[0,\infty)\to[0,\infty)$ satisfies the properties:

- $H\in C^{(\infty)}[0,\infty)$,
- $H(0)=0$,
- $H(t)>0 \quad$ for $\quad t>0$,
- $H(t)<H(t') \quad$ whenever $\quad 0<t<t'$.

The sequence $\{\varphi_j(n)\}_{j\in\mathbb{N}_0}$ is monotonously increasing for each $n\in\mathbb{N}_0$. The functions φ_j, $j\in\mathbb{N}_0$, define an $\mathcal{H}_{\sigma,\sigma}$-scaling function $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ by letting $(\Phi_j)^\wedge(n)=\varphi_j(n)$, $n\in\mathbb{N}_0$, provided that $\Phi_j(\cdot,\cdot)$, $j\in\mathbb{N}_0$, are admissible $\mathcal{H}_{\sigma,\sigma}$-kernel functions. It is not hard to see that

$$(\Phi_j*\Phi_j)(x,y)$$
$$=\sum_{n=0}^{\infty}\frac{(e^{-2^{-j}H(n)})^2}{A_n^2}\,\frac{2n+1}{4\pi\sigma^2}\left(\frac{\sigma^2}{|x|\,|y|}\right)^{n+1}P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right)$$

$$= \sum_{n=0}^{\infty} \frac{e^{-2^{-(j-1)}H(n)}}{A_n^2} \frac{2n+1}{4\pi\sigma^2} \left(\frac{\sigma^2}{|x|\,|y|}\right)^{n+1} P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right)$$
$$= \Phi_{j-1}(x,y) \tag{7.52}$$

holds for all $j \in \mathbb{N}$ and all $(x,y) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\sigma^{\text{ext}}}$. The scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ constitute an $\mathcal{H}_{\sigma,\sigma}$-multiresolution analysis. Altogether we find the following result for exponential $\mathcal{H}_{\sigma,\sigma}$-scaling functions: The family $\{P_j\}_{j\in\mathbb{N}_0}$ of operators $P_j : \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ (defined by $P_j(F) = \Phi_j^{(2)} * F$, $F \in \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$) forms an $\mathcal{H}_{\sigma,\sigma}$-contracting approximate identity (called the *exponential $\mathcal{H}_{\sigma,\sigma}$-contracting approximate identity*), i.e., the following properties are satisfied:

(i) P_j is a bounded linear operator for every $j \in \mathbb{N}_0$ and $P_\infty = I$ (identity),
(ii) $P_{j-1} = P_j P_j$ for all $j \in \mathbb{N}_0$,
(iii) $\lim_{j\to\infty} \|F - P_j(F)\|_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})} = 0$ for all $F \in \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$,
(iv) $\|P_j(F)\|_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})} \le \|F\|_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}$ for all $j \in \mathbb{N}_0$, $F \in \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$.

As examples we mention the *Abel–Poisson $\mathcal{H}_{\sigma,\sigma}$-contracting approximate identity* given by $H(t) = \alpha t$, $\alpha > 0$, and the *Gauss–Weierstraß $\mathcal{H}_{\sigma,\sigma}$-contracting approximate identity* given by $H(t) = \alpha t(t+1)$, $\alpha > 0$.

Remark 7.13. Non-bandlimited scaling functions become bandlimited ones by suitable truncation in momentum space. To be more specific, if $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ is a non-bandlimited $\mathcal{H}_{\sigma,\sigma}$-scaling function, then $\{\Gamma_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by $(\Gamma_j)^\wedge(n) = (\Phi_j)^\wedge(n)$ for $n \in [0, 2^j)$ and $(\Gamma_j)^\wedge(n) = 0$ for $n \in [2^j, \infty)$ represents a bandlimited $\mathcal{H}_{\sigma,\sigma}$-scaling function.

We now explain the connection between the solution in the scale spaces and the unfiltered solution.

Theorem 7.14. *Suppose that G is of class $\mathcal{H}_s^\Lambda(\overline{\Omega_\gamma^{\text{ext}}})$. Let $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ be the unique solution of $\Lambda F = G$. Then*

$$F_j = (\Phi_j^{(2)})^{\sigma,\sigma} * F \tag{7.53}$$

is the unique solution in $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ of the equation

$$\Lambda F_j = (\Phi_j^{(2)})^{\gamma,\gamma} * G \tag{7.54}$$

for every $j \in \mathbb{N}_0$. Furthermore, the limit relation

$$\lim_{J\to\infty} (\Phi_J^{(2)})^{\sigma,\sigma} * F = F \tag{7.55}$$

holds (in $\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}$-sense).

In the case of bandlimited scaling functions, the preceding theorem shows that the (scalar) SST or SGG problem is well posed: A unique solution always exists and due to the finite dimension of the scale spaces the solution is also stable. According to the multiscale approach the solution in the scale space is given by adding the solution of the corresponding detail spaces to the solution of

the scale space of a lower scale. Because of the limit relation given in Theorem 7.14 the filtered solutions converge to the unfiltered solution in the Sobolev space $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$. If we now turn to non-bandlimited scaling functions, the stability of the solution cannot be ensured, because the (scalar) SST or SGG problem is an exponentially ill-posed problem with unbounded inverse operator Λ^{-1}. In order to obtain a well-posed problem, we have to replace the inverse operator by an appropriate bounded operator, that is we have to use a regularization of Λ^{-1}.

Definition 7.15. A family of linear operators $S_j : \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, is called a *regularization* of Λ^{-1}, if it satisfies the following properties:

(i) S_j is bounded on $\mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$ for all $j \in \mathbb{N}_0$,
(ii) for any member $G \in im(\Lambda)$, the limit relation $\lim_{J\to\infty} S_J G = \Lambda^{-1}G$ holds (in $\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}$-sense).

The function $F_J = S_J G$ is called the *J-level regularization of* $\Lambda^{-1}G$. In our approach we want to represent the J-level regularization with the help of harmonic wavelets which guarantees that we can calculate the $J+1$-level regularization by adding the corresponding detail information to the J-level regularization. In order to formulate the multiscale regularization concept, we start with the definition of a generator of a regularization scaling function by modifying Definition 7.2.

Definition 7.16. A family $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ of sequences $\{\varphi_j(n)\}_{n\in\mathbb{N}_0}$ is called a *generator of a regularization scaling function with respect to* Λ^{-1}, if it satisfies the following requirements:

(i) $(\varphi_j(0))^2 = \frac{1}{\Lambda\hat{}(0)}$, for all $j \in \mathbb{N}_0$,
(ii) $(\varphi_j(n))^2 \le (\varphi_{j'}(n))^2$, for all $j, j' \in \mathbb{N}_0$ with $j \le j'$ and all $n \in \mathbb{N}$,
(iii) $\lim_{j\to\infty} (\varphi_j(n))^2 = \frac{1}{\Lambda\hat{}(n)}$, for all $n \in \mathbb{N}$.

Now we are able to define the decomposition and reconstruction regularization scaling functions in such a way that the corresponding convolutions lead to the J-level approximation of $\Lambda^{-1}G$, $G \in \text{im}(\Lambda)$.

Definition 7.17. Let $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ be a generator of a regularization scaling function with respect to Λ^{-1}. Then a family $\{{}^d\Phi_j^{\sigma,\gamma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of admissible $\mathcal{H}_{\sigma,\gamma}$-kernels given by

$$ {}^d\Phi_j^{\sigma,\gamma}(x,z) = \sum_{n=0}^{\infty} \varphi_j(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\sigma;x) H_{n,m}^s(\gamma;z), \tag{7.56} $$

$(x,z) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\gamma^{\text{ext}}}$, is called a *decomposition regularization* $\mathcal{H}_{\sigma,\gamma}$*-scaling function with respect to* Λ^{-1}, whereas a family $\{{}^r\Phi_j^{\sigma,\sigma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of admissible $\mathcal{H}_{\sigma,\sigma}$-kernels given by

$$ {}^r\Phi_j^{\sigma,\sigma}(x,y) = \sum_{n=0}^{\infty} \varphi_j(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\sigma;x) H_{n,m}^s(\sigma;y), \tag{7.57} $$

$(x,y) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\sigma^{\text{ext}}}$ is called a *reconstruction regularization $\mathcal{H}_{\sigma,\sigma}$-scaling function with respect to Λ^{-1}*.

Obviously, the regularization scaling functions fulfill

$$ {}^d\Phi_j^{\sigma,\gamma}(x,\cdot) \in \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}}), \quad x \in \overline{\Omega_\sigma^{\text{ext}}},\ j \in \mathbb{N}_0, \tag{7.58} $$

$$ {}^r\Phi_j^{\sigma,\sigma}(x,\cdot) \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}), \quad x \in \overline{\Omega_\sigma^{\text{ext}}},\ j \in \mathbb{N}_0. \tag{7.59} $$

As already stated, we obtain the following theorem:

Theorem 7.18. *Let $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ be a generator of a regularization scaling function with respect to Λ^{-1}. If we define the admissible $\mathcal{H}_{\sigma,\gamma}$-kernel $({}^r\Phi_j * {}^d\Phi_j)^{\sigma,\gamma}(\cdot,\cdot)$ by*

$$ ({}^r\Phi_j * {}^d\Phi_j)^{\sigma,\gamma}(x,z) = {}^r\Phi_j^{\sigma,\sigma}(x,\cdot) * {}^d\Phi_j^{\sigma,\gamma}(\cdot,z), \tag{7.60} $$

$(x,z) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\gamma^{\text{ext}}}$, *then*

$$ F_J = ({}^r\Phi_J * {}^d\Phi_J)^{\sigma,\gamma} * G, \quad G \in \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}}), \tag{7.61} $$

represents the J-level regularization of $\Lambda^{-1}G$.

If, in addition, $G \in \text{im}(\Lambda) = \mathcal{H}_s^\Lambda(\overline{\Omega_\gamma^{\text{ext}}})$, then

$$ \lim_{J\to\infty} \|F_J - \Lambda^{-1}G\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} = 0. \tag{7.62} $$

If we define the convolution operators $S_J : \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $J \in \mathbb{N}_0$, by

$$ S_J(G) = ({}^r\Phi_J * {}^d\Phi_J)^{\sigma,\gamma} * G, \tag{7.63} $$

and introduce the scale spaces$S_J(\text{im}(\Lambda))$ as follows

$$ S_J(\text{im}(\Lambda)) = \{({}^r\Phi_J * {}^d\Phi_J)^{\sigma,\gamma} * G \ :\ G \in \text{im}(\Lambda)\}, \tag{7.64} $$

the following theorem holds.

Theorem 7.19. *The scale spaces satisfy the following properties:*

(i) $S_0(\text{im}(\Lambda)) \subset \cdots \subset S_J(\text{im}(\Lambda)) \subset S_{J'}(\text{im}(\Lambda)) \subset \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $J \le J'$, *i.e., for any right-hand side $G \in \text{im}(\Lambda)$ of the (scalar) SST or SGG problem, all J-level regularizations with fixed parameter J are sampled in a scale space $S_J(\text{im}(\Lambda))$ with the above property,*

(ii) $\overline{\bigcup_{J=0}^\infty S_J(\text{im}(\Lambda))}^{\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}} = \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$.

A set of subspaces of $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ satisfying the conditions of Theorem 7.19 is called *regularization $\mathcal{H}_{\sigma,\gamma}$-multiresolution analysis (RMRA) of the (scalar) SST or SGG problem.*

We now turn to the definition of regularization wavelets following the procedure described in the case of regularization scaling functions. Obviously, we have to define decomposition and reconstruction regularization wavelets.

Definition 7.20. Let $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ be a generator of a regularization scaling function with respect to Λ^{-1}. Then the generating symbols $\{\psi_j(n)\}_{n\in\mathbb{N}_0}$ and $\{\tilde{\psi}_j(n)\}_{n\in\mathbb{N}_0}$ of the corresponding regularization wavelets are defined by the refinement equation (7.10). The admissible $\mathcal{H}_{\sigma,\gamma}$-kernel $\{{}^d\Psi_j^{\sigma,\gamma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by

$$ {}^d\Psi_j^{\sigma,\gamma}(x,z) = \sum_{n=0}^{\infty} \psi_j(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\sigma;x) H_{n,m}^s(\gamma;z), \tag{7.65}$$

$(x,z) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\gamma^{\text{ext}}}$ is called the *decomposition regularization* $\mathcal{H}_{\sigma,\gamma}$*-wavelet*, while the admissible $\mathcal{H}_{\sigma,\sigma}$-kernel $\{{}^r\tilde{\Psi}_j^{\sigma,\sigma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by

$$ {}^r\tilde{\Psi}_j^{\sigma,\sigma}(x,y) = \sum_{n=0}^{\infty} \tilde{\psi}_j(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\sigma;x) H_{n,m}^s(\sigma;y), \tag{7.66}$$

$(x,y) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\sigma^{\text{ext}}}$ is called the *reconstruction regularization* $\mathcal{H}_{\sigma,\sigma}$*-wavelet*.

We now define the convolution operators $T_j : \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, by

$$ T_j(G) = ({}^r\tilde{\Psi}_j * {}^d\Psi_j)^{\sigma,\gamma} * G, \quad G \in \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}}). \tag{7.67}$$

Obviously, due to the refinement equation, the operator S_{J+1} can be represented in the form

$$ S_{J+1} = S_0 + \sum_{j=0}^{J} T_j. \tag{7.68}$$

Thus, we now introduce the *detail spaces* $T_J(\text{im}(\Lambda))$ by

$$ T_J(\text{im}(\Lambda)) = \left\{ ({}^r\tilde{\Psi}_J * {}^d\Psi_J)^{\sigma,\gamma} * G \ : \ G \in \text{im}(\Lambda) \right\}. \tag{7.69}$$

The space $T_J(\text{im}(\Lambda))$ contains the detail information which has to be added in order to turn from the J-level regularization to the $J+1$-level regularization:

$$ S_{J+1}(\text{im}(\Lambda)) = S_J(\text{im}(\Lambda)) + T_J(\text{im}(\Lambda)). \tag{7.70}$$

In general, the sum is neither direct nor orthogonal.

Theorem 7.21. *Let* $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ *be a generator of a regularization scaling function with respect to* Λ^{-1}. *Suppose that* $\{\{\psi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$, $\{\{\tilde{\psi}_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ *are the generating symbols of the corresponding regularization wavelets. Furthermore, let* G *be of class* $\mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}})$. *Define the* regularization $\mathcal{H}_{\sigma,\gamma}$-wavelet transform at scale $j \in \mathbb{N}_0$ and position $x \in \overline{\Omega_\sigma^{\text{ext}}}$ *by*

$$ (\text{RWT})(G)(j;x) = {}^d\Psi_J^{\sigma,\gamma}(x,\cdot) * G, \quad G \in \mathcal{H}_s(\overline{\Omega_\gamma^{\text{ext}}}). \tag{7.71}$$

Then

$$ F_J = \left({}^r\Phi_0 * {}^d\Phi_0\right)^{\sigma,\gamma} * G + \sum_{j=0}^{J-1} {}^r\tilde{\Psi}_J^{\sigma,\sigma} * (RWT)(G)(j;\cdot) \tag{7.72}$$

is the J-level regularization of the (scalar) SST or SGG problem satisfying

$$\lim_{J\to\infty} \|F_J - \Lambda^{-1}G\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} = 0 \tag{7.73}$$

provided that $G \in \text{im}(\Lambda) = \mathcal{H}_s^\Lambda(\overline{\Omega_\gamma^{\text{ext}}})$.

Some examples of regularization wavelets and numerical calculations can be found in [19], where, in addition, all the above-mentioned theorems are proved.

7.2. Vectorial wavelet theory

We now give the extension of the scalar wavelet theory to the vectorial case. First we define vectorial scaling functions and wavelets. The reconstruction formula is the main result stating how the function can be split into a lowpass part and an infinite sum of bandpass parts. Then we solve the (vectorial) SST or SGG problem defining regularization wavelets. We use the notation $\hat{\Phi}_j^{(i)} \star \hat{\Phi}_j^{(i)} * f$ instead of $\hat{\Phi}_j^{(i)} \star (\hat{\Phi}_j^{(i)} * f)$, and $\hat{\Phi}_j \star \hat{\Phi}_j * f = \sum_{i=1}^3 \hat{\Phi}_j^{(i)} \star \hat{\Phi}_j^{(i)} * f^{(i)}$.

Definition 7.22. Let $\{\hat{\Phi}_j^{(i)}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be a family of admissible $h_{\sigma,\sigma}^{(i)}$-kernels, $i \in \{1,2,3\}$. Then the family $\{\mathcal{V}_j^{(i)}(\overline{\Omega_\sigma^{\text{ext}}})\}_{j\in\mathbb{N}_0}$ of scale spaces $\mathcal{V}_j^{(i)}(\overline{\Omega_\sigma^{\text{ext}}})$ defined by

$$\mathcal{V}_j^{(i)}(\overline{\Omega_\sigma^{\text{ext}}}) = \{\hat{\Phi}_j^{(i)} \star \hat{\Phi}_j^{(i)} * f \ : \ f \in h_s(\overline{\Omega_\sigma^{\text{ext}}})\}, \tag{7.74}$$

is called an $h_{\sigma,\sigma}^{(i)}$-*multiresolution analysis*, if the following properties are satisfied:

(i) $\mathcal{V}_0^{(i)}(\overline{\Omega_\sigma^{\text{ext}}}) \subset \cdots \subset \mathcal{V}_j^{(i)}(\overline{\Omega_\sigma^{\text{ext}}}) \subset \mathcal{V}_{j+1}^{(i)}(\overline{\Omega_\sigma^{\text{ext}}}) \subset \cdots \subset h_s^{(i)}(\overline{\Omega_\sigma^{\text{ext}}})$,

(ii) $\overline{\bigcup_{j\in\mathbb{N}_0} \mathcal{V}_j^{(i)}(\overline{\Omega_\sigma^{\text{ext}}})}^{\|\cdot\|_{h_s(\overline{\Omega_\sigma^{\text{ext}}})}} = h_s^{(i)}(\overline{\Omega_\sigma^{\text{ext}}})$.

Definition 7.23. Let $\{\hat{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be a family of admissible $h_{\sigma,\sigma}$-kernels. The set of scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ defined by

$$\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) = \{\hat{\Phi}_j \star \hat{\Phi}_j * f \ : \ f \in h_s(\overline{\Omega_\sigma^{\text{ext}}})\} \tag{7.75}$$

is called an $h_{\sigma,\sigma}$-multiresolution analysis, if $\{\mathcal{V}_j^{(i)}(\overline{\Omega_\sigma^{\text{ext}}})\}_{j\in\mathbb{N}_0}$ is an $h_s^{(i)}(\overline{\Omega_\sigma^{\text{ext}}})$-multiresolution analysis for every $i \in \{1,2,3\}$.

Our next purpose is to define scaling functions.

Definition 7.24. A family $\{\{\varphi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$ of sequences $\{\varphi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}$ is called a *generator of a scaling function of kind* i, $i \in \{1,2,3\}$, if it satisfies the following requirements:

(i) $(\varphi_j^{(i)}(0_i))^2 = 1$ for all $j \in \mathbb{N}_0$,

(ii) $(\varphi_j^{(i)}(n))^2 \le \left(\varphi_{j'}^{(i)}(n)\right)^2$ for all $j, j' \in \mathbb{N}_0$ with $j \le j'$ and all $n \in \mathbb{N}_{0_i+1}$,

(iii) $\lim_{j\to\infty} (\varphi_j^{(i)}(n))^2 = 1$ for all $n \in \mathbb{N}_{0_i+1}$.

Furthermore, the family $\{\{\{\varphi^{(i)}(n)\}_{i\in\{1,2,3\}}\}_{n\in\mathbb{N}_{0_i+1}}\}_{j\in\mathbb{N}_0}$ is called a *generator of a scaling function*, if $\{\{\varphi^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$ are generators of a scaling function of kind i, $i\in\{1,2,3\}$.

Based on the definition of a generator of a scaling function, we now introduce $h_{\sigma,\sigma}$-scaling functions.

Definition 7.25. A family $\{\hat{\Phi}_j^{(i)}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of $h^{(i)}$-kernels $\hat{\Phi}_j^{(i)}(\cdot,\cdot)$ defined by

$$\hat{\Phi}_j^{(i)\wedge}(n)=\varphi_j^{(i)}(n),\quad j\in\mathbb{N}_0,\ n\in\mathbb{N}_{0_i},\ i\in\{1,2,3\},$$

i.e.,

$$\hat{\Phi}_j^{(i)}(x,y)=\sum_{n=0_i}^{\infty}\varphi_j^{(i)}(n)\sum_{m=1}^{2n+1}H_{n,m}^{s}(\sigma;x)h_{n,m}^{(i)s}(\sigma;y),\quad x,y\in\overline{\Omega_\sigma^{\mathrm{ext}}},\tag{7.76}$$

is called an $h_{\sigma,\sigma}^{(i)}$-*scaling function*, if it satisfies the following properties:

(i) $\hat{\Phi}_j^{(i)}(\cdot,\cdot)$ is an admissible $h_{\sigma,\sigma}^{(i)}$-kernel for every $j\in\mathbb{N}_0$,

(ii) $\{\{\hat{\Phi}_j^{(i)\wedge}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$ constitutes a generator of a scaling function of kind i.

Furthermore, the family $\{\hat{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of $h_{\sigma,\sigma}$-kernels $\hat{\Phi}_j(\cdot,\cdot)$ is called an $h_{\sigma,\sigma}$-scaling function, if $\{\hat{\Phi}_j^{(i)}\}_{j\in\mathbb{N}_0}$ are $h_{\sigma,\sigma}^{(i)}$-scaling functions for $i\in\{1,2,3\}$.

The following approximation property can be derived.

Theorem 7.26. *Let $\{\hat{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $h_{\sigma,\sigma}$-scaling function. Then*

$$\lim_{j\to\infty}\|f-\hat{\Phi}_j\star\hat{\Phi}_j*f\|_{h_s(\overline{\Omega_\sigma^{\mathrm{ext}}})}=0\tag{7.77}$$

holds for all $f\in h_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$.

Definition 7.27. Let $\{\hat{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $h_{\sigma,\sigma}$-scaling function. Then $\{P_j\}_{j\in\mathbb{N}_0}$ with $P_j: h_s(\overline{\Omega_\sigma^{\mathrm{ext}}})\to h_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$ defined by

$$P_j(f)(x)=\hat{\Phi}_j\star\hat{\Phi}_j*f,\quad f\in h_s(\overline{\Omega_\sigma^{\mathrm{ext}}}),\ x\in\overline{\Omega_\sigma^{\mathrm{ext}}},\tag{7.78}$$

is called an $h_{\sigma,\sigma}$-*approximate identity.*

The kernel $\hat{\Phi}_0$ is called the mother kernel of the $h_{\sigma,\sigma}$-scaling function.

Theorem 7.28. *Let $\{\hat{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $h_{\sigma,\sigma}$-scaling function. Then $\{\mathcal{V}_j(\overline{\Omega_\sigma^{\mathrm{ext}}})\}_{j\in\mathbb{N}_0}$ defined in* (7.75) *forms an $h_{\sigma,\sigma}$-multiresolution analysis.*

We are now at the point to define the (primal/dual) wavelet with the help of the bilinear refinement equation.

Definition 7.29. Let $\{\hat{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $h_{\sigma,\sigma}$-scaling function. Then the families of $h_{\sigma,\sigma}$-kernels $\{\hat{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$, $\{\tilde{\hat{\Psi}}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by

$$(\hat{\Psi}_j)^{(i)\wedge}(n)=\psi_j^{(i)}(n),\quad j\in\mathbb{N}_0,\ n\in\mathbb{N}_{0_i},\ i\in\{1,2,3\},\tag{7.79}$$

$$(\tilde{\hat{\Psi}}_j)^{(i)\wedge}(n)=\tilde{\psi}_j^{(i)}(n),\quad j\in\mathbb{N}_0,\ n\in\mathbb{N}_{0_i},\ i\in\{1,2,3\},\tag{7.80}$$

are called (*primal*) $h_{\sigma,\sigma}$*-wavelet* and *dual* $h_{\sigma,\sigma}$*-wavelet*, respectively, if all $h_{\sigma,\sigma}$-kernels $\hat{\Psi}_j(\cdot,\cdot)$, $\tilde{\hat{\Psi}}_j(\cdot,\cdot)$, $j\in\mathbb{N}_0$, are admissible and the symbols $\{\psi_j^{(i)}(n)\}$, $\{\tilde{\psi}_j^{(i)}(n)\}$, in addition, satisfy the (vectorial) refinement equation

$$\tilde{\psi}_j^{(i)}(n)\psi_j^{(i)}(n)=(\varphi_{j+1}^{(i)}(n))^2-(\varphi_j^{(i)}(n))^2\tag{7.81}$$

for all $j\in\mathbb{N}_0$, $n\in\mathbb{N}_{0_i}$, $i\in\{1,2,3\}$.

The following equation can directly be seen:

$$(\varphi_{J+1}^{(i)}(n))^2=(\varphi_0^{(i)}(n))^2+\sum_{j=0}^{J}\tilde{\psi}_j^{(i)}(n)\psi_j^{(i)}(n),\quad J\in\mathbb{N}_0,\tag{7.82}$$

for all $n\in\mathbb{N}_{0_i}$. We now define the wavelet transform. To this end we let $\psi_{-1}^{(i)}(n)=\tilde{\psi}_{-1}^{(i)}(n)=\varphi_0^{(i)}(n)$ and $\hat{\Psi}_{-1}(\cdot,\cdot)=\tilde{\hat{\Psi}}_{-1}(\cdot,\cdot)=\hat{\Phi}_0(\cdot,\cdot)$ for $n\in\mathbb{N}_{0_i}$, $i\in\{1,2,3\}$. We remember that we have already defined the space $\mathcal{H}_s(\mathbb{N}_{-1}\times\overline{\Omega_\sigma^{\mathrm{ext}}})$ (see Eqs. (7.12)–(7.14))

Definition 7.30. Let $\{\hat{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_{-1}}$ be a (primal) $h_{\sigma,\sigma}$-wavelet. Then $(WT)^{(i)}: h_s(\overline{\Omega_\sigma^{\mathrm{ext}}})\to\mathcal{H}_s(\mathbb{N}_{-1}\times\overline{\Omega_\sigma^{\mathrm{ext}}})$ defined by

$$(WT)^{(i)}(f)(j;y)=(\hat{\Psi}_j^{(i)}*f)(y)\tag{7.83}$$

is called $h_{\sigma,\sigma}$-wavelet transform of kind i of f at position $y\in\overline{\Omega_\sigma^{\mathrm{ext}}}$ and scale $j\in\mathbb{N}_{-1}$.

As usual, we define the detail space $\mathcal{W}_j^{(i)}(\overline{\Omega_\sigma^{\mathrm{ext}}})$ at scale j by

$$\mathcal{W}_j^{(i)}(\overline{\Omega_\sigma^{\mathrm{ext}}})=\big\{\tilde{\hat{\Psi}}_j^{(i)}\star\hat{\Psi}_j^{(i)}*f\ :\ f\in h_s(\overline{\Omega_\sigma^{\mathrm{ext}}})\big\},\tag{7.84}$$

and

$$\mathcal{W}_j(\overline{\Omega_\sigma^{\mathrm{ext}}})=\big\{\tilde{\hat{\Psi}}_j\star\hat{\Psi}_j*f\ :\ f\in h_s(\overline{\Omega_\sigma^{\mathrm{ext}}})\big\}.\tag{7.85}$$

Theorem 7.31 (Vectorial Reconstruction Formula for the Outer Space). *Let the families* $\{\hat{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ *and* $\{\tilde{\hat{\Psi}}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$*, respectively, be a (primal)* $h_{\sigma,\sigma}$*-wavelet and its dual corresponding to an* $h_{\sigma,\sigma}$*-scaling function* $\{\hat{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$*. Then*

$$f=\sum_{j=-1}^{\infty}\tilde{\hat{\Psi}}_j\star\hat{\Psi}_j*f\tag{7.86}$$

holds for all $f\in h_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$ *(in* $\|\cdot\|_{h_s(\overline{\Omega_\sigma^{\mathrm{ext}}})}$*-sense).*

Our next purpose is to solve the (vectorial) SST or SGG problem with the help of bandlimited harmonic wavelets. First, we transfer the theory of $h^{(i)}_{\sigma,\sigma}$-scaling functions to the case of $h^{(i)}_{\alpha,\alpha}$-scaling functions $\hat{\Phi}^{(i),\alpha,\alpha}_j$ with $\alpha \geq \sigma$:

$$\hat{\Phi}^{(i),\alpha,\alpha}_j(x,y) = \sum_{n=0_i}^{\infty} \varphi^{(i)}_j(n) \sum_{m=1}^{2n+1} H^s_{n,m}(\alpha;x) h^{(i)s}_{n,m}(\alpha;y), \tag{7.87}$$

where

$$(\hat{\Phi}^{(i),\alpha,\alpha}_j)^\wedge(n) = \varphi^{(i)}_j(n). \tag{7.88}$$

Obviously, Theorem 7.26 can be directly transferred substituting σ by α. The scale spaces are defined in the following way:

$$\mathcal{V}^{(i)}_j(\overline{\Omega^{\text{ext}}_\alpha}) = \{\hat{\Phi}^{(i),\alpha,\alpha}_j \star \hat{\Phi}^{(i),\alpha,\alpha}_j * f \ : \ f \in h_s(\overline{\Omega^{\text{ext}}_\alpha})\}. \tag{7.89}$$

The system $\{\mathcal{V}^{(i)}_j(\overline{\Omega^{\text{ext}}_\alpha})\}$ of scale spaces forms a multiresolution analysis.

Theorem 7.32. *The restriction of the operator $\lambda^{(i)} : \mathcal{H}_s(\overline{\Omega^{\text{ext}}_\sigma}) \to h^{(i)}_s(\overline{\Omega^{\text{ext}}_\gamma})$ to a scale space $\mathcal{V}_j(\overline{\Omega^{\text{ext}}_\sigma})$, $j \in \mathbb{N}_0$, i.e.,*

$$\lambda^{(i)}|_{\mathcal{V}_j(\overline{\Omega^{\text{ext}}_\sigma})} : \mathcal{V}_j(\overline{\Omega^{\text{ext}}_\sigma}) \to \mathcal{V}^{(i)}_j(\overline{\Omega^{\text{ext}}_\gamma}), \tag{7.90}$$

is injective for $i = 1$, whereas in the case of $i \in \{2,3\}$ the Fourier coefficient of degree 0 cannot be recovered and the Fourier coefficients of degree $n \geq 1$ are uniquely defined. Moreover, we have the following results:

(i) *If the families $\{\{\varphi^{(i)}_j(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$, $i \in \{1,2,3\}$, and $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ are bandlimited (for example, $\varphi^{(i)}_j(n) = \varphi_j(n) = 0$ for all $n \geq 2^j$), then the restricted operator is even bijective (in the sense described above). To be more specific, for $g^{(i)} \in h^{(i)}_s(\overline{\Omega^{\text{ext}}_\gamma})$ the (in the case of $i = 2,3$ up to Fourier coefficients of degree 0) unique solution $F_j \in \mathcal{V}_j(\overline{\Omega^{\text{ext}}_\sigma})$, $j \in \mathbb{N}_0$, of the equation*

$$\lambda^{(i)} F_j = \hat{\Phi}^{(i),\gamma,\gamma}_j \star \hat{\Phi}^{(i),\gamma,\gamma}_j * g^{(i)} \tag{7.91}$$

is given by

$$F_j = \Phi^{\sigma,\sigma}_j * \Phi^{\sigma,\sigma}_j * Q, \tag{7.92}$$

where $Q \in \mathcal{H}_s(\overline{\Omega^{\text{ext}}_\sigma})$ is obtainable by

$$Q^\wedge(n,m) = \begin{cases} \frac{g^{(i)\wedge}(n,m)}{\lambda^{(i)\wedge}(n)}, & n \in [0_i, 2^j), \\ 0, & n \in [2^j, \infty). \end{cases} \tag{7.93}$$

(ii) *If the families $\{\{\varphi^{(i)}_j(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$, $i \in \{1,2,3\}$, and $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ are non-bandlimited, the equation*

$$\lambda^{(i)} F_j = \hat{\Phi}^{(i),\gamma,\gamma}_j \star \hat{\Phi}^{(i),\gamma,\gamma}_j * g^{(i)} \tag{7.94}$$

has a solution $F_j \in \mathcal{V}_j(\overline{\Omega^{\text{ext}}_\sigma})$ provided that $g^{(i)} \in h^{(i)\Lambda}_s(\overline{\Omega^{\text{ext}}_\gamma})$, where $h^{(i)\Lambda}_s(\overline{\Omega^{\text{ext}}_\gamma})$ is a suitable Sobolev space (see the Ph.D.-thesis [58] for more details). In this

case, the (in the case of $i = 2, 3$ up to Fourier coefficients of degree 0) unique solution is given by

$$F_j = \Phi_j^{\sigma,\sigma} * \Phi_j^{\sigma,\sigma} * Q, \tag{7.95}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ is obtainable in spectral language by

$$Q^\wedge(n,m) = \frac{g^{(i)\wedge}(n,m)}{\lambda^{(i)\wedge}(n)}, \tag{7.96}$$

$n = 0_i, \ldots; m = 1, \ldots, 2n+1$.

The following corollary shows that in the case of general operators $\lambda = \sum_{i=1}^3 \lambda^{(i)}$ we have to claim an additional assumption onto the function g.

Corollary 7.33. *The restriction of the operator $\lambda = \sum_{i=1}^3 \lambda^{(i)}$ to a scale space $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, i.e.,*

$$\lambda|_{\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})} : \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) \to \bigoplus_{i=1}^3 \mathcal{V}_j^{(i)}(\overline{\Omega_\gamma^{\text{ext}}}) \tag{7.97}$$

has, in general, no solution. Under the assumption $\varphi_j^{(i)}(n) = \varphi_j(n)$, $i \in \{1,2,3\}$, we have to claim, in addition, that

$$\frac{g^{(i)\wedge}(n,m)}{\lambda^{(i)\wedge}(n)} = \frac{g^{(l)\wedge}(n,m)}{\lambda^{(l)\wedge}(n)}, \tag{7.98}$$

with $i, l \in \{1,2,3\}$; $n = \max\limits_{i,l\in\{1,2,3\}} (0_i, 0_l), \ldots; m = 1, \ldots, 2n+1$.

Then the results in Theorem 7.32 can directly be transferred.

Note that according to Theorem 7.32 the restriction of a pseudodifferential operator of kind i to a scale space $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ is injective. Therefore, in the case of a pseudodifferential operator $\lambda = \sum_{i=1}^3 \lambda^{(i)}$ each pseudodifferential operator $\lambda^{(i)}$ leads to a unique solution. The additional assumption (7.98) is thus necessary, in order to guarantee that the pseudodifferential operators of kind i do not lead to different solutions.

With the help of the refinement equation (7.81) we now define the primal wavelets $\{\hat{\Psi}_j^{(i),\alpha,\alpha}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ and the dual wavelets $\{\tilde{\hat{\Psi}}_j^{(i),\alpha,\alpha}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ for $\alpha \geq \sigma$, $i \in \{1,2,3\}$:

$$\hat{\Psi}_j^{(i),\alpha,\alpha}(x,y) = \sum_{n=0_i}^{\infty} \psi_j^{(i)}(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\alpha;x) h_{n,m}^{(i)s}(\alpha;y), \tag{7.99}$$

$$\tilde{\hat{\Psi}}_j^{(i),\alpha,\alpha}(x,y) = \sum_{n=0_i}^{\infty} \tilde{\psi}_j^{(i)}(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\alpha;x) h_{n,m}^{(i)s}(\alpha;y), \tag{7.100}$$

where

$$(\hat{\Psi}_j^{(i),\alpha,\alpha})^\wedge(n) = \psi_j^{(i)}(n), \quad (\tilde{\hat{\Psi}}_j^{(i),\alpha,\alpha})^\wedge(n) = \tilde{\psi}_j^{(i)}(n). \tag{7.101}$$

The detail spaces are defined in canonical manner:

$$\mathcal{W}_j^{(i)}(\overline{\Omega_\alpha^{\text{ext}}}) = \{\hat{\Psi}_j^{(i),\alpha,\alpha} \star \tilde{\hat{\Psi}}_j^{(i),\alpha,\alpha} * f \ : \ f \in h_s(\overline{\Omega_\alpha^{\text{ext}}})\}. \tag{7.102}$$

Theorem 7.31 can be directly transferred by substituting the convolutions with respect to the sphere Ω_σ by the corresponding convolutions with respect to the sphere Ω_α. We now transfer Theorem 7.32 to the case of the detail spaces and get the following theorem, where we use the terms injectivity, bijectivity, and uniqueness in the same sense as before (i.e., up to Fourier coefficients of degree 0 in the case of $i = 2, 3$).

Theorem 7.34. *The restriction of the operator* $\lambda^{(i)} : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}})$ *to a detail space* $\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, *i.e.,*

$$\lambda^{(i)}|_{\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})} : \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{W}_j^{(i)}(\overline{\Omega_\gamma^{\text{ext}}}) \tag{7.103}$$

with $\psi_j(n) = \psi_j^{(i)}(n)$ *is injective. Moreover, we have the following results:*

(i) *If the families* $\{\{\varphi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$, $i \in \{1, 2, 3\}$, *and* $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ *are bandlimited (for example,* $\varphi_j^{(i)}(n) = \varphi_j(n) = 0$ *for all* $n \geq 2^j$*), then the restricted operator is even bijective. To be more specific, for* $g^{(i)} \in h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}})$ *the unique solution* $H_j \in \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, *of the equation*

$$\lambda^{(i)} H_j = \tilde{\hat{\Psi}}_j^{(i),\gamma,\gamma} \star \hat{\Psi}_j^{(i),\gamma,\gamma} * g^{(i)} \tag{7.104}$$

is given by

$$H_j = \tilde{\Psi}_j^{\sigma,\sigma} * \Psi_j^{\sigma,\sigma} * Q, \tag{7.105}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ *is obtainable by*

$$Q^\wedge(n,m) = \begin{cases} \frac{g^{(i)\wedge}(n,m)}{\lambda^{(i)\wedge}(n)}, & n \in [0_i, 2^{j+1}), \\ 0, & n \in [2^{j+1}, \infty). \end{cases} \tag{7.106}$$

(ii) *If the families* $\{\{\varphi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$, $i \in \{1, 2, 3\}$, *and* $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ *are non-bandlimited, the equation*

$$\lambda^{(i)} H_j = \tilde{\hat{\Psi}}_j^{(i),\gamma,\gamma} \star \hat{\Psi}_j^{(i),\gamma,\gamma} * g^{(i)} \tag{7.107}$$

has a solution $H_j \in \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$ *provided that the condition*

$$\sum_{n=0_i}^{\infty} \sum_{m=1}^{2n+1} \frac{g^{(i)\wedge}(n,m)}{\lambda^{(i)\wedge}(n)} < \infty \tag{7.108}$$

is satisfied for $g^{(i)} \in h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}})$. *In this case, the unique solution of the equation is given by*

$$H_j = \tilde{\Psi}_j^{\sigma,\sigma} * \Psi_j^{\sigma,\sigma} * Q, \tag{7.109}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ *is obtainable by*

$$Q^\wedge(n,m) = \frac{g^{(i)\wedge}(n,m)}{\lambda^{(i)\wedge}(n)}, \tag{7.110}$$

$n = 0_i, \ldots;\ m = 1, \ldots, 2n+1.$

Corollary 7.35. *The restriction of the operator* $\lambda = \sum_{i=1}^{3} \lambda^{(i)}$ *to a detail space* $\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, *i.e.,*

$$\lambda|_{\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})} : \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}}) \to \bigoplus_{i=1}^{3} \mathcal{W}_j^{(i)}(\overline{\Omega_\gamma^{\text{ext}}}) \tag{7.111}$$

has, in general, no solution. Under the assumption $\psi_j^{(i)}(n) = \psi_j(n)$ *and* $\tilde{\psi}_j^{(i)}(n) = \tilde{\psi}_j(n)$, $i \in \{1,2,3\}$, *we have to claim, in addition, that*

$$\frac{g^{(i)\wedge}(n,m)}{\lambda^{(i)\wedge}(n)} = \frac{g^{(l)\wedge}(n,m)}{\lambda^{(l)\wedge}(n)}, \tag{7.112}$$

with $i,l \in \{1,2,3\}$; $n = \max\limits_{i,l\in\{1,2,3\}} (0_i, 0_l), \ldots;\ m = 1, \ldots, 2n+1.$

Then the results in Theorem 7.34 *can be directly transferred.*

Up to now, we have summarized some results about the filtered solution, i.e., the solution when we restrict the operator to scale or detail spaces. In this case, we have injectivity (in the case of $i = 2,3$ up to Fourier coefficients of degree 0) for the operators $\lambda^{(i)}$, whereas in the case of general operators $\lambda = \sum_{i=1}^{3} \lambda^{(i)}$ we have to claim that (7.98) is valid. In the case of the unfiltered solution, we obtain the following theorem.

Theorem 7.36. *Let* $g^{(i)} \in h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}})$ *satisfy the condition* $g^{(i)} \in \text{im}(\lambda^{(i)})$, $i \in \{1,2,3\}$. *Then the unique solution* $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ (*in the case of* $i = 2,3$ *up to Fourier coefficients of degree* 0) *of the equation* $\lambda^{(i)}F = g^{(i)}$ *is given by*

$$F^\wedge(n,m) = \frac{g^{(i)\wedge}(n,m)}{\lambda^{(i)\wedge}(n)}, \tag{7.113}$$

$n = 0_i, \ldots;\ m = 1, \ldots, 2n+1$. *In the case of the operator* $\lambda = \sum_{i=1}^{3} \lambda^{(i)}$ *we have to claim, in addition, that* (7.112) *holds in order to guarantee the solvability.*

Last, we explain the connection between the solution in the scale spaces and the unfiltered solution.

Theorem 7.37. *Suppose that* $g^{(i)}$ *is of the class* $h_s^{(i)\Lambda}(\overline{\Omega_\gamma^{\text{ext}}})$. *Let* $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ *be the unique* (*in the case of* $i = 2,3$ *up to Fourier coefficients of degree* 0) *solution of* $\lambda^{(i)}F = g^{(i)}$. *Then*

$$F_j = (\Phi_j^{(2)})^{\sigma,\sigma} * F \tag{7.114}$$

is the unique solution in $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ *of the equation*

$$\lambda^{(i)}F_j = \hat{\Phi}_j^{(i),\gamma,\gamma} \star \hat{\Phi}_j^{(i),\gamma,\gamma} * g^{(i)} \tag{7.115}$$

for every $j \in \mathbb{N}_0$. *Furthermore, the limit relation*

$$\lim_{J\to\infty} (\Phi_J^{(2)})^{\sigma,\sigma} * F = F \tag{7.116}$$

holds (*in* $\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}$-*sense*).

The preceding theorem shows that in the case of bandlimited scaling functions the (vectorial) SST or SGG problem is well posed, because a unique solution always exists and due to the finite dimension of the scale spaces the solution is also stable. We now investigate the case of non-bandlimited scaling functions and it turns out that the stability cannot be ensured. The reason is that the (vectorial) SST or SGG problem is an exponentially ill-posed problem with unbounded inverse operator λ^{-1}. Therefore, we have to turn to regularization methods and replace the inverse operator by an appropriate bounded operator.

Definition 7.38. A family of linear operators $S_j^{(i)} : h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, is called a *regularization* of $(\lambda^{(i)})^{-1}$, $i \in \{1,2,3\}$, if it satisfies the following properties:

(i) $S_j^{(i)}$ is bounded on $h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}})$ for all $j \in \mathbb{N}_0$,

(ii) for any member $g^{(i)} \in im(\lambda^{(i)})$, the limit relation

$$\lim_{J\to\infty} S_J^{(i)} g^{(i)} = (\lambda^{(i)})^{-1} g^{(i)} \tag{7.117}$$

holds (in $\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}$-sense).

The operator $S : h_s(\overline{\Omega_\gamma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ given by $S|_{h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}})} = S_j^{(i)}$ is called a regularization of λ^{-1}.

The function $F_J = S_J g$ is called the *J-level regularization of* $\lambda^{-1} g$, whereas $F_J^{(i)} = S_J^{(i)} g^{(i)}$ is called the *J-level regularization of* $(\lambda^{(i)})^{-1} g$. Within our multiscale approach, we now represent the $(J+1)$-level regularization using the J-level regularization by adding the corresponding detail information. To this end, we first introduce a multiscale regularization concept starting with the definition of a generator of a regularization scaling function.

Definition 7.39. A family $\{\{\varphi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$ of sequences $\{\varphi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}$, $i \in \{1,2,3\}$, is called a *generator of a regularization scaling function with respect to* $(\lambda^{(i)})^{-1}$, if it satisfies the following requirements:

(i) $(\varphi_j^{(i)}(0_i))^2 = \frac{1}{\lambda^{(i)\wedge}(0_i)}$ for all $j \in \mathbb{N}_0$,

(ii) $(\varphi_j^{(i)}(n))^2 \le (\varphi_{j'}^{(i)}(n))^2$ for all $j, j' \in \mathbb{N}_0$ with $j \le j'$ and all $n \in \mathbb{N}_{0_i+1}$,

(iii) $\lim\limits_{j\to\infty} (\varphi_j^{(i)}(n))^2 = \frac{1}{\lambda^{(i)\wedge}(n)}$ for all $n \in \mathbb{N}_{0_i+1}$.

Furthermore, $\{\{\{\varphi_j^{(i)}(n)\}_{i\in\{1,2,3\}}\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$ is called a *generator of a regularization scaling function with respect to* λ^{-1}, if $(\lambda^{(i)})^{-1}$ is a generator of a regularization scaling function with respect to $(\lambda^{(i)})^{-1}$ for every $i = 1,2,3$.

We now define decomposition and reconstruction regularization scaling functions.

Definition 7.40. Let $\{\{\varphi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$ be a generator of a regularization scaling function with respect to $(\lambda^{(i)})^{-1}$. Then a family $\{{}^d\hat{\Phi}_j^{(i),\sigma,\gamma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of admissible $h_{\sigma,\gamma}^{(i)}$-kernels given by

$$ {}^d\hat{\Phi}_j^{(i),\sigma,\gamma}(x,z) = \sum_{n=0_i}^{\infty} \varphi_j^{(i)}(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\sigma;x) h_{n,m}^{(i)s}(\gamma;z), \tag{7.118} $$

$(x,z) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\gamma^{\text{ext}}}$, is called a *decomposition regularization $h_{\sigma,\gamma}^{(i)}$-scaling function with respect to* $(\lambda^{(i)})^{-1}$, whereas a family $\{{}^r\hat{\Phi}_j^{(i),\sigma,\sigma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of admissible $h_{\sigma,\sigma}^{(i)}$-kernels given by

$$ {}^r\hat{\Phi}_j^{(i),\sigma,\sigma}(x,y) = \sum_{n=0_i}^{\infty} \varphi_j^{(i)}(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\sigma;x) h_{n,m}^{(i)s}(\sigma;y), \tag{7.119} $$

$(x,y) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\sigma^{\text{ext}}}$, is called a *reconstruction regularization $h^{(i),\sigma,\sigma}$-scaling function with respect to* $(\lambda^{(i)})^{-1}$.

We obtain the following theorem:

Theorem 7.41. *Let $\{\{\varphi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$ be a generator of a regularization scaling function with respect to $(\lambda^{(i)})^{-1}$, $i \in \{1,2,3\}$. If we formally define*

$$ ({}^r\hat{\Phi}_j^{(i)} \star {}^d\hat{\Phi}_j^{(i)})^{\sigma,\gamma}(\cdot,\cdot) $$

by

$$ ({}^r\hat{\Phi}_j^{(i)} \star {}^d\hat{\Phi}_j^{(i)})^{\sigma,\gamma}(x,z) = {}^r\hat{\Phi}_j^{(i),\sigma,\sigma}(x,\cdot) \star {}^d\hat{\Phi}_j^{(i),\sigma,\gamma}(\cdot,z), \tag{7.120} $$

$(x,z) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\gamma^{\text{ext}}}$, *then*

$$ F_J^{(i)} = ({}^r\hat{\Phi}_J^{(i)} \star {}^d\hat{\Phi}_J^{(i)})^{\sigma,\gamma} * g^{(i)}, \quad g^{(i)} \in h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}}), \tag{7.121} $$

represents the J-level regularization of $(\lambda^{(i)})^{-1} g^{(i)}$. If, in addition, $g^{(i)} \in \text{im}(\lambda^{(i)})$, then

$$ \lim_{J\to\infty} \|F_J^{(i)} - (\lambda^{(i)})^{-1} g^{(i)}\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} = 0. \tag{7.122} $$

Furthermore,

$$ F_J = \sum_{i=1}^{3} ({}^r\hat{\Phi}_J^{(i)} \star {}^d\hat{\Phi}_J^{(i)})^{\sigma,\gamma} * g^{(i)}, \quad g = \sum_{i=1}^{3} g^{(i)} \in h_s(\overline{\Omega_\gamma^{\text{ext}}}), \tag{7.123} $$

represents the J-level regularization of $\lambda^{-1} g$. If, in addition, $g \in \text{im}(\lambda)$, then

$$ \lim_{J\to\infty} \|F_J - \lambda^{-1} g\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} = 0. \tag{7.124} $$

We now define the convolution operators $S_J^{(i)} : h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $J \in \mathbb{N}_0$, by

$$S_J^{(i)}(g^{(i)}) = ({}^r\hat{\Phi}_J^{(i)} \star {}^d\hat{\Phi}_J^{(i)})^{\sigma,\gamma} * g^{(i)}, \tag{7.125}$$

whereas the convolution operator $S_J : h_s(\overline{\Omega_\gamma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $J \in \mathbb{N}_0$, is given by

$$S_J(g) = \sum_{i=1}^{3} S_J^{(i)}(g^{(i)}). \tag{7.126}$$

Furthermore, we introduce the corresponding scale spaces $S_J^{(i)}(\text{im}(\lambda^{(i)}))$, $i \in \{1,2,3\}$, and $S_J(\text{im}(\lambda))$ as follows

$$S_J^{(i)}(\text{im}(\lambda^{(i)})) = \left\{ ({}^r\hat{\Phi}_J^{(i)} \star {}^d\hat{\Phi}_J^{(i)})^{\sigma,\gamma} * g^{(i)} \ : \ g^{(i)} \in \text{im}(\lambda^{(i)}) \right\}, \tag{7.127}$$

$$S_J(\text{im}(\lambda)) = \left\{ \sum_{i=1}^{3} ({}^r\hat{\Phi}_J^{(i)} \star {}^d\hat{\Phi}_J^{(i)})^{\sigma,\gamma} * g^{(i)} \ : \ g = \sum_{i=1}^{3} g^{(i)} \in \text{im}(\lambda) \right\}. \tag{7.128}$$

Theorem 7.42. *The scale spaces satisfy the following properties:*

(i) $S_0^{(i)}(\text{im}(\lambda^{(i)})) \subset \cdots \subset S_J^{(i)}(\text{im}(\lambda^{(i)})) \subset S_{J'}^{(i)}(\text{im}(\lambda^{(i)})) \subset \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $J \le J'$, *i.e., for any right-hand side* $g^{(i)} \in \text{im}(\lambda^{(i)})$ *of the* (*vectorial*) *SST or SGG problem, all J-level regularizations with fixed parameter J are sampled in a scale space* $S_J^{(i)}(\text{im}(\lambda^{(i)}))$ *with the above property,*

(ii) $\overline{\bigcup_{J=0}^{\infty} S_J^{(i)}(\text{im}(\lambda^{(i)}))}^{\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}} = \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$.

Obviously, Theorem 7.42 is also valid substituting $S_J^{(i)}$ by S_J which leads to the following corollary.

Corollary 7.43. *The scale spaces satisfy the following properties:*

(i) $S_0(\text{im}(\lambda)) \subset \cdots \subset S_J(\text{im}(\lambda)) \subset S_{J'}(\text{im}(\lambda)) \subset \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $J \le J'$, *i.e., for any right-hand side* $g \in \text{im}(\lambda)$ *of the* (*vectorial*) *SST or SGG problem, all J-level regularizations with fixed parameter J are sampled in a scale space* $S_J(\text{im}(\lambda))$ *with the above property,*

(ii) $\overline{\bigcup_{J=0}^{\infty} S_J(\text{im}(\lambda))}^{\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}} = \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$.

A set of subspaces of $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ satisfying the conditions of Corollary 7.43 is called *regularization* $h_{\sigma,\gamma}$*-multiresolution analysis* (*RMRA*) *of the* (*vectorial*) *SST or SGG problem.*

Definition 7.44. Let $\{\{\varphi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$ be a generator of a regularization scaling function with respect to $(\lambda^{(i)})^{-1}$. Then the generating symbols $\{\psi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}$, $\{\tilde{\psi}_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}$ of the corresponding regularization wavelets are defined by the refinement equation (7.81). The admissible $h_{\sigma,\gamma}^{(i)}$-kernels $\{{}^d\hat{\Psi}_j^{(i),\sigma,\gamma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by

$${}^d\hat{\Psi}_j^{(i),\sigma,\gamma}(x,z) = \sum_{n=0_i}^{\infty} \psi_j^{(i)}(n) \sum_{m=1}^{2n+1} H_{n,m}^{s}(\sigma;x) h_{n,m}^{(i)s}(\gamma;z), \tag{7.129}$$

$(x,z) \in \overline{\Omega_\sigma^{\rm ext}} \times \overline{\Omega_\gamma^{\rm ext}}$, are called the *decomposition regularization* $h_{\sigma,\gamma}^{(i)}$*-wavelets*, while the admissible $h_{\sigma,\sigma}^{(i)}$-kernels $\{{}^r\tilde{\hat{\Psi}}_j^{(i),\sigma,\sigma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by

$$ {}^r\tilde{\hat{\Psi}}_j^{(i),\sigma,\sigma}(x,y) = \sum_{n=0_i}^{\infty} \tilde{\psi}_j^{(i)}(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\sigma;x) h_{n,m}^{(i)s}(\sigma;y), \tag{7.130} $$

$(x,y) \in \overline{\Omega_\sigma^{\rm ext}} \times \overline{\Omega_\sigma^{\rm ext}}$ are called the *reconstruction regularization* $h_{\sigma,\sigma}^{(i)}$*-wavelets*.

We now define the convolution operators $T_j^{(i)} : h_s^{(i)}(\overline{\Omega_\gamma^{\rm ext}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$, $j \in \mathbb{N}_0$, $i = 1,2,3$, by

$$ T_j^{(i)}(g^{(i)}) = ({}^r\tilde{\hat{\Psi}}_j^{(i)} \star {}^d\hat{\Psi}_j^{(i)})^{\sigma,\gamma} * g^{(i)}, \quad g^{(i)} \in h_s^{(i)}(\overline{\Omega_\gamma^{\rm ext}}), \tag{7.131} $$

and the convolution operator $T_j : h_s(\overline{\Omega_\gamma^{\rm ext}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\rm ext}})$, $j \in \mathbb{N}_0$, by

$$ T_J(g) = \sum_{i=1}^{3} T_J^{(i)}(g^{(i)}). \tag{7.132} $$

Obviously, due to the refinement equation, the operators $S_{J+1}^{(i)}$ and S_{J+1} can be represented in the form

$$ S_{J+1}^{(i)} = S_0^{(i)} + \sum_{j=0}^{J} T_j^{(i)}, \tag{7.133} $$

$$ S_{J+1} = S_0 + \sum_{j=0}^{J} T_j. \tag{7.134} $$

Thus, we now introduce the *detail spaces* $T_J^{(i)}(\mathrm{im}(\lambda^{(i)}))$ and $T_J(\mathrm{im}(\lambda))$ by

$$ T_J^{(i)}(\mathrm{im}(\lambda^{(i)})) = \left\{ ({}^r\tilde{\hat{\Psi}}_J^{(i)} \star {}^d\hat{\Psi}_J^{(i)})^{\sigma,\gamma} * g^{(i)} \ : \ g^{(i)} \in \mathrm{im}(\lambda^{(i)}) \right\}, \tag{7.135} $$

$$ T_J(\mathrm{im}(\lambda)) = \left\{ \sum_{i=1}^{3} ({}^r\tilde{\hat{\Psi}}_J^{(i)} \star {}^d\hat{\Psi}_J^{(i)})^{\sigma,\gamma} * g^{(i)} \ : \ g = \sum_{i=1}^{3} g^{(i)} \in \mathrm{im}(\lambda) \right\}. \tag{7.136} $$

In terms of the multiscale concept, the space $T_J(\mathrm{im}(\lambda))$ contains the detail information which has to be added in order to turn from the J-level regularization to the $(J+1)$-level regularization:

$$ S_{J+1}(\mathrm{im}(\lambda)) = S_J(\mathrm{im}(\lambda)) + T_J(\mathrm{im}(\lambda)). \tag{7.137} $$

In general, the sum is neither direct nor orthogonal.

Theorem 7.45. *Let* $\{\{\varphi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$ *be a generator of a regularization scaling function with respect to* $(\lambda^{(i)})^{-1}$, $i \in \{1,2,3\}$. *Suppose that* $\{\{\psi_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$,

$\{\{\tilde{\psi}_j^{(i)}(n)\}_{n\in\mathbb{N}_{0_i}}\}_{j\in\mathbb{N}_0}$ *are the generating symbols of the corresponding regularization wavelets. Furthermore, let* $g^{(i)}$ *be of class* $h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}})$. *Define the* regularization $h_{\sigma,\gamma}^{(i)}$-wavelet transform at scale $j\in\mathbb{N}_0$ and position $x\in\overline{\Omega_\sigma^{\text{ext}}}$ *by*

$$(RWT)(g^{(i)})(j;x) = {}^d\hat{\Psi}_j^{(i),\sigma,\gamma}(x,\cdot) * g^{(i)}, \quad g^{(i)} \in h_s^{(i)}(\overline{\Omega_\gamma^{\text{ext}}}). \tag{7.138}$$

Then

$$F_J = ({}^r\hat{\Phi}_0^{(i)} \star {}^d\hat{\Phi}_0^{(i)})^{\sigma,\gamma} * h^{(i)} + \sum_{j=0}^{J-1} {}^r\tilde{\hat{\Psi}}_j^{(i),\sigma,\sigma} \star (RWT)(g^{(i)})(j;\cdot)$$

is the J-level regularization of the (vectorial) SST or SGG problem satisfying

$$\lim_{J\to\infty} \|F_J - (\lambda^{(i)})^{-1} g^{(i)}\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})} = 0 \tag{7.139}$$

provided that $g^{(i)} \in \text{im}(\lambda^{(i)})$.

7.3. Tensorial wavelet theory

The extension from vector to tensor theory is performed in this section. First, we define tensorial scaling functions and wavelets and give the reconstruction formula. The solution of the tensorial SGG problem is presented using regularization wavelets.

Definition 7.46. Let $\{\mathbf{\Phi}_j^{(i,k)}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$, $i,k\in\{1,2,3\}$, be a family of admissible $\mathbf{h}_{\sigma,\sigma}^{(i,k)}$-kernels. Then the family $\{\mathcal{V}_j^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})\}_{j\in\mathbb{N}_0}$ of scale spaces $\mathcal{V}_j^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$ defined by

$$\mathcal{V}_j^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}) = \{\mathbf{\Phi}_j^{(i,k)} \star \mathbf{\Phi}_j^{(i,k)} * \mathbf{f} \ : \ \mathbf{f}\in\mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})\}, \tag{7.140}$$

is called an $\mathbf{h}_{\sigma,\sigma}^{(i,k)}$*-multiresolution analysis*, if the following properties are satisfied:

(i) $\mathcal{V}_0^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}) \subset \cdots \subset \mathcal{V}_j^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}) \subset \mathcal{V}_{j+1}^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}) \subset \cdots \subset \mathbf{h}_s^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$,

(ii) $\overline{\bigcup_{j\in\mathbb{N}_0} \mathcal{V}_j^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})}^{\|\cdot\|_{\mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})}} = \mathbf{h}_s^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$.

Definition 7.47. Let $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be a family of admissible $\mathbf{h}_{\sigma,\sigma}$-kernels. The set of scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ defined by

$$\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) = \{\mathbf{\Phi}_j \star \mathbf{\Phi}_j * \mathbf{f} \ : \ \mathbf{f}\in\mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})\} \tag{7.141}$$

is called an $\mathbf{h}_{\sigma,\sigma}$-multiresolution analysis, if $\{\mathcal{V}_j^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})\}_{j\in\mathbb{N}_0}$ is an $\mathbf{h}_s^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$-multiresolution analysis for every $i,k\in\{1,2,3\}$.

We now define the scaling functions.

Definition 7.48. A family $\{\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ of sequences $\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}$ is called a *generator of a scaling function of kind* (i,k), $i,k\in\{1,2,3\}$, if it satisfies the following requirements:

(i) $(\varphi_j^{(i,k)}(0_{ik}))^2 = 1$, for all $j \in \mathbb{N}_0$,

(ii) $(\varphi_j^{(i,k)}(n))^2 \leq \left(\varphi_{j'}^{(i,k)}(n)\right)^2$, for all $j, j' \in \mathbb{N}_0$ with $j \leq j'$ and all $n \in \mathbb{N}_{0_{ik}+1}$,

(iii) $\lim_{j\to\infty} (\varphi_j^{(i,k)}(n))^2 = 1$, for all $n \in \mathbb{N}_{0_{ik}+1}$.

Furthermore, the family $\{\{\{\varphi^{(i,k)}(n)\}_{i,k\in\{1,2,3\}}\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ is called a generator of a scaling function, if $\{\{\varphi^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ are generators of a scaling function of kind (i,k), $i,k \in \{1,2,3\}$.

Based on the definition of a generator of a scaling function, we now introduce $\mathbf{h}_{\sigma,\sigma}$-scaling functions.

Definition 7.49. A family $\{\mathbf{\Phi}_j^{(i)}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of $\mathbf{h}^{(i,k)}$-kernels $\mathbf{\Phi}_j^{(i,k)}(\cdot,\cdot)$ defined by $\mathbf{\Phi}_j^{(i,k)\wedge}(n) = \varphi_j^{(i,k)}(n)$, $j \in \mathbb{N}_0$, $n \in \mathbb{N}_{0_{ik}}$, i.e.,

$$\mathbf{\Phi}_j^{(i,k)}(x,y) = \sum_{n=0_{ik}}^{\infty} \varphi_j^{(i,k)}(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\sigma;y)\mathbf{h}_{n,m}^{(i,k)s}(\sigma;x), \quad x,y \in \overline{\Omega_\sigma^{\text{ext}}}, \quad (7.142)$$

is called an $\mathbf{h}_{\sigma,\sigma}^{(i,k)}$-*scaling function*, if it satisfies the following properties:

(i) $\mathbf{\Phi}_j^{(i,k)}(\cdot,\cdot)$ is an admissible $\mathbf{h}_{\sigma,\sigma}^{(i,k)}$-kernel for every $j \in \mathbb{N}_0$,

(ii) $\{\{\mathbf{\Phi}_j^{(i,k)\wedge}(n)_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ constitutes a generator of a scaling function of kind (i,k).

Furthermore, the family $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of $\mathbf{h}_{\sigma,\sigma}$-kernels $\mathbf{\Phi}_j(\cdot,\cdot)$ is called an $\mathbf{h}_{\sigma,\sigma}$-scaling function, if $\{\mathbf{\Phi}_j^{(i,k)}\}_{j\in\mathbb{N}_0}$ are $\mathbf{h}_{\sigma,\sigma}^{(i,k)}$-scaling functions for $i,k \in \{1,2,3\}$.

As in the scalar and vectorial theory, the following approximation theorem is valid.

Theorem 7.50. *Let $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathbf{h}_{\sigma,\sigma}$-scaling function. Then*

$$\lim_{j\to\infty} \|\mathbf{f} - \mathbf{\Phi}_j \star \mathbf{\Phi}_j * \mathbf{f}\|_{\mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})} = 0 \quad (7.143)$$

holds for all $\mathbf{f} \in \mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})$.

Definition 7.51. Let $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathbf{h}_{\sigma,\sigma}$-scaling function. Then $\{P_j\}_{j\in\mathbb{N}_0}$ with $P_j : \mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})$ defined by

$$P_j(\mathbf{f})(x) = \mathbf{\Phi}_j \star \mathbf{\Phi}_j * \mathbf{f}, \quad \mathbf{f} \in \mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}}), \quad x \in \overline{\Omega_\sigma^{\text{ext}}}, \quad (7.144)$$

is called an $\mathbf{h}_{\sigma,\sigma}$-*approximate identity.*

The kernel $\mathbf{\Phi}_0$ is called the mother kernel of the $\mathbf{h}_{\sigma,\sigma}$-scaling function. We obtain the following theorem.

Theorem 7.52. *Let $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathbf{h}_{\sigma,\sigma}$-scaling function. Then $\{\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})\}_{j\in\mathbb{N}_0}$ given in* (7.141) *forms an $\mathbf{h}_{\sigma,\sigma}$-multiresolution analysis.*

The next purpose is to define the primal and dual wavelet with the help of the tensorial refinement equation.

Definition 7.53. Let $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathbf{h}_{\sigma,\sigma}$-scaling function. Then the families of $\mathbf{h}_{\sigma,\sigma}$-kernels $\{\mathbf{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$, $\{\tilde{\mathbf{\Psi}}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by

$$(\mathbf{\Psi}_j)^{(i,k)\wedge}(n) = \psi_j^{(i,k)}(n), \quad j\in\mathbb{N}_0,\ n\in\mathbb{N}_{0_{ik}},\ i,k\in\{1,2,3\}, \tag{7.145}$$

$$(\tilde{\mathbf{\Psi}}_j)^{(i,k)\wedge}(n) = \tilde{\psi}_j^{(i,k)}(n), \quad j\in\mathbb{N}_0,\ n\in\mathbb{N}_{0_{ik}},\ i,k\in\{1,2,3\}, \tag{7.146}$$

are called (*primal*) $\mathbf{h}_{\sigma,\sigma}$-*wavelet* and *dual* $\mathbf{h}_{\sigma,\sigma}$-*wavelet*, respectively, if all $\mathbf{h}_{\sigma,\sigma}$-kernels $\mathbf{\Psi}_j(\cdot,\cdot)$, $\tilde{\mathbf{\Psi}}_j(\cdot,\cdot)$, $j\in\mathbb{N}_0$, are admissible and the symbols $\{\psi_j^{(i,k)}(n)\}$, $\{\tilde{\psi}_j^{(i,k)}(n)\}$, in addition, satisfy the (tensorial) refinement equation

$$\tilde{\psi}_j^{(i,k)}(n)\psi_j^{(i,k)}(n) = (\varphi_{j+1}^{(i,k)}(n))^2 - (\varphi_j(n)^{(i,k)})^2 \tag{7.147}$$

for all $j\in\mathbb{N}_0$, $n\in\mathbb{N}_{0_{ik}}$, $i,k\in\{1,2,3\}$.

As a direct consequence we get the following equation:

$$(\varphi_{J+1}^{(i,k)}(n))^2 = (\varphi_0^{(i,k)}(n))^2 + \sum_{j=0}^{J}\tilde{\psi}_j^{(i,k)}(n)\psi_j^{(i,k)}(n), \quad J\in\mathbb{N}_0, \tag{7.148}$$

for all $n\in\mathbb{N}_{0_{ik}}$. We now define the wavelet transform. To this end we let $\psi_{-1}^{(i,k)}(n) = \tilde{\psi}_{-1}^{(i,k)}(n) = \varphi_0^{(i,k)}(n)$, for $n\in\mathbb{N}_{0_{ik}}$, $i,k\in\{1,2,3\}$, $\mathbf{\Psi}_{-1}(\cdot,\cdot) = \tilde{\mathbf{\Psi}}_{-1}(\cdot,\cdot) = \mathbf{\Phi}_0(\cdot,\cdot)$. We remember the space $\mathcal{H}(\mathbb{N}_{-1}\times\overline{\Omega_\sigma^{\text{ext}}})$ (see Eqs. (7.12)–(7.14)).

Definition 7.54. Let $\{\mathbf{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_{-1}}$ be a (primal) $\mathbf{h}_{\sigma,\sigma}$-wavelet. Then $(WT)^{(i,k)}: \mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{H}_s(\mathbb{N}_{-1}\times\overline{\Omega_\sigma^{\text{ext}}})$ defined by

$$(WT)^{(i,k)}(\mathbf{f})(j;y) = (\mathbf{\Psi}_j^{(i,k)} * \mathbf{f})(y) \tag{7.149}$$

is called $\mathbf{h}_{\sigma,\sigma}$-wavelet transform if kind (i,k) of $\mathbf{f}$ at position $y\in\overline{\Omega_\sigma^{\text{ext}}}$ and scale $j\in\mathbb{N}_{-1}$.

As usual, we define the detail space $\mathcal{W}_j^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}})$ at scale j by

$$\mathcal{W}_j^{(i,k)}(\overline{\Omega_\sigma^{\text{ext}}}) = \{\tilde{\mathbf{\Psi}}_j^{(i,k)} \star \mathbf{\Psi}_j^{(i,k)} * \mathbf{f} \ :\ \mathbf{f}\in\mathbf{h}({}_s\overline{\Omega_\sigma^{\text{ext}}})\}, \tag{7.150}$$

and

$$\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}}) = \{\tilde{\mathbf{\Psi}}_j \star \mathbf{\Psi}_j * \mathbf{f} \ :\ \mathbf{f}\in\mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})\}. \tag{7.151}$$

Theorem 7.55 (Tensorial Reconstruction Formula for the Outer Space). *Let the families* $\{\mathbf{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ *and* $\{\tilde{\mathbf{\Psi}}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$, *respectively, be a* (*primal*) $\mathbf{h}_{\sigma,\sigma}$-*wavelet and its dual corresponding to an* $\mathbf{h}_{\sigma,\sigma}$-*scaling function* $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$. *Then*

$$\mathbf{f} = \sum_{j=-1}^{\infty}\tilde{\mathbf{\Psi}}_j \star \mathbf{\Psi}_j * \mathbf{f} \tag{7.152}$$

holds for all $\mathbf{f}\in\mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})$ (*in* $\|\cdot\|_{\mathbf{h}_s(\overline{\Omega_\sigma^{\text{ext}}})}$-*sense*).

We now solve the (tensorial) SGG problem using regularization wavelets. First, we transfer the theory of $\mathbf{h}_{\sigma,\sigma}^{(i,k)}$-scaling functions to the general case of $\mathbf{h}_{\alpha,\alpha}^{(i,k)}$-scaling functions $\mathbf{\Phi}_j^{(i,k),\alpha,\alpha}$ with $\alpha \geq \sigma$:

$$\mathbf{\Phi}_j^{(i,k),\alpha,\alpha}(x,y) = \sum_{n=0_{ik}}^{\infty} \varphi_j^{(i,k)}(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\alpha;x)\mathbf{h}_{n,m}^{(i,k)s}(\alpha;y), \tag{7.153}$$

where

$$(\mathbf{\Phi}_j^{(i,k),\alpha,\alpha})^\wedge(n) = \varphi_j^{(i,k)}(n). \tag{7.154}$$

Theorem 7.50 can be directly transferred substituting σ by α. The scale spaces are defined in the following way:

$$\mathcal{V}_j^{(i,k)}(\overline{\Omega_\alpha^{\text{ext}}}) = \{\mathbf{\Phi}_j^{(i,k),\alpha,\alpha} \star \mathbf{\Phi}_j^{(i,k),\alpha,\alpha} * \mathbf{f} \ : \ \mathbf{f} \in \mathbf{h}_s(\overline{\Omega_\alpha^{\text{ext}}})\}. \tag{7.155}$$

The system $\{\mathcal{V}_j^{(i,k)}(\overline{\Omega_\alpha^{\text{ext}}})\}$ of scale spaces forms a multiresolution analysis.

Theorem 7.56. *The restriction of the operator $\boldsymbol{\lambda}^{(i,k)} : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}})$ to a scale space $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, i.e.,*

$$\boldsymbol{\lambda}^{(i,k)}|_{\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})} : \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{V}_j^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}}), \tag{7.156}$$

is injective for $(i,k) \in \{(1,1), (2,1), (3,1)\}$, whereas in the case of $(i,k) \in \{(1,2), (1,3), (2,3), (3,3)\}$ the Fourier coefficient of degree 0 cannot be recovered and the Fourier coefficients of degree $n \geq 1$ are uniquely defined. In the case of $(i,k) \in \{(2,2),(3,2)\}$ the Fourier coefficient of degree 0 and 1 cannot be recovered and the Fourier coefficients of degree $n \geq 2$ are uniquely defined (in the following text, injectivity, bijectivity and uniqueness is always used in this sense).

Moreover, we have the following results:

(i) *If the families $\{\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ and $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$, $i,k\in\{1,2,3\}$, are bandlimited (for example, $\varphi_j^{(i,k)}(n) = \varphi_j(n) = 0$ for all $n \geq 2^j$), then the restricted operator is even bijective (in the sense described above). To be more specific, for $\mathbf{g}^{(i,k)} \in \mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}})$ the unique solution $F_j \in \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, of the equation*

$$\boldsymbol{\lambda}^{(i,k)} F_j = \mathbf{\Phi}_j^{(i,k),\gamma,\gamma} \star \mathbf{\Phi}_j^{(i,k),\gamma,\gamma} * \mathbf{g}^{(i,k)} \tag{7.157}$$

is given by

$$F_j = \Phi_j^{\sigma,\sigma} * \Phi_j^{\sigma,\sigma} * Q, \tag{7.158}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ is obtainable by

$$Q^\wedge(n,m) = \begin{cases} \frac{\mathbf{g}^{(i,k)\wedge}(n,m)}{\boldsymbol{\lambda}^{(i,k)\wedge}(n)}, & n \in [0_{ik}, 2^j), \\ 0, & n \in [2^j, \infty). \end{cases} \tag{7.159}$$

(ii) *If the families* $\{\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$, $i,k\in\{1,2,3\}$, *and* $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ *are non-bandlimited, the equation*

$$\boldsymbol{\lambda}^{(i,k)}F_j = \boldsymbol{\Phi}_j^{(i,k),\gamma,\gamma} \star \boldsymbol{\Phi}_j^{(i,k),\gamma,\gamma} * \mathbf{g}^{(i,k)} \tag{7.160}$$

has a solution $F_j \in \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ *provided that* $\mathbf{g}^{(i,k)} \in \mathbf{h}_s^{(i,k)\Lambda}(\overline{\Omega_\gamma^{\text{ext}}})$, *where* $\mathbf{h}_s^{(i,k)\Lambda}(\overline{\Omega_\gamma^{\text{ext}}})$ *is an appropriate Sobolev space* (*see the Ph.D.-thesis* [58] *for more details*). *In this case, the unique solution of the equation is given by*

$$F_j = \Phi_j^{\sigma,\sigma} * \Phi_j^{\sigma,\sigma} * Q, \tag{7.161}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ *is obtainable in spectral language by*

$$Q^\wedge(n,m) = \frac{\mathbf{g}^{(i,k)\wedge}(n,m)}{\boldsymbol{\lambda}^{(i,k)\wedge}(n)}, \tag{7.162}$$

$n = 0_{ik}, \ldots;\ m = 1, \ldots, 2n+1$.

The following corollary shows that in the case of general operators $\boldsymbol{\lambda} = \sum_{i,k=1}^3 \boldsymbol{\lambda}^{(i,k)}$ we have to claim an additional assumption onto the function $\mathbf{g}$.

Corollary 7.57. *The restriction of the operator* $\boldsymbol{\lambda} = \sum_{i,k=1}^3 \boldsymbol{\lambda}^{(i,k)}$ *to a scale space* $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, *i.e.,*

$$\boldsymbol{\lambda}|_{\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})} : \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) \to \bigoplus_{i,k=1}^3 \mathcal{V}_j^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}}) \tag{7.163}$$

has, in general, no solution. Under the assumption $\varphi_j^{(i,k)}(n) = \varphi_j(n)$, $i,k \in \{1,2,3\}$, *we have to claim, in addition, that*

$$\frac{\mathbf{g}^{(i,k)\wedge}(n,m)}{\boldsymbol{\lambda}^{(i,k)\wedge}(n)} = \frac{\mathbf{g}^{(l,r)\wedge}(n,m)}{\boldsymbol{\lambda}^{(l,r)\wedge}(n)}, \tag{7.164}$$

with $i,k,l,r \in \{1,2,3\}$; $n = \max\limits_{i,k,l,r\in\{1,2,3\}}(0_{ik}, 0_{lr}), \ldots;\ m = 1, \ldots, 2n+1$. *Then the results in Theorem* 7.56 *can be directly transferred.*

With the help of the refinement equation (7.147) we now define the primal wavelets $\{\boldsymbol{\Psi}_j^{(i,k),\alpha,\alpha}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ and the dual wavelets $\{\tilde{\boldsymbol{\Psi}}_j^{(i,k),\alpha,\alpha}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ for $\alpha \geq \sigma$, $i,k \in \{1,2,3\}$:

$$\boldsymbol{\Psi}_j^{(i,k),\alpha,\alpha}(x,y) = \sum_{n=0_{ik}}^{\infty} \psi_j^{(i,k)}(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\alpha;x)\mathbf{h}_{n,m}^{(i,k)s}(\alpha;y), \tag{7.165}$$

$$\tilde{\boldsymbol{\Psi}}_j^{(i,k),\alpha,\alpha}(x,y) = \sum_{n=0_{ik}}^{\infty} \tilde{\psi}_j^{(i,k)}(n) \sum_{m=1}^{2n+1} H_{n,m}^s(\alpha;x)\mathbf{h}_{n,m}^{(i,k)s}(\alpha;y), \tag{7.166}$$

where

$$(\boldsymbol{\Psi}_j^{(i,k),\alpha,\alpha})^\wedge(n) = \psi_j^{(i,k)}(n), \quad (\tilde{\boldsymbol{\Psi}}_j^{(i,k),\alpha,\alpha})^\wedge(n) = \tilde{\psi}_j^{(i,k)}(n). \tag{7.167}$$

The detail spaces are defined in canonical manner:

$$\mathcal{W}_j^{(i,k)}(\overline{\Omega_\alpha^{\text{ext}}}) = \{\mathbf{\Psi}_j^{(i,k),\alpha,\alpha} \star \tilde{\mathbf{\Psi}}_j^{(i,k),\alpha,\alpha} * \mathbf{f} \ : \ \mathbf{f} \in \mathbf{h}_s(\overline{\Omega_\alpha^{\text{ext}}})\}. \tag{7.168}$$

Theorem 7.55 can be directly transferred by substituting the convolutions with respect to the sphere Ω_σ by the corresponding convolutions with respect to the sphere Ω_α. We now transfer Theorem 7.56 to the detail spaces and get the following theorem, where we use the terms injectivity, bijectivity, and uniqueness in the same sense as before.

Theorem 7.58. *The restriction of the operator* $\boldsymbol{\lambda}^{(i,k)} : \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}})$ *to a detail space* $\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, *i.e.,*

$$\boldsymbol{\lambda}^{(i,k)}|_{\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})}\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}}) \to \mathcal{W}_j^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}}) \tag{7.169}$$

with $\psi_j(n) = \psi_j^{(i,k)}(n)$ *is injective. Moreover, we have the following results:*

(i) *If the families* $\{\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ *and* $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$, $i,k\in\{1,2,3\}$, *are bandlimited (for example,* $\varphi_j^{(i,k)}(n) = \varphi_j(n) = 0$ *for all* $n \geq 2^j$*), then the restricted operator is even bijective. To be more specific, for* $\mathbf{g}^{(i,k)} \in \mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}})$ *the unique solution* $H_j \in \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, *of the equation*

$$\boldsymbol{\lambda}^{(i,k)} H_j = \tilde{\mathbf{\Psi}}_j^{(i,k),\gamma,\gamma} * \mathbf{\Psi}_j^{(i,k),\gamma,\gamma} * \mathbf{g}^{(i,k)} \tag{7.170}$$

is given by

$$H_j = \tilde{\Psi}_j^{\sigma,\sigma} * \Psi_j^{\sigma,\sigma} * Q, \tag{7.171}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ *is obtainable by*

$$Q^\wedge(n,m) = \begin{cases} \dfrac{\mathbf{g}^{(i,k)\wedge}(n,m)}{\boldsymbol{\lambda}^{(i,k)\wedge}(n)}, & n \in [0_{ik}, 2^{j+1}), \\ 0, & n \in [2^{j+1}, \infty). \end{cases} \tag{7.172}$$

(ii) *If the families* $\{\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$, $i,k\in\{1,2,3\}$, *and* $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ *are non-bandlimited, the equation*

$$\boldsymbol{\lambda}^{(i,,k)} H_j = \tilde{\mathbf{\Psi}}_j^{(i,k),\gamma,\gamma} \star \mathbf{\Psi}_j^{(i,k),\gamma,\gamma} * \mathbf{g}^{(i,k)} \tag{7.173}$$

has a solution $H_j \in \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$ *provided that the condition*

$$\sum_{n=0_{ik}}^{\infty} \sum_{m=1}^{2n+1} \frac{\mathbf{g}^{(i,k)\wedge}(n,m)}{\boldsymbol{\lambda}^{(i,k)\wedge}(n)} < \infty \tag{7.174}$$

is satisfied for $\mathbf{g}^{(i,k)} \in \mathbf{h_s}^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}})$. *In this case, the unique solution of the equation is given by*

$$H_j = \tilde{\Psi}_j^{\sigma,\sigma} * \Psi_j^{\sigma,\sigma} * Q, \tag{7.175}$$

where $Q \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ *is obtainable by*

$$Q^\wedge(n,m) = \frac{\mathbf{g}^{(i,k)\wedge}(n,m)}{\boldsymbol{\lambda}^{(i,k)\wedge}(n)}, \tag{7.176}$$

$n = 0_{ik}, \ldots;\ m = 1, \ldots, 2n+1.$

Furthermore, we have the following corollary.

Corollary 7.59. *The restriction of the operator* $\boldsymbol{\lambda} = \sum_{i,k=1}^{3} \boldsymbol{\lambda}^{(i,k)}$ *to a detail space* $\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, *i.e.,*

$$\boldsymbol{\lambda}|_{\mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}})} : \mathcal{W}_j(\overline{\Omega_\sigma^{\text{ext}}}) \to \bigoplus_{i,k=1}^{3} \mathcal{W}_j^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}}) \tag{7.177}$$

has, in general, no solution. Under the assumption

$$\psi_j^{(i,k)}(n) = \psi_j(n) \quad \textit{and} \quad \tilde{\psi}_j^{(i,k)}(n) = \tilde{\psi}_j(n), \quad i,k \in \{1,2,3\},$$

we have to claim, in addition, that

$$\frac{\mathbf{g}^{(i,k)\wedge}(n,m)}{\boldsymbol{\lambda}^{(i,k)\wedge}(n)} = \frac{\mathbf{g}^{(l,r)\wedge}(n,m)}{\boldsymbol{\lambda}^{(l,r)\wedge}(n)}, \tag{7.178}$$

with $i,k,l,r \in \{1,2,3\}$; $n = \max\limits_{i,k,l,r}(0_{ik}, 0_{lr}), \ldots;\ m = 1, \ldots, 2n+1$. *Then the results in Theorem* 7.58 *can be directly transferred.*

Up to now, we have summarized some results about the filtered solution, i.e., the solution when we restrict the operator to the scale or detail spaces. In this case, the injectivity for the operators $\boldsymbol{\lambda}^{(i,k)}$ could be proved, whereas in the case of general operators $\boldsymbol{\lambda} = \sum_{i,k=1}^{3} \boldsymbol{\lambda}^{(i,k)}$ we have to claim that (7.164) is valid. In the case of the unfiltered solution, we obtain the following theorem.

Theorem 7.60. *Let* $\mathbf{g}^{(i,k)} \in \mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}})$ *satisfy the condition* $\mathbf{g} \in \text{im}(\boldsymbol{\lambda}^{(i,k)})$, $i,k \in \{1,2,3\}$. *Then the unique solution* $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ *of the equation* $\boldsymbol{\lambda}^{(i,k)} F = \mathbf{g}^{(i,k)}$ *is given by*

$$F^\wedge(n,m) = \frac{\mathbf{g}^{(i,k)\wedge}(n,m)}{\boldsymbol{\lambda}^{(i,k)\wedge}(n)}, \tag{7.179}$$

$n = 0_{ik}, \ldots;\ m = 1, \ldots, 2n+1$. *In the case of the operator* $\boldsymbol{\lambda} = \sum_{i,k=1}^{3} \boldsymbol{\lambda}^{(i,k)}$ *we have to claim, in addition, that* (7.178) *holds in order to guarantee the solvability.*

Last, we explain the connection between the solution in the scale spaces and the unfiltered solution.

Theorem 7.61. *Suppose that* $\mathbf{g}^{(i,k)}$ *is of the class* $\mathbf{h}_s^{(i,k)\Lambda}(\overline{\Omega_\gamma^{\text{ext}}})$. *Let* $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ *be the unique solution of* $\boldsymbol{\lambda}^{(i,k)} F = \mathbf{g}^{(i,k)}$. *Then*

$$F_j = (\Phi_j^{(2)})^{\sigma,\sigma} * F \tag{7.180}$$

is the unique solution in $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ *of the equation*

$$\boldsymbol{\lambda}^{(i,k)} F_j = \boldsymbol{\Phi}_j^{(i,k),\gamma,\gamma} \star \boldsymbol{\Phi}_j^{(i,k),\gamma,\gamma} * \mathbf{g}^{(i,k)} \tag{7.181}$$

for every $j \in \mathbb{N}_0$. *Furthermore, the limit relation*

$$\lim_{J\to\infty} (\Phi_J^{(2)})^{\sigma,\sigma} * F = F \tag{7.182}$$

holds (in $\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}$*-sense).*

The preceding theorem shows that in the case of bandlimited scaling functions the (tensorial) SGG-problem is well posed, because a unique solution always exists and due to the finite dimension of the scale spaces the solution is also stable. We now investigate the case of non-bandlimited scaling functions, where the stability cannot be ensured and we have to use regularization methods.

Definition 7.62. A family of linear operators $S_j^{(i,k)} : \mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, is called a *regularization* of $(\boldsymbol{\lambda}^{(i,k)})^{-1}$, $i,k \in \{1,2,3\}$, if it satisfies the following properties:

(i) $S_j^{(i,k)}$ is bounded on $\mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}})$ for all $j \in \mathbb{N}_0$,

(ii) for any member $\mathbf{g}^{(i,k)} \in im(\boldsymbol{\lambda}^{(i,k)})$, the limit relation

$$\lim_{J\to\infty} S_J^{(i,k)} \mathbf{g}^{(i,k)} = (\boldsymbol{\lambda}^{(i,k)})^{-1} \mathbf{g}^{(i,k)} \tag{7.183}$$

holds (in $\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}$-sense).

The operator $S : \mathbf{h}_s(\overline{\Omega_\gamma^{\text{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ given by $S|_{\mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}})} = S_j^{(i,k)}$ is called a regularization of $\boldsymbol{\lambda}^{-1}$.

The function $F_J = S_J \mathbf{g}$ is called the *J-level regularization of* $\boldsymbol{\lambda}^{-1}\mathbf{g}$, whereas $F_J^{(i,k)} = S_J^{(i,k)} \mathbf{g}^{(i,k)}$ is called the *J-level regularization of* $(\boldsymbol{\lambda}^{(i,k)})^{-1}\mathbf{g}$. Within our multiscale approach, we now represent the $(J+1)$-level regularization using the J-level regularization by adding the corresponding detail information. To this end we first introduce a multiscale regularization concept starting with the definition of a generator of a regularization scaling function.

Definition 7.63. A family $\{\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ of sequences $\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}$, $i,k \in \{1,2,3\}$, is called a *generator of a regularization scaling function with respect to* $(\boldsymbol{\lambda}^{(i,k)})^{-1}$, if it satisfies the following requirements:

(i) $(\varphi_j^{(i,k)}(0_{ik}))^2 = \frac{1}{\boldsymbol{\lambda}^{(i,k)\wedge}(0_{ik})}$, for all $j \in \mathbb{N}_0$,

(ii) $(\varphi_j^{(i,k)}(n))^2 \le (\varphi_{j'}^{(i,k)}(n))^2$, for all $j, j' \in \mathbb{N}_0$ with $j \le j'$ and all $n \in \mathbb{N}_{0_{ik}+1}$,

(iii) $\lim_{j\to\infty} (\varphi_j^{(i,k)}(n))^2 = \frac{1}{(\boldsymbol{\lambda}^{(i,k)})^\wedge(n)}$, for all $n \in \mathbb{N}_{0_{ik}+1}$.

Furthermore, $\{\{\{\varphi_j^{(i,k)}(n)\}_{i,k\in\{1,2,3\}}\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ is called a *generator of a regularization scaling function with respect to* $\boldsymbol{\lambda}^{-1}$, if $(\boldsymbol{\lambda}^{(i,k)})^{-1}$ is a generator of a regularization scaling function with respect to $(\boldsymbol{\lambda}^{(i,k)})^{-1}$ for every $i,k = 1,2,3$.

We now define decomposition and reconstruction regularization scaling functions.

Definition 7.64. Let $\{\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ be a generator of a regularization scaling function with respect to $(\boldsymbol{\lambda}^{(i,k)})^{-1}$, $i,k\in\{1,2,3\}$.

Then a family $\{{}^d\boldsymbol{\Phi}_j^{(i,k),\sigma,\gamma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of admissible $\mathbf{h}_{\sigma,\gamma}^{(i,k)}$-kernels given by

$$ {}^d\boldsymbol{\Phi}_j^{(i,k),\sigma,\gamma}(x,z)=\sum_{n=0_{ik}}^{\infty}\varphi_j^{(i,k)}(n)\sum_{m=1}^{2n+1}H_{n,m}^s(\sigma;x)\mathbf{h}_{n,m}^{(i,k)s}(\gamma;z), \tag{7.184}$$

$(x,z)\in\overline{\Omega_\sigma^{\text{ext}}}\times\overline{\Omega_\gamma^{\text{ext}}}$, is called a *decomposition regularization* $\mathbf{h}_{\sigma,\gamma}^{(i,k)}$*-scaling function with respect to* $(\boldsymbol{\lambda}^{(i,k)})^{-1}$, whereas a family $\{{}^r\boldsymbol{\Phi}_j^{(i,k),\sigma,\sigma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of admissible $\mathbf{h}_{\sigma,\sigma}^{(i,k)}$-kernels given by

$$ {}^r\boldsymbol{\Phi}_j^{(i,k),\sigma,\sigma}(x,y)=\sum_{n=0_{ik}}^{\infty}\varphi_j^{(i,k)}(n)\sum_{m=1}^{2n+1}H_{n,m}^s(\sigma;x)\mathbf{h}_{n,m}^{(i,k)s}(\sigma;y), \tag{7.185}$$

$(x,y)\in\overline{\Omega_\sigma^{\text{ext}}}\times\overline{\Omega_\sigma^{\text{ext}}}$ is called a *reconstruction regularization* $\mathbf{h}_{\sigma,\sigma}^{(i,k)}$*-scaling function with respect to* $(\boldsymbol{\lambda}^{(i,k)})^{-1}$.

We obtain the following theorem:

Theorem 7.65. *Let* $\{\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ *be a generator of a regularization scaling function with respect to* $(\boldsymbol{\lambda}^{(i,k)})^{-1}$, $i,k\in\{1,2,3\}$. *If we formally define*

$$({}^r\boldsymbol{\Phi}_j^{(i,k)}\star{}^d\boldsymbol{\Phi}_j^{(i,k)})^{\sigma,\gamma}(\cdot,\cdot)$$

by

$$({}^r\boldsymbol{\Phi}_j^{(i,k)}\star{}^d\boldsymbol{\Phi}_j^{(i,k)})^{\sigma,\gamma}(x,z)={}^r\boldsymbol{\Phi}_j^{(i,k),\sigma,\sigma}(x,\cdot)\star{}^d\boldsymbol{\Phi}_j^{(i,k),\sigma,\gamma}(\cdot,z), \tag{7.186}$$

$(x,z)\in\overline{\Omega_\sigma^{\text{ext}}}\times\overline{\Omega_\gamma^{\text{ext}}}$, *then*

$$F_J^{(i,k)}=({}^r\boldsymbol{\Phi}_J^{(i,k)}\star{}^d\boldsymbol{\Phi}_J^{(i,k)})^{\sigma,\gamma}*\mathbf{g}^{(i,k)},\quad \mathbf{g}^{(i,k)}\in\mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\text{ext}}}), \tag{7.187}$$

represents the J-level regularization of $(\boldsymbol{\lambda}^{(i,k)})^{-1}\mathbf{g}^{(i,k)}$. *If, in addition,* $\mathbf{g}^{(i,k)}\in\text{im}(\boldsymbol{\lambda}^{(i,k)})$, *then*

$$\lim_{J\to\infty}\|F_J^{(i,k)}-(\boldsymbol{\lambda}^{(i,k)})^{-1}\mathbf{g}^{(i,k)}\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}=0. \tag{7.188}$$

Furthermore,

$$F_J=\sum_{i,k=1}^{3}({}^r\boldsymbol{\Phi}_J^{(i,k)}\star{}^d\boldsymbol{\Phi}_J^{(i,k)})^{\sigma,\gamma}*\mathbf{g}^{(i,k)},\quad \mathbf{g}=\sum_{i,k=1}^{3}\mathbf{g}^{(i,k)}\in\mathbf{h}_s(\overline{\Omega_\gamma^{\text{ext}}}), \tag{7.189}$$

represents the J-level regularization of $\boldsymbol{\lambda}^{-1}g$. *If, in addition,* $\mathbf{g}\in\text{im}(\boldsymbol{\lambda})$, *then*

$$\lim_{J\to\infty}\|F_J-\boldsymbol{\lambda}^{-1}\mathbf{g}\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}=0. \tag{7.190}$$

We define the convolution operators $S_J^{(i,k)} : \mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\mathrm{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$, $J \in \mathbb{N}_0$, by

$$S_J^{(i,k)}(\mathbf{g}^{(i,k)}) = ({}^r\mathbf{\Phi}_J^{(i,k)} \star {}^d\mathbf{\Phi}_J^{(i,k)})^{\sigma,\gamma} * \mathbf{g}^{(i,k)}, \tag{7.191}$$

whereas the convolution operator $S_J : \mathbf{h}_s(\overline{\Omega_\gamma^{\mathrm{ext}}}) \to \mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$, $J \in \mathbb{N}_0$, is given by

$$S_J(\mathbf{g}) = \sum_{i,k=1}^{3} S_J^{(i,k)}(\mathbf{g}^{(i,k)}). \tag{7.192}$$

Furthermore, we introduce the corresponding scale spaces $S_J^{(i,k)}(\mathrm{im}(\boldsymbol{\lambda}^{(i,k)}))$, $i,k \in \{1,2,3\}$, and $S_J(\mathrm{im}(\boldsymbol{\lambda}))$ as follows

$$S_J^{(i,k)}(\mathrm{im}(\boldsymbol{\lambda}^{(i,k)})) = \left\{ ({}^r\mathbf{\Phi}_J^{(i,k)} \star {}^d\mathbf{\Phi}_J^{(i,k)})^{\sigma,\gamma} * \mathbf{g}^{(i,k)} \ : \ \mathbf{g}^{(i,k)} \in \mathrm{im}(\boldsymbol{\lambda}^{(i,k)}) \right\}, \tag{7.193}$$

$$S_J(\mathrm{im}(\boldsymbol{\lambda})) = \left\{ \sum_{i,k=1}^{3} ({}^r\mathbf{\Phi}_J^{(i,k)} \star {}^d\mathbf{\Phi}_J^{(i,k)})^{\sigma,\gamma} * \mathbf{g}^{(i,k)} : \mathbf{g} = \sum_{i,k=1}^{3} \mathbf{g}^{(i,k)} \in \mathrm{im}(\boldsymbol{\lambda}) \right\}. \tag{7.194}$$

Theorem 7.66. *The scale spaces satisfy the following properties:*

(i) $S_0^{(i,k)}(\mathrm{im}(\boldsymbol{\lambda}^{(i,k)})) \subset \cdots \subset S_J^{(i,k)}(\mathrm{im}(\boldsymbol{\lambda}^{(i,k)})) \subset S_{J'}^{(i,k)}(\mathrm{im}(\boldsymbol{\lambda}^{(i,k)})) \subset \mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$, $J \le J'$, *i.e., for any right-hand side* $\mathbf{g}^{(i,k)} \in \mathrm{im}(\boldsymbol{\lambda}^{(i,k)})$ *of the* (*tensorial*) *SGG problem, all* J*-level regularizations with fixed parameter* J *are sampled in a scale space* $S_J^{(i,k)}(\mathrm{im}(\boldsymbol{\lambda}^{(i,k)}))$ *with the above property,*

(ii) $\overline{\bigcup_{J=0}^{\infty} S_J^{(i,k)}(\mathrm{im}(\boldsymbol{\lambda}^{(i,k)}))}^{\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})}} = \mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$.

Obviously, Theorem 7.66 is also valid substituting $S_J^{(i,k)}$ by S_J which leads to the following corollary.

Corollary 7.67. *The scale spaces satisfy the following properties:*

(i) $S_0(\mathrm{im}(\boldsymbol{\lambda})) \subset \cdots \subset S_J(\mathrm{im}(\boldsymbol{\lambda})) \subset S_{J'}(\mathrm{im}(\boldsymbol{\lambda})) \subset \mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$, $J \le J'$, *i.e., for any right-hand side* $\mathbf{g} \in \mathrm{im}(\boldsymbol{\lambda})$ *of the* (*tensorial*) *SGG problem, all* J*-level regularizations with fixed parameter* J *are sampled in a scale space* $S_J(\mathrm{im}(\boldsymbol{\lambda}))$ *with the above property,*

(ii) $\overline{\bigcup_{J=0}^{\infty} S_J(\mathrm{im}(\boldsymbol{\lambda}))}^{\|\cdot\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})}} = \mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$.

A set of subspaces of $\mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})$ satisfying the conditions of Corollary 7.67 is called *regularization* $h_{\sigma,\gamma}$*-multiresolution analysis* (*RMRA*) *of the* (*tensorial*) *SGG problem.*

Definition 7.68. Let $\{\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ be a generator of a regularization scaling function with respect to $\left(\boldsymbol{\lambda}^{(i,k)}\right)^{-1}$. Then the generating symbols

$$\{\tilde{\psi}_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}, \ \{\psi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}$$

of the corresponding regularization wavelets are defined by the refinement equation

(7.147). The admissible $\mathbf{h}_{\sigma,\gamma}^{(i,k)}$-kernels $\{{}^{d}\mathbf{\Psi}_{j}^{(i,k),\sigma,\gamma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by

$$ {}^{d}\mathbf{\Psi}_{j}^{(i,k),\sigma,\gamma}(x,z) = \sum_{n=0_{ik}}^{\infty} \psi_{j}^{(i,k)}(n) \sum_{m=1}^{2n+1} H_{n,m}^{s}(\sigma;x)\mathbf{h}_{n,m}^{(i,k)s}(\gamma;z), \tag{7.195} $$

$(x,z) \in \overline{\Omega_{\sigma}^{\text{ext}}} \times \overline{\Omega_{\gamma}^{\text{ext}}}$ are called the *decomposition regularization* $\mathbf{h}_{\sigma,\gamma}^{(i,k)}$*-wavelets*, while the admissible $\mathbf{h}_{\sigma,\sigma}^{(i,k)}$-kernels $\{{}^{r}\tilde{\mathbf{\Psi}}_{j}^{(i,k),\sigma,\sigma}(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by

$$ {}^{r}\tilde{\mathbf{\Psi}}_{j}^{(i,k),\sigma,\sigma}(x,y) = \sum_{n=0_{ik}}^{\infty} \tilde{\psi}_{j}^{(i,k)}(n) \sum_{m=1}^{2n+1} H_{n,m}^{s}(\sigma;x)\mathbf{h}_{n,m}^{(i,k)s}(\sigma;y), \tag{7.196} $$

$(x,y) \in \overline{\Omega_{\sigma}^{\text{ext}}} \times \overline{\Omega_{\sigma}^{\text{ext}}}$ are called the *reconstruction regularization* $\mathbf{h}_{\sigma,\sigma}^{(i,k)}$*-wavelets.*

We now define the convolution operators $T_{j}^{(i,k)} : \mathbf{h}_{s}^{(i,k)}(\overline{\Omega_{\gamma}^{\text{ext}}}) \to \mathcal{H}_{s}(\overline{\Omega_{\sigma}^{\text{ext}}})$, $j \in \mathbb{N}_0$, $i,k = 1,2,3$, by

$$ T_{j}^{(i,k)}(\mathbf{g}^{(i,k)}) = ({}^{r}\tilde{\mathbf{\Psi}}_{j}^{(i,k)} \star {}^{d}\mathbf{\Psi}_{j}^{(i,k)})^{\sigma,\gamma} * \mathbf{g}^{(i,k)}, \quad \mathbf{g}^{(i,k)} \in \mathbf{h}_{s}^{(i,k)}(\overline{\Omega_{\gamma}^{\text{ext}}}), \tag{7.197} $$

and the convolution operator $T_j : \mathbf{h}_{s}(\overline{\Omega_{\gamma}^{\text{ext}}}) \to \mathcal{H}_{s}(\overline{\Omega_{\sigma}^{\text{ext}}})$, $j \in \mathbb{N}_0$, by

$$ T_{J}(\mathbf{g}) = \sum_{i,k=1}^{3} T_{J}^{(i,k)}(\mathbf{g}^{(i,k)}). \tag{7.198} $$

Obviously, due to the refinement equation the operators $S_{J+1}^{(i,k)}$ and S_{J+1} can be represented in the form

$$ S_{J+1}^{(i,k)} = S_{0}^{(i,k)} + \sum_{j=0}^{J} T_{j}^{(i,k)}, \tag{7.199} $$

$$ S_{J+1} = S_{0} + \sum_{j=0}^{J} T_{j}. \tag{7.200} $$

Thus, we now introduce the *detail spaces* $T_{J}^{(i,k)}(\text{im}(\boldsymbol{\lambda}^{(i,k)}))$ and $T_{J}(\text{im}(\boldsymbol{\lambda}))$ by

$$ T_{J}^{(i,k)}(\text{im}(\boldsymbol{\lambda}^{(i,k)})) = \left\{ ({}^{r}\tilde{\mathbf{\Psi}}_{J}^{(i,k)} \star {}^{d}\mathbf{\Psi}_{J}^{(i,k)})^{\sigma,\gamma} * \mathbf{g}^{(i,k)} : \mathbf{g}^{(i,k)} \in \text{im}(\boldsymbol{\lambda}^{(i,k)}) \right\}, \tag{7.201} $$

$$ T_{J}(\text{im}(\boldsymbol{\lambda})) = \left\{ \sum_{i,k=1}^{3} ({}^{r}\tilde{\mathbf{\Psi}}_{J}^{(i,k)} \star {}^{d}\mathbf{\Psi}_{J}^{(i,k)})^{\sigma,\gamma} * \mathbf{g}^{(i,k)} : \mathbf{g} = \sum_{i=1}^{3} \mathbf{g}^{(i,k)} \in \text{im}(\boldsymbol{\lambda}) \right\}. \tag{7.202} $$

In terms of the multiscale concept, the space $T_{J}(\text{im}(\boldsymbol{\lambda}))$ contains the detail information which has to be added in order to turn from the J-level regularization to the $(J+1)$-level regularization:

$$ S_{J+1}(\text{im}(\boldsymbol{\lambda})) = S_{J}(\text{im}(\boldsymbol{\lambda})) + T_{J}(\text{im}(\boldsymbol{\lambda})). \tag{7.203} $$

In general, the sum is neither direct nor orthogonal.

Theorem 7.69. *Let* $\{\{\varphi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$ *be a generator of a regularization scaling function with respect to* $(\boldsymbol{\lambda}^{(i,k)})^{-1}$, $i,k\in\{1,2,3\}$. *Suppose that*

$$\{\{\psi_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0},\ \{\{\tilde{\psi}_j^{(i,k)}(n)\}_{n\in\mathbb{N}_{0_{ik}}}\}_{j\in\mathbb{N}_0}$$

are the generating symbols of the corresponding regularization wavelets. Furthermore, let $\mathbf{g}^{(i,k)}$ *be of the class* $\mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\mathrm{ext}}})$. *Define the* regularization $\mathbf{h}_{\sigma,\gamma}^{(i,k)}$-wavelet transform at scale $j\in\mathbb{N}_0$ and position $x\in\overline{\Omega_\sigma^{\mathrm{ext}}}$ *by*

$$(RWT)(\mathbf{g}^{(i,k)})(j;x)={}^d\boldsymbol{\Psi}_j^{(i,k),\sigma,\gamma}(x,\cdot)*\mathbf{g}^{(i,k)},\quad \mathbf{g}^{(i,k)}\in\mathbf{h}_s^{(i,k)}(\overline{\Omega_\gamma^{\mathrm{ext}}}).\tag{7.204}$$

Then

$$F_J=\left({}^r\boldsymbol{\Phi}_0^{(i,k)}\star{}^d\boldsymbol{\Phi}_0^{(i,k)}\right)^{\sigma,\gamma}*\mathbf{h}^{(i,k)}+\sum_{j=0}^{J-1}{}^r\tilde{\boldsymbol{\Psi}}_j^{(i,k),\sigma,\sigma}\star(RWT)(\mathbf{g}^{(i,k)})(j;\cdot)\tag{7.205}$$

is the J*-level regularization of the* (*tensorial*) *SGG problem satisfying*

$$\lim_{J\to\infty}\|F_J-(\boldsymbol{\lambda}^{(i,k)})^{-1}\mathbf{g}^{(i,k)}\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\mathrm{ext}}})}=0\tag{7.206}$$

provided that $\mathbf{g}^{(i,k)}\in\mathrm{im}(\boldsymbol{\lambda}^{(i,k)})$.

7.4. Combined outer harmonic and wavelet concept

In geodetic practice, there exists a variety of realizations of spherical harmonic models of the Earth's external gravitational potential. In [19] it is explained how to combine an outer harmonic model of fixed order m with a harmonic wavelet model. The justification for such a combined model is the fact that on the one hand the appropriate candidate for the approximation of the low frequency parts of the gravitational potential (i.e., global modeling) is a spherical harmonic (i.e., a multipole) model of moderate order m and on the other hand for the representation of the high frequency parts (i.e., local modeling) new wavelet techniques have to come into play (see also the investigations in spherical continuous wavelet theory [33, 34]).

Starting point of this model is the "refinement equation" (compare Eq. (7.10))

$$\tilde{\psi}_j(n)\psi_j(n)=(\varphi_{j+1}(n))^2-(\varphi_j(n))^2.$$

It is clear that $\tilde{\psi}_j(n)\psi_j(n)=0$ if and only if $(\varphi_{j+1}(n))^2=(\varphi_j(n))^2$. Due to condition (i) in Definition 7.2, the wavelet (or its dual) satisfy the mean value condition $\psi_j(0)=0$, i.e., it has to oscillate. For purposes of combined approximation we need, however, $(\varphi_{j+1}(n))^2=(\varphi_j(n))^2$ for all $n\in[0,\ldots,m]$. Under these assumptions it may be guaranteed that the wavelets constructed in this way have more vanishing moments and we call them *wavelets of order* m. In [19] the reconstruction formula for such wavelets is studied in more detail. The transition of the combined outer harmonic and wavelet concept to the vectorial and tensorial case is also easy to perform.

8. Bandlimited Runge–Walsh multiscale approximation

In the previous sections we developed several methods of wavelet approximation. We briefly reformulate the main results: Let $\{\Psi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathcal{H}_{\sigma,\sigma}$-wavelet corresponding to an $\mathcal{H}_{\sigma,\sigma}$-scaling function $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$. Then any potential $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ can be expressed by a multiscale approximation given by

$$\Phi_0^{(2)} * F + \sum_{j=0}^{J-1} \tilde{\Psi}_j * \Psi_j * F, \quad F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}). \tag{8.1}$$

For a numerical realization, the discretization of the $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-convolutions (i.e., the $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-inner products) occurring in the J-level wavelet approximation is necessary. For that purpose we observe that any $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-convolution is expressible as a bounded linear functional on $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$. Thus fully discretized wavelet approximation amounts to the problem of approximating a bounded linear functional (i.e., an $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-inner product) by a linear combination of known bounded linear functionals. In this context it should again be mentioned that following our nomenclature an $\mathcal{H}_0(\overline{\Omega_\sigma^{\text{ext}}})$-inner product can be identified with an ordinary integral over the sphere Ω_σ. Therefore, fully discretized $\mathcal{H}_0(\overline{\Omega_\sigma^{\text{ext}}})$-wavelet approximation can be organized appropriately by numerical integration (cubature) over the sphere Ω_σ. Looking at the inner products in our general $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-framework we are confronted with convolutions involving a pseudodifferential operator Λ with symbol $\Lambda^\wedge(n) = A_n$ for $n \in \mathbb{N}_0$. Their discretization requires the knowledge of linear (observational) functionals for the potential $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ under consideration. Usually, in gravitational field determination, these (observational) functionals are heterogeneous in nature. In addition, the approximate formulae have to be formulated in dependence on the scale parameter, since increasing space localization demands increasing data material.

All these requirements, however, do not lead to a unique procedure for discretizing $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-convolutions. Many variants of approximate formulae are reasonable and conceivable. In fact, the choice of a suitable method is essentially dependent on the purpose for which scaling functions and wavelets are used. Unfortunately, it turns out that each of the discretization methods has its own drawback. Nevertheless, a lot of approximation schemata for $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-convolutions can be found so that at least some of the requests can be fulfilled. As most important discretization rules we mention:

1. *Fast Fourier techniques and multipole techniques* (cf. [19, 39, 74]) are economical in time, but they are based on evaluation functionals on equiangular latitude-longitude grids. Thus the sample points are merely equidistributed on the (ϑ,φ)-parameter interval $[0,\pi]\times[0,2\pi]$ in Euclidean space $\mathbb{R}^2$, but not on a sphere.
2. *Polynomial* (*i.e., outer harmonic*) *exact approximation of bandlimited functions* is a well-established tool for application to bandlimited potentials of moderate degree (cf. [12, 28, 29, 54]). The problem is that the preliminary

work includes the solution of a linear system of equations (which is full-sized and tends to be ill conditioned for an increasing number of nodal points). However, it can be shown that (outer harmonics) exact approximation of bandlimited potentials can be used very efficiently (without *a priori* solving any linear system) on equiangular grids (cf. [23, 39]).

3. Another method for the approximate evaluation of $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-convolutions, which includes the exact approximation of bandlimited functions as a special case, is *harmonic spline exact best approximation* (cf. [12, 14, 19]). It can be applied appropriately for modeling the medium to short wavelength parts of a signal.
4. The *low discrepancy method* (cf. [21, 48]) represents an adequate tool if a great number of data is available, so that the solution of linear equations should be avoided. Sufficient accuracy can be guaranteed only if a high number of equidistributed data points are available. Thus it is of advantage for integrands of high complexity (e.g., short wavelength parts of a signal).

In what follows, it will be shown that both discretization techniques, i.e., outer harmonic and spline exact integration, lead to pyramid schemata adapted to the space localization properties of the potential we are interested in. To be more specific, the bandlimited variant of fast wavelet computation (based on the Shannon kernel and its modifications) can be based on outer harmonic exact formulae for the evaluation of $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$-inner products. It is proposed for the application to moderate phenomena of space localization (i.e., low-to-medium wavelength approximation) so that one can work with smaller data sets (cf. [31, 32]). In fast computation by bandlimited wavelets the number of wavelet coefficients is reduced, since they contain information of a more extended area. In addition, a certain spectral band is expressible exactly in terms of wavelets because of their bandlimited character. The non-bandlimited variant of fast wavelet evaluation (using non-bandlimited kernels such as Tikhonov, rational, exponential, and "locally supported" kernels (cf. [29, 31]) is meant for the application to seriously space localizing potentials (i.e., short wavelength approximation). In consequence, huge data sets can be handled since only a small subset of the data is needed for the purpose of numerical evaluation. On the other hand, a large number of wavelet coefficients is needed, since they only give local information related to a small area. Again, we are confronted with the drawback that large linear systems must be solved in an a priori step to obtain the weights in (spline exact) best approximation formulae. In the non-bandlimited case, however, panel clustering or sparse matrix techniques (cf. [23]) are efficiently applicable because of the strong space localization properties of the non-bandlimited kernel functions.

Next, the use of outer harmonic exact approximation will be discussed in more detail following [19]. A constructive version of the Runge–Walsh theorem will be developed in terms of bandlimited wavelets. The advantage is that when using bandlimited wavelets, we do not need the wavelet transform at all positions. It suffices to know a finite set of linear functionals for each scale J to evaluate

the wavelet transform exactly. In conclusion, each J-level wavelet approximation $\Phi_J^{(2)} * F$ can be expressed exactly as a finite sum.

Our concept using bandlimited wavelets is presented under the assumption that the families $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$, $\{\Psi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$, and $\{\tilde{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ consist of bandlimited kernels such that

$$\varphi_j(n) \neq 0, \quad n = 0, \ldots, 2^j - 1 \tag{8.2}$$

and

$$\varphi_j(n) = 0, \quad n = 2^j, 2^j + 1, \ldots. \tag{8.3}$$

In the following we use the notation

$$\mathcal{H}_{p,\ldots,q}(\overline{\Omega_\sigma^{\text{ext}}}) = Harm_{p,\ldots,q}(\overline{\Omega_\sigma^{\text{ext}}}). \tag{8.4}$$

Consequently, we have

$$\Phi_j(x,\cdot) \in \mathcal{H}_{0,\ldots,2^j-1}(\overline{\Omega_\sigma^{\text{ext}}}), \tag{8.5}$$

and

$$\Psi_j(x,\cdot), \tilde{\Psi}_j(x,\cdot) \in \mathcal{H}_{0,\ldots,2^{j+1}-1}(\overline{\Omega_\sigma^{\text{ext}}}) \tag{8.6}$$

for all $x \in \overline{\Omega_\sigma^{\text{ext}}}$. Thus the scale spaces and the detail spaces, respectively, fulfill the relations

$$\mathcal{V}_j = \mathcal{H}_{0,\ldots,2^j-1}(\overline{\Omega_\sigma^{\text{ext}}}), \quad \mathcal{W}_j \subset \mathcal{H}_{0,\ldots,2^{j+1}-1}(\overline{\Omega_\sigma^{\text{ext}}}). \tag{8.7}$$

Suppose now that there is known a set $\{v_1, \ldots, v_M\}$ of M values v_i, $i = 1, \ldots, M$, from a potential V (for example, the gravitational potential or the anomalous potential of the Earth) of class $Pot^{(0)}(\overline{\Sigma^{\text{ext}}})$ corresponding to linear (observational) functionals $\mathcal{L}_1, \ldots, \mathcal{L}_M$. Then an extended version of Helly's theorem (cf. [76]) tells us that, corresponding to the potential $V \in Pot^{(0)}(\overline{\Sigma^{\text{ext}}})$, there exists a member F (i.e., a Runge–Walsh approximation of the (anomalous) potential) of class $\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ such that $F|_{\overline{\Sigma^{\text{ext}}}}$ is in an $(\varepsilon/2)$-neighbourhood to V (understood in uniform topology on $\overline{\Sigma^{\text{ext}}}$) and $\mathcal{L}_i F = v_i$, $i = 1, \ldots, M$ (note that we may write more accurately $F_{0,\ldots,\infty}$ instead of F to indicate that all $Harm_n$-spaces generally contribute to the "nature" of F when the Earth's gravitational potential is required). Moreover, there exists an element $F_{0,\ldots,m}$ (i.e., a bandlimited approximation to the Runge–Walsh approximation) of class $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$ such that the restriction $F_{0,\ldots,m}|_{\overline{\Sigma^{\text{ext}}}}$ may be considered to be in $(\varepsilon/2)$-accuracy to $F|_{\overline{\Sigma^{\text{ext}}}}$ uniformly on $\overline{\Sigma^{\text{ext}}}$ and, in addition, $\mathcal{L}_i F_{0,\ldots,m} = \mathcal{L}_i F = v_i$, $i = 1, \ldots, M$. In other words, corresponding to a potential $V \in Pot^{(0)}(\overline{\Sigma^{\text{ext}}})$ there exists on $\overline{\Sigma^{\text{ext}}}$ a bandlimited potential in $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$, (namely, $F_{0,\ldots,m} \in \mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$) consistent with the original data in ε-accuracy (i.e., $v_i = \mathcal{L}_i F = \mathcal{L}_i F_{0,\ldots,m}$, $i = 1, \ldots, M$). This is the reason why we are interested in wavelet approximations of potentials $F_{0,\ldots,m}$ of class $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$ uniformly on $\overline{\Sigma^{\text{ext}}}$ from a finite set of functional values (note that, for the Earth's anomalous potential, the approximation consistent with the original data may be found in the class $\mathcal{H}_{2,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$ which is a subspace of $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$).

Our strategy is to represent $F_{0,\dots,m} \in \mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$ by a J-level approximation $\Phi_J^{(2)} * F_{0,\dots,m}$ with J chosen in such a way that $2^{J+1}-1 \geq m$ (note that $F_{0,\dots,m}$ coincides with $\Phi_{J+1}^{(2)} * F_{0,\dots,m}$ uniformly on $\overline{\Sigma^{\text{ext}}}$ in the case of Shannon wavelets). We want to express the J-level wavelet approximation $\Phi_{J+1}^{(2)} * F_{0,\dots,m}$ of the potential $F_{0,\dots,m}$(with $2^{J+1}-1 \geq m$) *exactly* only by use of the M values $v_1, \dots, v_M$ corresponding to the linear functionals $\mathcal{L}_1, \dots, \mathcal{L}_M$.

First, our purpose is to apply outer harmonic based approximation formulae. To this end, we introduce fundamental systems of bounded linear functionals and derive some approximation formulae. Consider the matrix

$$\mathbf{m} = \begin{pmatrix} \mathcal{L}_1 H_{0,1}(\sigma;\cdot) & \dots & \mathcal{L}_N H_{0,1}(\sigma;\cdot) \\ \vdots & & \vdots \\ \mathcal{L}_1 H_{m,2m+1}(\sigma;\cdot) & \dots & \mathcal{L}_N H_{m,2m+1}(\sigma;\cdot) \end{pmatrix} \tag{8.8}$$

associated to a system of $N \geq \sum_{n=0}^{m}(2n+1) = (m+1)^2$ (linearly independent) bounded linear functionals $\mathcal{L}_1, \dots, \mathcal{L}_N$ on $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$. According to well-known arguments of approximation theory, the matrix (8.8) is not of maximal rank for all systems $\{\mathcal{L}_1, \dots, \mathcal{L}_N\}$, $N \geq (m+1)^2$. However, it is clear from a well-known construction principle (see, for example, [19]) that there exist systems $\{\mathcal{L}_1, \dots, \mathcal{L}_N\}$ possessing a non-degenerate matrix (8.8).

Definition 8.1. A system $\{\mathcal{L}_1, \dots, \mathcal{L}_N\}$ of $N \geq (m+1)^2$ bounded linear functionals on $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ is called an $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-*fundamental system*, if the conditions $F \in \mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$ and $\mathcal{L}_i F = 0$, $i = 1, \dots, N$, imply $F = 0$.

From Definition 8.1 it is clear that the matrix (8.8) is of maximal rank $(m+1)^2$ if and only if $\{\mathcal{L}_1, \dots, \mathcal{L}_N\}$ is an $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system. Moreover, it should be noted that the addition theorem of outer harmonics gives us

$$\mathbf{m}^T\mathbf{m} = \begin{pmatrix} \mathcal{L}_1\mathcal{L}_1 K_{\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) & \dots & \mathcal{L}_1\mathcal{L}_N K_{\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) \\ \vdots & & \vdots \\ \mathcal{L}_N\mathcal{L}_1 K_{\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) & \dots & \mathcal{L}_N\mathcal{L}_N K_{\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) \end{pmatrix}.$$

The Gram matrix $\mathbf{m}^T\mathbf{m}$ is regular if and only if the system $\{\mathcal{L}_1, \dots, \mathcal{L}_N\}$ is an $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system. Moreover, it is clear that the property of $\{\mathcal{L}_1, \dots, \mathcal{L}_N\}$ of being an $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system, is independent of the choice of the $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-orthonormal basis.

For later use we introduce the following definition.

Definition 8.2. Let Ξ be a regular surface with $\Xi \subset \overline{\Omega_\sigma^{\text{ext}}}$.

Let $\{\mathcal{L}_1, \dots, \mathcal{L}_N\}$ be an $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system of Dirichlet functionals $\mathcal{L}_1, \dots, \mathcal{L}_N$ on $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ (i.e., $\mathcal{L}_i F = F(y_i)$ for $y_i \in \Xi$, $i = 1, \dots, N$ and all $F \in \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$). Then the associated system $\{y_1, \dots, y_N\}$ is called an $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-*Dirichlet-fundamental system on* Ξ.

Let $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\}$ be an $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system of Neumann functionals $\mathcal{L}_i$, $i = 1, \ldots, N$ (i.e., $\mathcal{L}_i F = (\lambda \cdot (\nabla F))(y_i)$ for $y_i \in \Xi$ and all $F \in \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$) with $\lambda : \Xi \to \mathbb{R}^3$ being a unit vector field satisfying $\inf_{x\in\Xi} \nu(x) \cdot \lambda(x) > 0$ (where ν denotes the outer normal). Then the system $\{y_1, \ldots, y_N\}$ is called an $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-*Neumann-fundamental system on* Ξ (*relative to* λ).

Let $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\}$ be an $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system in the sense of Definition 8.1. Suppose that F is a potential of class $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$. Furthermore, let P be an element of $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$ with the representation

$$P = \sum_{n=0}^{m} \sum_{l=1}^{2n+1} P^\wedge(n,l) H_{n,l}(\sigma;\cdot).$$

Then, for all solutions $a \in \mathbb{R}^N$, $a = (a_1, \ldots, a_N)^T$, of the linear system

$$\sum_{k=1}^{N} a_k \mathcal{L}_k H_{n,l}(\sigma;\cdot) = P^\wedge(n,l), \tag{8.9}$$

$n = 0, \ldots, m$; $l = 1, \ldots, 2n+1$, we find

$$P = \sum_{k=1}^{N} a_k \sum_{n=0}^{m} \sum_{l=1}^{2n+1} \left(\mathcal{L}_k H_{n,l}(\sigma;\cdot)\right) H_{n,l}(\sigma;\cdot). \tag{8.10}$$

Observing this fact we get the following theorem.

Theorem 8.3. *Let* $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\}$ *be an* $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$*-fundamental system of bounded linear functionals on* $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$. *Then the identity*

$$F * P = \sum_{k=1}^{N} a_k \mathcal{L}_k F - \sum_{k=1}^{N} a_k \mathcal{L}_k K_{\mathcal{H}_{m+1,\ldots,\infty}(\overline{\Omega_\sigma^{\text{ext}}})} * F$$

holds for all $F \in \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ *and all solutions* $a \in \mathbb{R}^N$, $a = (a_1, \ldots, a_N)^T$, *satisfying the linear system* (8.9).

By virtue of the Cauchy–Schwarz inequality it follows from Theorem 8.3 that the estimate

$$\begin{aligned} &\left| F * P - \sum_{k=1}^{N} a_k \mathcal{L}_k F \right| \\ &\quad \le \left(\sum_{k=1}^{N} \sum_{s=1}^{N} a_k a_s \mathcal{L}_k \mathcal{L}_s K_{\mathcal{H}_{m+1,\ldots,\infty}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) \right)^{1/2} \|F\|_{\mathcal{H}_{m+1,\ldots,\infty}(\overline{\Omega_\sigma^{\text{ext}}})} \end{aligned} \tag{8.11}$$

holds for all $F \in \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ and all solutions $a \in \mathbb{R}^N$, $a = (a_1, \ldots, a_N)^T$, satisfying (8.9). In particular, we have for $F \in \mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$

$$F * P = \sum_{k=1}^{N} a_k \mathcal{L}_k F, \tag{8.12}$$

since $\|F\|_{\mathcal{H}_{m+1,\ldots,\infty}(\overline{\Omega_\sigma^{\text{ext}}})} = 0$. But this shows us that

$$K_{\mathcal{H}_{0,\ldots,m}} * P = \sum_{k=1}^{N} a_k \mathcal{L}_k K_{\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) \tag{8.13}$$

holds for all $a \in \mathbb{R}^n$, $a = (a_1, \ldots, a_N)^T$, satisfying the linear equations (8.9).

Next we adopt a famous criterion due to [73] from Theorem 8.3.

Lemma 8.4. *The following statements are equivalent:*

(i) $\lim\limits_{N\to\infty} \sum\limits_{k=1}^{N} a_k \mathcal{L}_k H_{n,l}(\sigma;\cdot) = 0, \quad n = m+1, m+2, \ldots; \; l = 1, \ldots, 2n+1,$

(ii) $F * P = \lim\limits_{N\to\infty} \sum\limits_{k=1}^{N} a_k \mathcal{L}_k F, \quad F \in \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}}).$

As shown in [19], the definition of *fundamental systems* and *approximation formulae* leads us to exact approximation rules on $\mathcal{H}_{0,\ldots,2m}(\overline{\Omega_\sigma^{\text{ext}}})$-spaces. To this end we have to summarize shortly some results concerning interpolation by outer harmonics (see [19]).

We start mentioning the *Shannon sampling* theorem for the finite-dimensional space $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$.

Lemma 8.5. *Let F be in $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$. Assume that $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\}$ forms an $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-fundamental system. Then F can be reconstructed from its samples at the bounded linear functionals $\mathcal{L}_1, \ldots, \mathcal{L}_N$ by the following interpolation formula*

$$F(x) = \sum_{k=1}^{N} (\mathcal{L}_k F) P_k^N(x), \quad x \in \overline{\Omega_\sigma^{\text{ext}}},$$

where the "Lagrangians" $P_k^N \in \mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$, $k = 1, \ldots, N$, are given by

$$P_k^N = \sum_{l=1}^{N} w_{l,k}^N \mathcal{L}_l K_{\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot)$$

and the coefficients $w_{l,k}^N$ have to satisfy the linear equations

$$\sum_{l=1}^{N} w_{l,k}^N \mathcal{L}_i \mathcal{L}_l K_{\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) = \delta_{i,k},$$

$i, k = 1, \ldots, N$.

Next we come to some aspects on numerical integration on the sphere. Theorem 8.3 allows as special cases the following variants.

Lemma 8.6 (Koksma–Hlawka formula of approximation order 0). *Let F be of class $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ with $\{A_n\}$ being summable in the sense of Definition* 3.3. *Assume that*

$\{y_1^N, \ldots, y_N^N\}$ is a subset of points on Ω_σ. Then the integral formula

$$\begin{aligned}&\frac{1}{4\pi\sigma^2}\int_{\Omega_\sigma} F(y)\, d\omega(y)\\ &\quad = \sum_{k=1}^{N} w_k^N F(y_k^N) - \sum_{k=1}^{N} w_k^N \left(K_{\mathcal{H}_{1,\ldots,\infty}(\overline{\Omega_\sigma^{\mathrm{ext}}})}(\cdot\,, y_k^N), F\right)_{\mathcal{H}_{1,\ldots,\infty}(\overline{\Omega_\sigma^{\mathrm{ext}}})}\end{aligned} \tag{8.14}$$

holds for all $w^N = (w_1^N, \ldots, w_N^N)^T$ with $\sum_{k=1}^N w_k^N = 1$ (e.g., $w_k^N = 1/N$).

Lemma 8.7 (Koksma–Hlawka formula of approximation order m). *Let F be a member of class $\mathcal{H}(\overline{\Omega_\sigma^{\mathrm{ext}}})$ with $\{A_n\}$ being summable in the sense of Definition* 3.3. *Assume that $\{y_1^N, \ldots, y_N^N\} \subset \overline{\Omega_\sigma^{\mathrm{ext}}}$ is an $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\mathrm{ext}}})$-Dirichlet-fundamental system, i.e., a pointset on the sphere Ω_σ such that*

$$\begin{pmatrix} K_{\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\mathrm{ext}}})}(y_1^N, y_1^N) & \ldots & K_{\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\mathrm{ext}}})}(y_1^N, y_N^N) \\ \vdots & & \vdots \\ K_{\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\mathrm{ext}}})}(y_N^N, y_1^N) & \ldots & K_{\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\mathrm{ext}}})}(y_N^N, y_N^N) \end{pmatrix}$$

is regular. Then the integral formula

$$\begin{aligned}&\frac{1}{4\pi\sigma^2}\int_{\Omega_\sigma} F(y)\, d\omega(y)\\ &\quad = \sum_{k=1}^{N} w_k^N F(y_k^N) - \sum_{k=1}^{N} w_k^N \left(K_{\mathcal{H}_{m+1,\ldots,\infty}(\overline{\Omega_\sigma^{\mathrm{ext}}})}(\cdot\,, y_k^N), F\right)_{\mathcal{H}_{m+1,\ldots,\infty}(\overline{\Omega_\sigma^{\mathrm{ext}}})}\end{aligned} \tag{8.15}$$

holds for all $w^N = (w_1^N, \ldots, w_N^N)^T$, satisfying

$$\sum_{l=1}^{N} w_l^N = 1, \tag{8.16}$$

$$\sum_{l=1}^{N} w_l^N H_{n,k}(\sigma; y_l^N) = 0, \quad n = 1, \ldots, m,\ k = 1, \ldots, 2n+1. \tag{8.17}$$

Finally we are interested in an extension of the Koksma–Hlawka formula for spherical integrals (see Lemma 8.7) to $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\mathrm{ext}}})$-inner products. To this end we understand the summable sequence $\{A_n\}$ generating the reference space $\mathcal{H}(\overline{\Omega_\sigma^{\mathrm{ext}}})$ to be the symbol of a pseudodifferential operator A with $AH_{n,k}(\sigma;\cdot) = A^\wedge(n)H_{n,k}(\sigma;\cdot) = A_n H_{n,k}(\sigma;\cdot)$ for all $n \in \mathbb{N}_0$; $k = 1, \ldots, 2n+1$. Then the framework of the space $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\mathrm{ext}}})$ tells us that

$$\begin{aligned}F * P &= \sum_{n=0}^{m}\sum_{k=1}^{2n+1} F^\wedge(n,k)P^\wedge(n,k)\\ &= \int_{\Omega_\sigma} (AF)(y)(AP)(y)\, d\omega(y)\end{aligned} \tag{8.18}$$

holds for all $F \in \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ and $P \in \mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$. Moreover, we see that

$$\int_{\Omega_\sigma} (AF)(y)(AP)(y)\, d\omega(y) = \int_{\Omega_\sigma} F(y)(A^2P)(y)\, d\omega(y).$$

Clearly, A^2P is a member of class $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$ (as defined in the foregoing). Assuming F to be of class $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$, $F(A^2P)|_{\overline{\Omega_\sigma^{\text{ext}}}}$ is the product of two elements of class $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$, hence, $F(A^2P)|_{\overline{\Omega_\sigma^{\text{ext}}}}$ is a member of class $\mathcal{H}_{0,\dots,2m}(\overline{\Omega_\sigma^{\text{ext}}})$. In connection with Lemma 8.7 this leads us to the following result.

Lemma 8.8. *Let F and P be elements of class $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$.*

Assume that $\{y_1^N, \dots, y_N^N\} \subset \Omega_\sigma$ is an $\mathcal{H}_{0,\dots,2m}(\overline{\Omega_\sigma^{\text{ext}}})$-Dirichlet-fundamental system on Ω_σ (with $N \geq (2m+1)^2$). Then the identity

$$F * P = \sum_{k=1}^{N} w_k^N F(y_k^N)(A^2P)(y_k^N)$$

holds for all $w^N = (w_1^N, \dots, w_N^N)^T$ satisfying

$$\sum_{l=1}^{N} w_l^N = 1, \tag{8.19}$$

$$\sum_{l=1}^{N} w_l^N H_{n,k}(\sigma; y_l^N) = 0, \quad n = 1, \dots, 2m;\ k = 1, \dots, 2n+1. \tag{8.20}$$

In particular, we have

$$K_{\mathcal{H}_{0,\dots,m}} * F = \sum_{k=1}^{N} w_k^N F(y_k^N) K_{Harm_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot, y_k^N).$$

Lemma 8.8 is an essential tool for the development of "tree algorithms" (pyramid schemata) in bandlimited harmonic wavelet theory.

Lemma 8.9. *Let the system $\{y_1^M, \dots, y_M^M\} \subset \Omega_\sigma$, $M = (2m+1)^2$, define an $\mathcal{H}_{0,\dots,2m}(\overline{\Omega_\sigma^{\text{ext}}})$-Dirichlet-fundamental system. Furthermore, suppose that $P_{0,\dots,m}$, $Q_{0,\dots,m}$, respectively, are elements of class $\mathcal{H}_{0,\dots,m}(\Omega_\sigma)$. Then the identity*

$$P_{0,\dots,m} * Q_{0,\dots,m} = \sum_{n=1}^{M} b_n^M P_{0,\dots,m}(y_n^M)(A^2Q)_{0,\dots,m}(y_n^M) \tag{8.21}$$

holds for all weights $b_1^M, \dots, b_M^M$ satisfying

$$\sum_{r=1}^{M} b_r^M K_{\mathcal{H}_{0,\dots,2m}(\overline{\Omega_\sigma^{\text{ext}}})}(y_i^M, y_r^M) = \int_{\Omega_\sigma} K_{\mathcal{H}_{0,\dots,2m}(\overline{\Omega_\sigma^{\text{ext}}})}(y_i^M, x)\, d\omega(x), \quad i = 1, \dots, M. \tag{8.22}$$

Furthermore, we have the following results.

Lemma 8.10. *Let* $\{\mathcal{L}_1^M, \ldots, \mathcal{L}_M^M\}$, $M = (m+1)^2$, *be an* $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$*-fundamental system, and suppose that* $P_{0,\ldots,m}$ *and* $Q_{0,\ldots,m}$ *are members of* $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$*. Then the identity*

$$P_{0,\ldots,m} * Q_{0,\ldots,m} = \sum_{n=0}^{m} \sum_{k=1}^{2n+1} \sum_{r=1}^{M} d_r^{n,k} \left(Q_{0,\ldots,m} * H_{n,k}(\sigma;\cdot)\right) \mathcal{L}_r^M P_{0,\ldots,m} \tag{8.23}$$

holds for all weights $d_1^{n,k}, \ldots, d_M^{n,k}$*;* $n = 0, \ldots, m$*;* $k = 1, \ldots, 2n+1$*, satisfying the linear equations*

$$\sum_{r=1}^{M} d_r^{n,k} \mathcal{L}_r^M H_{l,i}(\sigma;\cdot) = \delta_{n,l}\delta_{k,i},$$

$l = 0, \ldots, m$*;* $i = 1, \ldots, 2l+1$.

In order to reduce the number of weights in our approximation rules we formulate the following lemma.

Lemma 8.11. *Under the assumptions of Lemma* 8.10*, the formula*

$$Q_{0,\ldots,m} * P_{0,\ldots,m} = \sum_{r=1}^{M} d_r^M \mathcal{L}_r^M P_{0,\ldots,m} \tag{8.24}$$

holds for all weights $d_1^M, \ldots, d_M^M$ *satisfying the linear equations*

$$\begin{aligned} &\sum_{r=1}^{M} d_r^M \mathcal{L}_i^M \mathcal{L}_r^M K_{\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) \\ &\quad = \sum_{n=0}^{m} \sum_{k=1}^{2n+1} \left(\mathcal{L}_i^M H_{n,k}(\sigma;\cdot)\right) Q_{0,\ldots,m} * H_{n,k}(\sigma;\cdot) = \mathcal{L}_i^M Q_{0,\ldots,m}, \end{aligned} \tag{8.25}$$

$i = 1, \ldots, M$.

It should be mentioned that on the one hand the number of integration weights is reduced, but on the other hand the integration weights depend on $Q_{0,\ldots,m}$. Other variants of discretization rules have been presented by W. Freeden and W. Schneider [30] which allow different aspects of approximation. In this work, however, we restrict ourselves to the above results (more explicitly, Lemma 8.9, Lemma 8.10, Lemma 8.11) based on linear systems of $\mathcal{O}(M)$-dimension.

In what follows the Runge concept is of basic interest. Once again, it tells us that to any potential $V \in Pot^{(0)}(\overline{\Sigma^{\text{ext}}})$ (for example, the Earth's gravitational potential) there exists a function F (namely, a Runge–Walsh approximation) harmonic in $\Omega_\sigma^{\text{ext}}$ and being regular at infinity in the sense that the absolute error becomes arbitrarily small on the whole space $\overline{\Sigma^{\text{ext}}}$. In this formulation as we already mentioned, the Runge–Walsh theorem is a pure existence theorem. It guarantees only the existence of an approximating potential and does not provide a method to find it. The theorem merely describes the theoretical background of approximating a potential by another one defined on a larger harmonicity domain. The

results developed now, however, enable us to derive a constructive version of the Runge–Walsh theorem by means of a J-level wavelet approximation when the potential F we are looking for is assumed to be a member of class $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})|_{\overline{\Sigma^{\rm ext}}}$ (note that $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})|_{\overline{\Sigma^{\rm ext}}}$is a uniformly dense subset of $Pot^{(0)}(\overline{\Sigma^{\rm ext}})$). Essential tools of our considerations are the approximation formulae formulated above.

Theorem 8.12. *Let $\{\mathcal{L}_1^M,\dots,\mathcal{L}_M^M\}$, $M=(m+1)^2$, be an $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\rm ext}})$-fundamental system. Furthermore, suppose that $\{y_1^{M_j},\dots,y_{M_j}^{M_j}\}\subset\Omega_\sigma$, $M_j=(2m_j+1)^2$, define $\mathcal{H}_{0,\dots,2m_j}(\overline{\Omega_\sigma^{\rm ext}})$-Dirichlet-fundamental systems for $j=0,\dots,J$. Moreover, assume that from a potential $F_{0,\dots,m}\in Harm_{0,\dots,m}(\overline{\Omega_\sigma^{\rm ext}})$ there are known the data $\mathcal{L}_i^M F_{0,\dots,m}=v_i$, $i=1,\dots,M$. Then, under our assumption of bandlimited wavelets, the fully discrete J-level wavelet approximation of $F_{0,\dots,m}$ reads as follows:*

(α)
$$\begin{aligned}\Phi_J^{(2)}*F_{0,\dots,m}&=\sum_{n=1}^{M_0}b_n^0\sum_{k=0}^{m}\sum_{l=1}^{2k+1}\sum_{s=1}^{M}d_s^{k,l}A_k^2\varphi_0(k)v_sH_{k,l}(\sigma;y_n^{M_0})\Phi_0(y_n^{M_0},\cdot)\\&\quad+\sum_{j=0}^{J-1}\sum_{n=1}^{M_j}b_n^j\sum_{k=0}^{m}\sum_{l=1}^{2k+1}\sum_{s=1}^{M}d_s^{k,l}A_k^2\psi_j(k)v_sH_{k,l}(\sigma;y_n^{M_j})\tilde{\Psi}_j(y_n^{M_j},\cdot),\end{aligned}\tag{8.26}$$

where the weights $d_1^{k,l},\dots,d_M^{k,l}$; $k=0,\dots,m$; $l=1,\dots,2k+1$, satisfy the linear equations

$$\sum_{s=1}^{M}d_s^{k,l}\mathcal{L}_s^MH_{n,i}(\sigma;\cdot)=\delta_{n,k}\delta_{i,l},\tag{8.27}$$

$n=0,\dots,m$; $i=1,\dots,2n+1$, and $b_1^j,\dots,b_{M_j}^j$; $j=0,\dots,J$, satisfy the linear equations

$$\sum_{n=1}^{M_j}b_n^jK_{\mathcal{H}_{0,\dots,2m_j}(\overline{\Omega_\sigma^{\rm ext}})}(y_i^{M_j},y_n^{M_j})=\int_{\Omega_\sigma}K_{\mathcal{H}_{0,\dots,2m_j}(\overline{\Omega_\sigma^{\rm ext}})}(y_i^{M_j},x)\,d\omega(x),\tag{8.28}$$

$i=1,\dots,M_j$.

(β)
$$\Phi_J^{(2)}*F_{0,\dots,m}=\sum_{n=1}^{M_0}b_n^0\sum_{s=1}^{M}\tilde{d}_s^{0,n}v_s\Phi_0(y_n^{M_0},\cdot)+\sum_{j=0}^{J-1}\sum_{n=1}^{M_j}b_n^j\sum_{s=1}^{M}d_s^{j,n}v_s\tilde{\Psi}_j(y_n^{M_j},\cdot),\tag{8.29}$$

where the weights $\tilde{d}_1^{0,n},\dots,\tilde{d}_M^{0,n}$; $n=1,\dots,M_0$, satisfy the linear equations

$$\sum_{s=1}^{M}\tilde{d}_s^{0,n}\mathcal{L}_i^M\mathcal{L}_s^MK_{Harm_{0,\dots,m}(\overline{\Omega_\sigma^{\rm ext}})}(\cdot,\cdot)=\mathcal{L}_i^M(A^2\Phi_0)(y_n^{M_0},\cdot),\tag{8.30}$$

$i = 1, \ldots, M$, and the weights $d_1^{j,n}, \ldots, d_M^{j,n}$; $j = 0, \ldots, J$; $n = 1, \ldots, M_j$, satisfy

$$\sum_{s=1}^{M} d_s^{j,n} \mathcal{L}_i^M \mathcal{L}_s^M K_{Harm_{0,\ldots,m(\overline{\Omega_\sigma^{\text{ext}}})}}(\cdot,\cdot) = \mathcal{L}_i^M (A^2 \Psi_j)(y_n^{M_j}, \cdot), \tag{8.31}$$

$i = 1, \ldots M$, and the coefficients $b_1^j, \ldots, b_{M_j}^j; j = 0, \ldots, J$ satisfy the linear system (8.28).

It should be remarked that a great number of linear systems must be solved in an a priori step. But if we look carefully we realize that we are always confronted with the same coefficient matrix. Having inverted the coefficient matrix once, all weights for numerical integration can be obtained by a matrix-vector multiplication and stored elsewhere (in an *a priori* step for computation). In addition, it should be mentioned that the solution of the linear systems determining the weights of the reconstruction step (8.28) can be avoided completely if we place the knots for numerical integration of the wavelet coefficients for each detail step $j = 0, \ldots, J-1$ on a special longitude-latitude grid on the sphere Ω_σ. The corresponding set of integration weights for reconstruction purposes are explicitly available without solving any linear system (for more details concerning numerical integration the reader is referred, e.g., to a paper due to Driscoll Healy [8]).

Until now the linear (observational) functionals have not been specified in more detail in our bandlimited wavelet approach presented above. In fact, the different types of linear functionals enable us to develop three important variants of wavelet approximation in the reality of gravitational potential determination:

(1) *Terrestrial-only Multiscale Approximation.* The linear functionals are understood to represent gravity observations (function values and/or derivatives) related to locations on the Earth's surface. If the data material is homogeneous, i.e., the linear functionals are all of the same type, terrestrial-only approximation reduces to the wavelet solution of a boundary-value problem of potential theory from discretely given data.

(2) *Spaceborne-only Multiscale Approximation.* In this case the linear functionals are understood to represent data measured by spacecraft in locations of $\overline{\Omega_\gamma^{\text{ext}}}$. As result we get a spaceborne-only approximation.

In practice, however, we are confronted with the situation that terrestrial, airborne as well as spaceborne data are available in gravitational potential determination (cf. [1, 2, 19, 22, 32, 45, 46, 50, 60, 62, 63, 66, 71, 75]). As a matter of fact, there are some areas on the continents (for example, some parts of Australia, Europe, and North-America), where the gravity field has been surveyed in much detail. Thus it is reasonable that such areas may be used for the verification or the calibration of the results obtained from spaceborne data.

(3) *Combined Multiscale Approximation.* Linear functionals representing terrestrial, airborne, and spaceborne observations are taken into account, i.e., numerical computation is required for a heterogeneous data set.

8.1. Runge–Walsh wavelet approximation of classical boundary value problems corresponding to regular surfaces

The wavelet representations (Theorem 8.12) of a bandlimited potential from a given finite set of linear functionals admit a variety of applications. The list includes the following examples of classical boundary value problems:

(i) *Dirichlet Problem.* First we are interested in the wavelet approximation $\Phi_J^{(2)} * F_{0,\dots,m}$ of the solution of the exterior Dirichlet problem

$$F_{0,\dots,m}|_{\overline{\Sigma^{\text{ext}}}} \in Harm_{0,\dots,m}(\overline{\Sigma^{\text{ext}}}), \quad F_{0,\dots,m}|_{\Sigma} = G_{0,\dots,m}.$$

under the knowledge of the $M = (m+1)^2$ boundary data

$$v_i = \mathcal{L}_i^M F_{0,\dots,m} = F_{0,\dots,m}(x_i^M) = G_{0,\dots,m}(x_i^M), \quad i = 1,\dots,M.$$

Theorem 8.13. *Under the assumptions of Theorem* 8.12 *the fully discrete J-level wavelet approximation of the solution of the exterior Dirichlet problem $F_{0,\dots,m}|_{\overline{\Sigma^{\text{ext}}}} \in Harm_{0,\dots,m}(\overline{\Sigma^{\text{ext}}})$, $(F_{0,\dots,m})|_{\Sigma} = G_{0,\dots,m}$ reads as follows:*

(α)
$$\begin{aligned}\Phi_J^{(2)} * F_{0,\dots,m} &= \sum_{n=1}^{M_0} b_n^0 \sum_{k=0}^{m} \sum_{l=1}^{2k+1} \sum_{s=1}^{M} d_s^{k,l} A_k^2 \varphi_0(k) G_{0,\dots,m}(x_s^M) H_{k,l}(\sigma; y_n^{M_0}) \Phi_0(y_n^{M_0}, \cdot) \\ &\quad + \sum_{j=0}^{J-1} \sum_{n=1}^{M_j} b_n^j \sum_{k=0}^{m} \sum_{l=1}^{2k+1} \sum_{s=1}^{M} d_s^{k,l} A_k^2 \psi_j(k) G_{0,\dots,m}(x_s^M) H_{k,l}(\sigma; y_n^{M_j}) \tilde{\Psi}_j(y_n^{M_j}; \cdot)\end{aligned} \tag{8.32}$$

(β)
$$\begin{aligned}\Phi_J^{(2)} * F_{0,\dots,m} &= \sum_{m=1}^{M_0} b_n^0 \sum_{s=1}^{M} \tilde{d}_s^{0,n} G_{0,\dots,m}(x_s^M) \Phi_0(y_n^{M_0}, \cdot) \\ &\quad + \sum_{j=0}^{J-1} \sum_{n=1}^{M_j} b_n^j \sum_{s=1}^{M} d_s^{j,n} G_{0,\dots,m}(x_s^M) \tilde{\Psi}_j(y_n^{M_j}, \cdot).\end{aligned} \tag{8.33}$$

The formulae (α), (β) of Theorem 8.13 are especially valid on the regular (Earth's) surface Σ, i.e., we automatically obtain by $\Phi_J^{(2)} * F_{0,\dots,m}|_{\Sigma}$ a J-level wavelet approximation of the "boundary function" $F_{0,\dots,m}|_{\Sigma} = G_{0,\dots,m}$ (by applying Shannon wavelets we even know that $\Phi_J^{(2)} * F_{0,\dots,m} = F_{0,\dots,m}$). In other words, a wavelet representation of a (bandlimited) function on regular surfaces has been found from a discrete data set of function values.

By treating non-bandlimited potentials $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $s > 1$, the developed integration formulae are only valid in approximate sense. To be more concrete, if $\Phi_J^{(2)} * F$ denotes the J-level wavelet approximation we actually calculate an approximation $\Phi_J^{(2)} * F_{0,\dots,m}$ by performing the numerical integration methods in (α), (β) of Theorem 8.13. Since this approximation also is harmonic in Σ^{ext} the biggest absolute error between $\Phi_J^{(2)} * F$ and its numerical approximation $\Phi_J^{(2)} * F_{0,\dots,m}$ is attained at the boundary Σ. Thus, the numerical error can be estimated by the use of the following theorem (cf. [29, 30]).

Theorem 8.14. *Let F satisfy $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}}), F|_\Sigma = G$, $s > 1$. Furthermore, assume that $X_M^\Sigma = \{x_1^M, \dots, x_M^M\} \subset \Sigma$, $M = (m+1)^2$, is an $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-Dirichlet-fundamental system on Σ. Then, for any $Q \in \mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$, we have*

$$\left| \int_{\Omega_\sigma} F(x)Q(x)\, d\omega(x) - \sum_{r=1}^{M} d_r G(x_r^M) \right| \leq \frac{C}{m^{s-1}} \left(\sum_{r=1}^{M} |d_r^M| \right) \|F\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}, \tag{8.34}$$

where C is a constant depending only on s and $d_1^M, \dots, d_M^M$ are the weights of the integration rule.

(ii) *Neumann Problem.* Now we are interested in the wavelet approximation $\Phi_J^{(2)} * F_{0,\dots,m}$ of the solution of the oblique Neumann problem

$$F_{0,\dots,m}|_{\overline{\Sigma^{\text{ext}}}} \in Harm_{0,\dots,m}(\overline{\Sigma^{\text{ext}}}), \quad \frac{\partial F_{0,\dots,m}}{\partial \lambda} = G_{0,\dots,m},$$

under the knowledge of the $M = (m+1)^2$ boundary data

$$v_i = \mathcal{L}_i^M F_{0,\dots,m} = \frac{\partial F_{0,\dots,m}}{\partial \lambda}(x_i^M) = G_{0,\dots,m}(x_i^M), \quad i = 1, \dots, M,$$

where $\lambda : \Sigma \to \mathbb{R}^3$ is a $C^{[1,\rho)}$-unit vector field (such that $0 < \rho < 1$ for $\lambda \neq \nu$ and $\rho = 0$ for $\lambda = \nu$) forming an angle with the outer normal ν satisfying

$$\inf_{x \in \Sigma} \nu(x) \cdot \lambda(x) > 0 \tag{8.35}$$

at any point of Σ.

Note that the boundedness of the linear functionals of the oblique derivative on Σ follows from well-known arguments (cf. [16, 18, 20]).

For the decomposition step we need in contrast to the Dirichlet problem an integration method in terms of oblique derivatives on Σ. From our results we obtain a fully discrete wavelet approximation for the solution of the exterior Neumann problem.

Theorem 8.15. *Let $X_M^\Sigma = \{x_1^M, \dots, x_M^M\} \subset \Sigma$, $M = (m+1)^2$, be an $\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-Neumann-fundamental system on Σ. Furthermore, let $X_{M_j} = \{y_1^{M_j}, \dots, y_{M_j}^{M_j}\}$, $M_j = (2m_j+1)^2$, be $\mathcal{H}_{0,\dots,2m_j}(\overline{\Omega_\sigma^{\text{ext}}})$-Neumann-fundamental systems on Ω_σ for $j = 0, \dots, J$. Moreover, assume that from a function $F_{0,\dots,m} \in \mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})$ there are known the oblique derivatives $G_{0,\dots,m} = (\partial F_{0,\dots,m}/\partial\lambda)$ at all points of X_M^Σ. Then, under our assumption of bandlimited wavelets, the fully discrete J-level wavelet approximation of the solution of the exterior Neumann problem $F_{0,\dots,m} \in$*

$\mathcal{H}_{0,\dots,m}(\overline{\Sigma^{\text{ext}}})$, $(\partial F_{0,\dots,m})/\partial\lambda = G_{0,\dots,m}$ *reads as follows:*

(α) $\Phi_J^{(2)} * F_{0,\dots,m}$

$$\begin{aligned}&= \sum_{n=1}^{M_0} b_n^0 \sum_{k=0}^{m}\sum_{l=1}^{2k+1}\sum_{s=1}^{M} d_s^{k,l} A_k^2 \varphi_0(k) G_{0,\dots,m}(x_s^M) H_{k,l}(\sigma; y_n^{M_0}) \Phi_0(y_n^{M_0},\cdot)\\ &\quad + \sum_{j=0}^{J-1}\sum_{n=1}^{M_j} b_n^j \sum_{k=0}^{m}\sum_{l=1}^{2k+1}\sum_{s=1}^{M} d_s^{k,l} A_k^2 \psi_j(k) G_{0,\dots,m}(x_s^M) H_{k,l}(\sigma; y_n^{M_j}) \tilde{\Psi}_j(y_n^{M_j},\cdot),\end{aligned} \tag{8.36}$$

where the weights $d_1^{k,l},\dots,d_M^{k,l}$; $k=0,\dots,m$; $l=1,\dots,2k+1$ *have to satisfy the linear equations*

$$\sum_{s=1}^{M} d_s^{k,l} \frac{\partial H_{n,i}(\sigma; x_s^M)}{\partial\lambda} = \delta_{n,k}\delta_{i,l}, \quad n=0,\dots,\ m; i=1,\dots,2n+1,$$

and $b_1^j,\dots,b_M^j$, $j=0,\dots,J$ *must satisfy the linear equations* (8.28).

(β)

$$\begin{aligned}&\Phi_J^{(2)} * F_{0,\dots,m}\\ &\quad = \sum_{n=1}^{M_0} b_n^0 \sum_{s=1}^{M} \tilde{d}_s^{0,n} G_{0,\dots,m}(x_s^M) \Phi_0(y_n^{M_0},\cdot)\\ &\qquad + \sum_{j=0}^{J-1}\sum_{n=1}^{M_j} b_n^j \sum_{s=1}^{M} d_s^{j,n} G_{0,\dots,m}(x_s^M) \tilde{\Psi}_j(y_n^{M_j},\cdot),\end{aligned} \tag{8.37}$$

where the weights $\tilde{d}_1^{0,n},\dots,\tilde{d}_M^{0,n}$; $n=1,\dots,M_0$, *have to satisfy the linear equations*

$$\sum_{s=1}^{M} \tilde{d}_s^{0,n} \frac{\partial}{\partial\lambda_{y_i^M}} \frac{\partial}{\partial\lambda_{y_s^M}} K_{\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) = \frac{\partial}{\partial\lambda_{y_i^M}} (A^2\Phi_0)(y_n^{M_0},\cdot), \tag{8.38}$$

$i=1,\dots,M$, *and the weights* $d_1^{j,n},\dots,d_M^{j,n}$; $j=0,\dots,J$; $n=1,\dots,M_j$, *must satisfy*

$$\sum_{s=1}^{M} d_s^{j,n} \frac{\partial}{\partial\lambda_{y_i^M}} \frac{\partial}{\partial\lambda_{y_s^M}} K_{\mathcal{H}_{0,\dots,m}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot) = \frac{\partial}{\partial\lambda_{y_i^M}} (A^2\Psi_j)(y_n^{M_j},\cdot), \tag{8.39}$$

$i=1,\dots,M$, *and* $b_1^j,\dots,b_{M_j}^j$; $j=0,\dots,J$, *satisfy the linear equations* (8.28).

The formulae (α), (β) of Theorem 8.15 are especially valid on Σ. Thus, we obtain by $\partial(\Phi_J^{(2)} * F_{0,\dots,m})/\partial\lambda$ a J-level wavelet approximation of $G_{0,\dots,m} = \partial F_{0,\dots,m}/\partial\lambda$.

In order to examine the error in the integration formulae when we turn over to non-bandlimited potentials we finally mention the following theorem.

Theorem 8.16. *Let F satisfy $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $\frac{\partial F}{\partial \lambda} = G$, $s > 2$. Furthermore, let $X_M^\Sigma = \{x_1^M, \ldots, x_M^M\} \subset \Sigma$, $M = (m+1)^2$, be an $\mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$-Neumann-fundamental system on Σ. Then, for any $Q \in \mathcal{H}_{0,\ldots,m}(\overline{\Omega_\sigma^{\text{ext}}})$, we have*

$$\left| \int_{\Omega_\sigma} F(x)Q(x)\, d\omega(x) - \sum_{r=1}^{M} d_r^M G(x_r^M) \right| \leq \frac{C}{m^{s-2}} \left(\sum_{r=1}^{M} |d_r^M| \right) \|F\|_{\mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})}, \tag{8.40}$$

where C is a constant depending only on s and $d_1^M, \ldots, d_M^M$ are the weights of the integration rule.

Hence, by treating non-bandlimited potentials $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$, $s > 2$, we obtain in similarity to the Dirichlet case a J-level wavelet approximation by performing the numerical rules as indicated by (α), (β) of Theorem 8.15, and the numerical errors can be estimated using Theorem 8.16.

Remark 8.17. The existence of all types of fundamental systems to be needed in our preceding approximation rules is guaranteed by a well-known induction procedure (as described, for example in [21, 24, 57]. Furthermore, more detailed remainder estimates for the integration formulae can be found in [28]).

8.2. Pyramid schemata based on outer harmonic exact approximation

Our purpose now is to use two variants of exact (outer harmonic) approximation to derive tree algorithms, i.e., pyramid schemata for fast evaluation of bandlimited potentials. Without loss of generality, we assume that $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$, $\{\Psi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$, and $\{\tilde{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ are families of bandlimited kernels satisfying the conditions (8.2) and (8.3). Variant 1 is based on the ideas of Lemma 8.8 using evaluation (i.e. Dirichlet functionals) on a sphere, while Variant 2 is based on the Shannon sampling Theorem 8.5 in terms of linear functionals. Both variants are particularly suitable for application to medium wavelength parts of a signal (potential). As shown in [19], Variant 2 can be extended to non-bandlimited potentials. This variant is therefore also suitable for the transition from medium to short wavelength parts of a signal (potential).

Variant 1. The key ideas of our first discretization method using outer harmonic exact approximation formulae are based on the following observations:

(1) For some suitably large J, the scale space $\mathcal{V}_{J+1}(\overline{\Omega_\sigma^{\text{ext}}}) = \mathcal{H}_{0,\ldots,2^{J+1}-1}(\overline{\Omega_\sigma^{\text{ext}}})$ is "sufficiently close" to $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$. Consequently, for each potential $F \in \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$, there exists a bandlimited potential of class $\mathcal{V}_{J+1}(\overline{\Omega_\sigma^{\text{ext}}})$ such that the error between F and $\Phi_{J+1}^{(2)} * F$ (understood in $\|\cdot\|_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}$-topology) is negligible. This is the reason why the input data $v_l^{N_J}$, $l = 1, \ldots, N_J$, are assumed to be given from a potential of class $\mathcal{V}_{J+1}(\overline{\Omega_\sigma^{\text{ext}}})$ (for the remainder of this subsection).

(2) For $j = 0, \dots, J$, the generating coefficients $b_l^{N_j}$ and nodal points $y_l^{N_j} \in \Omega_\sigma$ of the exact outer harmonic formulae of order $2^{j+2} - 2 (= 2 \cdot (2^{j+1} - 1))$ (cf. Lemma 8.8) are determined such that

$$K_{\mathcal{H}_{0,\dots,2^j-1}(\overline{\Omega_\sigma^{\rm ext}})} * P = \sum_{l=1}^{N_j} b_l^{N_j} K_{Harm_{0,\dots,2^j-1}(\overline{\Omega_\sigma^{\rm ext}})}(\cdot, y_l^{N_j}) P(y_l^{N_j})$$

holds for all $P \in \mathcal{H}_{0,\dots,2^j-1}(\overline{\Omega_\sigma^{\rm ext}})$ with $N_j \geq ((2^{j+2} - 2) + 1)^2 = (2^{j+2} - 1)^2$. The coefficients $b_l^{N_j}$ may be calculated from the linear equations

$$\begin{aligned} &\sum_{l=1}^{N_j} b_l^{N_j} K_{\mathcal{H}_{0,\dots,2^{j+2}-2}(\overline{\Omega_\sigma^{\rm ext}})}(y_i^{N_j}, y_l^{N_j}) \\ &\qquad = \frac{1}{4\pi\sigma^2} \int_{\Omega_\sigma} K_{\mathcal{H}_{0,\dots,2^{j+2}-2}(\overline{\Omega_\sigma^{\rm ext}})}(x, y_i^{N_j})\, d\omega(x), \end{aligned} \tag{8.41}$$

$i = 1, \dots, N_j$, in an a priori step and stored elsewhere.

Our goal is to show that all convolutions occurring in the J-level wavelet approximation of a bandlimited potential (of order $2^{J+1} - 1$) can be evaluated *exactly* by means of outer harmonic approximation formulae. As a matter of fact, what we realize is the following *pyramid scheme*: Starting from a sufficiently large J, there exist vectors $a^{N_j} \in \mathbb{R}^{N_j}$, $j = 0, \dots, J$ (being, of course, dependent on the potential $F \in \mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ under consideration) such that the following statements hold true:

(i) For $j = 0, \dots, J$, all wavelet coefficients can be calculated via the formulae

$$(WT)(F)(j; \cdot) = \sum\nolimits_{i=1}^{N_j} a_i^{N_j} \Psi_j(\cdot, y_i^{N_j}).$$

(ii) The vectors $a^j \in \mathbb{R}^{N_j}$ are obtainable from $a^{j+1} \in \mathbb{R}^{N_{j+1}}$ by recursion:

$$a_i^{N_j} = b_i^{N_j} \sum\nolimits_{l=1}^{N_{j+1}} a_l^{N_{j+1}} K_{Harm_{0,\dots,2^{j+1}-1}(\overline{\Omega_\sigma^{\rm ext}})}(y_i^{N_j}, y_l^{N_{j+1}}),$$

$i = 1, \dots, N_j$.

(iii) The vectors satisfy, in addition, the identities

$$\Phi_{j+1}^{(2)} * F = \sum_{i=1}^{N_j} a_i^{N_j} \Phi_{j+1}^{(2)}(\cdot, y_i^{N_j})$$

and

$$(\tilde{\Psi}_j * \Psi_j) * F = \sum_{i=1}^{N_j} a_i^{N_j} (\tilde{\Psi}_j * \Psi_j)(\cdot, y_i^{N_j}).$$

Our considerations are divided into two parts, viz. the initial step concerning the scale level J and the pyramid step establishing the recursion relation.

The Initial Step. Observing the exact (outer harmonic) formulae we obtain from Lemma 8.8 for all potentials $F \in \mathcal{V}_{J+1}(\overline{\Omega_\sigma^{\text{ext}}}) = \mathcal{H}_{0,\dots,2^{J+1}-1}(\overline{\Omega_\sigma^{\text{ext}}})$

$$K_{\mathcal{H}_{0,\dots,2^{J+1}-1}(\overline{\Omega_\sigma^{\text{ext}}})} * F = \sum_{l=1}^{N_J} b_l^{N_J} F(y_l^{N_J}) K_{Harm_{0,\dots,2^{J+1}-1}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot, y_l^{N_J}).$$

It follows that $a^{N_J} \in \mathbb{R}^{N_J}$, $a^{N_J} = (a_1^{N_J}, \dots, a_{N_J}^{N_J})^T$, given by

$$a_l^{N_J} = b_l^{N_J} F(y_l^{N_J}) = b_l^{N_J} v_l^{N_J}, \quad l = 1, \dots, N_J, \tag{8.42}$$

satisfies the equation

$$K_{\mathcal{H}_{0,\dots,2^{J+1}-1}(\overline{\Omega_\sigma^{\text{ext}}})} * F = \sum_{i=1}^{N_J} a_i^{N_J} K_{Harm_{0,\dots,2^{J+1}-1}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot, y_i^{N_J}).$$

Note that the coefficients $a_i^{N_J}$ are dependent on F. Again Lemma 8.8 now implies the following result.

Lemma 8.18. *Let F be of class $\mathcal{V}_{J+1}(\overline{\Omega_\sigma^{\text{ext}}}) = \mathcal{H}_{0,\dots,2^{J+1}-1}(\overline{\Omega_\sigma^{\text{ext}}})$. Suppose that $K(\cdot,\cdot)$ is (an $\mathcal{H}_{\sigma,\sigma}$-kernel) such that $K^\wedge(n) = 0$ for all $n > 2^{J+1} - 1$. Then the coefficients* (8.42) *satisfy the equation*

$$K * F = \sum_{i=1}^{N_J} a_i^{N_J} A^2 K(\cdot, y_i^{N_J}).$$

It should be noted that

$$A^2 K(x,y) = \sum_{n=0}^{2^{J+1}-1} A_n^2 K^\wedge(n) \sum_{k=1}^{2n+1} H_{n,k}^*(\sigma; x) H_{n,k}(\sigma; y) \tag{8.43}$$

for all $(x,y) \in \overline{\Omega_\sigma^{\text{ext}}} \times \overline{\Omega_\sigma^{\text{ext}}}$. Furthermore, the vector a^{N_J} is independent of the choice of the $\mathcal{H}_{\sigma,\sigma}$-kernel $K(\cdot,\cdot)$.

As special cases we obtain from Lemma 8.18 the following identities:

$$\Phi_{J+1} * F = \sum_{i=1}^{N_J} a_i^{N_J} A^2 \Phi_{J+1}(\cdot, y_i^{N_J}), \tag{8.44}$$

$$(\Phi_{J+1} * \Phi_{J+1}) * F = \sum_{i=1}^{N_J} a_i^{N_J} A^2 (\Phi_{J+1} * \Phi_{J+1})(\cdot, y_i^{N_J}), \tag{8.45}$$

and

$$\Psi_J * F = \sum_{i=1}^{N_J} a_i^{N_J} A^2 \Psi_J(\cdot, y_i^{N_J}), \tag{8.46}$$

$$(\tilde{\Psi}_J * \Psi_J) * F = \sum_{i=1}^{N_J} a_i^{N_J} A^2 (\tilde{\Psi}_J * \Psi_J)(\cdot, y_i^{N_J}). \tag{8.47}$$

The Pyramid Step. An essential tool for the pyramid step is the following lemma.

Lemma 8.19. *Let F be of class $\mathcal{V}_{J+1}(\overline{\Omega_\sigma^{\text{ext}}})$. Suppose that $K(\cdot,\cdot)$ is an $\mathcal{H}_{\sigma,\sigma}$-kernel with $K^\wedge(n) = 0$ for all $n > 2^J - 1$. Then the vector $a^{N_{J-1}} \in \mathbb{R}^{N_{J-1}}$, $a^{N_{J-1}} = (a_1^{N_{J-1}}, \ldots, a_{N_{J-1}}^{N_{J-1}})^T$, given by*

$$a_i^{N_{J-1}} = b_i^{N_{J-1}}(K_{\mathcal{H}_{0,\ldots,2^J-1}(\overline{\Omega_\sigma^{\text{ext}}})} * F)(y_i^{N_{J-1}}), \quad i = 1, \ldots, N_{J-1},$$

satisfies the equation

$$K * F = \sum_{i=1}^{N_{J-1}} a_i^{N_{J-1}} A^2 K(\cdot, y_i^{N_{J-1}}).$$

Suppose that $K(\cdot,\cdot)$ satisfies the assumption of Lemma 8.19. Looking at our foregoing results we notice that there are two ways of discretizing an $\mathcal{H}$-convolution $K * F$. On the one hand we obtain from Lemma 8.18

$$K * F = \sum_{i=1}^{N_J} a_i^{N_J} A^2 K(\cdot, y_i^{N_J}) \tag{8.48}$$

with coefficients $a_1^{N_J}, \ldots, a_{N_J}^{N_J}$ given by

$$a_i^{N_J} = b_i^{N_J} F(y_i^{N_J}) = b_i^{N_J} v_i^{N_J}, \quad i = 1, \ldots, N_J. \tag{8.49}$$

It is remarkable that the coefficients are independent of the choice of the kernel $K(\cdot,\cdot)$. As particularly important case we mention

$$K_{\mathcal{H}_{0,\ldots,2^J-1}(\overline{\Omega_\sigma^{\text{ext}}})} * F = \sum_{i=1}^{N_J} a_i^{N_J} K_{Harm_{0,\ldots,2^J-1}(\overline{\Omega_\sigma^{\text{ext}}})}(y_i^{N_J}, \cdot). \tag{8.50}$$

On the other hand, we are able to deduce from Lemma 8.19 that

$$K * F = \sum_{i=1}^{N_{J-1}} a_i^{N_{J-1}} A^2 K(\cdot, y_i^{N_{J-1}}) \tag{8.51}$$

with coefficients $a_1^{N_{J-1}}, \ldots, a_{N_{J-1}}^{N_{J-1}}$ given by

$$a_i^{N_{J-1}} = b_i^{N_{J-1}}(K_{\mathcal{H}_{0,\ldots,2^J-1}(\overline{\Omega_\sigma^{\text{ext}}})} * F)(y_i^{N_{J-1}}), \tag{8.52}$$

$i = 1, \ldots, N_{J-1}$. Inserting (8.50) into (8.52) we find

$$a_i^{N_{J-1}} = b_i^{N_{J-1}} \sum_{l=1}^{N_J} a_l^{N_J} K_{Harm_{0,\ldots,2^J-1}(\overline{\Omega_\sigma^{\text{ext}}})}(y_i^{N_{J-1}}, y_l^{N_J}) \tag{8.53}$$

for $i = 1, \ldots, N_{J-1}$. In other words, the coefficients $a_i^{N_{J-1}}$ can be calculated recursively. Moreover, the coefficients are independent of the special choice of the kernel

$K(\cdot,\cdot)$. This finally leads us to the following discretization of the $\mathcal{H}$-convolutions

$$\Phi_J * F = \sum_{i=1}^{N_{J-1}} a_i^{N_{J-1}} A^2 \Phi_J(\cdot, y_i^{N_{J-1}}), \tag{8.54}$$

$$(\Phi_J * \Phi_J) \ *_{\mathcal{H}} \ F = \sum_{i=1}^{N_{J-1}} a_i^{N_{J-1}} A^2 (\Phi_J * \Phi_J)(\cdot, y_i^{N_{J-1}}), \tag{8.55}$$

and

$$\Psi_{J-1} * F = \sum_{i=1}^{N_{J-1}} a_i^{N_{J-1}} A^2 \Psi_{J-1}(\cdot, y_i^{N_{J-1}}), \tag{8.56}$$

$$(\tilde{\Psi}_{J-1} * \Psi_{J-1}) * F = \sum_{i=1}^{N_{J-1}} a_i^{N_{J-1}} A^2 (\tilde{\Psi}_{J-1} * \Psi_{J-1})(\cdot, y_i^{N_{J-1}}). \tag{8.57}$$

In conclusion, we end up with the following pyramid scheme for the decomposition of a potential F:

$$\begin{array}{ccccccccc} F & \longrightarrow & a^{N_J} & \longrightarrow & a^{N_{J-1}} & \longrightarrow & \cdots & \longrightarrow & a^{N_0} \\ & & \downarrow & & \downarrow & & & & \downarrow \\ & & (WT)(F)(J;\cdot) & & (WT)(F)(J-1;\cdot) & & & & (WT)(F)(0;\cdot). \end{array}$$

The reconstruction of the wavelet coefficients can be performed as described before via the formula

$$\begin{aligned} R_j(F) &= \tilde{\Psi}_j * (WT)(F)(j;\cdot) \\ &= \sum_{i=1}^{N_j} b_i^{N_j} (WT)(F)(j; y_i^{N_j}) A^2 \tilde{\Psi}_j(\cdot, y_i^{N_j}). \end{aligned} \tag{8.58}$$

This leads us to the following scheme:

$$\begin{array}{ccccccccc} (WT)(F)(0; y_i^{N_0}) & & & & (WT)(F)(1; y_i^{N_1}) & & & & \\ \downarrow & & & & \downarrow & & & & \\ R_0(F) & & & & R_1(F) & & & & \\ & \searrow & & & & \searrow & & & \\ P_0(F) & \rightarrow & + & & P_1(F) & \rightarrow & + & \cdots & . \end{array}$$

According to our approach the wavelet transform $(WT)(F)(j;\cdot)$ is given by the coefficients $a_1^{N_j}, \ldots, a_{N_j}^{N_j}$. This also enables us to reconstruct the potential only by use of the coefficients $a_i^{N_j}$, rather than calculating the wavelet coefficients of F:

$$R_j(F) = \sum_{i=1}^{N_j} a_i^{N_j} A^2 (\tilde{\Psi}_j * \Psi_j)(\cdot, y_i^{N_j}).$$

Thus the decomposition and reconstruction, respectively, can be simplified as follows:

$$F \to a^{N_J} \to a^{N_{J-1}} \to \cdots \to a^{N_0}$$

and

$$\begin{array}{ccccccccccccc} a^{N_0} & & & & a^{N_1} & & & & a^{N_2} & & & & \\ \downarrow & & & & \downarrow & & & & \downarrow & & & & \\ R_0(F) & & & & R_1(F) & & & & R_2(F) & & & & \\ & \searrow & & & & \searrow & & & & \searrow & & & \\ P_0(F) & \to & + & \to & P_1(F) & \to & + & \to & P_2(F) & \to & + & \to \cdots . \end{array}$$

That means the reconstruction of the potential is not performed with $\tilde{\Psi}_j$. Instead we have used the $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$-convolution $\tilde{\Psi}_j * \Psi_j$. Of particular significance is that the vectors a^{N_j} do not depend on the special choice of the bandlimited scaling function. As a matter of fact, we are able to reconstruct the potential with respect to different types of wavelets just by use of the vectors a^{N_j}.

Remark 8.20. The critical point of our pyramid scheme is the determination of the coefficients $b_l^{N_j}$, $j = 0, \ldots, J$, from the linear system (8.41) which provides outer harmonic exactness up to the order $2^{j+2} - 2$. It should be mentioned that the solution of this linear system can be avoided completely if we place the knots for each detail step $j = 0, \ldots, J$ on a spherical longitude-latitude grid on the sphere Ω_σ. The corresponding set of weights is explicitly available without solving any linear system from results due to [8].

Variant 2. In what follows we use outer harmonic exact approximation (Lemma 8.5) to develop a bandlimited variant of the pyramid scheme based on the Shannon sampling theorem. Our approach consists of the following steps:

(i) According to our bandlimited wavelet approach the (reference) Sobolev space $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ is subdivided by a nested sequence of 2^{2j}-dimensional scale spaces $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ as follows: $\cdots \subset \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) \subset \mathcal{V}_{j+1}(\overline{\Omega_\sigma^{\text{ext}}}) \subset \cdots \subset \mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$.

(ii) $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, $j \in \mathbb{N}_0$, can be identified with the set

$$\mathcal{H}_{0,\ldots,2^j-1}(\overline{\Omega_\sigma^{\text{ext}}}) = \mathcal{H}(\{A_n/(\varphi_j(n))^2\}; \overline{\Omega_\sigma^{\text{ext}}}),$$

and $\Phi_j^{(4)}(\cdot,\cdot)$ is the uniquely determined reproducing kernel in $(\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}), (\cdot,\cdot)_{\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})})$ with $(\cdot,\cdot)_{\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})}$ given by

$$(\cdot,\cdot)_{\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})} = (\cdot,\cdot)_{\mathcal{H}(\{A_n/(\varphi_j(n))^2\}; \overline{\Omega_\sigma^{\text{ext}}})}.$$

(iii) For each $j \in \mathbb{N}_0$, consider sequences $\{\mathcal{L}_1^{N_j}, \ldots, \mathcal{L}_{N_j}^{N_j}\}$ of $N_j \geq 2^{2j}$ (linearly independent) bounded linear functionals on $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ such that

$$\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) = \text{span}\left(\mathcal{L}_1^{N_j}\Phi_j^{(4)}(\cdot,\cdot), \ldots, \mathcal{L}_{N_j}^{N_j}\Phi_j^{(4)}(\cdot,\cdot)\right).$$

Then it also follows that

$$\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) = \text{span}\left(\mathcal{L}_1^{N_j}\Phi_j^{(2)}(\cdot,\cdot), \ldots, \mathcal{L}_{N_j}^{N_j}\Phi_j^{(2)}(\cdot,\cdot)\right).$$

(iv) $\mathcal{V}_j(\overline{\Omega_\sigma^{\rm ext}})$, $j \in \mathbb{N}_0$, can be identified with the set $\mathcal{H}(\{A_n/\varphi_j(n)\};\overline{\Omega_\sigma^{\rm ext}})$, and $\Phi_j^{(2)}(\cdot,\cdot)$ is the reproducing kernel in $\left(\mathcal{V}_j(\overline{\Omega_\sigma^{\rm ext}}),(\cdot,\cdot)_{\mathcal{V}_j^{(1/2)}(\overline{\Omega_\sigma^{\rm ext}})}\right)$ with $(\cdot,\cdot)_{\mathcal{V}_j^{(1/2)}(\overline{\Omega_\sigma^{\rm ext}})}$ defined by

$$(\cdot,\cdot)_{\mathcal{V}_j^{(1/2)}(\overline{\Omega_\sigma^{\rm ext}})} = (\cdot,\cdot)_{\mathcal{H}(\{A_n/\varphi_j(n)\};\overline{\Omega_\sigma^{\rm ext}})}.$$

The key idea of our fast evaluation method using the Shannon sampling theorem in terms of linear functionals is based on the following observations:

(1) For some suitably large J, the scale space $\mathcal{V}_J(\overline{\Omega_\sigma^{\rm ext}})$ is "sufficiently close" to $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$. Consequently, for each $F \in \mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$, there exists a function of class $\mathcal{V}_J(\overline{\Omega_\sigma^{\rm ext}})$ such that the error between F and $\Phi_J^{(2)} * F$ (understood in $\|\cdot\|_{\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})}$-topology) is negligible. This is the reason why the input data $v_k^{N_J} = \mathcal{L}_k^{N_J} F$, $k = 1,\dots,N_J$, are assumed to be of a potential F of class $\mathcal{V}_J(\overline{\Omega_\sigma^{\rm ext}})$ for the remainder of this subsection.

(2) For $j = 0,\dots,J$, consider sequences $\{\mathcal{L}_1^{N_j},\dots,\mathcal{L}_{N_j}^{N_j}\}$ of $N_j \geq 2^{2j}$ (linearly independent) bounded linear functionals on $\mathcal{H}(\overline{\Omega_\sigma^{\rm ext}})$ such that

$$\mathcal{V}_j(\overline{\Omega_\sigma^{\rm ext}}) = \mathcal{H}_{0,\dots,2^j-1}(\overline{\Omega_\sigma^{\rm ext}}) = \operatorname{span}\left(\mathcal{L}_1^{N_j}\Phi_j^{(2)}(\cdot,\cdot),\dots,\mathcal{L}_{N_j}^{N_j}\Phi_j^{(2)}(\cdot,\cdot)\right).$$

In an *a priori* step the coefficients $w_{l,k}^{N_j}$ have to be determined from the systems of linear equations (see Lemma 8.5)

$$\sum_{l=1}^{N_j} w_{l,k}^{N_j}\mathcal{L}_i^{N_j}\mathcal{L}_l^{N_j}\Phi_j^{(2)}(\cdot,\cdot) = \delta_{i,k}, \quad i,k = 1,\dots,N_j,$$

$j = 0,\dots,J$, and can be stored elsewhere. Looking carefully at the linear systems, it can be recognized that the coefficients $w_{l,k}^{N_j}$ do not depend on the particular function F under consideration, but only on the chosen linear functionals and pointsets.

Next our considerations are divided into two parts, viz. the initial step concerning the scale level J and the pyramid step establishing the recursion relation.

The Initial Step. The exact approximation

$$J_{N_J}S = \sum_{i=1}^{N_J} a_i^{N_J}\mathcal{L}_i^{N_J}S, \quad S \in \mathcal{V}_J(\overline{\Omega_\sigma^{\rm ext}}),$$

to the bounded linear functionals $\mathcal{L}$ on $\mathcal{V}_J(\overline{\Omega_\sigma^{\rm ext}})$ defined by

$$\mathcal{L}S = (S,F)_{\mathcal{V}_J^{(1/2)}(\overline{\Omega_\sigma^{\rm ext}})} = S *_{\mathcal{V}_J^{(1/2)}} F, \quad S \in \mathcal{V}_J(\overline{\Omega_\sigma^{\rm ext}}),\ F \in \mathcal{V}_J(\overline{\Omega_\sigma^{\rm ext}}),$$

is given by

$$a_i^{N_J} = \sum_{k=1}^{N_J} w_{i,k}^{N_J}\mathcal{L}\mathcal{L}_k^{N_J}\Phi_J^{(2)}(\cdot,\cdot), \quad i = 1,\dots,N_J.$$

Note that in order to clarify the convolution we use a lower index at the symbol "$*$" in the following text if necessary. In accordance with our assumption $F \in \mathcal{V}_J(\overline{\Omega_\sigma^{\text{ext}}})$ and the reproducing property of $\Phi_J^{(2)}(\cdot,\cdot)$ in $\mathcal{V}_J^{(1/2)}(\overline{\Omega_\sigma^{\text{ext}}})$ we see that $\Phi_J^{(2)} \ *_{\mathcal{V}_J^{(1/2)}} F = F$. Thus we find

$$a_i^{N_J} = \sum_{k=1}^{N_J} w_{i,k}^{N_J} (\mathcal{L}_k^{N_J} \Phi_J^{(2)}(\cdot,\cdot) \ *_{\mathcal{V}_J^{(1/2)}} F) = \sum_{k=1}^{N_J} w_{i,k}^{N_J} \mathcal{L}_k^{N_J} F = \sum_{k=1}^{N_J} w_{i,k}^{N_J} v_k^{N_J}$$

for $i = 1, \ldots, N_J$. This leads us to the following conclusion.

Lemma 8.21. *If F is a member of class $\mathcal{V}_J(\overline{\Omega_\sigma^{\text{ext}}})$, then the identity*

$$S \ *_{\mathcal{V}_J^{(1/2)}} \ F = \sum_{i=1}^{N_J} a_i^{N_J} \mathcal{L}_i^{N_J} S$$

holds for all $S \in \mathcal{V}_J(\overline{\Omega_\sigma^{\text{ext}}})$.

Lemma 8.21 immediately enables us to formulate the following lemma.

Lemma 8.22. *Let F be a member of class $\mathcal{V}_J(\overline{\Omega_\sigma^{\text{ext}}})$, then the identity*

$$K * F = \sum_{i=1}^{N_J} a_i^{N_J} \mathcal{L}_i^{N_J} K(\cdot,\cdot)$$

holds for all $\mathcal{H}_{\sigma,\sigma}$-kernels $K(\cdot,\cdot)$ with $K\hat{}(n) = 0$ for $n = 2^J, ^J + 1, \ldots$.

The next theorem clarifies the remarkable consequences for our wavelet concept.

Theorem 8.23. *Under the assumptions of Lemma* 8.22 *we have*

$$\Phi_J * F = \sum_{i=1}^{N_J} a_i^{N_J} \mathcal{L}_i^{N_J} \Phi_J(\cdot,\cdot), \tag{8.59}$$

$$(\Phi_J * \Phi_J) * F = \sum_{i=1}^{N_J} a_i^{N_J} \mathcal{L}_i^{N_J} (\Phi_J * \Phi_J)(\cdot,\cdot), \tag{8.60}$$

and

$$\Psi_{J-1} * F = \sum_{i=1}^{N_J} a_i^{N_J} \mathcal{L}_i^{N_J} \Psi_{J-1}(\cdot,\cdot), \tag{8.61}$$

$$(\tilde{\Psi}_{J-1} * \Psi_{J-1}) * F = \sum_{i=1}^{N_J} a_i^{N_J} \mathcal{L}_i^{N_J} (\tilde{\Psi}_{J-1} * \Psi_{J-1})(\cdot,\cdot). \tag{8.62}$$

In conclusion, the vector $a^{N_J} = (a_1^{N_J}, \ldots, a_{N_J}^{N_J})^T \in \mathbb{R}^{N_J}$ does <u>not</u> depend on the special choice of the $\Phi_J^{(2)}(\cdot,\cdot)$-kernel in $\mathcal{V}_J(\overline{\Omega_\sigma^{\text{ext}}})$. Wavelet transform, lowpass, and bandpass filter can be computed by use of the same set of coefficients.

The Pyramid Step. This step provides an algorithm such that $a^{N_J} \in \mathbb{R}^{N_J}$ serves as starting vector for $a^{N_j} \in \mathbb{R}^{N_j}$, $j = 0, \ldots, J-1$, which fulfill the following properties:

(i) The vectors a^{N_j} satisfy

$$\Phi_j^{(2)} * F = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} \Phi_j^{(2)}(\cdot, \cdot),$$

$j = 0, \ldots, J$.

(ii) The wavelet transforms are given by

$$\Psi_{j-1} * F = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} \Psi_{j-1}(\cdot, \cdot),$$

$j = 1, \ldots, J$.

(iii) The vector a^{N_j} is obtainable from $a^{N_{j+1}}$, $j = 0, \ldots, J-1$, by recursion.

In the remainder of this section the properties (i), (ii) and (iii) are described in more detail. The exact approximations J_{N_j}, $j = 0, \ldots, J-1$,

$$J_{N_j} S = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} S, \quad S \in \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$$

to the bounded linear functional $\mathcal{L}$ on $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$ defined by

$$\mathcal{L}S = S \ *_{\mathcal{V}_j^{(1/2)}} \ (\Phi_j^{(2)} \ *_{\mathcal{H}} \ F), \quad S \in \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}),\ F \in \mathcal{V}_J(\overline{\Omega_\sigma^{\text{ext}}}),$$

(note that $\Phi_j^{(2)} \ *_{\mathcal{H}} \ F \in \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$) are given by the coefficients

$$a_l^{N_j} = \sum_{i=1}^{N_j} w_{l,i}^{N_j} \mathcal{L}_i^{N_j} \Phi_j^{(2)}(\cdot, \cdot), \quad l = 1, \ldots, N_j.$$

Consequently it is easily seen that for $l = 1, \ldots, N_j$

$$a_l^{N_j} = \sum_{i=1}^{N_j} w_{l,i}^{N_j} \mathcal{L}_i^{N_j} (\Phi_j^{(2)}(\cdot, \cdot) * F).$$

Thus we obtain the following lemma.

Lemma 8.24. *If F is a member of class $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, then the identity*

$$S \ *_{\mathcal{V}_j^{(1/2)}} \ (\Phi_j^{(2)} \ *_{\mathcal{H}} \ F) = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} S$$

holds for all $S \in \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$. In particular,

$$\Phi_j^{(2)} \ *_{\mathcal{H}} \ F = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} \Phi_j^{(2)}(\cdot, \cdot).$$

By the same arguments as given in the last subsection we obtain the following lemma.

Lemma 8.25. *Let F be a function of class $\mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}})$, then the identity*

$$K * F = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} K(\cdot,\cdot)$$

holds for all $\mathcal{H}_{\sigma,\sigma}$-kernels $K(\cdot,\cdot)$ with $K^\wedge(n) = 0$, $n = 2^j, 2^j+1, \ldots$.

Finally we get the following results.

Theorem 8.26. *Under the assumptions of Lemma* 8.25 *we have*

$$\Phi_j * F = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} \Phi_j(\cdot,\cdot),$$

$$(\Phi_j * \Phi_j) * F = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} (\Phi_j * \Phi_j)(\cdot,\cdot),$$

and

$$\Psi_{j-1} * F = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} \Psi_{j-1}(\cdot,\cdot),$$

$$(\tilde{\Psi}_{j-1} * \Psi_{j-1}) * F = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} (\tilde{\Psi}_{j-1} * \Psi_{j-1})(\cdot,\cdot).$$

From Theorem 8.26 we are able to deduce that

$$\Phi_{J-1}^{(2)} * F = \sum_{i=1}^{N_{J-1}} a_i^{N_{J-1}} \mathcal{L}_i^{N_{J-1}} \Phi_{J-1}^{(2)}(\cdot,\cdot), \tag{8.63}$$

where

$$a_l^{N_{J-1}} = \sum_{i=1}^{N_{J-1}} w_{l,i}^{N_{J-1}} \mathcal{L}_i^{N_{J-1}} (\Phi_{J-1}^{(2)}(\cdot,\cdot) * F). \tag{8.64}$$

On the other hand, by virtue of Lemma 8.22, we have

$$\Phi_{J-1}^{(2)} * F = \sum_{i=1}^{N_J} a_i^{N_J} \mathcal{L}_i^{N_J} \Phi_{J-1}^{(2)}(\cdot,\cdot). \tag{8.65}$$

Combining (8.64) and (8.65) we obtain

$$a_l^{N_{J-1}} = \sum_{i=1}^{N_{J-1}} \sum_{k=1}^{N_J} w_{l,i}^{N_{J-1}} a_k^{N_J} \mathcal{L}_i^{N_{J-1}} \mathcal{L}_k^{N_J} \Phi_{J-1}^{(2)}(\cdot,\cdot) \tag{8.66}$$

for $l = 1, \ldots, N_{J-1}$. Assuming the sets $\{\mathcal{L}_1^{N_j}, \ldots, \mathcal{L}_{N_j}^{N_j}\}$ to be hierarchical, i.e., $\mathcal{L}_i^{N_j} = \mathcal{L}_i^{N_{j+1}}$, $i = 1, \ldots, N_j$; $j = 0, \ldots, J-1$, and observing the symmetry of the

matrix $(w_{l,i}^{N_{J-1}})$ we gain a reduction of computational costs as follows:

$$\begin{aligned} a_l^{N_{J-1}} &= \sum_{i=1}^{N_{J-1}} \sum_{k=1}^{N_J} w_{i,l}^{N_{J-1}} a_k^{N_J} \mathcal{L}_i^{N_{J-1}} \mathcal{L}_k^{N_J} \Phi_{J-1}^{(2)}(\cdot,\cdot) \\ &= \sum_{i=1}^{N_{J-1}} \sum_{k=1}^{N_{J-1}} w_{i,l}^{N_{J-1}} a_k^{N_J} \mathcal{L}_i^{N_{J-1}} \mathcal{L}_k^{N_{J-1}} \Phi_{J-1}^{(2)}(\cdot,\cdot) \\ &\quad + \sum_{i=1}^{N_{J-1}} \sum_{k=N_{J-1}+1}^{N_J} w_{i,l}^{N_{J-1}} a_k^{N_J} \mathcal{L}_i^{N_{J-1}} \mathcal{L}_k^{N_J} \Phi_{J-1}^{(2)}(\cdot,\cdot) \\ &= a_l^{N_J} + \sum_{i=1}^{N_{J-1}} \sum_{k=N_{J-1}+1}^{N_J} w_{i,l}^{N_{J-1}} a_k^{N_J} \mathcal{L}_i^{N_{J-1}} \mathcal{L}_k^{N_J} \Phi_{J-1}^{(2)}(\cdot,\cdot). \end{aligned}$$

The recursion relation (8.66) leads us to the following *decomposition scheme*:

$$\begin{array}{ccccccc} F & \to & a^{N_J} & \to & a^{N_{J-1}} & \to \cdots & a^{N_0} \\ & & \downarrow & & \downarrow & & \downarrow \\ & & (WT)(F)(J;\cdot) & & (WT)(F)(J-1;\cdot) & & (WT)(F)(0;\cdot). \end{array}$$

The bandpass filter $R_j(F)$ can be deduced from the formula

$$R_j(F) = \tilde{\Psi}_j * (WT)(F)(j;\cdot) = \sum_{i=1}^{N_j} a_i^{N_j} \mathcal{L}_i^{N_j} (\tilde{\Psi}_j * \Psi_j)(\cdot,\cdot). \tag{8.67}$$

This allows the following *reconstruction scheme* of F:

$$\begin{array}{ccccccccccc} a^{N_0} & & & & a^{N_1} & & & & a^{N_2} & & \\ \downarrow & & & & \downarrow & & & & \downarrow & & \\ R_0(F) & & & & R_1(F) & & & & R_2(F) & & \\ & \searrow & & & & \searrow & & & & \searrow & \\ P_0(F) & \longrightarrow & + & \longrightarrow & P_1(F) & \longrightarrow & + & \longrightarrow & P_2(F) & \longrightarrow & + \quad \longrightarrow \quad \cdots . \end{array}$$

We have seen that the vectors a^{N_j} do *not* depend on the special choice of the scaling function $\{\Phi_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$. In other words, we are able to reconstruct a function with respect to different wavelets just by the knowledge of the vectors a^{N_j}.

Let us finally make some comments concerning the pyramid schemata:

(1) In signal processing a variant of the pyramid scheme is known as subband coding. This technique was originally studied before wavelet theory. The decomposition step consists of applying a lowpass and a bandpass filter followed by downsampling; the reconstruction consists of upsampling followed by filtering.

(2) Any bandlimited potential can be reconstructed exactly via the pyramid scheme by use of bandlimited wavelets (see also [67]). In this case spline exact approximation coincides with polynomial (i.e., outer harmonic) exact approximation. The scale and detail spaces are finite-dimensional so that the

detail information of a potential is only determined by a finite number of wavelet coefficients for each scale.

(3) In case of evaluation functionals and (radial) derivatives at certain points on a sphere $\Omega_r, r \geq \sigma$, the numerical effort can be drastically reduced by three integration procedures on the sphere. The first method is to use gridded pointsystems and then to apply FFT-techniques (cf. the Ph.D.-thesis [74]). The second technique is to use a suitable Gauss-quadrature rule in north-south direction. The third method is to apply the idea of fast summation and panel clustering (cf. [23, 39]). For more details concerning numerical integration on the sphere the reader is referred to [21, 44].

(4) The pyramid scheme provides a powerful tool in interpreting and constructing lowpass and bandpass filters. The wavelets localize in space and frequency. This makes wavelets particularly useful for data compression. Compression techniques aim at reducing storage requirements and speeding up read or write operations to or from disks. For the compression scheme we are ready to accept an error as long as the quality after compression is acceptable.

(5) Another application is, that for the evaluation of a potential or its derivatives at a point, only wavelet coefficients close to the point have to be taken into account. This enables us to observe local features of the geopotential in a global model.

Example. In the foregoing we have seen that bandlimited harmonic wavelets provide "building blocks" that enable fast decorrelation of geopotential data. Next we are interested in discussing the concept of multiresolution analysis from practical point of view. To be more specific, the multiresolution analysis "looks at" the Earth's gravitational potential through a microscope, whose resolution gets finer and finer. Thus it associates to the gravitational potential a sequence of smoothed versions, labelled by the scale parameter. This aspect is illustrated by the figures below for the (bandlimited) EGM96 model. The computation has been performed on the basis of the CP-wavelets following Variant 1.

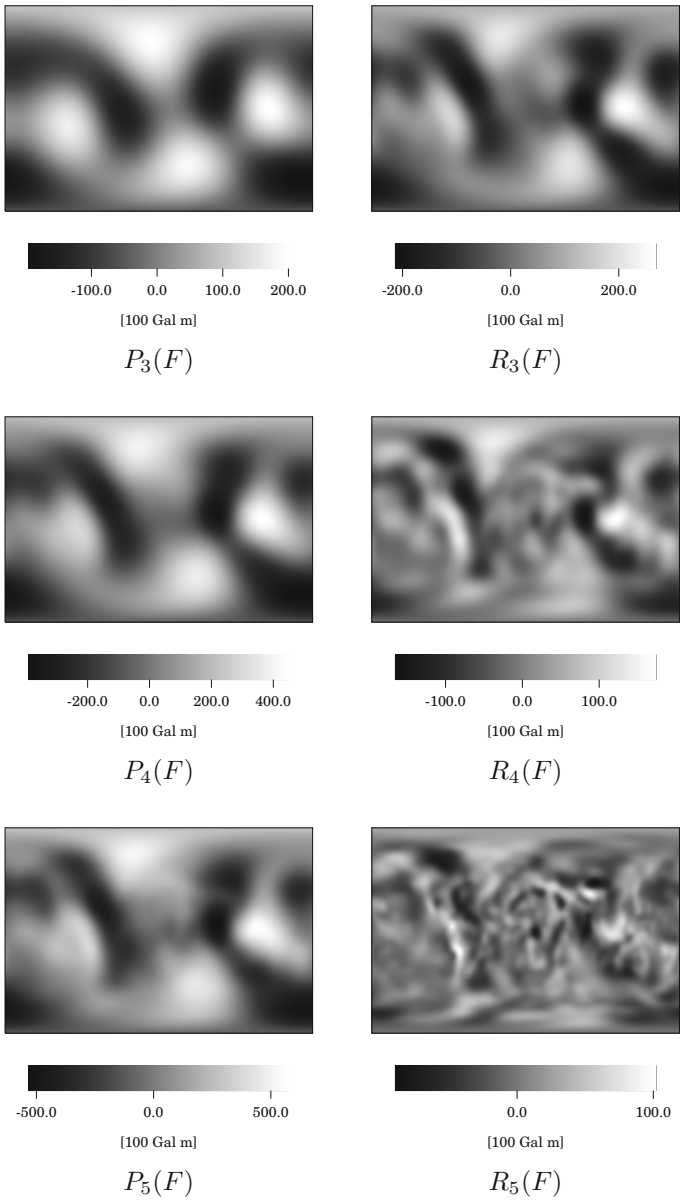


FIGURE 8.1. EGM96 CP-wavelet representation at height 0 km.

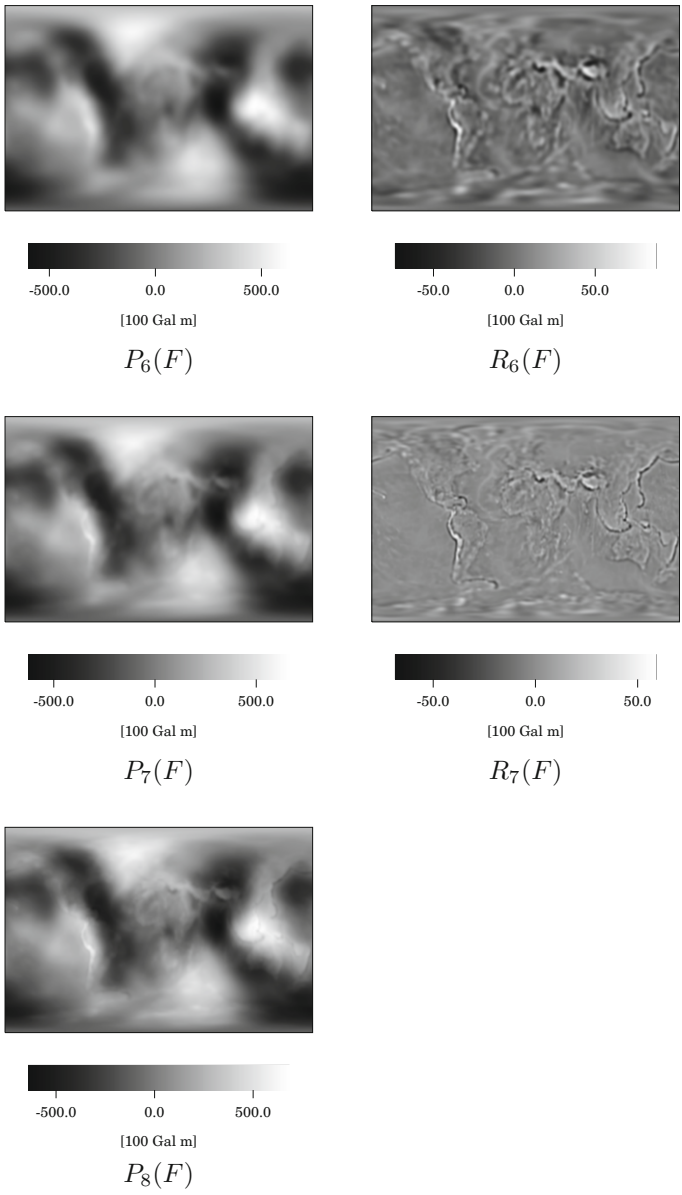


FIGURE 8.2. EGM96 CP-wavelet representation at height 0 km (cont.).

9. Illustrations of Meissl schemata

In this section we derive Meissl schemata for the SST and SGG operators (cf. [20, 32, 58]). In our contribution we focus on the gravitational potential, but obviously, the results are also valid for the disturbing potential.

9.1. Meissl schemata based on outer harmonic framework

We start from the scalar Fourier expansion of the gravitational potential V in terms of outer harmonics

$$V(x) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} V^{\wedge}(n,m) H^s_{n,m}(\sigma;\cdot). \tag{9.1}$$

If the observables are given both at minimum satellites altitude γ and at minimum Earth's radius σ (see Figure 3.1), the symbols of the pseudodifferential operators for the SST and SGG problem can be arranged in a Meissl scheme. The symbols at the arrows indicate how the Fourier coefficients of degree n change at the transition form one quantity to another. In order to avoid confusion the corresponding basis functions are also given. In the case of radial derivatives we remember that the basis system $H_{n,m}$ fulfills

$$H_{n,m}(\sigma;\cdot)|_{\Omega_\sigma} = (1/R) Y_{n,m}. \tag{9.2}$$

Therefore, we get the Meissl scheme for radial derivatives given in Figure 9.1.

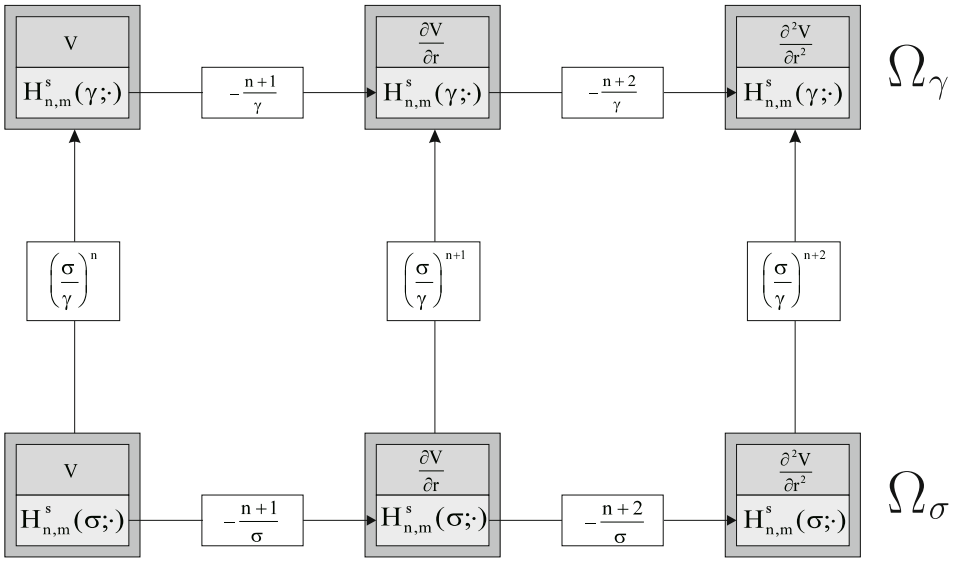


FIGURE 9.1. Meissl scheme for radial derivatives.

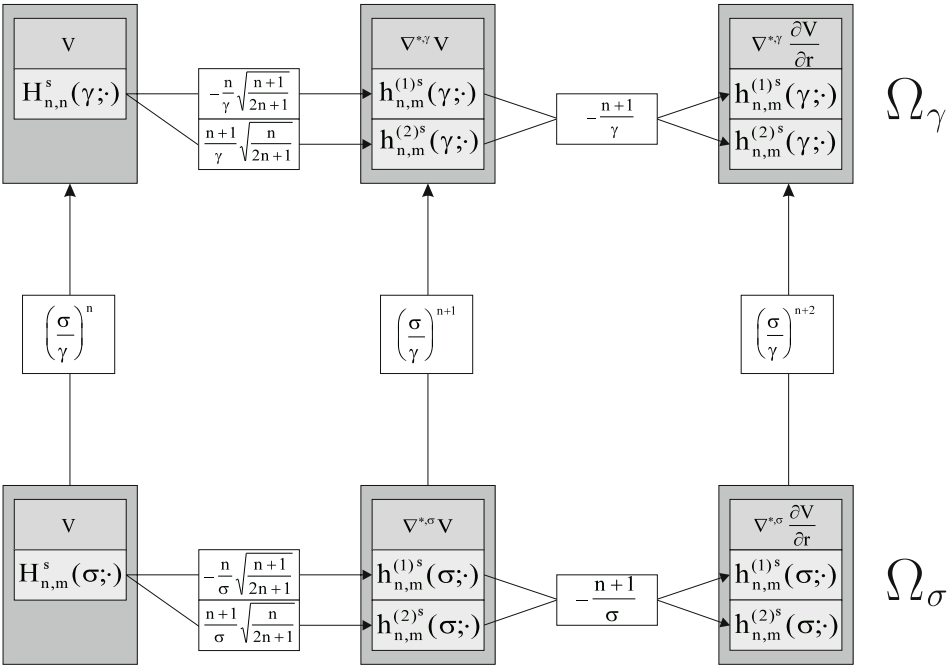


FIGURE 9.2. Meissl scheme for first-order tangential derivatives and second-order mixed derivatives.

If vectorial observables are investigated, we need that

$$o^{(2)}Y_{n,m} = -n\sqrt{\frac{n+1}{2n+1}}\tilde{y}^{(1)}_{n,m} + (n+1)\sqrt{\frac{n}{2n+1}}\tilde{y}^{(2)}_{n,m}, \tag{9.3}$$

which yields the Meissl schemata in Figures 9.2 and 9.3.

Finally, in the case of second-order tangential derivatives $(\nabla^* \otimes \nabla^*)$ we calculate

$$\nabla^* \otimes \nabla^* \tilde{y}^{(1)}_{n,m} = \rho^{(1,1)}_n \frac{n+1}{2n+3}\tilde{\mathbf{y}}^{(1,1)}_{n,m} + \rho^{(2,1)}_n \frac{n+2}{2n+3}\tilde{\mathbf{y}}^{(2,1)}_{n,m} + \rho^{(2,2)}_n \frac{2(n+1)}{(2n+1)(2n-1)}\tilde{\mathbf{y}}^{(2,2)}_{n,m} \tag{9.4}$$

and

$$\begin{aligned}\nabla^* \otimes \nabla^* \tilde{y}^{(2)}_{n,m} = {} & \tau^{(1,1)}_n (-1)\frac{2n(n+1)}{(2n+1)(2n+3)}\tilde{\mathbf{y}}^{(1,1)}_{n,m} + \tau^{(1,2)}_n \frac{n-1}{(2n-1)(2n+1)}\tilde{\mathbf{y}}^{(1,2)}_{n,m} \\ & + \tau^{(2,1)}_n \frac{2n(n+2)}{(2n+3)(2n+1)}\tilde{\mathbf{y}}^{(2,1)}_{n,m} + \tau^{(2,2)}_n (-1)\frac{n}{(2n-1)(2n+1)}\tilde{\mathbf{y}}^{(2,2)}_{n,m},\end{aligned} \tag{9.5}$$

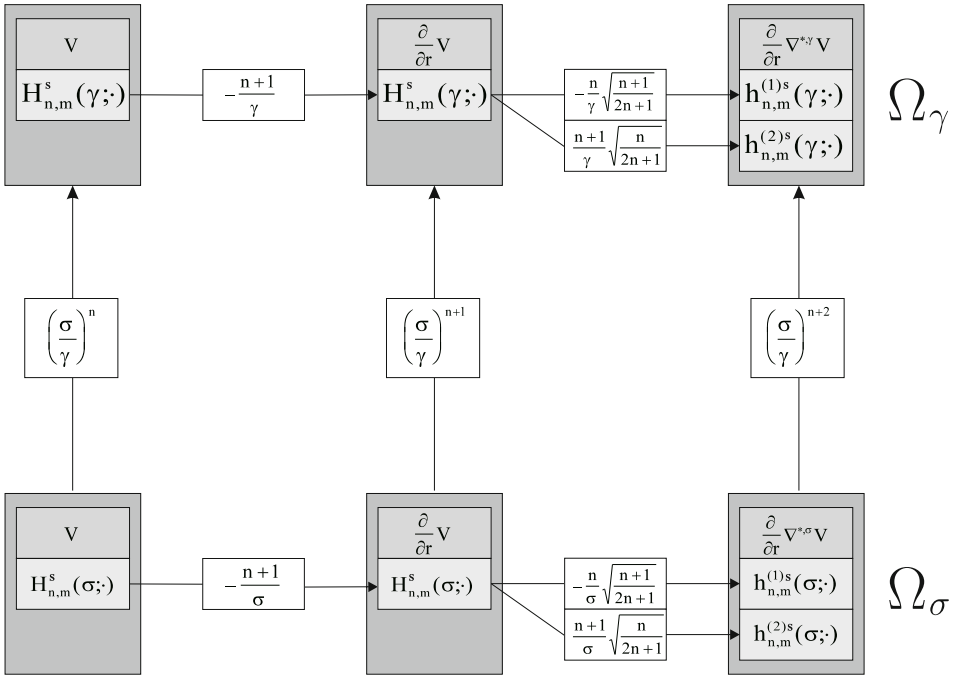


FIGURE 9.3. Meissl scheme for first-order radial derivatives and second-order mixed derivatives.

where the constants $\rho_n^{(i,k)}$ and $\tau_n^{(i,k)}$ are given by

$$\rho_n^{(i,k)} = \sqrt{\frac{\nu_n^{(i,k)}}{(2n+1)(n+1)}}, \tag{9.6}$$

$$\tau_n^{(i,k)} = \sqrt{\frac{\nu_n^{(i,k)}}{(2n+1)n}}. \tag{9.7}$$

In conclusion, we get the Meissl scheme for first- and second-order tangential derivatives (see Figure 9.4).

9.2. Meissl schemata based on kernel function framework

In order to derive Meissl schemata based on kernel functions we want to recapitulate the convolutions which are used in this section (see Table 4).

Our point of departure is the description of a function $F \in \mathcal{H}_s(\overline{\Omega_\sigma^{\text{ext}}})$ in terms of outer harmonics

$$F(x) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} F^\wedge(n,m) H^s_{n,m}(\sigma; x), \tag{9.8}$$

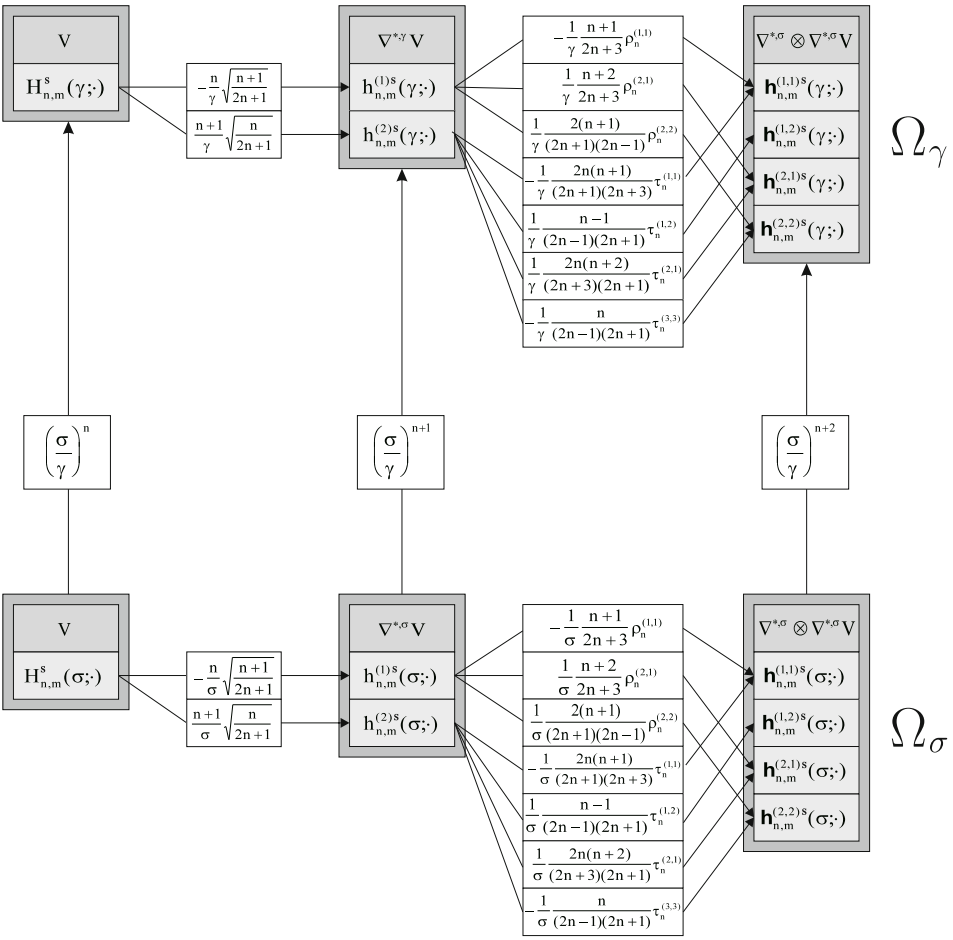


FIGURE 9.4. Meissl scheme for first- and second-order tangential derivatives.

$x \in \overline{\Omega_\sigma^{\text{ext}}}$, and we first derive the kernel functions corresponding to the SST and SGG operators.

Scalar SST and SGG Operators

The SST and SGG operators are given by the convolution equation

$$\Lambda F(x) = (K^\Lambda)^{\sigma,\gamma}(\cdot, x) * F, \quad x \in \overline{\Omega_\gamma^{\text{ext}}}, \tag{9.9}$$

where the symbol of the kernel $(K^\Lambda)^{\sigma,\gamma}$ is given by

$$(K^\Lambda)^\wedge(n) = \Lambda^\wedge(n) = \begin{cases} \left(\frac{\sigma}{\gamma}\right)^n \frac{n+1}{\gamma}, & n = 0, 1, \ldots \quad \text{for SST}, \\ \left(\frac{\sigma}{\gamma}\right)^n \frac{(n+1)(n+2)}{\gamma^2}, & n = 0, 1, \ldots \quad \text{for SGG}. \end{cases} \tag{9.10}$$

$K * F$ $= \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} K^{\wedge}(n) F^{\wedge}(n,m) H^s_{n,m}(\gamma;\cdot)$	$F, K(\cdot,y) \in \mathcal{H}_s(\overline{\Omega^{\text{ext}}_\sigma})$
$k^{(i)} * f$ $= \sum_{n=0_i}^{\infty} \sum_{m=1}^{2n+1} k^{(i)\wedge}(n) f^{(i)\wedge}(n,m) H^s_{n,m}(\gamma;\cdot)$	$f, k^{(i)}(\cdot,y) \in h^{(i)}_s(\overline{\Omega^{\text{ext}}_\sigma})$
$k \star F$ $= \sum_{i=1}^{3} \sum_{n=0_i}^{\infty} \sum_{m=1}^{2n+1} k^{(i)\wedge}(n) F^{\wedge}(n,m) h^{(i)s}_{n,m}(\gamma;\cdot)$	$F \in \mathcal{H}_s(\overline{\Omega^{\text{ext}}_\sigma})$, $k(\cdot,y) \in h_s(\overline{\Omega^{\text{ext}}_\sigma})$
$\mathbf{k}^{(i,k)} * \mathbf{f}$ $= \sum_{n=\tilde{0}_{ik}}^{\infty} \sum_{m=1}^{2n+1} \mathbf{k}^{(i,k)\wedge}(n) \mathbf{f}^{(i,k)\wedge}(n,m) H^s_{n,m}(\gamma;\cdot)$	$\mathbf{f}$, $\mathbf{k}^{(i,k)}(\cdot,y)$ $\in \mathbf{h}^{(i,k)}_s(\overline{\Omega^{\text{ext}}_\sigma})$
$\mathbf{k} \star F$ $= \sum_{i,k=1}^{3} \sum_{n=\tilde{0}_{ik}}^{\infty} \sum_{m=1}^{2n+1} \mathbf{k}^{(i,k)\wedge}(n) F^{\wedge}(n,m) \mathbf{h}^{(i,k)s}_{n,m}(\gamma;\cdot)$	$F \in \mathcal{H}_s(\overline{\Omega^{\text{ext}}_\sigma})$, $\mathbf{k}(\cdot,y) \in \mathbf{h}_s(\overline{\Omega^{\text{ext}}_\sigma})$

TABLE 4. List of the convolutions.

Vectorial SST and SGG Operators

In the vectorial case we have

$$\lambda F(x) = (k^{\lambda})^{\sigma,\gamma}(\cdot,x) \star F, \quad x \in \overline{\Omega^{\text{ext}}_\gamma}, \tag{9.11}$$

with the symbol $(k^{\lambda})^{(i)\wedge}(n)$ given by

$$(k^{\lambda})^{(1)\wedge}(n) = \lambda^{(1)\wedge}(n) = \begin{cases} -\left(\frac{\sigma}{\gamma}\right)^n \frac{n}{\gamma} \sqrt{\frac{n+1}{2n+1}}, & n = 1,2,\ldots \quad \text{for SST,} \\ -\left(\frac{\sigma}{\gamma}\right)^{n+1} \frac{n(n+1)}{\gamma^2} \sqrt{\frac{n+1}{2n+1}}, & n = 1,2,\ldots \quad \text{for SGG,} \end{cases} \tag{9.12}$$

and

$$(k^{\lambda})^{(2)\wedge}(n) = \lambda^{(2)\wedge}(n) = \begin{cases} \left(\frac{\sigma}{\gamma}\right)^n \frac{n+1}{\gamma} \sqrt{\frac{n}{2n+1}}, & n = 1,2,\ldots \quad \text{for SST,} \\ \left(\frac{\sigma}{\gamma}\right)^{n+1} \frac{(n+1)^2}{\gamma^2} \sqrt{\frac{n}{2n+1}}, & n = 1,2,\ldots \quad \text{for SGG,} \end{cases} \tag{9.13}$$

and $(k^{\lambda})^{(3)\wedge}(n) = 0$.

Tensorial SGG Operator

This operator is given by

$$\boldsymbol{\lambda} F(x) = (\mathbf{k}^{\boldsymbol{\lambda}})^{\sigma,\gamma}(\cdot,x) \star F, \quad x \in \overline{\Omega^{\text{ext}}_\gamma}, \tag{9.14}$$

where we have the symbol

$$(\mathbf{k}\,\boldsymbol{\lambda})^{(i,k)\wedge}(n) = \boldsymbol{\lambda}^{(i,k)\wedge}(n) = \begin{cases} \left(\frac{\sigma}{\gamma}\right)^n \frac{1}{\gamma^2}\frac{n(n+1)}{(2n+1)(2n+3)}\sqrt{\nu_n^{(1,1)}}, & (i,k)=(1,1), \\ \left(\frac{\sigma}{\gamma}\right)^n \frac{1}{\gamma^2}\frac{-(n+1)(n-1)}{((2n-1)(2n+1)}\sqrt{\nu_n^{(1,2)}}, & (i,k)=(1,2), \\ \left(\frac{\sigma}{\gamma}\right)^n \frac{1}{\gamma^2}\frac{-n(n+2)}{(2n+3)(2n+1)}\sqrt{\nu_n^{(2,1)}}, & (i,k)=(2,1), \\ \left(\frac{\sigma}{\gamma}\right)^n \frac{1}{\gamma^2}\frac{n(n+1)(n+2)}{(2n-1)(2n+1)}\sqrt{\nu_n^{(2,2)}}, & (i,k)=(2,2), \\ 0, & \text{else.} \end{cases} \tag{9.15}$$

Upward Continuation Operators

The kernels of the (scalar) upward continuation operators K_U, $K_{U'}$, and $K_{U''}$ are given by

$$K_U(x,y) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\left(\frac{\sigma}{\gamma}\right)^n H_{n,m}^s(\gamma;x)H_{n,m}^s(\sigma;y), \tag{9.16}$$

$$K_{U'}(x,y) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\left(\frac{\sigma}{\gamma}\right)^{n+1} H_{n,m}^s(\gamma;x)H_{n,m}^s(\sigma;y), \tag{9.17}$$

$$K_{U''}(x,y) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\left(\frac{\sigma}{\gamma}\right)^{n+2} H_{n,m}^s(\gamma;x)H_{n,m}^s(\sigma;y). \tag{9.18}$$

The upward continuation operators for vector and tensor fields can be introduced in the same way by use of the vectorial and tensorial basis functions $h_{n,m}^{s(i)}$ and $\mathbf{h}_{n,m}^{s(i,k)}$, $i,k \in \{1,2,3\}$.

The Meissl schemata for the scalar/vectorial/tensorial wavelets can now be derived as follows:

Scalar Meissl Scheme. From the reconstruction formula in the scalar case (7.9) we get

$$F(x) = \sum_{j=-1}^{\infty}\tilde{\Psi}_j * (WT)(F)(j;x) = \sum_{j=-1}^{\infty}(\tilde{\Psi}_j * \Psi_j * F)(x), \tag{9.19}$$

$x \in \overline{\Omega_\sigma^{\text{ext}}}$, whereas

$$\frac{\partial F}{\partial r}(x) = \sum_{j=-1}^{\infty}\left(\tilde{\Psi}_j * \Psi_j * K_{\frac{\partial}{\partial r}}^{\sigma} * F\right)(x), \tag{9.20}$$

where the kernel of the first radial derivative $K_{\frac{\partial}{\partial r}}^{\sigma}$ on the sphere Ω_σ is given by

$$K_{\frac{\partial}{\partial r}}^{\sigma}(x,y) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\left(-\frac{n+1}{\sigma}\right)H_{n,m}^s(\sigma;x)H_{n,m}^s(\sigma;y). \tag{9.21}$$

The same calculation for the second radial derivative $\frac{\partial^2}{\partial r^2}$ leads to

$$\frac{\partial^2 F}{\partial r^2}(x) = \sum_{j=-1}^{\infty} \left(\tilde{\Psi}_j * \Psi_j * K^R_{\frac{\partial^2}{\partial r^2}} * F \right)(x), \tag{9.22}$$

where $K^{\sigma}_{\frac{\partial^2}{\partial r^2}}$ is given by

$$\begin{aligned} K^{\sigma}_{\frac{\partial^2}{\partial r^2}}(x,y) &= \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \frac{(n+1)(n+2)}{\sigma^2} H^s_{n,m}(\sigma;x) H^s_{n,m}(\sigma;y) \\ &= \left(K^{\sigma}_{\frac{\partial}{\partial r}} * \tilde{K}^{\sigma}_{\frac{\partial}{\partial r}} \right)(x,y), \end{aligned} \tag{9.23}$$

and the kernel $\tilde{K}^{\sigma}_{\frac{\partial}{\partial r}}$ is given by

$$\tilde{K}^{\sigma}_{\frac{\partial}{\partial r}}(x,y) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \left(-\frac{n+2}{\sigma} \right) H^s_{n,m}(\sigma;x) H^s_{n,m}(\sigma;y). \tag{9.24}$$

Therefore, we get the Meissl scheme shown in Figure 9.5.

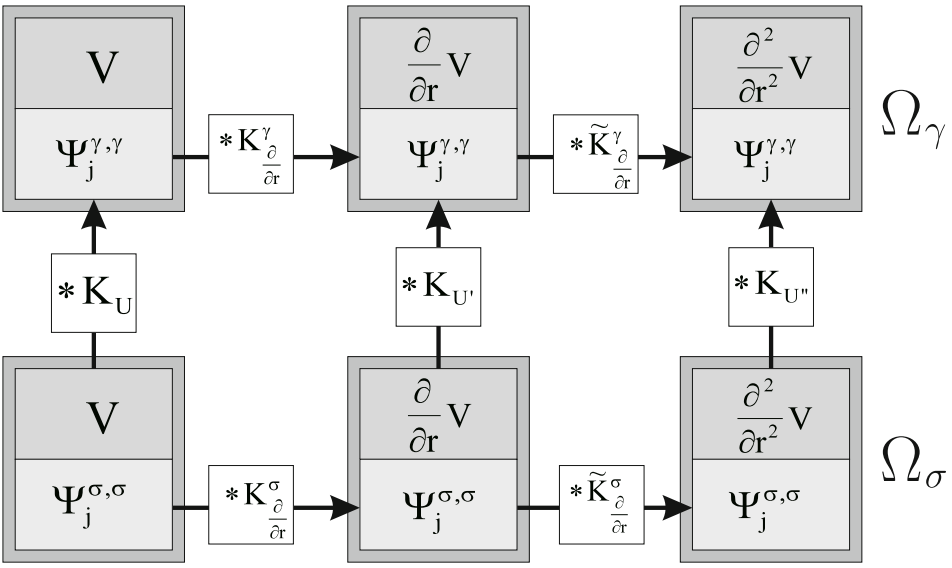


FIGURE 9.5. Meissl scheme for kernel functions (scalar case).

Scalar/Vectorial Meissl Scheme. The extension the the case of vectorial operators is straightforward:

$$o^{(2),\sigma} F(x) = \sum_{j=-1}^{\infty} \sum_{i=1}^{2} \left(\tilde{\Psi}_j^{(i)} \star \Psi_j^{(i)} * \left(k^{\sigma,(i)}_{o^{(2),\sigma}} \star F \right) \right)(x), \tag{9.25}$$

where the kernel functions $k^{\sigma,(i)}_{o^{(2),\sigma}}$ are given by

$$k^{\sigma,(1)}_{o^{(2),\sigma}}(x,y) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\left(-\frac{n}{\sigma}\right)\sqrt{\frac{n+1}{2n+1}}h^{s(1)}_{n,m}(\sigma;x)H^s_{n,m}(\sigma;y), \tag{9.26}$$

$$k^{\sigma,(2)}_{o^{(2),\sigma}}(x,y) = \sum_{n=1}^{\infty}\sum_{m=1}^{2n+1}\frac{n+1}{\sigma}\sqrt{\frac{n}{2n+1}}h^{s(2)}_{n,m}(\sigma;x)H^s_{n,m}(\sigma;y). \tag{9.27}$$

In the SGG case we calculate

$$o^{(2),\sigma}\frac{\partial F}{\partial r}(x) = \sum_{j=-1}^{\infty}\sum_{i=1}^{2}\left(\tilde{\Psi}^{(i)}_j \star \Psi^{(i)}_j * \left(k^{\sigma,(i)}_{o^{(2)}\frac{\partial}{\partial r}} \star F\right)\right)(x), \tag{9.28}$$

where the kernels $k^{\sigma,(i)}_{o^{(2),\sigma}\frac{\partial}{\partial r}}$ are given by

$$\begin{aligned} k^{\sigma,(1)}_{o^{(2),\sigma}\frac{\partial}{\partial r}}(x,y) &= \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\left(-\frac{n+1}{\sigma}\right)\frac{n}{\sigma}\sqrt{\frac{n+1}{2n+1}}h^{s(1)}_{n,m}(\sigma;x)H^s_{n,m}(\sigma;y) \\ &= \left(k^{\sigma,(1)}_{o^{(2),\sigma}} \star K^{\sigma}_{\frac{\partial}{\partial r}}\right)(x,y), \end{aligned} \tag{9.29}$$

$$\begin{aligned} k^{\sigma,(2)}_{o^{(2),\sigma}\frac{\partial}{\partial r}}(x,y) &= \sum_{n=1}^{\infty}\sum_{m=1}^{2n+1}\frac{n+1}{\sigma}\frac{n+1}{\sigma}\sqrt{\frac{n}{2n+1}}h^{s(2),\sigma}_{n,m}(\sigma;x)H^s_{n,m}(\sigma;y) \\ &= \left(k^{\sigma,(2)}_{o^{(2)}} \star K^{\sigma}_{\frac{\partial}{\partial r}}\right)(x,y). \end{aligned} \tag{9.30}$$

Summing up, we finally get the Meissl schemata given in Figures 9.6 and 9.7 for the vector approach.

Scalar/Vectorial/Tensorial Meissl Scheme. We get

$$\begin{aligned} &\nabla^{*,\sigma}\otimes\nabla^{*,\sigma}F(x) \\ &\quad= \sum_{j=-1}^{\infty}\sum_{\substack{(i,k)\in \\ \{(1,1),(1,2),(2,1),(2,2)\}}}\left(\tilde{\mathbf{\Psi}}^{(i,k)}_j \star \mathbf{\Psi}^{(i,k)}_j * \left(\mathbf{k}^{\sigma,(i,k)}_{\nabla^{*,\sigma}\otimes\nabla^{*,\sigma}} \star F\right)\right)(x), \end{aligned} \tag{9.31}$$

where the kernel functions $\mathbf{k}^{\sigma,(i,k)}_{\nabla^{*,\sigma}\otimes\nabla^{*,\sigma}}$ are given by

$$\mathbf{k}^{\sigma,(1,1)}_{\nabla^{s,\sigma}\otimes\nabla^{*,\sigma}}(x,y) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\sqrt{\nu^{(1,1)}_n}\frac{n(n+1)}{\sigma^2(2n+1)(2n+3)}\mathbf{h}^{s(1,1)}_{n,m}(\sigma;x)H^s_{n,m}(\sigma;y), \tag{9.32}$$

$$\mathbf{k}^{\sigma,(1,2)}_{\nabla^{*,\sigma}\otimes\nabla^{*,R}}(x,y) = \sum_{n=1}^{\infty}\sum_{m=1}^{2n+1}\left(-\sqrt{\nu^{(1,2)}_n}\right)\frac{(n-1)(n+1)}{\sigma^2(2n-1)(2n+1)}\mathbf{h}^{s(1,2)}_{n,m}(\sigma;x)H^s_{n,m}(\sigma;y), \tag{9.33}$$

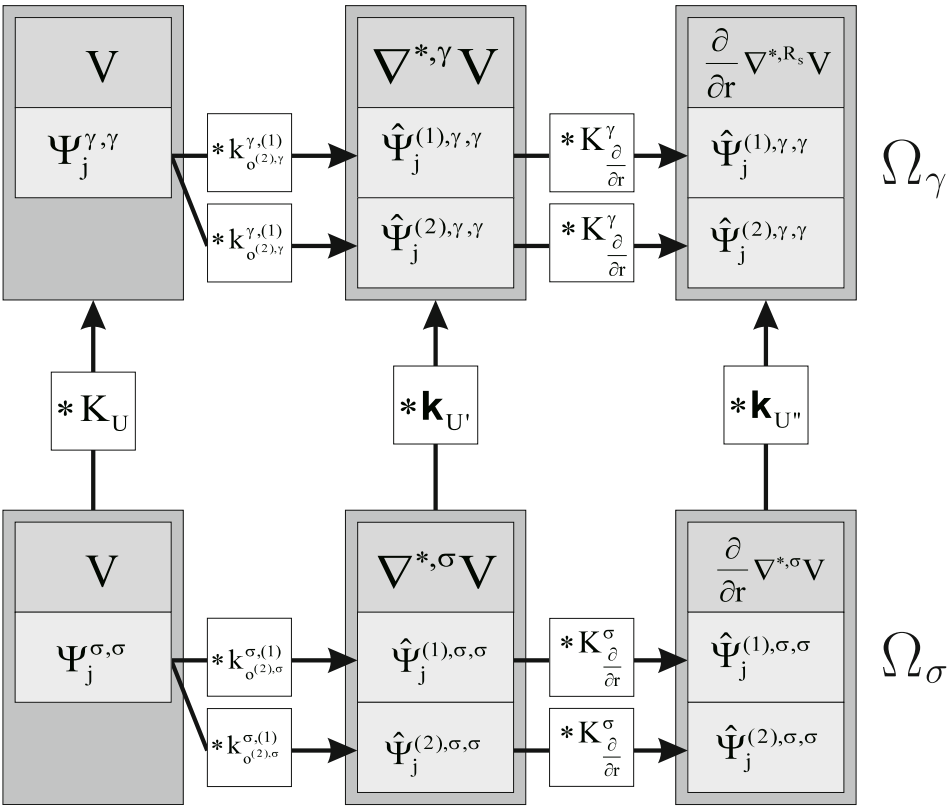


FIGURE 9.6. Meissl scheme for kernel functions (scalar/vectorial case).

$$\mathbf{k}^{\sigma,(2,1)}_{\nabla^{*,\sigma}\otimes\nabla^{*,\sigma}}(x,y)=\sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\left(-\sqrt{\nu_n^{(2,1)}}\right)\frac{n(n+2)}{\sigma^2(2n+1)(2n+3)}\mathbf{h}^{s(2,1)}_{n,m}(\sigma;x)H^s_{n,m}(\sigma;y), \tag{9.34}$$

$$\mathbf{k}^{\sigma,(2,2)}_{\nabla^{*,\sigma}\otimes\nabla^{*,\sigma}}(x,y)=\sum_{n=2}^{\infty}\sum_{m=1}^{2n+1}\sqrt{\nu_n^{(2,2)}}\,\frac{n(n+1)(n+2)}{\sigma^2(2n-1)(2n+1)}\mathbf{h}^{s(2,2)}_{n,m}(\sigma;x)H^s_{n,m}(\sigma;y). \tag{9.35}$$

Note that the kernels $\mathbf{k}^{\sigma,(i,k)}_{\nabla^{*,\sigma}\otimes\nabla^{*,\sigma}}$, $(i,k)\in\{(1,1),(1,2),(2,1),(2,2)\}$ can be split into $\mathbf{k}^{\sigma,(i,k)}_{\nabla^{*,\sigma}\otimes\nabla^{*,\sigma}}=\sum_{l=1}^{2}\mathbf{k}^{\sigma,(i,k),(l)}_{\nabla^{*,\sigma}}\star k^{\sigma,(l)}_{o^{(2),\sigma}}$, where the kernels $\mathbf{k}^{\sigma,(i,k)(l)}_{\nabla^{*,\sigma}}$ are given by

$$\mathbf{k}^{\sigma,(1,1),(1)}_{\nabla^{*,\sigma}}=\sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\frac{n+1}{\sigma(2n+3)}\rho_n^{(1,1)}\mathbf{h}^{s(1,1)}_{n,m}(\sigma;x)h^{s(1)}_{n,m}(\sigma;y), \tag{9.36}$$

$$\mathbf{k}^{\sigma,(1,1),(2)}_{\nabla^{*,\sigma}}=\sum_{n=1}^{\infty}\sum_{m=1}^{2n+1}\left(-\frac{2n(n+1)}{\sigma(2n+1)(2n+3)}\right)\tau_n^{(1,1)}\mathbf{h}^{s(1,1)}_{n,m}(\sigma;x)h^{s(2)}_{n,m}(\sigma;y), \tag{9.37}$$

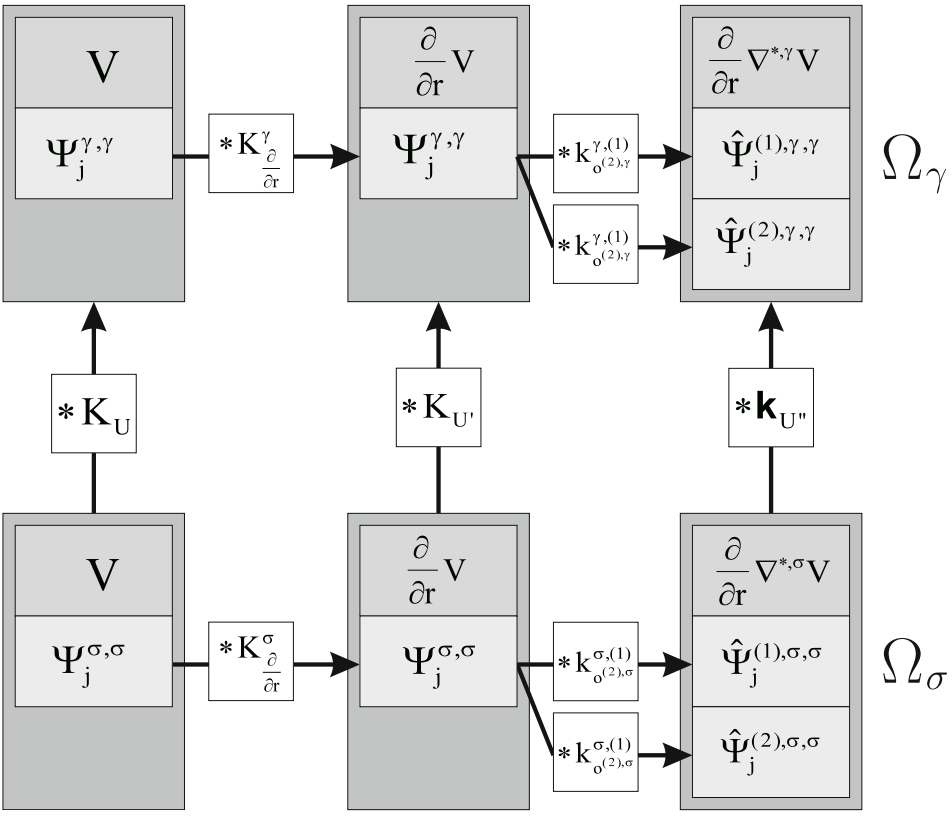


FIGURE 9.7. Meissl scheme for kernel functions (scalar/vectorial case).

$$\mathbf{k}^{\sigma,(1,2),(1)}_{\nabla^*,\sigma} = 0, \tag{9.38}$$

$$\mathbf{k}^{\sigma,(1,2),(2)}_{\nabla^*,\sigma} = \sum_{n=1}^{\infty}\sum_{m=1}^{2n+1} \frac{n-1}{\sigma(2n-1)(2n+1)} \tau_n^{(1,2)} \mathbf{h}^{s(1,2)}_{n,m}(\sigma;x) h^{s(2)}_{n,m}(\sigma;y), \tag{9.39}$$

$$\mathbf{k}^{\sigma,(2,1),(1)}_{\nabla^*,\sigma} = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1} \frac{n+2}{\sigma(2n+3)} \rho_n^{(2,1)} \mathbf{h}^{s(2,1)}_{n,m}(\sigma;x) h^{s(1)}_{n,m}(\sigma;y), \tag{9.40}$$

$$\mathbf{k}^{\sigma,(2,1),(2)}_{\nabla^*,\sigma} = \sum_{n=1}^{\infty}\sum_{m=1}^{2n+1} \frac{2n(n+2)}{\sigma(2n+3)(2n+1)} \tau_n^{(2,1)} \mathbf{h}^{s(2,1)}_{n,m}(\sigma;x) h^{s(2)}_{n,m}(\sigma;y), \tag{9.41}$$

$$\mathbf{k}^{\sigma,(2,2),(1)}_{\nabla^*,\sigma} = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1} \frac{2(n+2)}{\sigma(2n-1)(2n+1)} \rho_n^{(2,2)} \mathbf{h}^{s(2,2)}_{n,m}(\sigma;x) h^{s(1)}_{n,m}(\sigma;y), \tag{9.42}$$

$$\mathbf{k}^{\sigma,(2,2),(2)}_{\nabla^*,\sigma} = \sum_{n=2}^{\infty}\sum_{m=1}^{2n+1} \left(-\frac{n}{\sigma(2n-1)(2n+1)} \right) \tau_n^{(2,2)} \mathbf{h}^{s(2,2)}_{n,m}(\sigma;x) h^{s(2)}_{n,m}(\sigma;y). \tag{9.43}$$

The convolution of the kernel $\mathbf{k}^{\sigma,(i,k),(l)} = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \mathbf{k}^{\sigma,(i,k),(l)\wedge}(n)\mathbf{h}_{n,m}^{s(i,k)} h_{n,m}^{s(l)}$ and the vector field $f^{(l)} \in h(\overline{\Omega_\sigma^{\text{ext}}})$ is given by

$$\mathbf{k}^{\sigma,(i,k),(l)} \star f^{(l)} = \sum_{n=\tilde{0}_{ik}}^{\infty} \sum_{m=1}^{2n+1} \mathbf{k}^{\sigma}, (i,k), (l) \wedge (n) f^{(l)\wedge_h}(n,m)\mathbf{h}_{n,m}^{s(i,k)}(\sigma;\cdot). \quad (9.44)$$

Thus, we get the Meissl scheme given in Figure 9.8.

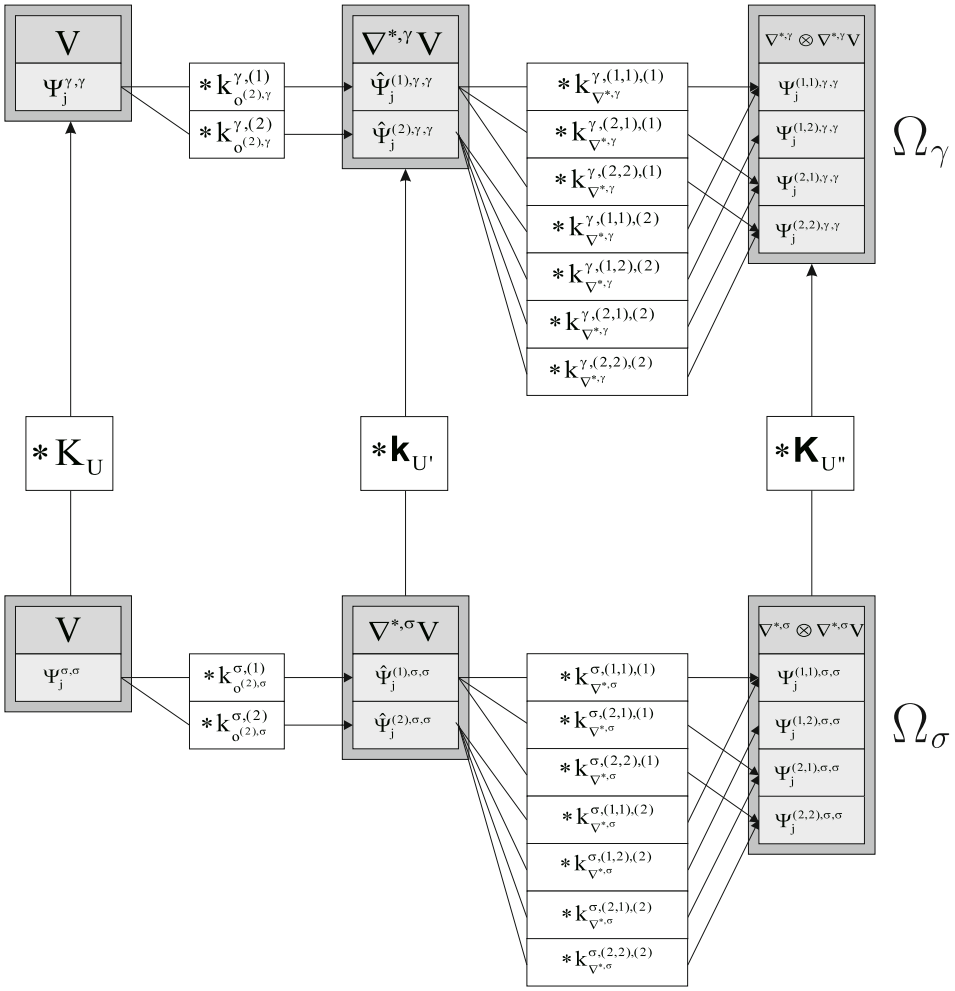


FIGURE 9.8. Meissl scheme for kernel functions (scalar/vectorial/tensorial case). (Note that the tensor-2 wavelets could not be written in bold letter for technical reasons.)

10. Conclusions

As already pointed out, accurate knowledge of the gravitational potential of the Earth is required in order to solve, for example, problems in geodesy, navigation, oceanography, solid Earth physics, and exploration geophysics. In physical geodesy it is the essential pre-stage of geoid computation. Earlier it was envisaged that the gravitational potential could be determinable as a solution of a boundary value problem. The classical problem was the Stokes problem, the boundary values were the gravity anomalies, for which the hitherto unrealistic assumption of global (terrestrial) coverage was required. But today we are confronted with the situation where also other quantities give information about the Earth's gravity potential, for example, gravity disturbance vector or second-order gradients of the disturbance potential from air- and spacecraft. In recent years the geometric shape of the Earth, continents and ocean surface, became measurable with unprecedented precision, due to the enormous progress of space methods like GNSS, VLBI, SLR, and satellite altimetry. The mathematical connection between the gravitational data within a georelevant geometry is the integrated concept. Usually, this concept is formulated in the framework of a reproducing kernel Hilbert space $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ consisting of potentials harmonic down to an internal (Runge) sphere Ω_σ. Mathematically, the gravitational (anomalous) potential of the Earth is assumed to be an element of such a space $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$. In the Hilbert space $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ any element may be represented by its expansion with respect to a complete system of kernel expressions $\mathcal{L}_i K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot)$ related to (linear) observables $\mathcal{L}_i$ on $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$. Because of the reproducing kernel structure imposed on $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$, orthonormalization of a finite system $\{\mathcal{L}_i K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot)\}_{i=1,\ldots,N}$ is equivalent to the spline problem of finding the minimum norm interpolant in the associated $\|\cdot\|_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}$-metric. When using minimum norm interpolation (or smoothing), however, the normal equation matrix $(\mathcal{L}_i\mathcal{L}_k K_{\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})}(\cdot,\cdot))_{i,k=1,\ldots,N}$ is in general a full matrix, reflecting the certain status of decorrelation guaranteed by the reproducing kernel (covariance function) under consideration. This problem causes numerical difficulties which may to a certain extend be overcome by several techniques (for example, fast summation, panel clustering, etc.). But the numerical obstacles are the main reasons why approximation methods of the Earth's gravitational field determination based on spline procedures could not keep pace with the increasing flow of observational information. In other words, the serious drawback of spline approximation is that there is no efficient transition from global to local modeling by only using one kernel (covariance) function with (fixed) space/momentum localization property.

The power of harmonic wavelets lies in the fact that kernel functions with variable space/momentum localization come into use according to a suitable dilation process. By using a sequence of more and more kernels reflecting the various levels of space/momentum localization the reference Sobolev space $\mathcal{H}(\overline{\Omega_\sigma^{\text{ext}}})$ is decomposed into a nested sequence of approximating subspaces

$$\cdots \mathcal{V}_j(\overline{\Omega_\sigma^{\text{ext}}}) \subset \mathcal{V}_{j+1}(\overline{\Omega_\sigma^{\text{ext}}}) \subset \cdots$$

reflecting the different stages of decorrelation. In doing so, harmonic wavelets may be used as mathematical means for breaking up a complicated function (such as the Earth's gravitational potential) into many simple pieces at different scales and positions. This allows multiresolution analysis and compression of data. The particular efficiency of wavelets is caused by the property that only a few wavelet coefficients in the wavelet table are needed in areas where the gravitational potential is "smooth", whereas stronger resolution of a complicated pattern is settled by a zooming-in capability. Wavelets offer canonical tools for combined terrestrial, airborne, and spaceborne data management under realistic assumptions imposed on the geometry of the Earth's surface and the "orbital configuration". Fast computation becomes available in form of tree algorithms. This enables gravitational potential determination with millions of data. Thus harmonic (regularization) wavelets are particularly important for inverse multiscale modeling of spaceborne data. In a subsequent step geoid computation can be based on a highly accurate gravitational potential derived from a homogeneous set of spaceborne data combined with terrestrial and/or airborne data.

For inverse multiscale modeling of spaceborne data two different ways of wavelet regularization are available, namely bandlimited truncated singular value decomposition and non-bandlimited regularization using, e.g., Tikhonov, rational, exponential, and "locally supported" kernels. In accordance with the uncertainty principle the different constituting elements of regularization may be explained as follows: *Non-bandlimited regularization wavelets* tend to be extremely space localizing. Thus huge data sets of irregular distribution can be handled since only data in a small neighborhood, whose size is determined by the particular choice of the wavelet type, is needed for the purpose of evaluating the wavelet coefficients. On the other hand, a large number of wavelet coefficients depending on the choice of the wavelet for the regularization is needed, since the wavelet coefficients only give local information of a small neighborhood. It appears that non-bandlimited regularization is an appropriate tool of local gravity surveys for oil and mineral exploration. However, little practical work has been done yet in this application area for non-synthetic data sets, although the use of linear functionals allows a very promising combination of terrestrial and/or airborne data within a unified setup in terms of wavelets. Moreover, fast summation techniques and panel clustering is adequately applicable in pyramid schemata.

Bandlimited regularization wavelets show more moderate phenomena of space localization so that one can work with smaller data sets in numerical evaluation. In consequence, the number of wavelet coefficients can be reduced, since they contain information of a more extended area. Moreover, a certain spectral band can be expressed exactly in terms of wavelets because of their bandlimited character even when the airborne data are combined with terrestrial information. Pyramid schemata can be based on exact (outer harmonic) approximation. In conclusion, dependent on the space/momentum character of the bandlimited wavelets inverse multiscale gravity modeling of spaceborne data can be handled successfully by multiresolution analysis.

Finally, it should be pointed out that our approach is given within a spherical context. Geodesists sometimes believe that ellipsoidal reference surfaces in combination with ellipsoidal harmonics might be the better choice. No doubt, an ellipsoidally reflected multiscale formulation is mathematically interesting and geodetically relevant. However, its numerical realization is by far more complicated than the spherical oriented variant chosen for our study here. As a matter of fact, Meissl schemata are involved with gravitational quantities not including the centrifugal influence. In this case, however, Runge–Walsh methods corresponding to Runge–Walsh (Bjerhammar) spheres form an adequate alternative which, in the opinion of the authors, is superior when numerical purposes come into play because of the much more efficient and economical structure inherent in spherical framework. Even better, Runge–Walsh procedures are not only applicable for ellipsoidal reference surfaces, but also for geometrically complicated reference surfaces such as telluroid, or (co)geoid.

Acknowledgment. The authors thank the "Federal Ministry for Economic Affairs and Energy, Berlin" and the "Project Management Jülich" for funding the project "SPE" (funding reference number 0324016).

11. Appendix A: List of basic gravity field quantities

The list of this appendix essentially follows [ESA1]. It provides an introductory collection of quantities used in classical geodesy that could not be explained throughout the paper:

Definition	**Observation method**
Gravity potential, W: Sum of the gravitational and the centrifugal potential.	Differences between values in two points observed by levelling.
Equipotential surface: Surface where the gravity potential is equal to a constant.	Points on one surface determined regionally by tide-gauges, which define the regional mean sea level.
Height datum: The equipotential surface best agreeing with local mean sea level.	Mean sea-level calculated from tide-gauges for a specific time period.
Geoid: The equipotential surface which agrees with global mean sea level.	The equipotential surface which agrees with a global set of tide-gauges and leveling bench-marks.

Gravity: Magnitude of gradient of the gravity potential at Earth's surface and of the gravitational potential in the outer space.	Observed by absolute (e.g., free fall experiment) or relative (as a difference) spring gravimeter.
Gravity gradient: Derivatives of the gravity vector, i.e., second-order derivatives of W.	Certain linear combinations measured by torsion-balance at Earth's surface, by difference between accelerometers in space (gradiometry).
Mean Earth Ellipsoid: Ellipse rotated around the ε^3-axis, with center at the Earth's gravity center.	Surface which gives best fit to mean sea-level, and which has centrum in the gravity centre.
Height above ellipsoid: Height above mean Earth ellipsoid measured along the normal to ellipsoid.	Observed indirectly by GPS from cartesian coordinates.
Geoid height: Height of a point on the geoid above the reference ellipsoid.	Observed by GPS at tide-gauge or at leveling point.
Orthometric height: Height from geoid measured along a plumb-line (often height above mean sea-level).	Observed by leveling and converted to metric units by dividing with gravity.
GNSS: A satellite navigation system with global coverage.	GNSS: GPS, GLONASS, Galileo or Beidou.
Gravity anomaly: A model gravity potential with a reference ellipsoid as an equipotential surface is used to calculate normal gravity (needed is latitude and orthometric height).	It is a value derived by subtracting measured and normal gravity. The normal gravity is calculated in a point with the ellipsoidal height put equal to the orthometric height.

12. Appendix B: List of basic units in gravitational field theory

Units and orders in gravity field theory are the following: The gravity is expressed in $\mathrm{m/s^2}$ or in milligal (1 mgal$= 10^{-5}$ $\mathrm{m/s^2}$); the mean Earth gravity is about 981 000 mgal, and varies from 978 100 mgal to 983 200 mgal from equator to pole due to the Earth's flattening and rotation. Deviations due to density inhomogenities, mountain ridges, etc. range from tens to hundreds of milligals. On the other hand, the excursions of the geoid, measured from the mean Earth ellipsoid, amount to about -105 and $+90$ meters. Gravity gradients are expressed in Eőtvős ($1E = 10^{-9}$ $\mathrm{s^2}$). The largest component is the vertical gravity gradient, being on Earth's surface of about $3000E$ (gravity changes by $3 \cdot 10^{-6}$ $\mathrm{m/s^2}$ per meter of elevation). The horizontal components are approximately half this size, mixed gradients are below $100E$ for the normal field. Gravity gradient anomalies can be much larger and reach about $1000E$ in mountainous areas (for more details see, for example, [R4]).

	SI units	**traditional**
gravity		
	$10^{-2} ms^{-2}$	1 Gal
	$10^{-5} ms^{-2}$	1 mGal
	$10^{-8} ms^{-2}$	1μ Gal
gravity potential		
	$10 m^2 s^{-2}$	1 kGal $\cdot m$
gravity gradients		
	$10^{-9} s^{-2}$	$1E$

References

[1] Arabelos, D., Tscherning, C.C. (1995) *Regional Recovery of the Gravity Field from SGG and Gravity Vector Data Using Collocation.* J. Geophys. Res., **100**, B 11, 22009–22015.

[2] Arabelos, D., Tscherning, C.C. (1998) *Calibration of Satellite Gradiometer Data Aided by Ground Gravity Data.* J. of Geodesy, **72**, 617–625.

[3] Aronszajn, N. (1950) *Theory of Reproducing Kernels.* Trans. Am. Math. Soc., **68**, 337–404.

[4] Augustin, M., Freeden, W., Nutz, H. (2018) *About the Importance of the Runge–Walsh Concept for Physical Geodesy.* In: Freeden, W., Nashed , M.Z. (Eds.), Handbook of Mathematical Geodesy, this volume, Springer International Publishing, 517–560.

[5] Cui, J., Freeden, W., Witte, B. (1992) *Gleichmäßige Approximation mittels sphärischer Finite-Elemente und ihre Anwendung auf die Geodäsie*. Z. f. Vermessungswes., ZfV, **117**, 266–278.

[6] Daubechies, I. (1992) Ten Lectures on Wavelets. SIAM.

[7] Davis, P.J. (1963) Interpolation and Approximation. Blaisdell Publishing Company.

[8] Driscoll, J.R., Healy, R.M. (1994) Computing Fourier Transforms and Convolutions on the 2-Sphere. Adv. Appl. Math., **15**, 202–250.

[9] ESA (1996) *The Nine Candidate Earth Explorer Missions*. Publications Division ESTEC, Nordwijk, SP-1196 (1).

[10] ESA (1998) European Views on Dedicated Gravity Field Missions: GRACE and GOCE. ESD–MAG–REP–CON–001.

[11] ESA (1999) Gravity Field and Steady-State Ocean Circulation Mission. ESTEC, Nordwijk, ESA SP-1233 (1).

[12] Freeden, W. (1979) Über eine Klasse von Integralformeln der Mathematischen Geodäsie. Veröff. Geod. Inst. RWTH Aachen, Heft 27.

[13] Freeden, W. (1980) On the Approximation of External Gravitational Potential with Closed Systems of (Trial) Functions, *Bull. Geod.*, **54**, 1–20.

[14] Freeden, W. (1981) *On Approximation by Harmonic Splines*. Manuscr. Geod., **6**, 193–244.

[15] Freeden, W. (1982) *Interpolation and Best Approximation by Harmonic Spline Functions*. Boll. Geod. Sci. Aff., **1**, 105–120.

[16] Freeden, W. (1982) *On Spline Methods in Geodetic Approximation Problems*. Math. Meth. in the Appl. Sci., **4**, 382–396.

[17] Freeden, W. (1983) Least Squares Approximation by Linear Combinations of (Multi-) Poles, Dept. Geod. Science, 344, The Ohio State University, Columbus.

[18] Freeden, W. (1987) *Harmonic Splines for Solving Boundary Value Problems of Potential Theory*. In: Algorithms for Approximation (J.C. Mason, M.G. Cox, Eds.), The Institute of Mathematics and its Applications, Conference Series, vol. 10, Clarendon Press, Oxford, 507–529.

[19] Freeden, W. (1999) Multiscale Modelling of Spaceborne Geodata, B.G. Teubner, Stuttgart, Leipzig.

[20] Freeden, W., Gerhards, C. (2013) Geomathematically Oriented Potential Theory. CRC Press, Taylor & Francis, Boca Raton.

[21] Freeden, W., Gervens, T., Schreiner, M. (1998) Constructive Approximation on the Sphere (With Applications to Geomathematics). Oxford Science Publications, Clarendon.

[22] Freeden, W., Glockner, O., Thalhammer, M. (1999) *Multiscale Gravitational Field Recovery from GPS-Satellite-to-Satellite Tracking*. Studia Geoph. Geod, **43**, 229–264.

[23] Freeden, W., Glockner, O., Schreiner, M. (1998) *Spherical Panel Clustering and Its Numerical Aspects*. J. of Geodesy, **72**, 586–599.

[24] Freeden, W., Gutting, M. (2013) Special Functions of Mathematical (Geo-)Physics. Birkhäuser, Basel.

[25] Freeden, W., Kersten, H. (1981) *The Geodetic Boundary Value Problem Using the Known Surface of the Earth.* Veröff. Geod. Inst. RWTH Aachen, Heft 29.

[26] Freeden, W., Mason, J.C. (1990) *Uniform Piecewise Approximation on the Sphere.* In: Algorithms for Approximation II (J.C. Mason, M.G. Cox, Eds.), Chapman and Hall Mathematics, 320–333.

[27] Freeden, W., Nutz, H.. (2011) *Satellite Gravity Gradiometry as Tensorial Inverse Problem.* Int. J. Geomath, **2**, 177–218.

[28] Freeden, W., Reuter, R. (1982) *Remainder Terms in Numerical Integration Formulas of the Sphere.* Internat. Series Numeric. Math., **61**, 151–170.

[29] Freeden, W., Schneider, F. (1998a) Wavelet Approximation on Closed Surfaces and Their Application to Boundary Value Problems of Potential Theory, *Math. Meth. Appl. Sci.*, **21**: 129–165.

[30] Freeden, W., Schneider, F. (1998b) *An Integrated Wavelet Concept of Physical Geodesy.* J. of Geodesy, **72**, 259–281.

[31] Freeden, W., Schreiner, M. (1997) *Orthogonal and Non-orthogonal Multiresolution Analysis, Scale Discrete and Exact Fully Discrete Wavelet Transform on the Sphere.* Constr. Approx. **14**, 493–515.

[32] Freeden, W., Schreiner, M. (2009) *Spherical Functions of Mathematical Geosciences – A Scalar, Vectorial, and Tensorial Setup.* Springer, Heidelberg.

[33] Freeden, W., Windheuser, U. (1996) Spherical Wavelet Transform and Its Discretization, *Adv. Comput. Math.*, **5**, 51–94.

[34] Freeden, W., Windheuser, U. (1997) Combined Spherical Harmonic and Wavelet Expansion – A Future Concept in Earth's Gravitational Determination, *Appl. Comput. Harm. Anal.*, **4**, 1–37

[35] Gerhards, C. (2011) Spherical Multiscale Methods in Terms of Locally Supported Wavelets: Theory and Application to Geomagnetic Modeling. Ph.D.-thesis, University of Kaiserslautern.

[36] Grafarend, E.W., Klapp, M., Martinec, Z. (2015) *Spacetime Modelling of the Earth's Gravity Field by Ellipsoidal Harmonics.* In: Handbook of Geomathematics, 2nd. edition (W. Freeden, M.Z. Nashed, T. Sonar, Eds.), Springer, 381–496.

[37] Groten, E. (1979) Geodesy and the Earth's Gravity Field I, II. Dümmler.

[38] Gurtin, ME (1971) Theory of Elasticity, Handbuch der Physik, **6**.

[39] Gutting, M. (2008) Fast Multipole Methods for Oblique Derivative Problems. Ph.D.-thesis, University of Kaiserslautern, Geomathematics Group, Shaker, Aachen.

[40] Gutting, M. (2012) *Fast Multipole Accelerated Solution of the Oblique Derivative Boundary Value Problem.* Int. J. Geomath. **3**(2), 233–252.

[41] Gutting, M. (2015) *Fast Spherical/Harmonic Spline Modeling.* In: Handbook of Geomathematics, 2nd. edition (W. Freeden, M.Z. Nashed, T. Sonar, Eds.), Springer, 2711–2746.

[42] Gutting, M. (2018) *Parameter Choices for Fast Harmonic Spline Approximation.* In: Freeden, W., Nashed , M.Z. (Eds.), Handbook of Mathematical Geodesy, this volume, Springer International Publishing, 605–639

[43] Heiskanen, W.A., Moritz, H. (1967) Physical Geodesy. W.H. Freeman and Company.

[44] Hesse, K., Sloan, I.H., Womersly, R.S. (2015) *Numerical Integration on the Sphere.* In: Handbook of Geomathematics, 2nd. edition (W. Freeden, M.Z. Nashed, T. Sonar, Eds.), Springer, 2711–2746

[45] Ilk, K.H., Reigber, C., Rummel, R. (1978) *The Use of Satellite-to-Satellite Tracking for Gravity Parameter Recovery.* Proc. of the European Workshop on Space Oceanography, Navigation and Geodynamics (SONG), ESA SP–137.

[46] Ilk, K.H., Sigl, R., Thalhammer, M. (1990) *Regional Gravity Field Recovery from SST* (*GPS-Aristoteles*)*, from Gradiometer Measurements and Their Combination.* In CIGAR II, WP's 4.30, 4.40, 4.50.

[47] Kellogg, O.D. (1929) Foundations of Potential Theory. Frederick Ungar Publishing Company.

[48] Klug, M. (2014) Integral Formulas and Discrepancy Estimates Using the Fundamental Solution to the Beltrami Operator on Regular Surfaces. Ph.D.-thesis, University of Kaiserslautern, Geomathematics Group.

[49] Krarup, T. (1969) A Contribution to the Mathematical Foundation of Physical Geodesy. Publ. of the Danish Geodetic Institute, No. 44, Copenhagen.

[50] Kusche, J. (1998): *Regional adaptive Schwerefeldmodellierung für SST-Analysen.* In: Progress in Geodetic Science at GW98, (W. Freeden ed.), Shaker, 266–273.

[51] Laur, H., Liebig, V. (2015): *Earth Observation Satellite Missions and Data Access.* In: Handbook of Geomathematics, 2nd. edition (W. Freeden, M.Z. Nashed, T. Sonar, Eds.), Springer, 147–170.

[52] Magnus, W., Oberhettinger, F., Soni, R.P. (1966) Formulas and Theorems for the Special Functions of Mathematical Physics. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Bd. 52, Springer.

[53] Meissl, P.A. (1971) A Study of Covariance Functions Related to the Earth's Disturbing Potential. The Ohio State University, Department of Geodetic Science and Surveying, Columbus, OSU Report No. 152.

[54] Michel, V. (1999) A Multiscale Method for the Gravimetry Problem: Theoretical and Numerical Aspects of Harmonic and Anharmonic Modelling. Ph.D.-thesis, University of Kaiserslautern, Geomathematics Group.

[55] Moritz, H. (1980) Advanced Physical Geodesy. Herbert Wichmann Verlag, Karlsruhe, Abacus Press, Tunbridge, Wells, Kent.

[56] Moritz, H., Sünkel, H. (1978) Approximation Methods in Geodesy. H. Wichman Verlag, Karlsruhe.

[57] Müeller, C. (1966) Spherical Harmonics. *Lecture Notes in Mathematics*, **17**, Springer-Verlag

[58] Nutz, H. (2002) A Unified Setup of Gravitational Field Observables. Ph.D.-thesis, University of Kaiserslautern, Geomathematics Group, Shaker, Aachen.

[59] Pizetti, P. (1910) Sopra il calcoba tesrico delle deviazioni del geoide dall'ellisoide. Att. A. Acad. Sci. Torino, **46**, 331–350.

[60] Rummel, R. (1975) Downward Continuation of Gravity Information from Satellite to Satellite Tracking or Satellite Gradiometry in Local Areas. Dept. Geodetic Science, **221**, The Ohio State University, Columbus.

[61] Rummel, R. (1979) Determination of the Short-wavelength Components of the Gravity Field from Satellite-to-Satellite Tracking or Satellite Gradiometry. *Manuscr. Geod.*, **4**, 107–148.

[62] Rummel, R. (1986) *Satellite Gradiometry.* In: Lecture Notes in Earth Sciences 7, Mathematical and Numerical Techniques in Physical Geodesy (H. Sünkel, ed.), 318–363, Springer.

[63] Rummel, R. (1997) Spherical spectral properties of the earth's gravitational potential and its first and second derivatives. In Sansò, S. and Rummel, R., Eds., Geodetic Boundary Value Problems in View of the One Centimeter Geoid. Volume 65 of Lecture Notes in Earth Science, pages 359–404. Springer, Berlin, Heidelberg.

[64] Rummel, R., van Gelderen, M. (1992) Spectral Analysis of the Full Gravity Tensor, *Geophys. J. Int.*, **111**, 159–169

[65] Rummel, R, van Gelderen, M (1995) Meissl Scheme – Spectral Characteristics of Physical Geodesy. *Manuscr. Geod.*, **20**, 379–385.

[66] Rummel, R., van Gelderen, M., Koop, R., Schrama, E., Sanso, F., Brovelli, M., Miggliaccio, F., Sacerdote, F. (1993) Spherical Harmonic Analysis of Satellite Gradiometry. Netherlands Geodetic Commission, New Series, No. 39.

[67] Schreiner, M. (1997) *Locally Supported Kernels for Spherical Spline Interpolation.* J. of Approx. Theory, **89**, 172–194.

[68] Stokes, G.G. (1849) *On the Variation of Gravity on the Surface of the Earth.* Trans. Cambridge Phil. Soc., **8**, 672–695.

[69] Svensson, S.L. (1983) *Pseudodifferential Operators – a New Approach to the Boundary Value Problems of Physical Geodesy.* Manuscr. Geod., **8**, 1–40.

[70] Taylor, A.E., Lay, D.C. (1980) Introduction to Functional Analysis, 2nd edition. New York, Chichester, Brisbane, Toronto.

[71] Thalhammer, M. (1995) Regionale Gravitationsfeldbestimmung mit zukünftigen Satellitenmissionen (SST und Gradiometrie). Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften, Reihe C, Dissertation. Heft Nr. 437.

[72] Tscherning, C.C., Rapp, R.H. (1974) Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations, and Deflections of the Vertical Implies by Anomaly Degree-Variance Models. Dept. Geod. Sci., **208**, The Ohio State University, Columbus.

[73] Weyl, H. (1916) Über die Gleichverteilung von Zahlen mod. Eins. Math. Ann., **77**, 313–352.

[74] Windheuser, U. (1995) Sphärische Wavelets: Theorie und Anwendung in der Physikalischen Geodäsie. Ph.D.-thesis, University of Kaiserslautern, Geomathematics Group, Shaker, Aachen.

[75] Xu, P.L., Rummel, R. (1994) *A Simulation Study of Smoothness Methods in Recovery of Regional Gravity Fields.* Geophys. J. Int., **117**, 472–486.

[76] Yamabe, H. (1959) *On an Extension of Helly's Theorem.* Osaka Math. J., **2**, 15–22.

[77] Yosida, K (1978) Functional Analysis, Springer-Verlag, Berlin, Heidelberg, New York.

Willi Freeden and Helga Nutz
Geomathematics Group
University of Kaiserslautern
MPI-Gebäude, Paul-Ehrlich-Str. 26
D-67663 Kaiserslautern, Germany

Handbook of Mathematical Geodesy
Geosystems Mathematics, 459–489

The Analysis of the Geodetic Boundary Value Problem: State and Perspectives

Fernando Sansò

Abstract. The geodetic boundary value problem is mathematically a free-boundary, oblique derivative boundary value problem for the Laplace operator. The solution of the problem is the determination of the shape of the Earth and of its gravity field. The analysis of such a problem, specially for its non-linear formulation, is hard, so it started only in 1976 with a paper by L. Hörmander [13].

Since then the research has continued for both the non-linear and the linearized version, till recent years. In this article the author tries to give an overview on the subject, including a new result for the so-called Simple Molodensky Problem.

Keywords. Geodetic boundary value problem,linearization, scalar and vector variants, Molodensky problem.

Overview

In §1 the formulation of the vector GBVP is recalled, especially because it is historically its first formulation. However the section introduces the reader to the characteristic quasi non-uniqueness of the solution of the GBVP, due to a quasi-invariance of the data under translation of the solution.

In §2 the scalar formulation of the GBVP is introduced and its version under a partial Legendre transform is presented, what is called the GBVP in Marussi space. The new formulation is a fixed boundary oblique-derivative BVP for a certain non-linear partial differential equation only recently published. A theorem of existence and uniqueness in Hölder spaces, derived from intermediate Schauder estimates, is recalled. The result is nice in that on data we put regularity requirements which are very close to a realistic model of the Earth.

In §3 a careful linearization of the problem is performed, showing that the resulting boundary operator can be written in the form of a simple spherical operator plus a perturbation. The problem defined using only the spherical part of the boundary operator is known as the Simple Molodensky Problem (SMP) [15, 20].

In §4 the SMP is analyzed in suitable Sobolev spaces, obtaining a new unconditional theorem of existence and uniqueness, even for Lipschitz domains.

In §5 the above theorem is carried over to the linearized GBVP, by a simple perturbation technique. This is achieved at the cost of putting geometrical constraints on the boundary.

In §6 final considerations and some items that would be interesting to investigate, are reported. Finally the author would like to warn the reader that throughout the text many times the same letter C is used to define constants that can actually have different values. This notwithstanding, the context shall be clear.

1. Introduction to the Vector Geodetic Boundary Value Problem

The Geodetic Boundary Value Problem (GBVP) is basically to determine the figure of the Earth from as many as possible (and realistic) measurements on the gravity field and the least possible knowledge of the geometry.

After one century of evolution, across XIX and XX centuries, from the early G. Stokes research to the publication of the book by S. Molodensky, Eremiev and Yurkina [19], the problem has got a more rigorous formulation in the sixties of the last century, thanks to the work of A. Marussi and T. Krarup, the first interested in finding a suitable system of coordinates, (called by him *Intrinsic Coordinates* [16]) that would simplify from the roots the free-boundary characteristic of the GBVP, the second interested in a rigorous formulation of the GBVP in its non-linear form, in order to perform at least a correct linearization process [15, 20].

In this way we got the first definition of the vector, non-linear GBVP, based on the following hypotheses: assume that the Earth is a stationary body B, with a figure (the boundary S of B), constant in time and in uniform rotation (with constant angular velocity ω) around an axis having a fixed direction in an inertial reference system, as well as with respect to the body of the Earth itself; assume further that B is the support of a stationary mass distribution, that produces a stationary Newtonian field with potential $V(\mathbf{x})$ called *gravitational potential*, that is also in uniform rotation, with angular velocity ω, as seen from an inertial reference system; therefore we can define a *gravity potential*

$$W(\mathbf{x}) = V(\mathbf{x}) + \frac{1}{2}\omega^2\rho^2$$
$$(\rho = \text{ distance of } \mathbf{x} \text{ from the rotation axis, that we take as } Z \text{ coordinate axis, i.e., } \rho^2 = X^2 + Y^2) \tag{1.1}$$

the gradient of which

$$\mathbf{g}(\mathbf{x}) = \nabla W(\mathbf{x}) = \nabla V + \omega^2 \mathbf{x}_e$$
$$(\mathbf{x}_e = \text{projection on the } (X, Y) \text{ plane of } \mathbf{x}) \tag{1.2}$$

is the gravity field, namely the force field per unit mass felt by a point mass at rest in a reference system rotating with the Earth (a so-called *Terrestrial Reference System*).

At every point of S we assume to be able to perform a gravimetric measurement providing

$$g(\mathbf{x}) = |\nabla W| \ , \ \mathbf{x} \in S \ ; \tag{1.3}$$

moreover we assume that the direction of the vertical

$$\mathbf{n}(\mathbf{x}) = -\frac{\mathbf{g}(\mathbf{x})}{g(\mathbf{x})} \ , \tag{1.4}$$

is also observable by astrogeodetic measurements. The unit vector $\mathbf{n}$ is usually parameterized by 2 angles $\Sigma = (\Lambda, \Phi)$, the astrogeodetic longitude and latitude, that identify the direction of $\mathbf{n}$ in the terrestrial reference system. In addition we assume that, by combining leveling networks with gravimetry, we come to know the potential $W(\mathbf{x}), \forall \mathbf{x} \in S$.

If we accept the above hypotheses, we arrive at the following definition of the Vector GBVP.

Definition 1.1 (VGBVP). Given

$$\mathbf{g}(\mathbf{x}) = -g(\mathbf{x})\mathbf{n}(\mathbf{x}) \ , \ W(\mathbf{x}) \quad \mathbf{x} \in S \tag{1.5}$$

to find a surface

$$S \equiv \{\mathbf{x} = \mathbf{x}(\Sigma)\} \tag{1.6}$$

and a regular harmonic potential $V(\mathbf{x})$

$$\Delta V(\mathbf{x}) = 0 \ , \ \mathbf{x} \in \Omega \equiv (B \cup S)^c \ , \tag{1.7}$$

such that

$$W(\Sigma) = V(\mathbf{x}) + \frac{1}{2}\omega^2 \ (X^2 + Y^2)\big|_{\mathbf{x}=\mathbf{x}(\Sigma)} \ . \tag{1.8}$$

$$\mathbf{g}(\Sigma) = g(\Sigma)\mathbf{n}(\Sigma) \equiv \nabla V + \omega^2 \mathbf{x}_e\big|_{\mathbf{x}=\mathbf{x}(\Sigma)} \tag{1.9}$$

As one can see the problem is a free boundary, non-linear BVP for the Laplace operator.

A first remark on this formulation is that the centrifugal terms in (1.9), (1.8) are small compared to the main gravitational part. For instance $\max|\omega^2 \mathbf{x}_e| \sim$ 3 Gal (1 Gal = 1 $\mathrm{cm\,sec}^{-2}$), as opposed to $|\nabla V| \sim 10^3$ Gal. Even more, if we compute centrifugal potential and acceleration by using a reference figure, like the Earth ellipsoid, the residual unknown part goes down to $10^{-5} \div 10^{-6}$ times the main terms.

It is then only natural to think of analyzing first a rotation-free VGBVP, and to send the solution of (1.3), (1.9)(1.8) to a subsequent iteration process. So we take as a reference the problem of the following Definition 1.2.

Definition 1.2 (rotation-free VGBVP). To find a surface $S \equiv \{\mathbf{x} = \mathbf{X}(\Sigma)\}$ and a potential $V(\mathbf{x})$ such that

$$\Delta V(\mathbf{x}) = 0 \qquad \mathbf{x} \in \Omega \text{ (the exterior of } S) \tag{1.10}$$

$$V(\mathbf{x}) = \frac{\mu}{r} + O\left(\tfrac{1}{r^2}\right) \qquad \text{(regularity at infinity)} \tag{1.11}$$

$$V(\mathbf{x})|_S = V[\mathbf{x}(\Sigma)] = W(\Sigma) \tag{1.12}$$

$$\nabla V(\mathbf{x})|_S = (\nabla V)[\mathbf{x}(\Sigma)] = \mathbf{g}(\Sigma)\ ; \tag{1.13}$$

here μ stems for $\mu = \mathrm{GM}$, where G is Newton's universal constant and M is the mass of the Earth.

Let us underline that solving (1.10), (1.11), (1.12), (1.13) in a suitably linearized form, is known to provide solutions good at the limit of present day errors in the observations and even beyond. So the conceptual importance of the nonlinear GBVP is to have theorems that guarantee not only the existence and the uniqueness of the solution but also its continuous dependence on the data, in such a way that we do not introduce unwanted simplifications in passing to the linearized version.

One feature that emerges clearly in the rotation-free VGBVP is its invariance under translation.

It is in fact easy to verify that given a solution $S \equiv \{\mathbf{x} = \mathbf{X}(\Sigma)\}, V(\mathbf{x})(\mathbf{x} \in \Omega)$ of (1.10) to (1.13), the following family

$$S' \equiv \{\mathbf{x} = \mathbf{X}(\Sigma) + \mathbf{c}\}, V'(\mathbf{x}) = V(\mathbf{x} - \mathbf{c}) \quad (\forall \mathbf{x} \in \Omega' \equiv \Omega + \mathbf{c}) \tag{1.14}$$

provides a solution too, for every constant vector $\mathbf{c}$.

Indeed the original formulation was hiding this fact, because the small centrifugal terms are breaking the invariance in the X and Y directions, leaving unaltered only that along Z. Nevertheless it is clear that centrifugal terms give second-order variations to data, under an equatorial shift, when this last is small. Therefore the problem is naturally weakly sensitive to such parameters that will then be determined at most with large errors; a feature this that we want to avoid. So we want first of all to find conditions that assure uniqueness of the solution. We claim that such conditions can be to fix the origin of the coordinates at the barycenter of the masses; as it is known this corresponds to the asymptotic formula (see [28])

$$V(\mathbf{x}) = \frac{\mu}{r} + O\left(\frac{1}{r^3}\right)\ , \tag{1.15}$$

meaning that $V(\mathbf{x})$ has not a first-order harmonic component, when r runs far away from the masses. This is equivalent to say that we will search for a solution in a space of harmonic functions of co-dimension 3. This indeed puts three constraints on the spaces of data $W(\Sigma), \nabla W(\Sigma)$. Following Hörmander instead

of formulating this requirement on the data we rather augment the space of unknowns by introducing 3 unknown constants into the boundary relation, namely we substitute (1.12) with the relation

$$V(\mathbf{x})|_S = W(\Sigma) + \mathbf{a} \cdot \mathbf{v}(\mathbf{x})|_S \tag{1.16}$$

where $\mathbf{a}$ is a 3D vector and $V(\mathbf{x})$ a triple of functions that guarantee the satisfaction of (1.15). The simplest choice can be

$$v_k(\mathbf{x}) = \frac{Y_{ik}(x)}{r^2} \div \frac{x_k}{r^3} \tag{1.17}$$

although other choices can be done.

So the final formulation of the VGBVP, under a non-rotating condition, is:

$$\begin{cases} \Delta V = 0 & \text{in } \Omega \\ V|_S = V(\Sigma) + \mathbf{a} \cdot \mathbf{v}(\mathbf{x}) \\ \nabla V|_S = \mathbf{g}(\Sigma) \\ V = \frac{\mu}{r} + O\left(\frac{1}{r^3}\right) & (r \to \infty) \end{cases} \tag{1.18}$$

A first result on the analysis of (1.18) is found in the seminal paper by L. Hörmander [13]. Basically the result is that a solution $V(\mathbf{x})$ exists in Hölder space $H_{2,\lambda}(\Omega), \mathbf{X}(\Sigma) \in H_{2,\lambda}(\sigma)$, ($\sigma$ the unit sphere), if $W(\Sigma), \mathbf{g}(\Sigma)$ are close enough in $H_{2,\lambda}(\Omega)$ to spherical counterparts, i.e., $W_0(\Sigma), \mathbf{g}_0(\Sigma)$, satisfying the relation

$$\mu|\mathbf{g}_0(\Sigma)| = V(\Sigma)^2 \ . \tag{1.19}$$

This is obtained with the application of hard implicit function theorems (see [21]). A simpler but effective approach, the so-called Gravity Space approach [25, 26], makes use of the Legendre transformation, where

$$\mathbf{g} = \nabla V(\mathbf{x}) \tag{1.20}$$

becomes a new coordinate system and a new adjoint potential ψ, defined by

$$\psi = \mathbf{g} \cdot \mathbf{x} - V \tag{1.21}$$

is introduced.

It is then easy to see that S is mapped to a fixed surface S_g, Ω is mapped to the interior Ω_g of S_g and the Laplace equation for V becomes

$$\begin{aligned} &Tr\Psi^{-1} = 0 \ , \\ &\Psi = \left[\frac{\partial^2 \psi}{\partial g_i \partial g_k}\right] \ . \end{aligned} \tag{1.22}$$

One then verifies that, symmetrically to (1.20), the relation $\mathbf{x} = \nabla\psi(\mathbf{g})$ holds, so that the boundary condition becomes

$$\mathbf{g} \cdot \nabla\psi - \psi|_{S_g} = V(\Sigma) \ . \tag{1.23}$$

Furthermore the uniqueness constraint (1.15) becomes

$$\psi = \mu^{1/2} g^{1/2} + O(g^{3/2}) \quad (g \to 0) \ ; \tag{1.24}$$

suitably modifying (1.23), with the addition of 3 unknown constants, as

$$g\frac{\partial\psi}{\partial\psi} - \psi|_{S_g} = V(\Sigma) + \mathbf{c}\cdot\mathbf{g}(\Sigma) \tag{1.25}$$

one gets a perfect balance and existence, and uniqueness of $\psi(\mathbf{g})$ in $H_{2,\lambda}$ follows in a suitable neighborhood of the spherical solution $\psi_0 = \mu^{1/2}g^{1/2}$. Once ψ is retrieved, one gets S from the already mentioned relation

$$\mathbf{x}(\Sigma) = \nabla\psi|_{S_g} \ . \tag{1.26}$$

Better results can even be obtained by using the so-called *intermediate Schauder estimates* [8, 23].

But it is time now to go to a different BVP, which is more adherent to the physical situation of observable quantities.

2. From the vector to the scalar GBVP (SGBVP)

The formulation of the VGBVP, as for any mathematical model, requires a good deal of abstraction. First of all the gravity field in continental areas has been surveyed in a rather inhomogeneous way, concerning both the spatial resolution and the accuracy of data.

For instance Antarctica and Africa are only marginally covered by gravimetric measurements, while South America has a very uneven gravity data set. On the other hand the vast area of the ocean has only a very poor coverage of direct measurements, although a complex elaboration of satellite observations and oceanographic models can provide a significant data set in terms of both resolution and accuracy (see O. Andersen in [28], Chapt. 9).

But it is the data set of astrogeodetic observation, i.e., those defining the direction of the vertical $\mathbf{n}$ (see (1.4)), that is so quantitatively and qualitatively poor, to make the formulation of the VGBVP too far away from reality.

Fortunately, the physical structure of the Earth gravity field helps us in finding a more realistic model. As a matter of fact, the variations of g in vertical direction are roughly two orders of magnitude larger than in any horizontal direction. This reflects the prevailing horizontal layering of the bulky density of the body of the Earth. This means also that even an imprecise knowledge of the horizontal position of the measurement point is acceptable without introducing too large errors in the determination of the gravity potential.

This circumstance has suggested geodesists (see [23, 24]) to give a different formulation of the GBVP where the boundary is "free" only in the vertical direction.

Introducing ellipsoidal coordinates $(\sigma, h) \equiv (\lambda, \varphi, h)$ (longitude, latitude, ellipsoidal height) (see [28], §1.11) we can formulate the scalar GBVP as follows:

to find $S \equiv \{h = H(\sigma)\}$ and $V(\mathbf{x}) = V(\sigma, h)$ such that

$$\begin{cases} \Delta V = 0 & \text{in } \Omega \equiv \{h \geq H(\sigma)\} \\ V|_S = W(\sigma) - \frac{1}{2}\omega^2(X^2(\sigma) + Y^2(\sigma)) \\ |\nabla V + \omega^2(X(\sigma)\mathbf{e}_X + Y(\sigma)\mathbf{e}_Y)|_S = g(\sigma) \\ V(\mathbf{x}) = O\left(\frac{1}{r}\right) \ . \end{cases} \tag{2.1}$$

This problem is scalar precisely in the sense that of the unknown point P on S we know the projection on the ellipsoid, i.e., σ, but we do not know the scalar $H(\sigma)$. Such lack of information is compensated by giving both $W(\sigma)$ and $g(\sigma)$ at P.

Following the same reasonings as in §1, we can eliminate, with a very minor error, the dependence on the rotational potential too, so arriving at the somewhat simpler formulation

$$\begin{cases} \Delta V = 0 & \text{in } \Omega \equiv \{h \geq H(\sigma)\} \\ V|_S = W(\sigma) \\ |\nabla V|_S = g(\sigma) \\ V = O\left(\frac{1}{r}\right) \ . \end{cases} \tag{2.2}$$

An interesting historical remark is that, although (2.2) and (1.10)–(1.13) are two different problems, they have been considered as one problem only, in geodetic literature, until the difference was clearly highlighted in [24]. The reason why this could happen is that both problems, linearized and posed under the so-called spherical approximation (see [28] §2.6) are reduced to the same problem, known as the *Simple Molodensky Problem*, that we shall treat later in §4.

This remark is useful to us, to understand that it is convenient, though not strictly necessary, to reformulate (2.2) in the same fashion as (1.18), namely adding 3 scalar unknowns (the vector $\mathbf{a}$) and putting a stricter constraint on the asymptotic behaviour of $V(\mathbf{x})$, i.e.,

$$\begin{cases} \Delta V = & \text{in } \Omega \\ V|_S = W(\sigma) + \mathbf{a} \cdot \mathbf{v}(\mathbf{x}) \\ |\nabla V|_S = g(\sigma) \\ v = \dfrac{\mu}{r} + O\left(\frac{1}{r^3}\right) & r \to \infty \ . \end{cases} \tag{2.3}$$

As in (1.16), (1.17) the simplest choice of $\mathbf{v}(\mathbf{x})$ is

$$\mathbf{c}(\mathbf{x}) = \left.\frac{\mathbf{x}}{r^3}\right|_S \ . \tag{2.4}$$

The SGBVP has been first formulated by F. Sacerdote and F. Sansò [24], where the idea of a partial Legendre transform has been proposed. A significant step forward in the analysis of the problem has been done in [23]. Also recently in [27] a variant of the aforementioned paper has been introduced and we shall shortly summarize it hereafter.

In this context one has to mention the work of Seitz and Heck [30], who gave a numerical evidence of the superior characteristics of the SGBVP on the VGBVP. Basically the concept is to swap the ordinary spherical coordinate r with the potential V, considered, together with $\sigma = (\lambda, \varphi)$, as a radial coordinate in a *Marussi space*. The name is because Marussi has proposed long ago to use the potential, in this case the gravity potential, as *intrinsic coordinate* [16]. This is best done, first by imagining to put $r(V, \sigma)$ as the solution of the implicit equation

$$V = V(r, \sigma) \ , \tag{2.5}$$

and then considering the field variable

$$Y(V, \sigma) = \frac{1}{r(V, \sigma)} \ . \tag{2.6}$$

Putting

$$\begin{cases} Y' = \frac{\partial Y}{\partial V} \ , & Y'' = \frac{\partial^2 Y}{\partial V^2} \\ Y_\sigma = \nabla_\sigma Y(V, \sigma), & Y_{\sigma\sigma} = \Delta_\sigma Y(V, \sigma) \ , \end{cases} \tag{2.7}$$

by exploiting the implicit function theorem one arrives at the following field equation (see [27])

$$(Y^2 + |Y_\sigma|^2)Y'' - 2Y'Y_\sigma \cdot Y'_\sigma + Y'^2 Y_{\sigma\sigma} = 0 \ . \tag{2.8}$$

Considering in Marussi's space the known surface

$$S_M \equiv \{V \equiv W(\sigma)\} \tag{2.9}$$

and its internal domain

$$\Omega_M \equiv \{(V, \sigma); V \leq W(\sigma)\} \tag{2.10}$$

we have that (2.8), translating the first of (2.3), has to hold in Ω_M. Let us for the moment disregard the additional term in the second of (2.3), $\mathbf{a}\cdot\mathbf{v}$, that we shall add again when the formulation of the SGBVP in Marussi space will be completed. So the second of (2.3) is just used to define S_M. The third of (2.3), further considering the relations

$$g^2 = \left(\frac{\partial V}{\partial r}\right)^2 + \frac{1}{r^2}|\nabla_\sigma V|^2 \tag{2.11}$$

$$Y'V_\sigma + Y_\sigma \equiv 0 \ , \tag{2.12}$$

can be written as

$$Y'^2 g^2(\sigma) = Y^2(Y^2 + |Y_\sigma|^2) \tag{2.13}$$

to hold on S_M.

Finally the fourth of (2.3) becomes

$$Y = \mu^{-1}V + O(V^3) \quad (V \to 0) \ . \tag{2.14}$$

A first remark is that the field

$$Y = \mu^{-1}V \ , \tag{2.15}$$

corresponding to the spherical field

$$V = \frac{\mu}{r} \ , \ g = \frac{\mu}{r^2} \ , \tag{2.16}$$

is indeed a solution of (2.8), it satisfies the boundary relation (2.13), when

$$g^2(\sigma) = \mu^{-2} W^4(\sigma) \ , \tag{2.17}$$

as well as the asymptotic condition (2.14). Therefore it makes sense to search for a form of the SGBVP, which highlights that the actual field $Y(V, \sigma)$ is the sum of a spherical field plus a perturbation. This is achieved posing

$$Y = \mu^{-1} V (1 + y) \ , \tag{2.18}$$

where y will be considered a quantity small of the first order. Physical estimations show that y, which is an a-dimensional quantity, is between 10^{-3} and 10^{-2}, so confirming our guess. Substitution of (2.18) into (2.8) and (2.13) leads to the BVP

$$\begin{cases} \Delta_M y = F_2(y) + F_3(y) \\ V y' - y = h + b_2(y) + b_3(y) + b_4(y) \end{cases} \tag{2.19}$$

where

$$\Delta_M = \frac{\partial^2}{\partial V^2} + \frac{2}{V}\frac{\partial}{\partial V} + \frac{1}{r^2}\Delta_{\sigma\sigma} \ , \tag{2.20}$$

$$k = 1, 2 \qquad F_k(y) = \text{non-linear differential operators homogenous of degree } k \text{ in } y \tag{2.21}$$

$$k = 1, 2, 3 \qquad b_k(y) = \text{non-linear differential boundary operators homogeneous of degree } k \text{ in } y \tag{2.22}$$

$$h(\sigma) = \frac{1}{2}\left[1 - \frac{\mu^2 g^2(\sigma)}{W^4(\sigma)}\right] \ . \tag{2.23}$$

The explicit form of $F_k(y)$ and $b_k(y)$ can be found in [27].

Note has to be taken that since (2.17) holds, up to quantities small of the first order, $h(\sigma)$ has to be a first-order small quantity too.

Since y has to be small, it is natural that if we want to obtain a local invertibility theorem for (2.19), around $y = 0$, one has to look at its linearized version, namely

$$\begin{cases} \Delta_M y = f \\ V y' - y = h \ ; \end{cases} \tag{2.24}$$

this will help us to assess the augmented form of (2.19), by introducing a number of unknown constants and the same number of additional conditions.

Indeed (2.24) is a regular oblique derivative BVP (remember that by hypothesis Ω_M is a starshaped domain) and we expect the Fredholm alternative to hold for it.

It is immediate to see that the null space of (2.24) is just the family of linear functions in $\mathbf{v} \equiv V\mathbf{e}_\sigma$($\mathbf{e}$ the unit vector pointing to σ on the unit sphere), i.e.,

$$y = \mathbf{c} \cdot \mathbf{v} \ . \tag{2.25}$$

The conditions corresponding to the introduction of the unknown $\mathbf{c} \in R^3$, are derived naturally from (2.14), i.e.,

$$Y = \mu^{-1} V + O(V^3) \ .$$

We make here the realistic hypothesis that $\mu = GM$ is in fact known, so that (2.14) translated for y becomes

$$y = O(V^2) \ , \tag{2.26}$$

i.e., y has to satisfy four conditions at the origin, namely

$$y(0) = 0 \ , \ \nabla y(0) = 0 \ . \tag{2.27}$$

Correspondingly we have to introduce in (2.24) not only the unknown vector $\mathbf{c}$ but also another unknown constant c_0. In this way we arrive at the formulation

$$\begin{cases} \Delta_M y = f \\ V y' - y = h + c_0 + \mathbf{c} \cdot \mathbf{v} \\ y(0) = 0 \\ \nabla y(0) = 0 \ . \end{cases} \tag{2.28}$$

Instructed by this discussion on the linearized SGBVP, we stipulate as well that the non-linear problem will have the augmented form

$$\begin{cases} \Delta_M y = F_2(y) + F_3(y) \\ V y' - y = h + b_2(y) + b_3(y) + b_4(y) + c_0 + \mathbf{c} \cdot \mathbf{v} \\ y(0) = 0 \\ \nabla y(0) = 0 \end{cases} \tag{2.29}$$

A choice now has to be done of the space where we want to find y. In [27] the choice of the complement of $\{c_0 + \mathbf{c} \cdot \mathbf{v} \ ; \ c_0 \in R, \mathbf{x} \in R^3\}$ in the Hölder space $H_{2,\lambda}(\Omega_\mu)$ has been done and a suitable local theorem of existence and uniqueness of $y, c_0, \mathbf{c}$, has been derived. This though requires that $\Omega_M \in H_{2,\lambda}(\sigma)$ and $h \in H_{1,\lambda}(\sigma)$ and indeed that $\| \ h \ \|_{1,\lambda}$ is suitably small. To prove the continuity of the Fréchet derivative of the non-linear part of (2.29) is a technical but not difficult task, that we leave to the reader (see [23, 27]). On the other hand a crucial step is to prove the uniqueness for the homogenous version of (2.28). This is easily achieved by putting

$$z = V y' - y \ , \ (\mathbf{v} \in \Omega_M) \tag{2.30}$$

and observing that $\Delta y = 0 \Rightarrow \Delta z = 0$.

On the other hand $y(0) = 0, \nabla y(0) = 0$ implies, through (2.30), that $z(0) = 0$, $\nabla z(0) = 0$ too. So we must have

$$\begin{cases} \Delta z = 0 \\ z|_{S_M} = c_0 + \mathbf{x} \cdot \mathbf{v} \end{cases} \Rightarrow z = c_0 + \mathbf{c} \cdot \mathbf{v} \quad (\text{in } \Omega_M), \tag{2.31}$$

but then

$$z(0) = c_0 = 0 \qquad \nabla z(0) = \mathbf{c} = 0 \ .$$

Now (2.30) implies that

$$V^{-2}(Vy' - y) = 0 \ , \ \frac{\partial}{\partial V}\frac{y}{V} = 0 \ , \tag{2.32}$$

i.e.,

$$y = a(\sigma)V \qquad (\text{in } \Omega_M) \tag{2.33}$$

But since it has to be $\Delta y = 0$, the function (2.33) can only have the linear form

$$y = \mathbf{a} \cdot \mathbf{V}$$

and finally, by the condition

$$\nabla y(0) = 0 \tag{2.34}$$

we get $\mathbf{a} = 0$, i.e., $y \equiv 0$. Uniqueness is therefore proved.

Remark 2.1 (on natural regularity conditions of data $W(\sigma), g(\sigma)$). By following strictly the procedure presented in [23] one can extend the above result to the very interesting case that

$$\Omega_M \in H_{1,\lambda}(\sigma) \quad (\text{i.e., } W(\sigma) \in H_{1,\lambda}(\sigma)) \tag{2.35}$$

and

$$g(\sigma) \in H_{1,\lambda}(\sigma) \tag{2.36}$$

as well. This is derived by the so-called *intermediate Schauder estimates* [8]. Indeed in this case one has not λ Hölder continuous second derivatives up to the boundary, yet $y \in H_{1,\lambda}(\overline{\Omega}_M)$, meaning basically that (recall (2.6), (2.18))

$$r(\sigma) = \mu W^{-1}(\sigma)[1 + y(\sigma)]^{-1} \ ,$$

i.e., the unknown surface of the Earth, is a $H_{1,\lambda}(\sigma)$ function too. This might be an idealization, whereas the result of getting a surface $H_{0,1}(\sigma)$ (i.e., a Lipschitz surface) could be closer to reality. Yet such a strong result is till lacking for the fully non-linear theory.

So assuming that a regularized Earth surface $S \in H_{1,\lambda}(\sigma)$ be acceptable we see that (2.35) and (2.36) become *natural regularity conditions*.

In fact we know that the actual gravitational potential $V(r, \sigma)$ is a Newtonian potential, generated by a bounded mass density. This implies $V \in H^{2,p}(R_n)$, $\forall p \geq 2$ (see [18]) and, in view of the well-known Sobolev embedding theorems (see [1], Chapter V) we have $V \in H_{1,\lambda}(R_n)$ too. So the trace of V on a surface $S \in H_{1,\lambda}(\sigma)$ is expected to be $H_{1,\lambda}(\sigma)$, while the trace of $|\nabla V|$ on S will be in $H_\lambda(\sigma)$.

This concludes the discussion and the analysis of the non-linear SGBVP. We pass now to a linearization of the original problem (2.1) to get the standard linear version of the so-called Molodensky Problem [15, 20], for which more general and precise results will be obtained.

3. Linearizing the SGBVP in geometry space

We proceed now to the linearization of the SGBVP in its original form (2.1), namely formulated as a free boundary BVP in terms of the geometrical coordinates $(\sigma, h) \equiv (\lambda, \varphi, h)$. Let us recall that the first rigorous linearization has been performed by T. Krarup for the VGBVP (see [15]). This results in an oblique derivative BVP with a small difference in the direction along which the oblique derivative has to be taken with respect to the SGBVP. For the vector GBVP the direction is the so-called isozenithal, while for the scalar GBVP it becomes, as we shall see in this section, the normal vertical direction. The two are very close to one another and, as a byproduct of the analysis of the next sections, we will see that there is stability of the solution for perturbations even of the direction of the oblique derivative. This explains why, although conceptually different, the two problems have provided in the past almost identical solutions. The focus of this and next sections is to provide existence, uniqueness and stability of the solution in a suitable Sobolev space when the data are basically in $L^2(\sigma)$ and the approximate boundary, $\{h = \widetilde{H}(\sigma)\}$, is a starshaped Lipschitz function.

In order to linearize the problem (2.1) we need first of all approximate "values" for our unknowns, namely $H(\sigma)$ and $W(\mathbf{x})$.

We start form the second and we stipulate that the "approximate" potential is just the so-called *normal potential*

$$U(\mathbf{x}) = \widetilde{V}(\mathbf{x}) + \frac{1}{2}\omega^2(X^2 + Y^2) \ ,$$

where $\widetilde{V}(\mathbf{x})$ is harmonic and such that on an ellipsoid $\mathcal{E}$ of revolution, with given radius a and eccentricity e, $U(\mathbf{x}) = U_0$ constant. The radius a and the eccentricity e are adapted to the mean radius and flattening of the Earth. In other words $\widetilde{V}(\mathbf{x})$ is such that

$$\Delta\widetilde{V}(\mathbf{x}) = 0 \qquad \text{outside } \mathcal{E}$$
$$\widetilde{V}(\mathbf{x})|_{\mathcal{E}} = U_0 - \frac{1}{2}\omega^2(X^2 + Y^2)|_{\mathcal{E}} \ .$$

The value U_0 is chosen according to the relation (cf. [28], part I, §1.9).

$$U_0 = \frac{GM}{E}\text{arctg}\frac{E}{b} + \frac{1}{3}\omega^2 a^2 \ , \ (b = \sqrt{1-e^2}a, E = \sqrt{a^2 - b^2}) \ ,$$

where G is the universal constant of Newton's law and M is chosen to be equal to the mass of the Earth. With such a choice, the anomalous potential $T(\mathbf{x}) = W(\mathbf{x}) - U(\mathbf{x})$ has no centrifugal component, i.e., it is harmonic outside S, and when the coordinate system is placed with the origin in the barycentre and the Z axis along the rotation axis, in addition it satisfies the asymptotic relation

$$T(\mathbf{x}) = O\left(\frac{1}{|\mathbf{x}|^3}\right) \ . \tag{3.1}$$

Note that it is empirically known that T is 5 orders of magnitude smaller than W and therefore it is well suited for a linearization purpose.

As for the first unknown, i.e., $H(P)$, we take an approximate surface $\widetilde{S}$, called *telluroid*, defined by the so-called *Marussi mapping*, namely the point P, of unknown ellipsoidal height h_P, but of known ellipsoidal coordinates (λ_P, φ_P), is mapped to P' which is the point along the normal $\vec{\nu}$ to the ellipsoid through P, where the condition

$$U(\sigma, \widetilde{H}(\sigma)) = W(\sigma) \tag{3.2}$$

is satisfied.

An important remark is that, since $U(\mathbf{x})$ is a very smooth function, harmonic and therefore also real analytic even inside $\mathcal{E}$, down to a disk in the equatorial plane of radius ea, the regularity of $\widetilde{H}$ depends strictly on the regularity of $W(\sigma)$. In particular it is not difficult to see that if $W(\sigma)$ is a Lipschitz function, thanks to the implicit function theorem, $\widetilde{H}(\sigma)$ is Lipschitz too, i.e., its horizontal (i.e., parallel to $\mathcal{E}$) gradient is bounded, so that the surface $\widetilde{S}$ itself is Lipschitz and satisfies a cone condition. A property this which is intimately related to the theory of Sobolev solutions of elliptic equations (see [14, 17, 18]).

Now notice that

$$W(\sigma) = W(P) = U(\sigma, H(\sigma)) + T(\sigma, H(\sigma)) \equiv U(\sigma, \widetilde{H}(\sigma)) \ . \tag{3.3}$$

Let us put

$$\delta H = H(\sigma) - \widetilde{H}(\sigma) = h_P - h_{P'} = \zeta(\sigma) \ , \tag{3.4}$$

a quantity known in geodesy as height anomaly; moreover we denote by $\vec{\gamma}$ the vector

$$\vec{\gamma}(\sigma, h) = \nabla U(\sigma, h) \ , \tag{3.5}$$

called the *normal gravity vector*, and we set

$$\gamma(\sigma, h) = |\vec{\gamma}(\sigma, h)| \ , \tag{3.6}$$

the modulus of normal gravity.

Returning to (3.3) we have, retaining only first-order quantities,

$$W(\sigma) \cong U(\sigma, \widetilde{H}) + \zeta \vec{\nu} \cdot \vec{\gamma}(\sigma, \widetilde{H}) + T(\sigma, \widetilde{H}) \ ; \tag{3.7}$$

namely, defining the vector

$$\widetilde{\mathbf{n}} = -\frac{\vec{\gamma}}{\gamma} \ , \tag{3.8}$$

pointing in the direction of the normal vertical, and using again (3.3), we get

$$\zeta = \frac{T}{\vec{\nu} \cdot \widetilde{\mathbf{n}} \gamma} \ . \tag{3.9}$$

Since one can prove that, everywhere on $\widetilde{S}$,

$$\vec{\nu} \cdot \widetilde{\mathbf{n}} = \cos \widetilde{\delta} \ , \ \widetilde{\delta} < 10^{-5}$$

we see that

$$|\vec{\nu} \cdot \widetilde{\mathbf{n}} - 1| < 0.510^{-10} \ ,$$

a quantity that, multiplied by $\frac{T}{\gamma}$, is absolutely irrelevant. So (3.9) can be written as

$$\zeta = \frac{T}{\gamma} \tag{3.10}$$

called in literature the *Bruns's relation*.

To simplify the writing of next formulas we shall often use the symbol H_σ instead of $H(\sigma)$ and so forth for functions of σ; this should not be confused with the horizontal gradient $\nabla_\sigma H(\sigma)$, etc.

We now go onto the linearization of the boundary function $g(\sigma)$. One has, putting $\mathcal{U} = \left[\frac{\partial \vec{\gamma}}{\partial \mathbf{x}}\right]$, the matrix of the second derivatives of U,

$$\begin{aligned} g(\sigma) &= |\mathbf{g}(\sigma, H_\sigma)| = |\nabla W| = |\vec{\gamma}(\sigma, H_\sigma) + \nabla T(\sigma, H_\sigma)| \\ &\cong |\vec{\gamma}(\sigma, \widetilde{H}_\sigma) + \mathcal{U}\vec{\nu}\zeta + \nabla T(\sigma, \widetilde{H}_\sigma)| \\ &\cong \gamma(\sigma, \widetilde{H}_\sigma) + \frac{\vec{\gamma}}{\gamma} \cdot (\mathcal{U}\vec{\nu}\zeta + \nabla T) \\ &= \gamma(\sigma, \widetilde{H}_\sigma) - (\widetilde{\mathbf{n}} \cdot \mathcal{U}\vec{\nu})\zeta - \widetilde{\mathbf{n}} \cdot \nabla T \ . \end{aligned} \tag{3.11}$$

After defining the free air gravity anomaly

$$\Delta g = g(P) - \gamma(P') = g(\sigma, H_\sigma) - \gamma(\sigma, \widetilde{H}_\sigma) \tag{3.12}$$

and recalling (3.10), we can write (3.11) in the form

$$-\widetilde{\mathbf{n}} \cdot \nabla T - \frac{\widetilde{\mathbf{n}} \cdot \mathcal{U}\vec{\nu}}{\gamma} T = \Delta g \ , \tag{3.13}$$

where $\Delta g = \Delta g(\sigma)$ is indeed a known datum on $\widetilde{S}$. We reconduct (3.13) to a more usual form. First we note that

$$\widetilde{\delta} \cong |\widetilde{\mathbf{n}} - \vec{\nu}| < 10^{-5}$$

and, since[1]

$$|\nabla T| < 10^2 \text{mGal} \ ,$$

we have too

$$|(\widetilde{\mathbf{n}} - \vec{\nu}) \cdot \nabla T| < 10^{-3} \text{mGal} = 1\mu \text{ Gal}$$

which is a limit value for the error of the most precise absolute gravimeters.

Therefore we put

$$-\widetilde{\mathbf{n}} \cdot \nabla T \cong -\vec{\nu} \cdot \nabla T = -\frac{\partial T}{\partial h} \ . \tag{3.14}$$

Moreover, let us compute

$$\nabla\gamma \quad = \quad \frac{1}{2\gamma}\nabla\gamma^2 = \frac{1}{2\gamma}\nabla(\vec{\gamma} \cdot \vec{\gamma}) = \frac{1}{\gamma}\mathcal{U}\vec{\gamma} = -\mathcal{U}\widetilde{\mathbf{n}} \ .$$

[1]The gravity modulus is measured in Gal units (1Gal= 1 cm s^{-2}); in these units g as well as γ range around 10^3 Gal on S. Moreover it is 1mGal $= 10^{-3}$ Gal, 1μ Gal $= 10^{-6}$ Gal.

Taking the scalar product with $\vec{\nu}$ and recalling that $\mathcal{U}$ is a symmetric matrix, we get

$$\vec{\nu} \cdot \nabla\gamma = \frac{\partial \gamma}{\partial h} = -\vec{\nu} \cdot \mathcal{U}\widetilde{\mathbf{n}} = -\widetilde{\mathbf{n}} \cdot \mathcal{U}\vec{\nu} \ . \tag{3.15}$$

Finally, returning to (2.14), we obtain

$$-\frac{\partial T}{\partial h} + \frac{\frac{\partial \gamma}{\partial h}}{\gamma} T = \Delta g \ , \tag{3.16}$$

which is the ordinary form of the boundary condition of the LSMP. Summarizing, the linearized SGBVP at this stage is defined by: given $\widetilde{S}$ and $\Delta g(\sigma)$ on $\widetilde{S}$, find T such that

$$\begin{cases} \Delta T = 0 & \text{outside } \widetilde{S} \\ -\frac{\partial T}{\partial h} + \frac{\frac{\partial \gamma}{\partial h}}{\gamma} T \Big|_{\widetilde{S}} = \Delta g(\sigma) & \text{on } \widetilde{S} \\ T = O\left(\frac{1}{|\mathbf{x}|^3}\right) & |\mathbf{x}| \to \infty \ , \end{cases} \tag{3.17}$$

and then compute

$$\zeta(\sigma) = \frac{T(\sigma)}{\gamma} \Rightarrow H(\sigma) = \widetilde{H}(\sigma) + \zeta(\sigma) \ . \tag{3.18}$$

However we have already learnt that to balance the lack of first degree harmonics in T we have to add to Δg a linear combination of three suitable functions. The same reasoning can be extended to the zero degree term requiring the addition of another unknown constant in the second of (3.17). If we can assume that some asymptotic coefficients $\{\overline{T}_{\ell m}\}$ of a harmonic expansion of T, valid outside some sphere $\overline{S}$ strictly enclosing $\widetilde{S}$, are known, we can even extend the above reasoning to degrees $\ell > 1$.

This is the case, thanks to the space technology which provides us with models of anomalous potential like

$$\overline{T} = \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} \overline{T}_{\ell m} \left(\frac{\overline{R}}{r}\right)^{\ell+1} Y_{\ell m}(\sigma) \ . \tag{3.19}$$

with $Y_{\ell m}(\sigma)$ the ordinary set of spherical harmonic functions.

The lower the L, the lower is the error (also called commission error ([28], §3.8) that is introduced by (3.19). However, for low values of L, the over all error is so tiny that it can be quite safely accepted. For instance for $L = 24$ the overall mean commission error in terms of

$$\Delta\overline{g} = -\frac{\partial \overline{T}}{\partial h} + \frac{\frac{\partial \gamma}{\partial h}}{\gamma} \overline{T} \tag{3.20}$$

has a standard deviation of $\sigma \cong 1\mu$ Gal, which is certainly negligible.

On the other hand, we will see that the introduction of (3.19) as additional known data, will give us more freedom for the geometric constraints we will be forced to put on $\widetilde{S}$.

So we decide to further modify (3.17) in the following way: we put

$$T = \overline{T} + u \ , \ \Delta g = \Delta \overline{g} + \overline{f} \ ,$$

moreover we introduce suitable functions $\overline{\psi}_{jk}(\mathbf{x})$, that will be specified in §4, and unknown constants a_{jk}, with the ordering $0 \le j \le L \ , \ -j \le k \le j$, like that of degrees and order of spherical harmonics, and we state the problem

$$\begin{cases} \Delta u = 0 \qquad \text{outside } \widetilde{S} \\ -\frac{\partial u}{\partial h} + \frac{\frac{\partial \gamma}{\partial h}}{\gamma} u \Big|_{\widetilde{S}} = \overline{f}(\sigma) + \sum_{j=0}^{L} \sum_{k=-j}^{j} a_{jk} \overline{\psi}_{jk}(\mathbf{x})\big|_{\widetilde{S}} \quad \text{on } \widetilde{S} \\ u = O\left(\frac{1}{|\mathbf{x}|^{L+2}}\right) \ , \quad |\mathbf{x}| \to \infty \end{cases} \tag{3.21}$$

Note has to be taken that the harmonic potential u will have a series representation, outside a large sphere $\overline{S}$, lacking the first L degrees, i.e., starting from degree $L+1$; whence the asymptotic condition in the third relation of (3.21).

The last step we take in this section is to transform the boundary conditions, namely the second of (3.21), into a very convenient perturbative form.

This will be derived from the estimates contained in the following Proposition, based on elementary considerations on ellipsoidal geometry and the normal gravity formulas. A detailed proof is given in ([29] and [28], §15.2).

Proposition 3.1. *Let us put $(\lambda, \varphi) = \sigma$ for the spherical longitude and latitude of a point P and call $\mathbf{e}_r(\sigma), \vec{\nu}(\sigma)$ the radial unit vector and the ellipsoidal normal unit vector at P respectively; then one has*

$$\vec{\varepsilon}(\sigma) = \vec{\nu}(\sigma) - \mathbf{e}_r(\sigma) \ , \ \varepsilon_+ = \max_{\sigma} |\vec{\varepsilon}(\sigma)| \tag{3.22}$$

$$\varepsilon_+ \cong e^2 \ , \ (e^2 \cong 6,7 \cdot 10^{-3}) \tag{3.23}$$

where e is the eccentricity of the Earth ellipsoid $\mathcal{E}$. Moreover, let us put

$$R(\sigma) = |\mathbf{x}(\sigma)| \ , \ \{\mathbf{x}(\sigma) \in \widetilde{S}\} \tag{3.24}$$

and

$$\eta(\sigma) = R(\sigma) \frac{\partial \gamma}{\partial h}(\sigma) + 2 \ ; \ \eta_+ = \max_{\sigma} |\eta(\sigma)| \tag{3.25}$$

then one has

$$\eta_+ \cong 2e^2 \ . \tag{3.26}$$

With such estimates in mind we can write

$$\frac{\partial u}{\partial h} = \vec{\nu} \cdot \nabla u = \mathbf{e}_r \cdot \nabla u + \vec{\varepsilon} \cdot \nabla u = \frac{\partial u}{\partial r} + \vec{\varepsilon} \cdot \nabla u \ .$$

For the sake of brevity we shall use in the rest of the work the notation

$$\frac{\partial u}{\partial r} = u' \ , \tag{3.27}$$

so that the above relation can be written

$$\frac{\partial u}{\partial h} = u' + \vec{\varepsilon} \cdot \nabla u \ . \tag{3.28}$$

Furthermore we have, by using definition (3.27),

$$\frac{\frac{\partial\gamma}{\partial h}}{\gamma}u = \frac{\gamma'}{\gamma}u = -\frac{2}{r}u + \frac{\eta}{r}u \ . \tag{3.29}$$

So, multiplying the second of (3.21) by $r = R(\sigma)$, changing sign and redefining

$$f(\sigma) = -R_\sigma \overline{f}(\sigma) \ , \ \psi_{jk}(\mathbf{x}) = -r\overline{\psi}_{jk}(\mathbf{x}) \ ,$$

we get finally

$$ru' + 2u + (r\vec{\varepsilon}\cdot\nabla u - \eta u)|_{\widetilde{S}} = f(\sigma) + \sum a_{jk}\psi_{jk}\Big|_{\widetilde{S}} . \tag{3.30}$$

We already note that, according to our estimates (3.23), (3.26), the differential operator

$$D = r\vec{\varepsilon}\cdot\nabla - \eta \tag{3.31}$$

satisfies the inequality

$$|Du| \le \varepsilon_+ r|\nabla u| + \eta_+|u| \ , \tag{3.32}$$

so that it can be considered as a perturbation of the main boundary operator $D_0 = \left(r\frac{\partial}{\partial r} + 2\right)$.

Summarizing, we can finally state the Linearized Scalar Molodensky Problem in the following modified, perturbative form:

$$\begin{cases} \Delta u = 0 & \text{outside } \widetilde{S} \\ ru' + 2u = f + \displaystyle\sum_{i=0}^{L}\sum_{x=-j}^{j} a_{jk}\psi_{jk} - Du & \text{on } \widetilde{S} \\ u = O\left(\frac{1}{|\mathbf{x}|^{L+2}}\right) \ , & |\mathbf{x}| \to \infty \ . \end{cases} \tag{3.33}$$

In (3.33) the functions $\psi_{jk}(\mathbf{x})$ will be chosen in the next section, and will depend ultimately only on the shape of $\widetilde{S}$, while L will be left a free parameter to be fixed later on, to get convenient conditions for the stability of the solution.

Note that in (3.33) the unknowns are the potential u and the $(L+1)^2$ constants $\{a_{jk}\}$, which in the minimum case reduce to $(1+1)^2 = 4$ unknowns.

The analysis of the problem (3.33) will occupy us in the next two sections, following the ideas of [28], §15.4.

For a more general treatment of the oblique derivative problem, one can consult the recent book [6].

4. The analysis of the Simple Molodensky Problem

We define the Simple Molodensky Problem as the linearized SGBVP (3.21) when the $\frac{\partial}{\partial h}$ is approximated by $\frac{\partial}{\partial r}$ and $(\gamma)^{-1}\cdot\frac{\partial\gamma}{\partial h}$ is computed by using the spherical expression

$$\gamma^{-1}\frac{\partial\gamma}{\partial h} \sim \left(\frac{\mu}{r^2}\right)^{-1}\left(-2\frac{\mu}{r^3}\right) = -\frac{2}{r} \ . \tag{4.1}$$

This leads to the boundary relation (3.30) and then to a problem like (3.33), where however the perturbative term is disregarded, namely

$$\begin{cases} \Delta u = 0 & \text{in } \widetilde{\Omega} \\ ru' + 2u = f + \Sigma a_{jk}\psi_{jk} & \text{on } \widetilde{S} \\ u = O\left(\frac{1}{|\mathbf{x}|^{2+2}}\right) & |\mathbf{x}| \to \infty \ ; \end{cases} \tag{4.2}$$

this is called in geodetic literature the *Simple Molodensky Problem* ([15, 20]).

To study the existence, uniqueness and specially stability of the solution of (4.2) in a suitable Sobolev space, is the task of this section, a first step to pass then to the same work for the solution of (3.33), obtained by elementary perturbation techniques. A first goal of the section is to prove that (4.2) has a unique stable solution in H_1, i.e., a space of harmonic functions with L^2 gradient on the boundary $\widetilde{S}$, when the following hypotheses are fulfilled

$$\widetilde{S} \equiv \{r = R_\sigma\} \text{ is Lipschitz} \tag{4.3}$$

or

$$|\nabla_\sigma R_\sigma| \le C \quad \sigma \text{ a.e.} \tag{4.4}$$

and

$$f \in L^2(\sigma) \tag{4.5}$$

or

$$\int_\sigma f^2(\sigma) d\sigma < +\infty \ . \tag{4.6}$$

Under such hypotheses we shall prove that the solution of (4.2) satisfies

$$\| \ u \ \|_{H_1} \le C \ \| \ f \ \|_{L^2_\sigma} \tag{4.7}$$

and

$$j = 0, 1, \ldots, L \ , \ |k| \le j \ , \ |a_{jk}| \le C \ \| \ f \ \|_{L^2_\sigma} \ . \tag{4.8}$$

Unfortunately however we know that the constant C exists finite, but we do not know its value, so that the result appears to be too weak to establish a perturbation theory for (3.33). Then we shall derive a more restrictive theorem, where we find the value of C but we have to suppose that a model of the anomalous potential, complete up to some degree and order L, is known and the inclination of $\widetilde{S}$ with respect to $\mathbf{e}_r$ is bounded above by some suitable angle $I_+ < \pi/2$. This result will then be easily carried over to the analysis of (3.33).

Preliminary to the development of the analysis above described, is the proof of 4 propositions.

Proposition 4.1. *Let the hypotheses* (4.4) *and* (4.6) *be satisfied. Then there is one and only one u harmonic in Ω such that*

$$f(\sigma) = \lim_{r \to R_\sigma} u(r, \sigma) \qquad \sigma \text{ a.e.} \ ; \tag{4.9}$$

in addition, if we call $\overline{R}$ any Brillouin radius for $\widetilde{S}$, i.e.,

$$\overline{R} > R_+ = \sup_\sigma \{R_\sigma\} \ , \tag{4.10}$$

and we put

$$M(u,\sigma) = \sup_{R_\sigma < r \leq \overline{R}} |u(r,\sigma)| \tag{4.11}$$

we have

$$\| M(u,\sigma) \|^2_{L^2_\sigma} = \frac{1}{4\pi} \int M(u,\sigma)^2 d\sigma \leq C^2 \| f \|^2_{L^2_\sigma} , \tag{4.12}$$

where the constant C in (4.12) *is depending only on the shape of $\widetilde{S}$.*

Proof. This is a generalization of the Fatou theorem ([3]), adapted to the present configuration, proved by B.E.J. Dahlberg, and reported in [14], §3. The theory is essentially based on Hardy–Littlewood maximal inequalities. So we shall not reproduce the proof, but we rather send the interested reader to the mentioned paper. Here, we note only that (4.12), by applying a simple approximation of f by a continuous function, implies as well that if S_δ is a family of uniformly Lipschitz surfaces, $S_\delta = \{r = R_{\delta,\sigma} = R_\delta + h_\sigma \ ; 0 < h_\sigma \leq \delta\}, \sup_\sigma |\nabla R_{\delta,\sigma}| \leq C$, then for any u harmonic in Ω attaining boundary values $f(\sigma) \in L^2_\sigma$, one has too

$$\lim_{\delta \to 0} \int_\sigma |u(R_{\delta,\sigma},\sigma) - f(\sigma)|^2 d\sigma = 0 . \tag{4.13}$$

□

Remark 4.1. We notice that the first proof of existence of harmonic functions, or more generally of solutions of Poisson's equation, when boundary data on $\widetilde{S}$ are in $L^2(\widetilde{S})$ is due to Cimmino ([5]). However the approach of this author requires the boundary $\widetilde{S}$ to be of class $C^2(\sigma)$ so as to guarantee that it has bounded curvature. A different proof can be given, much simpler than Hardy–Littlewood theory, where the boundary $\widetilde{S}$ is required only to be of class $H_{1+\lambda}(\sigma)$. This constitutes thus a generalization of Cimmino's result, though it is not as general as Dahlberg's theorem.

Remark 4.2. As a further remark we observe that a consequence of Proposition 4.1 is that we can define a Hilbert space of harmonic functions in $\widetilde{\Omega}$, that we shall call H_0, which is just the isometric image of L^2_σ, through the solution of the corresponding Dirichlet problem, namely

$$\begin{cases} u \in H_0(\widetilde{S}) \leftrightarrow f(\sigma) = u|_{\widetilde{S}} \in L^2_\sigma \\ \| u \|^2_0 \equiv \| f \|_{L^2_\sigma} \ . \end{cases} \tag{4.14}$$

Note should be taken that with the above definition,

$$\| u \|^2_0 = \int_\sigma u^2(R_\sigma,\sigma) d\sigma , \tag{4.15}$$

which is not exactly equal to $\int_{\widetilde{S}} u^2(R_\sigma,\sigma) dS$, although it is equivalent to the more common norm when $\widetilde{S}$ is a starshaped Lipschitz domain, because (see [28], §15.1)

$$\begin{cases} dS = R_\sigma J d\sigma , \\ J = (\cos I)^{-1} , \ \cos I = \mathbf{n} \cdot \mathbf{e}_r , \quad 1 \leq J \leq A < +\infty . \end{cases} \tag{4.16}$$

Proposition 4.2. *There are functions* $\{\psi_{jk}\} \in H_0$ *such that*

$$\forall u \in H_0 \ , \langle \psi_{jk} \ , \ u \rangle_0 = \overline{u}_{jk} \equiv \frac{1}{4\pi} \int u(\overline{R}, \sigma) Y_{jk}(\sigma) d\sigma \ , \tag{4.17}$$

that are linearly independent in H_0.

Moreover, if we call

$$V_L \equiv \{\psi_{jk} \ ; \ 0 \le i \le L \ , \ |k| \le j\} \tag{4.18}$$

and

$$H_{0L} = V_L^{\perp} \equiv \{u \in H_0 \ ; \langle \psi_{jk} \ , \ u \rangle_0 = 0 \quad \psi_{jk} \in V_L\} \tag{4.19}$$

then we have

$$u \in H_{0L} \Leftrightarrow u = O\left(\frac{1}{r^{L+2}}\right) \ . \tag{4.20}$$

Proof. That $\{\psi_{jk}\}$ exist is just an application of the Riesz theorem when we realize that, (4.10) being satisfied, we have from (4.17)

$$|\langle \psi_{jk} \ u \rangle_0| \le \left(\frac{1}{4\pi} \int u(\overline{R}, \sigma)^2 d\sigma\right)^{1/2} \le C \parallel u \parallel_0 \ . \tag{4.21}$$

Moreover $\{\psi_{jk}\}$ are linearly independent because, if for some constants $\{c_{jk}\}$

$$\sum_{j=0}^{L} \sum_{k=-j}^{j} c_{jk} \psi_{jk} \equiv 0$$

then, by exploiting the definition (4.17) and the orthogonality of surface spherical harmonics, $\forall \ell \le L, |m| \le \ell$,

$$0 = \sum_{j=0}^{L} \sum_{k=-j}^{j} c_{jk} \left\langle \psi_{jk} \ , \ Y_{\ell m} \left(\frac{R_\sigma}{r}\right)^{\ell+1} \right\rangle_0 = c_{\ell m} \ .$$

Finally (4.20) holds because the very definition (4.19) yields the double implication

$$u = O\left(\frac{1}{r^{L+2}}\right) \Leftrightarrow \{\overline{u}_{\ell m} = 0 \ , \ \ell \le L \ \ |m| \le \ell\} \Leftrightarrow u \in H_{0L} \ . \qquad \square$$

Proposition 4.3. *Let us consider the modified Dirichlet problem for all* $L \ge 0$,

$$\begin{cases} \Delta u = 0 & \text{in } \widetilde{\Omega} \ , \\ u|_{\widetilde{S}} = f + \displaystyle\sum_{j=0}^{L} \sum_{k=-j}^{j} a_{ik} \psi_{jk} & \text{on } \widetilde{\Omega} \ ; \end{cases} \tag{4.22}$$

when $f \in L^2_\sigma$, *then* (4.22) *has one and only one solution* $u \in H_{0L}$, $\{a_{jk}\} \in R^{(L+1)^2}$, *such that*

$$\parallel u \parallel_0 \le \parallel f \parallel_{L^2_\sigma} \tag{4.23}$$

and

$$|a_{jk}| \le C \parallel f \parallel_{L^2_\sigma} \ . \tag{4.24}$$

Proof. Let us define $\widetilde{u}$ such that

$$\begin{cases} \Delta\widetilde{u} = 0 & (\text{in } \widetilde{\Omega}) \\ \widetilde{u}|_{\widetilde{S}} = f \ , \end{cases} \tag{4.25}$$

which exists and satisfies

$$\| \widetilde{u} \|_0 \equiv \| f \|_{L^2_\sigma} \tag{4.26}$$

in force of Remark 4.2. Note that in general $\widetilde{u} \notin H_{0L}$, not even for $L = 0$, because functions $u \in H_{0,L=0}$ have the asymptotic behaviour $u = O\left(\frac{1}{r^2}\right)$.

Since the second of (4.22) is equivalent to $u \in H_{0L}$, (4.22) itself is just equivalent to solving

$$\widetilde{u} = u - \sum_{j=0}^{L} \sum_{k=-j}^{j} a_{jk}\psi_{jk} \tag{4.27}$$

with

$$u \in H_{0L} = V_L^{\perp} \text{ and } \Sigma a_{jk}\psi_{jk} \in V_L \ . \tag{4.28}$$

So the solution of (4.27), introducing the orthogonal projector

$$P_L \equiv \text{ orthogonal projector on } V_L \ .$$

is just

$$\sum_{\ell=0}^{L} \sum_{k=-j}^{j} a_{jk}\psi_{jk} = -P_L\widetilde{u} \ , \tag{4.29}$$

$$u = (I - P_L)\widetilde{u} \ . \tag{4.30}$$

In turn, since $\{\psi_{jk}\}$ are linearly independent, (4.29) implies

$$|a_{jk}| \leq C \| P_L\widetilde{u} \|_0 \leq C \| \widetilde{u} \|_0 = C \| f \|_{L^2_\sigma}$$

which is (4.24); moreover (4.30) implies

$$\| u \|_0 \leq \| \widetilde{u} \|_0 \equiv \| f \|_{L^2_\sigma} \ .$$

which is (4.23). □

Finally we shall prove a proposition that introduces a generalized version of energy integrals already used in Geodesy for the analysis of the GBVP ([13, 27, 28]).

Proposition 4.4. *Let $u \in H_{0L}$ and $\nabla u \in H_{0L+1}$, let α be a real number such that*

$$\alpha < 2L + 4 \ , \tag{4.31}$$

then the following identity holds

$$\begin{aligned} \int_{\widetilde{S}} d\sigma R_\sigma^{\alpha+2} |\nabla u|^2 = 2(\alpha - 1) \int_{\widetilde{\Omega}} r^{\alpha-1} u'^2 d\Omega \\ - \alpha \int_{\widetilde{\Omega}} d\Omega r^{\alpha-1} |\nabla u|^2 + 2 \int_{\widetilde{S}} d\widetilde{S} R_\sigma^{\alpha} u' u_n \ . \end{aligned} \tag{4.32}$$

Proof. Note that $\int_{\widetilde{S}} d\sigma$ in the left-hand side of (4.33) means that $|\nabla u|^2$ is computed on $\widetilde{S}$, i.e. with $r = R_\sigma$, and the resulting function integrated in $d\sigma$ on the unit sphere. Similar is the situation of the right-hand side where we can put

$$d\widetilde{S} = R_\sigma^2 J d\sigma$$

and $u' u_n$ has again to be computed on $\widetilde{S}$. To prove (4.32) one can start from the differential identity

$$\nabla \cdot (r^\alpha u' \nabla u) = (\alpha - 1) r^{\alpha-1} u'^2 + r^{\alpha-1} |\nabla u|^2 + \frac{1}{2} r^\alpha \frac{\partial}{\partial r} |\nabla u|^2 \ . \tag{4.33}$$

By integrating (4.33) in $\widetilde{\Omega}$ and using Gauss theorem and the integration by parts in dr, one derives (4.32). The condition (4.31) is necessary in order that all integrals be convergent and integrals on large spheres tend to zero when the radius tends to infinity.

We only mention here, for some possible future use, that (4.32) can be further elaborated, trying to bring at least in part the $\int d\Omega r^{\alpha-1} |\nabla u|^2$ on the surface $\widetilde{S}$; so we arrive at the identity

$$\int_{\widetilde{S}} d\sigma R_\sigma^{\alpha+2} |\nabla u|^2 = 2(\alpha - 1) \int_{\widetilde{\Omega}} r^{\alpha-1} u'^2 d\Omega + \int_{\widetilde{S}} R_\sigma^{\alpha-1} (2 R_\sigma u' + \alpha u) u_n d\widetilde{S} \tag{4.34}$$

$$- \frac{\alpha(\alpha-1)}{2} \int_{\widetilde{S}} R_\sigma^\alpha u^2 d\sigma - \frac{\alpha^2(\alpha-1)}{2} \int_{\widetilde{\Omega}} d\Omega r^{\alpha-3} u^2 \ . \qquad \square$$

We are ready now to go to the fundamental theorem of this section.

Theorem 4.1. *Let $\widetilde{S}$ and f satisfy* (4.4) *and* (4.5) *respectively; then there is one and only one solution u of the SMP* (4.2)*, such that $u \in H_1 \cap H_{0L}$ for all $L \geq 1$ and*

$$\| u \|_1 \leq C \| f \|_{L^2_\sigma} \ , \tag{4.35}$$

where the constant C depends only on the shape of $\widetilde{S}$, namely on the Lipschitz constant of $\widetilde{S}$.

Proof. Let us call v the unique solution of the generalized Dirichlet problem

$$\begin{cases} \Delta v = 0 & \text{in } \widetilde{\Omega} \\ v|_{\widetilde{S}} = f + \displaystyle\sum_{j=0}^{L} \sum_{k=-j}^{j} a_{jk} \psi_{jk} & \text{on } \widetilde{S} \\ v \in H_{0L} \ , \end{cases} \tag{4.36}$$

which exists and is unique on the basis of Proposition 4.3, in particular when $L \geq 1$.

If you assume that (4.2) has one solution in $H_1 \cap H_{0L}$, then indeed

$$v = r u' + 2u \tag{4.37}$$

is the solution of (4.6). That v is harmonic in $\widetilde{\Omega}$ is a simple direct computation; that $v \in H_{0L} (L \geq 1)$ descends from the fact that $u \in H_1$ and that, for $r \geq \overline{R}$,

v can be expressed in spherical harmonics, so that (4.37) implies the well-known Stokes relation

$$\overline{v}_{jk} = -\frac{(j-1)}{\overline{R}}\overline{u}_{jk} \ , \quad (j \geq L+1) \tag{4.38}$$

and therefore $\overline{v}_{jk} = 0$ for $j \leq L$ if the same is true for $\overline{u}_{jk}$. Note that the condition $L > 1$ has to be set up exactly to allow a one-to-one correspondence between $\overline{v}_{jk}$ and $\overline{u}_{jk}$, for all $j \geq L+1$.

Now we have to reverse the above statement.

Namely given v satisfying (4.36) we start defining u satisfying (4.37), i.e.,

$$u(r,\sigma) = -\frac{1}{r^2}\int_r^{+\infty} v(s,\sigma)sds \ . \tag{4.39}$$

That u is harmonic in $\widetilde{\Omega}$ is verified by direct calculation of Δu in spherical coordinates; that $u = O\left(\frac{1}{r^{L+2}}\right)$, when $L \geq 1$, comes from (4.38) and the fact that $\overline{v}_{jk} \equiv 0$ for $j \leq L$. that $u \in H_0$ we verify by the following reasoning. Let us put

$$\begin{aligned} (R_\sigma \leq r \leq \overline{R}) \ , \quad u(r,\sigma) &= -\frac{1}{r^2}\int_r^{\overline{R}} v(s,\sigma)sds - \frac{1}{r^2}\int_{\overline{R}}^{+\infty} v(s,\sigma)sds \\ &\equiv -\frac{1}{r^2}u_1(r,\sigma) - \frac{1}{r^2}u_2(\sigma) \ ; \end{aligned} \tag{4.40}$$

since R_σ is bounded above and below, we need only to prove that $u_1(R_\sigma,\sigma), u_2(\sigma) \in L^2_\sigma$. On the other hand, recalling (4.11), (4.12), (4.23),

$$\begin{aligned} \int d\sigma u_1(R_\sigma,\sigma)^2 &\leq \int d\sigma \int_{R_\sigma}^{\overline{R}} v^2(s,\sigma)s^2 ds \cdot (\overline{R} - R_\sigma) \\ &\leq C \parallel v \parallel_0^2 \leq C \parallel f \parallel_{L^2_\sigma}^2 \end{aligned} \tag{4.41}$$

Moreover we have

$$\begin{aligned} u_2(\sigma) = \int_{\overline{R}}^{+\infty} v(s,\sigma)sds &= \int_{\overline{R}}^{+\infty} \sum_{j=(L+1)} \overline{v}_{jk}\left(\frac{\overline{R}}{s}\right)^{j+1} Y_{jk}(\sigma)sds \\ &= \overline{R}^2 \sum_{j=L+1} \frac{\overline{v}_{jk}}{j-1} Y_{jk}(\sigma) \ , \end{aligned}$$

so that, recalling again (4.11), (4.12), (4.23),

$$\parallel u_2(\sigma) \parallel_{L^2_\sigma} \leq \left(\overline{R}^2 \Sigma \overline{v}_{jk}^2\right)^{1/2} = \overline{R}\left(\frac{1}{4\pi}\int v(\overline{R},\sigma)^2 d\sigma\right)^{1/2} \leq C \parallel f \parallel_{L^2_\sigma} \ . \tag{4.42}$$

We note that combining (4.41) and (4.42) one has

$$\parallel u \parallel_0 \leq C \parallel f \parallel_{L^2_\sigma} \ . \tag{4.43}$$

On the other hand, since u satisfies (4.37), we obviously have

$$\parallel u' \parallel_0 = \parallel \frac{1}{r}(v - 2u) \parallel_0 \leq C \parallel f \parallel_{L^2_\sigma} \ . \tag{4.44}$$

Now we finally prove that $u \in H_1$. In this respect we note that a natural norm in H_1 is

$$\| u \|_1 = \| |\nabla u| \|_0 = \int_{\widetilde{S}} |\nabla u|^2 d\sigma \ ; \tag{4.45}$$

that this is a true norm is obvious because $\| u \|_1 = 0$ implies $|\nabla u| = 0$ on $\widetilde{S}$ and then by the maximum principle

$$\sup_{\widetilde{\Omega}} |\nabla u|^2 \leq \sup_{\widetilde{S}} |u_x|^2 + \sup_{\widetilde{S}} |u_y|^2 + \sup_{\widetilde{S}} |u_z|^2 = 0 \ ,$$

so that u is constant in $\widetilde{\Omega}$ and being regular at infinity, has to be zero in the whole $\widetilde{\Omega}$.

Now we can apply Proposition 4.4 with $\alpha = 0$ and $L \geq 1$, getting

$$\int d\sigma R_\sigma^2 |\nabla u|^2 = -2 \int_{\widetilde{\Omega}} r^{-1} u'^2 d\Omega + 2 \int d\widetilde{S} u' u_n \ ; \tag{4.46}$$

recalling (4.16) and putting

$$J_+ = \sup J \ , \ R_- = \inf R_\sigma > 0 \ , \ R_+ = \sup R_\sigma \ ,$$

we find then

$$\begin{aligned} R_-^2 \int d\sigma |\nabla u|^2 &\leq 2R_+^2 \int d\sigma J |u'| |u_n| \\ &\leq 2R_+^2 J_+ \left(\int d\sigma u'^2 \right)^{1/2} \left(\int d\sigma u_n^2 \right)^{1/2} \\ &\leq 2R_+^2 J_+ \left(\int d\sigma u'^2 \right)^{1/2} \left(\int d\sigma |\nabla u|^2 \right)^{1/2} \ . \end{aligned} \tag{4.47}$$

Combining (4.47) and (4.44) we readily arrive at

$$\| u \|_1 \leq C \| f \|_{L^2_\sigma} \ . \tag{4.48}$$

Strictly speaking we can apply the above reasoning only if we know a priori that $u \in H_1$, which is a condition for the validity of Proposition 4.4. Yet we can easily circumvent the difficulty by the following approximation process.

Let us put

$$u_\lambda(r, \sigma) = u(\lambda r, \sigma) \tag{4.49}$$

with $\lambda \geq 1$; it is obvious that this is a family of functions harmonic in $\widetilde{\Omega}$ and that in particular the trace of $u_\lambda(r, \sigma)$ on $\widetilde{S}$ is for $\lambda > 1$ a smooth function, i.e.,

$$u_\lambda(R_\sigma, \sigma) = u(\lambda R_\sigma, \sigma)$$

which is bounded in σ and even with a bounded $|\nabla_\sigma u_\lambda(R_\sigma, \sigma)|$, since R_σ is Lipschitz. So the function

$$\lambda > 1 \ , \ |\nabla u_\lambda|^2 \big|_{\widetilde{S}} = u'(\lambda R_\sigma, \sigma)^2 + \frac{1}{R_\sigma^2} |u_\sigma(\lambda R_\sigma, \sigma)|^2$$

is certainly bounded and therefore in L^2_σ, i.e., $u_\lambda \in H_1, \ \forall \lambda > 1$.

Let us note for the sake of clarity that indeed $\nabla_\sigma u(\lambda R_\sigma, \sigma) \neq u_\sigma(\lambda R_\sigma, \sigma)$ since the second function is just $u_\sigma(\lambda R_\sigma, \sigma) = \nabla_\sigma u(\lambda r, \sigma)|_{r=R_\sigma}$.

So (4.47) can be legitimately applied to the difference $u_\lambda - u_{\lambda'}$, i.e.,

$$\| u_\lambda - u_{\lambda'} \|_1 \leq C \| u'_\lambda - u'_{\lambda'} \|_0 \ ; \tag{4.50}$$

on the other hand we have already noted that $u' \in H_0$, so that $u'_\lambda \to u'$ in L^2_σ.

Therefore from (4.50), if we take a sequence $\lambda_n \to 1$ we see that u_{λ_n} is a Cauchy sequence in H_1 and then $u_{\lambda_u} \to \overline{u} \in H_1$. But we already know that $u_{\lambda_n} \to u$ in H_0 and therefore it has to be $u = \overline{u}$, namely $u \in H_1$. So our conclusion (4.48) holds. □

In principle at this point one could be content with the result of Theorem 4.1 considering that indeed a perturbation of the boundary operator $r\frac{\partial}{\partial r} + 2$, continuous in H_1, can be treated claiming that when the perturbation is "small enough", then a solution still exists and is unique in H_1. Yet, since we do not know the value of C in (4.48), we are not able to specify how small should actually be the perturbation. So we need to find a value for C. However, now that we know that $u \in H_1$, we can resort to the generalized energy identities of Proposition 4.4, but this time with $\alpha = 1$ and $L \geq 1$, as already done in ([27, 29]). In this case (4.32) reads

$$\begin{aligned} \int_{\widetilde{S}} d\sigma R^3_\sigma |\nabla u|^2 &= -\int_{\widetilde{\Omega}} d\Omega |\nabla u|^2 + 2\int_{\widetilde{S}} d\sigma J R^3_\sigma u' u_n \\ &= \int_{\widetilde{S}} d\sigma J R^2_\sigma (u u_n + 2R_\sigma u' u_n) \end{aligned} \tag{4.51}$$

On the other hand, on $\widetilde{S}$

$$R_\sigma u' = -2u + f \ ,$$

so that (4.51) becomes

$$\int_{\widetilde{S}} d\sigma R^3_\sigma |\nabla u|^2 = \int_{\widetilde{S}} d\sigma J R^2_\sigma (2f u_n - 3u u_n) \ ; \tag{4.52}$$

we are ready now to prove the following Theorem 4.2.

Theorem 4.2. *Let us put*

$$R_- = \inf R_\sigma, R_+ = \sup R_\sigma \ , \ k = \frac{R_+}{R_-} \ , \ \delta R = R_+ - R_- \ ,$$

and also

$$\begin{aligned} J_+ &= \sup J(\sigma) = \frac{1}{\cos I_+} \\ I_+ &= \sup I = \sup \arccos \mathbf{n} \cdot \mathbf{e}_r \\ \mathbf{n}(\sigma) &= \ \textit{normal to } \widetilde{S} \textit{ at } (R_\sigma, \sigma) \ ; \end{aligned}$$

then, if the condition

$$3J_+k^3C_{0L} = 3J_+^2k^3\left(\frac{\delta R}{R_-} + \frac{4}{2L+5}\right) < 1 \tag{4.53}$$

is satisfied, one has

$$\| u \|_1 \leq C_{1L} \| f \|_0 \tag{4.54}$$

with

$$C_{1L} = \frac{1}{R_-}\frac{2J_+k^2}{1-3J_+k^3C_{0L}} \ . \tag{4.55}$$

Proof. The proof comes from a combination of (4.52) and the following lemma.

Lemma 4.1. *Let u be harmonic in $\widetilde{\Omega}$ and such that $u \in H_1 \cap H_{0L}$, then*

$$\| u \|_0 \leq R_+C_{0L} \| u \|_1 \ , \tag{4.56}$$

Proof. Let us call $u_+(\sigma) = u(R_+, \sigma)$; then one can write

$$u_\sigma = u_\sigma - u_+ + u_+ \ , \qquad \| u \|_0 \leq \| u - u_+ \|_0 + \| u_+ \|_0 \ . \tag{4.57}$$

Now consider in $\Omega_+ \equiv \{r \geq R_+\}$ the identities

$$u_+(\sigma) = \sum_{u=L+1}^{+\infty}\sum_{m=-n}^{u} u_{+nm}Y_{nm}(\sigma)\left(\frac{R_+}{r}\right)^{n+1} \ , \tag{4.58}$$

$$\int_{\Omega_+} u'(r,\sigma)^2 d\Omega = 4\pi R_+ \sum_{u=L+1}^{+\infty}\sum_{m=-n}^{n} u_{+nm}^2\frac{(n+1)^2}{2n+1} \ , \tag{4.59}$$

Since

$$(n \geq L+1) \ , \quad \frac{(n+1)^2}{2n+1} \geq \frac{2L+5}{4}$$

and

$$\| u_+ \|_0^2 = 4\pi\Sigma u_{+nm}^2 \ ,$$

from (4.59) we deduce

$$\| u_+ \|_0^2 \leq \frac{1}{R_+}\frac{4}{2L+5}\int_{\Omega_+} u'(r,\sigma)^2 d\Omega \ . \tag{4.60}$$

Moreover

$$|u - u_+|^2 = \left|\int_{R_\sigma}^{R_+} u'(s,\sigma)sd\sigma\right|^2 \leq \left(\frac{1}{R_\sigma} - \frac{1}{R_+}\right)\int_{R_\sigma}^{R_+} u'^2 s^2 ds \ ,$$

so that, noting that $\frac{1}{R_\sigma} - \frac{1}{R_+} \leq \frac{\delta R}{R_+R_-}$,

$$\| u - u_+ \|_o^2 \leq \frac{\delta R}{R_+R_-}\int_{\delta\Omega} u'^2 d\Omega \ , \tag{4.61}$$

where $\delta\Omega \equiv \{(r,\sigma), R_\sigma \leq r \leq R_+\}$.

From (4.60), (4.61), used in (4.57) by applying the Cauchy–Schwarz inequality and the Gauss theorem, we get

$$\begin{aligned} \| u \|_0^2 &\leq \left[\frac{1}{R_+}\left(\frac{\delta R}{R_-}+\frac{4}{2L+5}\right)\right]\left(\int_{\delta\Omega} u'^2 d\Omega + \int_{\Omega_+} u'^2 d\Omega\right) \\ &\leq \left[\frac{1}{R_+}\left(\frac{\delta R}{R_-}+\frac{4}{2L+5}\right)\right]\int_{\Omega} \nabla u^2 d\Omega \\ &= \left[\frac{1}{R_+}\left(\frac{\delta R}{R_-}+\frac{4}{2L+5}\right)\right] R_+^2 J_+ \| u \|_0 \| u_n \|_0 \\ &\leq R_+ C_{0L} \| u \|_0 \| u \|_1 \ . \end{aligned}$$

Simplifying by $\| u \|_0$ we get (4.56) and the lemma is proved. □

Now we can go back to (4.52) and applying the Schwarz inequality we get

$$R_-^3 \| u \|_1^2 \leq 2R_+^2 J_+ \| f \|_0 \| u \|_1 + 3R_+^2 J_+ \| u \|_0 \| u \|_1 \ . \tag{4.62}$$

Simplifying and using (4.56) we obtain

$$\frac{R_+^2}{R_-^3} \| u \|_1 \leq 2J_+ \| f \|_0 + 3k^3 J_+ C_{0L} \| u \|_1 \ . \tag{4.63}$$

It is then clear that, if (4.53) is satisfied, (4.54) holds too. □

We shall discuss in the next section the geometric meaning of (4.53) together with the parallel condition holding for the complete linearized SGBVP. Here we just summarize the results of this section by commenting that the SMP has been proved to have a stable solution in $H_1 \cap H_{0L}$, unconditionally for any Lipschitzian $\widetilde{S}$ and data in L^2_σ, and that a specific evaluation of the majorization constant has obliged us to introduce the restrictive condition (4.53).

5. Analysis of the linearized SGBVP

We can turn now to the analysis of the BVP (3.33), that we summarize in the following Theorem.

Theorem 5.1. *Assume that the hypotheses of Theorem* 4.1 *hold true and that* (4.53) *is verified, so that* C_{1L} *of* (4.55) *is positive constant. Assume further on that the condition*

$$C_{1L} R_+ (\varepsilon_+ + \eta_+ C_{0L}) < 1 \tag{5.1}$$

is satisfied too. Then the problem (3.33) *has a unique solution* $u \in H_1 \cap H_{0L}$, $\{a_{jk}\} \in R^{(L+1)^2}$, *such that*

$$\| u \|_1 \leq C_{2L} \| f \|_0 \tag{5.2}$$

where

$$C_{2L} = \frac{C_{1L}}{1 - C_{1L} R_+ (\varepsilon_+ + \eta_+ C_{0L})} \ . \tag{5.3}$$

Proof. Note that the problem (3.33) is equivalent to the solution of a SMP with known term $f - Du$.

Therefore from Theorems (4.1) and (4.2) we know that

$$\| u \|_1 \leq C_{1L} \| f - Du \|_0 \tag{5.4}$$

$$|a_{jk}| \leq C \| f - Du \|_0 \ . \tag{5.5}$$

Recalling (3.32) and (4.56) we observe that indeed

$$\begin{aligned} \| Du \|_0 &\leq \varepsilon_+ R_+ \| \nabla u \|_0 + \eta_+ \| u \|_0 \\ &\leq \varepsilon_+ R_+ \| u \|_1 + \eta_+ R_+ C_{0L} \| u \|_1 \qquad (5.6) \\ &= R_+(\varepsilon_+ + \eta_+ C_{0L}) \| u \|_1 \qquad (5.7) \end{aligned}$$

Therefore, concentrating on (5.4), we find

$$\| u \|_1 \leq C_{1L} \| f \|_0 + C_{1L} R_+ (\varepsilon_+ + \eta_+ C_{0L}) \| u \|_1 \ . \tag{5.8}$$

It is then clear that if (5.1) is true, then (5.8) yields (5.2).

This indeed guarantees uniqueness and stability of the solution. As for existence however, it is enough to observe that if S is the solution operator of the functional part of the SMP, then we can write

$$u = Sf - SDu \tag{5.9}$$

and the condition (5.1) is exactly implying that SD, as an operator from $H_1 \cap H_{0L}$ into itself, is a contraction. Therefore the existence of u is a consequence of the well-known theorem on contractions [9, 31]. □

Remark 5.1. Now that we have defined two conditions for the stability constants of the SMP and of the linearized SGBVP, namely (4.58) and (5.1), it is interesting to investigate their functional and geometrical significance. As we see there are two parameters to play with, namely $J_+ = (\cos I_+)^{-1}$ and L. The third constant

$$k = \frac{R_+}{R_-} = 1 + \frac{\delta R}{R_-}$$

is in fact fixed by the global geometry of the Earth surface. If we take

$$\begin{aligned} R_- &= b = \ 6356.91 \text{ km} \\ R_+ &= a + H_c = \ 6384.52 \ \text{ km} \end{aligned}$$

where

$$H_c = \ 6.27 \ \text{ km}$$

is the height of the Chimborazo, a high mountain close to the equator, the tip of which is probably the furthest point of the surface from the center of the Earth, we get

$$\frac{\delta R}{R_-} = \ 4.34 \cdot 10^{-3} \ .$$

Other considerations of the geometric meaning of functional properties of the GBVP can be found in ([11]). Now if we want to be able to treat our BVP for boundaries with inclinations up to 60°, one has to put $J_+ = 2$.

With such a constraint one verifies that with $L = 23$

$$3J_+k^3C_{023} = \ 0.995 \ < 1$$

while

$$3J_+k^2C_{022} = \ 1.033 \ < 1$$

This means that if we want to be sure that the SMP has a stable solution we have to subtract from the data the contribution of a global model complete up to degree 23, what, as commented in §3, can be realistically done with a negligible error.

Similarly, one can work with the second condition that, recalling Theorem 3.1, can be written

$$C_{1L}R_+(\varepsilon_+ + \eta_+C_{0L}) = \frac{2J_+k^3}{1 - 3J_+k^3C_{0L}}e^2(1 + 2C_{0L}) < 1 \ .$$

So for $L = 23$ we obtain

$$C_{122}R_+(\varepsilon_+ + \eta_+C_{023}) = \ 6.545 \ > 1$$

and the condition (5.1) is not satisfied.

However, already for $L = 24$ one has

$$C_{124}R_+(\varepsilon_+ + \eta_+C_{024}) = \ 0.874 \ < 1$$

meaning that, subtracting a model complete up to degree 24, the linearized SGBVP becomes tractable for telluroids with inclination up to 60°. This is certainly a realistic and nice result.

6. Conclusions and open questions

Summarizing the state of the art of the analysis of the GBVP we can agree on the following statements:

- the non-linear problem, when the rotational terms are treated only as perturbations, has a satisfactory theory, especially in that data are required to have a natural degree of regularity, namely $W(\sigma) \in H_{1,\lambda}, g(\sigma) \in H_{0,\lambda}$,
- for the SMP we have a quite nice result, requiring only that the telluroid $\widetilde{S}$ be starshaped and Lipschitz, while gravity anomalies on $\widetilde{S}$ are only required to be in L^2_σ,
- for the linearized GBVP one has to force the theory of the solution of the SMP, in order to get a quantitative determination of the constants involved therein. Then the linear GBVP has a theorem of existence uniqueness and stability with reasonable requirements for the regularity of data, but enforcing stronger conditions both geometrical (limited inclination of the telluroid $\widetilde{S}$) and on the data (need of a certain realistic number of asymptotic coefficients of the potential).

The present theory could therefore be improved in some aspects:

- studying the numerical values of the constants involved in the non-linear theory, maybe generalizing it to other spaces than weighted Hölder spaces,
- to try to develop a theory like that of the SMP, but in ellipsoidal, rather than in spherical approximation, as in [12],
- to try to develop an $H_{1,p}$ theory for the linearized GBVP, that could allow to significantly relax the constraints on the inclination of $\widetilde{S}$.

In this respect, let us remark that it is the free boundary character of the GBVP, reflected in the term $2u$ in the boundary operator, that imposes more severe constraints. An analogous theory for the fixed boundary GBVP requires much weaker conditions to obtain a theorem of existence uniqueness and stability in H_1. In fact it is only required that $L = 1$, a standard condition for the anomalous potential, and inclinations can be up to 89°, what is even too much for a reasonable global model of the surface of the Earth.

References

[1] Adams R.A. Sobolev Spaces. Pure and Applied Mathematics, Academic Press, New York, San Francisco London, 1975.

[2] Anderson O. In: Sansò F., Sideris M. Geoid determination: theory and methods, Chapter 9, Part II. Springer-Verlag Berlin-Heidelberg (2013).

[3] Axler S., Bourdon P., Ramsey W. Harmonic Function Theory. Springer-Verlag (2001).

[4] Banach S. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. Fundamenta Mathematicae (1922).

[5] Cimmino G. Spazi hilbertiani di funzioni armoniche e questioni connesse. in: Equazioni lineari alle derivate parziali. UMI Roma (1955).

[6] Freeden W., Gerhards C. Geomathematically oriented potential theory. Chapman & Hall/CRC Press, Taylor Francis, Boca Raton (2013).

[7] Friedman A. Variational principles and Free Boundary Value Problems. John Wiley & Sons, New York (1982).

[8] Gilbarg D., Trudinger N.S. Elliptic Partial Differential Equations of Second Order. Classics in Mathematics, Springer, Berlin, Heidelberg (2001).

[9] Granas A., Dugundji J. Fixed Point Theory. Springer Monographs in Mathematics, vol. 41 (2003).

[10] Grothaus M., Raskop T. Oblique Stochastic Boundary Value Problems. Handbook of GeoMath, pp. 1052–1076 (2010).

[11] Holota P. Coerciveness of the linear gravimetric boundary value problem and a geometric interpretation. J. of Geod. 71:640–651 (1997).

[12] Holota P., Nesvadba O. Reproducing kernel Hilbert Space for the exterior of an ellipsoid and the method of successive approximations in solving geodetic boundary value problems. VIII Hotine Marussi Symp. on Math. Geod. Rome, June 2013. Springer (IAG Symp. Series, in press).

[13] Hörmander L. The Boundary Value Problems in Physical Geodesy. Arch. Prot. Mech. Anal., n. 62, pp. 51–52 (1976).

[14] Jerison D.S., Kenig C. Boundary Value Problems on Lipschitz Domains. MAA Studies in Math. v. 23, pp. 1–68 (1982).

[15] Krarup T. Mathematical Foundation of Geodesy. Springer, Berlin, Heidelberg (2006).

[16] Marussi A. Intrinsic geodesy. Springer, Berlin (1985).

[17] McLean W. Strongly Elliptic Systems and Boundary Integral Equations. Cambridge University Press (2000).

[18] Miranda C. Partial Differential Equations of Elliptic Type. Springer-Verlag, Berlin-Heidelberg (1970).

[19] Molodensky M.S., Eremeev V.F., Yurkina M.I. Methods for the study of the gravitational field of the Earth. Transl. Russian Israel Program for Scient. Trans. Jerusalem (1960).

[20] Moritz H. Advanced Physical Geodesy, Wichmann Verlag, Karlsruhe (1980).

[21] Moser J. A rapidly convergent iteration method and non-linear differential equations. Acc. Sc. Norm. Sup. Pisa, v. 20. pp. 265–315 (1966).

[22] Pavlis N.K. In: Sansò F., Sideris M. Geoid determination: theory and methods, Chapter 6, Part II. Springer-Verlag Berlin-Heidelberg (2013).

[23] Otero J., Sansò F. An analysis of the scalar geodetic boundary value problems with natural regularity results. J. of Geod., v. 73, pp. 437–435 (1999).

[24] Sacerdote F., Sansò F. The Scalar Boundary Value Problem of Physical Geodesy, Man. Geod. 11, pp. 15–28 (1986).

[25] Sansò F. Discussion on the existence and uniqueness of the solution of Molodensky's problem in gravity space. Rend. Acc. Naz. Lincei (Sc. Fisiche, Matematiche, Naturali) V. 61, S. 8 (1976).

[26] Sansò F. The Geodetic Boundary Value Problem in Gravity Space. Memorie Acc. Lincei V. 14, S. 8, n. 3 (1977).

[27] Sansò F. Geodetic Boundary Value Problem. Handbook of Geomathematics, Springer-Verlag, London (2014).

[28] Sansò F., Sideris M. Geoid determination: theory and methods. Springer-Verlag Berlin-Heidelberg (2013).

[29] Sansò F., Venuti G. On the Explicit Determination of stability Constants for the Linearized Geodetic Boundary Value Problems. J. Geod. v. 82, pp. 909–916 (2008).

[30] Seitz K., Heck B. Effects of Non-linearity in the Geodetic Boundary Value Problems. Deutsche Geodätische Kommission, Reihe A, Heft Nr. 109, Verlag der Bayerischen Ac. der Wissenschaften, Munich (1993).

[31] Yoshida K. Functional Analysis. Classics in Mathematics, Springer-Verlag, Berlin, Heidelberg (1980).

Fernando Sansò
Politecnico di Milano – DICA
P.zza Leonardo da Vinci, 32
I-20133 Milano, Italy
e-mail: fernando.sanso@polimi.it

Handbook of Mathematical Geodesy
Geosystems Mathematics, 491–516

Oblique Stochastic Boundary Value Problem

Martin Grothaus and Thomas Raskop

Abstract. Aim of this note is to report the current state of the analysis for weak solutions to oblique boundary problems for the Poisson equation. In this paper, as well deterministic as stochastic inhomogeneities are treated and existence and uniqueness results for corresponding weak solutions are presented. We consider the problem for inner bounded and outer unbounded domains in $\mathbb{R}^n$. Main tools for the deterministic inner problem are a Poincaré inequality and some analysis for Sobolev spaces on submanifolds, in order to use the Lax–Milgram Lemma. The Kelvin transformation enables us to translate the outer problem to a corresponding inner problem. Thus we can define a solution operator by using the solution operator of the inner problem. The extension to stochastic inhomogeneities is done with help of tensor product spaces of a probability space with the Sobolev spaces from the deterministic problems. We can prove a regularization result which shows that the weak solution fulfills the classical formulation for smooth data. A Ritz–Galerkin approximation method for numerical computations is available. Finally, we show that the results are applicable to geomathematical problems.

Keywords. Oblique derivation problem, Ritz–Galerkin approximation, stochastic extensions.

1. Introduction

The main subject of this article are existence results for solutions to oblique boundary problems for the Poisson equation. We start with the deterministic problems. The Poisson equation in the domain Σ is given by

$$\Delta u = f,$$

and the oblique boundary condition by

$$\langle \underline{a}, \nabla u \rangle + bu = g.$$

Oblique derivative problems represent classical topics in the context of mathematical geodesy. Thus, the editors and the publisher have decided to include this chapter, although an earlier version has been published in "Handbook of Geomathematics, Vol. 3, 2285–2315, 2015".

This condition is called regular if the equation

$$|\langle \underline{a}, \nu \rangle| > C > 0,$$

holds on $\partial\Sigma$ for a constant $0 < C < \infty$. The problem is called outer problem if the Poisson equation has to hold on an outer domain $\Sigma \subset \mathbb{R}^n$. This is a domain Σ, having the representation $\Sigma = \mathbb{R}^n \backslash \overline{D}$ where $0 \in D$ is a bounded domain. Consequently, $\partial\Sigma$ divides the Euclidian space $\mathbb{R}^n$ into a bounded domain D, called inner domain, and an unbounded domain Σ, called outer domain. A problem defined on a bounded domain is called inner problem. A classical solution corresponding to continuous $\underline{a}$, b, g and f of the oblique boundary problem for the Poisson equation is a function $u \in C^2(\Sigma) \cap C^1(\overline{\Sigma})$ which fulfills the first two equations. For the outer problem u must be regular at infinity, i.e., $u(x) \to 0$ for $|x| \to \infty$. Existence and uniqueness results for a classical solution to regular oblique boundary problems for the Poisson equation are already available, see, e.g., [11, 13] or [16]. In order to allow very weak assumptions on boundary, coefficients and inhomogeneities, we are interested in weak solutions from Sobolev spaces of one times weakly differentiable functions. When facing the deterministic problems, we have to distinguish the inner and the outer setting. The reason is that a Poincaré inequality, namely

$$\int_\Sigma \langle \nabla u, \nabla u \rangle \, d\lambda^n + \int_{\partial\Sigma} u^2 dH^{n-1} \geq C \left(\int_\Sigma u^2 \, d\lambda^n + \int_\Sigma \langle \nabla u, \nabla u \rangle \, d\lambda^n \right),$$

for all $u \in H^{1,2}(\Sigma)$, is only available for bounded Σ. Thus we can only use the Lax–Milgram Lemma for the inner problem in order to gain a solution operator. For the outer problem we use the Kelvin transformation to transform the unbounded domain Σ to a bounded domain Σ^K via

$$\Sigma^K := \left\{ \frac{x}{|x|^2} \middle| x \in \Sigma \right\} \cup \{0\}.$$

Additionally we transform coefficients as well as inhomogeneities and end up with a inner problem, which posses a unique weak solution v. Finally we transform this function to the outer space by

$$u(x) := \frac{1}{|x|^{n-2}} v\left(\frac{x}{|x|^2} \right),$$

for all $x \in \Sigma$. This u is then the weak solution to the outer problem and it can be shown that, in the case of existence, u is the classical solution. Additionally the transformations are continuous and consequently the solution depends continuous on the data. Before we go on with stochastic inhomogeneities and stochastic weak solutions we want to mention that we have to assume a regular inner problem, while we have a transformed regularity condition for the outer problem resulting from the transformations. Going to a stochastic setting we have to introduce the spaces of stochastic functions. These are constructed as the tensor product of $L^2(\Omega, dP)$, with a suitable probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and the Sobolev spaces used in the deterministic theory. They are again Hilbert spaces and we have isomorphisms to Hilbert space-valued random variables. For the stochastic inner problem we

again employ the Lax–Milgram Lemma, while in the outer setting we define the solution operator pointwisely for almost all $\omega \in \Omega$. For all solutions, deterministic as well as stochastic, a Ritz–Galerkin approximation method is available. Finally we give some examples from Geomathematics, where stochastic inhomogeneities are implemented. Proofs for the results presented in this report are given in [9] and [10]. The examples are taken from [3] and [7]. We want to mention that the articles [15] as well as [17] also deal with solutions to oblique boundary-value problem.

2. Scientifically relevant domains and function spaces

In this article we consider boundary value problems for the Poisson equation. This means we are searching for a function which satisfies the Poisson equation in a subset Σ of $\mathbb{R}^n$ and an additional condition on the boundary $\partial\Sigma$ of this set.

$$\begin{aligned} \Delta u &= f, \quad \text{in } \Sigma, \\ \langle \underline{a}, \nabla u \rangle + bu &= g \quad \text{on } \partial\Sigma, \end{aligned}$$

f and g are called inhomogeneities, $\underline{a}$ and b are called coefficients and such a function u is then called solution. Our analysis is motivated by problems from Geomathematics. Here oblique boundary problems arise frequently, because in general the normal of the Earth's surface does not coincide with the direction of the gravity vector. Therefore, the oblique boundary condition is more suitable then a Neumann boundary condition. For details see [3] or [12]. We are dealing with two different types of sets Σ, namely bounded and outer $C^{m,\alpha}$-domains, which are introduced by the following definition. In particular the outer problem is of major interest for applications.

Definition 1. $\partial\Sigma \subset \mathbb{R}^n$ is called a $C^{m,\alpha}$-surface , $m \in \mathbb{N}$ and $0 \leq \alpha \leq 1$ and Σ is called a bounded $C^{m,\alpha}$-domain , if and only if

- Σ is a bounded subset of $\mathbb{R}^n$ which is a domain, i.e., open and connected,
- There exists an open cover $(U_i)_{i=1,\ldots,N}$ of $\partial\Sigma$ and corresponding $C^{m,\alpha}$-diffeomorhisms $\Psi_i : B_1^{\mathbb{R}^n}(0) \to U_i$, $i = 1, \ldots, N$, such that

$$\begin{aligned} \Psi_i : B_1^0(0) &\to U_i \cap \partial\Sigma, \\ \Psi_i : B_1^+(0) &\to U_i \cap \Sigma, \\ \Psi_i : B_1^-(0) &\to U_i \cap \mathbb{R}^n \backslash \overline{\Sigma}, \end{aligned}$$

  where $B_1^{\mathbb{R}^n}(0)$ denotes the open unit ball in $\mathbb{R}^n$, i.e., all $x \in \mathbb{R}^n$ with $|x| < 1$. $B_1^0(0)$ denotes the set of all $x \in B_1^{\mathbb{R}^n}(0)$ with $x_n = 0$, $B_1^+(0)$ denotes the set of all $x \in B_1^{\mathbb{R}^n}(0)$ with $x_n > 0$ and $B_1^-(0)$ denotes the set of all $x \in B_1^{\mathbb{R}^n}(0)$ with $x_n < 0$.

On the other hand Σ is called an outer $C^{m,\alpha}$-domain, if and only if $\Sigma \subset \mathbb{R}^n$ is open, connected and representable as $\Sigma := \mathbb{R}^n \backslash \overline{D}$, where D is a bounded $C^{m,\alpha}$-domain such that $0 \in D$. Ψ_i is called $C^{m,\alpha}$-diffeomorphism if and only if it is bijective,

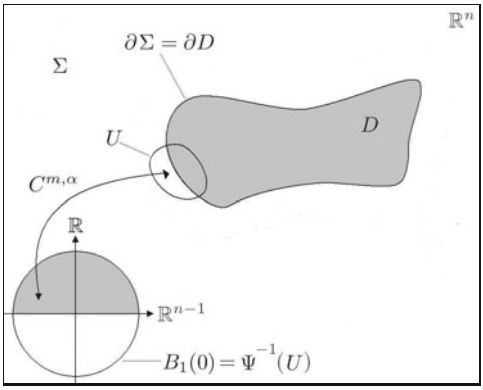


FIGURE 1. $C^{m,\alpha}$-surface

$(\Psi_i)_j \in C^{m,\alpha}\left(\overline{B_1^{\mathbb{R}^n}(0)}\right)$, $\left(\Psi_i^{-1}\right)_j \in C^{m,\alpha}\left(\overline{U_i}\right)$, $j = 1, \ldots, n$ and we have for the determinant of the Jacobian Matrix of Ψ_i, $\mathrm{Det}(\mathrm{D}\Psi_i) \neq 0$ in $\overline{B_1^{\mathbb{R}^n}(0)}$.

In Figure 1, such a $C^{m,\alpha}$-surface is illustrated. For this definition and further details see, e.g., [6]. The definition is independent of the mappings chosen. $\partial\Sigma$ is a compact closed and doublepointfree $(n-1)$-dimensional $C^{m,\alpha}$-submanifold. The outer unit normal vector ν is a C^{m-1}-vector field. Furthermore, we find a C^∞-partition of $(w_i)_{1\leq i\leq N}$ on $\partial\Sigma$ corresponding to the open cover $(U_i)_{1\leq i\leq N}$, provided by Alt [2]. H^{n-1} denotes the $(n-1)$-dimensional Hausdorff measure on $\partial\Sigma$ and λ^n the Lebesgue measure in $\mathbb{R}^n$. Throughout this paper we assume at least a Lipschitz boundaries, i.e., $C^{0,1}$-boundaries $\partial\Sigma$. Then we have $\nu \in L^\infty(\partial\Sigma; \mathbb{R}^n)$. Note that some geomathematical relevant examples are even C^∞-surfaces, e.g., a sphere or an ellipsoid. We will see in the following chapters that the case of bounded and outer domains have to be treated differently, because the unboundedness causes problems which do not occur in the bounded setting. Non the less, we are searching in both cases for solutions under as weak assumptions as possible. More precisely we are searching for solutions in Sobolev spaces for inhomogeneities from Banach space duals of Sobolev spaces. These spaces are introduced in the following.

Definition 2. Let Σ be a bounded $C^{0,1}$-domain and $r \in \mathbb{N}$. We define

$$H^{r,2}(\Sigma) := \left\{F : \Sigma \to \mathbb{R} | \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n} F \in L^2(\Sigma) \text{ for all } \alpha_1 + \cdots + \alpha_n \leq r\right\},$$

$$\|F\|_{H^{r,2}(\Sigma)} := \left(\sum_{|\alpha|=0}^{r} \sum_{i=1}^{N} \|\partial^\alpha F\|^2_{L^2(\Sigma)}\right)^{1/2}.$$

Let Σ be an outer $C^{0,1}$-domain and $\varrho_1, \varrho_2, \varrho_3$ be continuous, positive functions defined on $\overline{\Sigma}$. We define

$$L^2_{\varrho_1}(\Sigma) := \left\{F : \Sigma \to \mathbb{R} | F \text{ is measurable with} \int_\Sigma F^2(x)\varrho_1^2(x) d\lambda^n(x) < \infty\right\},$$

$$H^{1,2}_{\varrho_1,\varrho_2}(\Sigma) := \left\{F \in L^2_{\varrho_1}(\Sigma) | \partial_i F \in L^2_{\varrho_2}(\Sigma), 1 \le i \le n\right\},$$
$$H^{2,2}_{\varrho_1,\varrho_2,\varrho_3}(\Sigma) := \left\{F \in L^2_{\varrho_1}(\Sigma) | \partial_i F \in L^2_{\varrho_2}(\Sigma) \text{ and } \partial_i\partial_j F \in L^2_{\varrho_3}(\Sigma), 1 \le j, i \le n\right\},$$
$$\|F\|_{L^2_{\varrho_1}(\Sigma)} := \left(\int_\Sigma F^2(x)\varrho_1^2(x)d\lambda^n(x)\right)^{1/2},$$
$$\|F\|_{H^{1,2}_{\varrho_1,\varrho_2}(\Sigma)} := \left(\|F\|^2_{L^2_{\varrho_1}(\Sigma)} + \sum_{i=1}^n \|\partial_i F\|^2_{L^2_{\varrho_2}(\Sigma)}\right)^{1/2},$$
$$\|F\|_{H^{2,2}_{\varrho_1,\varrho_2,\varrho_3}(\Sigma)} := \left(\|F\|^2_{L^2_{\varrho_1}(\Sigma)} + \sum_{i=1}^n\left(\|\partial_i F\|^2_{L^2_{\varrho_2}(\Sigma)} + \sum_{j=1}^n \|\partial_i\partial_j F\|^2_{L^2_{\varrho_3}(\Sigma)}\right)\right)^{1/2}.$$

Let $\partial\Sigma$ be a $C^{0,1}$-surface and $(w_i)_{1\le i\le N}$ be the C^∞-partition of unity of $\partial\Sigma$ corresponding to the open cover from Definition 1. For a function F defined on $\partial\Sigma$ we obtain a function $\theta_i F$ defined on $\mathbb{R}^{n-1}$ by:

$$\theta_i F(y) := \begin{cases} (w_i F)(\Psi_i(y,0)) & y \in B_1^{\mathbb{R}^{n-1}}(0), \\ 0 & \text{otherwise.} \end{cases}$$

Let now $\partial\Sigma$ be a $C^{m,1}$-surface, $m \in \mathbb{N}$. Furthermore let $s \in \mathbb{R}$, $r \in \mathbb{N}$, with $s < m+1$ and $0 \le r \le m$. Then we define

$$H^{s,2}(\partial\Sigma) := \left\{F : \partial\Sigma \to \mathbb{R} | \theta_i F \in H^{s,2}(\mathbb{R}^{n-1}), 1 \le i \le N\right\},$$
$$H^{r,\infty}(\partial\Sigma) := \left\{F : \Sigma \to \mathbb{R} | \theta_i F \in H^{r,\infty}(\mathbb{R}^{n-1}), 1 \le i \le N\right\},$$
$$\|F\|_{H^{s,2}(\partial\Sigma)} := \left(\sum_{i=1}^N \|\theta_i F\|^2_{H^{s,2}(\mathbb{R}^{n-1})}\right)^{1/2},$$
$$\|F\|_{H^{r,\infty}(\partial\Sigma)} := \max_{0\le|s|\le r, 1\le i\le N}\left\{\operatorname{ess\,sup}_{B_1^{\mathbb{R}^{n-1}}(0)}\left(|\partial_1^{s_1}\cdots\partial_{n-1}^{s_{n-1}}\theta_i F|\right)\right\},$$

where $H^{0,p}(\partial\Sigma)$ is identical with $L^p(\partial\Sigma)$, $p \in \{2,\infty\}$. The spaces $H^{2,2}_{\varrho_1,\varrho_2,\varrho_3}(\Sigma)$, $H^{1,2}_{\varrho_1,\varrho_2}(\Sigma)$, $L^2_{\varrho_1}(\Sigma)$, $H^{s,2}(\partial\Sigma)$ and $H^{r,2}(\Sigma)$ are Hilbert spaces, while the spaces $H^{r,\infty}(\partial\Sigma)$ are Banach spaces with respect to the norms given above, see, e.g., [1] or[5].

The spaces $H^{s,2}(\mathbb{R}^{n-1})$ are defined via the Fourier transformation. Differentiation in the definition above has to be understood in sense of weak differentiation. The definition of the spaces on $\partial\Sigma$ above is independent from the choice of $(U_i)_{1\le i\le N}$, $(w_i)_{1\le i\le N}$ and $(\Psi_i)_{1\le i\le N}$. It is left to introduce the spaces $\left(H^{s,2}(\partial\Sigma)\right)'$ on a $C^{m,1}$-surface $\partial\Sigma$, $0 \le s < m+1$. We do this as follows. Identify each function $F \in L^2(\partial\Sigma)$ with a linear continuous functional on $H^{s,2}(\partial\Sigma)$, defined by

$$F(G) := \int_{\partial\Sigma} F(x)\cdot G(x)\, dH^{n-1}(x),$$

for all $G \in H^{s,2}(\partial\Sigma)$. Then $\left(H^{s,2}(\partial\Sigma)\right)'$ is defined as

$$\left(H^{s,2}(\partial\Sigma)\right)' := \overline{L^2(\partial\Sigma)}\Big|_{\|\cdot\|_{\left(H^{s,2}(\partial\Sigma)\right)'}},$$

where

$$\|F\|_{\left(H^{s,2}(\partial\Sigma)\right)'} := \sup_{G\in H^{s,2}(\partial\Sigma)} \frac{|F(G)|}{\|G\|_{H^{s,2}(\partial\Sigma)}}.$$

In this way we end up with the space $H^{-s,2}(\partial\Sigma)$ defined in the previous definition. We get the following chain of rigged Hilbert spaces, called Gelfand triple.

$$H^{s,2}(\partial\Sigma) \subset L^2(\partial\Sigma) \subset H^{-s,2}(\partial\Sigma),$$

densely and continuously. Additionally we have for the duality product

$${}_{H^{-s,2}(\partial\Sigma)}\langle F, G\rangle_{H^{s,2}(\partial\Sigma)} = \int_{\partial\Sigma} F(x)\cdot G(x)\, dH^{n-1}(x),$$

for all $F \in L^2(\partial\Sigma)$. Analogously, we introduce the Gelfand triples

$$H^{1,2}(\Sigma) \subset L^2(\Sigma) \subset \left(H^{1,2}(\Sigma)\right)',$$

for bounded $C^{0,1}$-domains and

$$H^{1,2}_{|x|^2,|x|^3}(\Sigma) \subset L^2_{|x|^2}(\Sigma) \subset \left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)',$$

for outer $C^{0,1}$-domains.

3. Poincaré inequality as key issue for the inner problem

In this chapter we will show how to derive a weak formulation for the deterministic inner regular oblique boundary value problem defined on bounded $C^{1,1}$-domains Σ. The corresponding weak solution will obviously coincide with the classical solution in the case of existence. First we derive a weak formulation, then a Poincaré inequality for the Sobolev space $H^{1,2}(\Sigma)$ allows us to apply the Lax–Milgram Lemma in order to provide a solution operator. At next we translate a regularization result for the Neumann boundary value problem to the oblique boundary value problem. Finally a Ritz–Galerkin method allows us to approximate the weak solutions with help of numerical calculations. We proceed with the stochastic extensions. This means we introduce the stochastic function spaces for the inhomogeneities and corresponding solutions with help of the tensor product. Then the results for the deterministic problem can be easily extended to the stochastic setting. The chapter is divided into five sections according to the described approach.

3.1. The weak formulation

In this section we present the theory of weak solutions to the regular oblique boundary problem for the Poisson equation for inner domains. Although the weak problem can be formulated for bounded $C^{0,1}$-domains, in order to prove the existence of a unique weak solution we need at least a bounded $C^{1,1}$-domain. Consequently we assume $\Sigma \subset \mathbb{R}^n$ throughout this section to be such a domain, if not stated otherwise. At first we give the definition of the regular oblique boundary problem together with the definition of the classical solution.

Definition 3. Let Σ be a bounded $C^{1,1}$-domain, $f \in C^0(\Sigma)$, $g, b \in C^0(\partial\Sigma)$ and $\underline{a} \in C^0(\Sigma; \mathbb{R}^n)$ be given, such that

$$|\langle \underline{a}(x), \nu(x)\rangle| > C_1 > 0, \tag{1}$$

for all $x \in \partial\Sigma$, where $0 < C_1 < \infty$. Finding a function $u \in C^2(\Sigma) \cap C^1(\overline{\Sigma})$ such that

$$\begin{aligned} \Delta u &= f \quad \text{in } \Sigma, \\ \langle \underline{a}, \nabla u\rangle + bu &= g \quad \text{on } \partial\Sigma, \end{aligned}$$

is called inner regular oblique boundary problem for the Poisson equation and u is called classical solution.

Because of the condition (1) the problem is called regular. It just means that the vector field $\underline{a}$ is non-tangential to $\partial\Sigma$ for all $x \in \partial\Sigma$. Now we derive the weak formulation. The fundamental theorem of the calculus of variations gives

$$\Delta u = f \quad \text{in } \Sigma$$

if and only if

$$\int_\Sigma \eta \Delta u \, d\lambda^n = \int_\Sigma \eta f \, d\lambda^n \quad \text{for all } \eta \in C_0^\infty(\Sigma)$$

if and only if

$$\int_\Sigma \eta \Delta u \, d\lambda^n = \int_\Sigma \eta f \, d\lambda^n \quad \text{for all } \eta \in C^\infty(\overline{\Sigma}).$$

Additionally on Σ the following Green formula is valid

$$\int_\Sigma \varphi \Delta\psi \, d\lambda^n + \int_\Sigma \langle \nabla\varphi, \nabla\psi\rangle \, d\lambda^n = \int_{\partial\Sigma} \varphi \frac{\partial\psi}{\partial\nu} \, dH^{n-1},$$

for all $\psi \in C^2(\Sigma) \cap C^1(\overline{\Sigma})$ and $\varphi \in C^\infty(\overline{\Sigma})$. This yields for a classical solution

$$\int_{\partial\Sigma} \eta \frac{\partial u}{\partial\nu} \, dH^{n-1} - \int_\Sigma \langle \nabla\eta, \nabla u\rangle \, d\lambda^n = \int_\Sigma \eta f \, d\lambda^n,$$

for all $\eta \in C^\infty(\overline{\Sigma})$. Now we transform the boundary condition

$$\langle \underline{a}, \nabla u\rangle + bu = g \quad \text{on} \quad \partial\Sigma,$$

to the form

$$\langle \underline{a}, \nu\rangle \frac{\partial}{\partial\nu} u + \langle \underline{a} - \langle (\underline{a}, \nu)\nu) \cdot \nabla_{\partial\Sigma} u\rangle + bu = g \quad \text{on} \quad \partial\Sigma.$$

Using equation (1) we divide by $\langle \underline{a}, \nu \rangle \neq 0$ to get the equivalent boundary condition

$$\frac{\partial}{\partial \nu} u + \left\langle \left(\frac{\underline{a}}{\langle \underline{a}, \nu \rangle} - \nu \right), \nabla_{\partial\Sigma} u \right\rangle + \frac{b}{\langle \underline{a}, \nu \rangle} u = \frac{g}{\langle \underline{a}, \nu \rangle} \quad \text{on} \quad \partial\Sigma.$$

Plugging this condition into the equation above, we get the following formulation of the regular oblique boundary problem for the Poisson equation which is equivalent to the formulation given in Definition 3. We want to find a function $u \in C^2(\Sigma) \cap C^1(\overline{\Sigma})$ such that

$$\int_{\partial\Sigma} \eta \left(\frac{g}{\langle \underline{a}, \nu \rangle} - \frac{b}{\langle \underline{a}, \nu \rangle} u - \left\langle \frac{\underline{a}}{\langle \underline{a}, \nu \rangle} - \nu, \nabla_{\partial\Sigma} u \right\rangle \right) dH^{n-1} - \int_\Sigma \langle \nabla \eta, \nabla u \rangle \, d\lambda^n - \int_\Sigma \eta f \, d\lambda^n = 0 \quad \text{for all } \eta \in C^\infty(\overline{\Sigma}).$$

The transformation of the boundary term is shown in Figure 2.

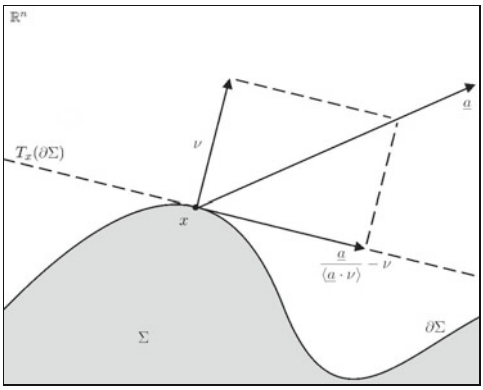


FIGURE 2. Transformation of the oblique boundary condition

Finally, we are weakening the assumptions on data, coefficients, test function and solution. We give the weak formulation of the inner regular oblique boundary problem to the Poisson equation, summarized in the following definition.

Definition 4. Let Σ be a bounded $C^{1,1}$-domain, $\underline{a} \in H^{1,\infty}(\partial\Sigma; \mathbb{R}^n)$ fulfilling condition (1), $b \in L^\infty(\partial\Sigma)$, $g \in H^{-\frac{1}{2},2}(\partial\Sigma)$ and $f \in \left(H^{1,2}(\Sigma)\right)'$. We want to find a function $u \in H^{1,2}(\Sigma)$ such that

$$\begin{aligned} {}_{H^{\frac{1}{2},2}(\partial\Sigma)} \left\langle \eta, \frac{g}{\langle \underline{a}, \nu \rangle} \right\rangle_{H^{-\frac{1}{2},2}(\partial\Sigma)} - \sum_{i=1}^n {}_{H^{\frac{1}{2},2}(\partial\Sigma)} \left\langle \eta \frac{\underline{a}_i}{\langle \underline{a}, \nu \rangle} - \nu_i, (\nabla_{\partial\Sigma} u)_i \right\rangle_{H^{-\frac{1}{2},2}(\partial\Sigma)} \\ - \int_\Sigma (\nabla \eta \cdot \nabla u) \, d\lambda^n - \int_{\partial\Sigma} \eta \frac{b}{\langle \underline{a}, \nu \rangle} u \, dH^{n-1} - {}_{H^{1,2}(\Sigma)} \langle \eta, f \rangle_{(H^{1,2}(\Sigma))'} = 0, \end{aligned}$$

for all $\eta \in H^{1,2}(\Sigma)$. Then u is called a weak solution of the inner regular oblique boundary problem for the Poisson equation.

3.2. Existence and uniqueness results for the weak solution

It is possible to prove the following existence and uniqueness result for the weak solution to the deterministic inner oblique boundary value problem for the Poisson equation.

Theorem 5. *Let Σ be a bounded $C^{1,1}$-domain, $\underline{a} \in H^{1,\infty}(\partial\Sigma;\mathbb{R}^n)$, fulfilling condition* (1)*, and $b \in L^\infty(\partial\Sigma)$ such that:*

$$\operatorname*{ess\,inf}_{\partial\Sigma}\left(\frac{b}{\langle \underline{a},\nu\rangle}-\frac{1}{2}\mathrm{div}_{\partial\Sigma}\left(\frac{\underline{a}}{\langle \underline{a},\nu\rangle}-\nu\right)\right)>0. \tag{2}$$

Then for all $f \in \left(H^{1,2}(\Sigma)\right)'$ and $g \in H^{-\frac{1}{2},2}(\partial\Sigma)$ there exists one and only one weak solution $u \in H^{1,2}(\Sigma)$ of the inner regular oblique boundary problem for the Poisson equation. Additionally we have for a constant $0 < C_2 < \infty$:

$$\|u\|_{H^{1,2}(\Sigma)} \le C_2\left(\|f\|_{(H^{1,2}(\Sigma))'}+\|g\|_{H^{-\frac{1}{2},2}(\partial\Sigma)}\right).$$

In the proof we apply the Lax–Milgram Lemma, which gives us a unique $u \in H^{1,2}(\Sigma)$ fulfilling the variational equation

$$F(\eta)=a(\eta,u),$$

for all $\eta \in H^{1,2}(\Sigma)$, provided we have that F and a are continuous and additionally a is a coercive bilinear form. F and a can be obtained easily from the weak formulation as

$$\begin{aligned}F(\eta)&=\underset{H^{\frac{1}{2},2}(\partial\Sigma)}{}\left\langle \eta,\frac{g}{\langle \underline{a},\nu\rangle}\right\rangle_{H^{-\frac{1}{2},2}(\partial\Sigma)} -{}_{H^{1,2}(\Sigma)}\langle \eta,f\rangle_{(H^{1,2}(\Sigma))'},\\ a(\eta,u)&=\sum_{i=1}^{n}\underset{H^{\frac{1}{2},2}(\partial\Sigma)}{}\left\langle \eta\frac{\underline{a}_i}{\langle \underline{a},\nu\rangle}-\nu_i,(\nabla_{\partial\Sigma}u)_i\right\rangle_{H^{-\frac{1}{2},2}(\partial\Sigma)}\\ &\quad+\int_\Sigma(\nabla\eta\cdot\nabla u)\,d\lambda^n+\int_{\partial\Sigma}\eta\frac{b}{\langle \underline{a},\nu\rangle}u\,dH^{n-1}.\end{aligned}$$

The continuity can be shown by some results about the Sobolev spaces occurring in the weak formulation. In order to prove that a is coercive, i.e., $|a(u,u)| \ge C_3\|u\|^2_{H^{1,2}(\Sigma)}$, the Poincaré inequality

$$\int_\Sigma\langle\nabla F,\nabla F\rangle\,d\lambda^n+\int_{\partial\Sigma}F^2\,dH^{n-1}\ge C_4\left(\int_\Sigma F^2\,d\lambda^n+\int_\Sigma\langle\nabla F,\nabla F\rangle\,d\lambda^n\right),$$

which is valid for all $F \in H^{1,2}(\Sigma)$ and a constant $0 < C_4 < \infty$, is indispensable. Finally, the condition

$$\operatorname*{ess\,inf}_{\partial\Sigma}\left(\frac{b}{\langle \underline{a},\nu\rangle}-\frac{1}{2}\mathrm{div}_{\partial\Sigma}\left(\frac{\underline{a}}{\langle \underline{a},\nu\rangle}-\nu\right)\right)>0,$$

is also essential to ensure the coercivity of a. Condition (2) can be transformed into the equivalent form

$$\langle \underline{a},\nu\rangle b>\frac{1}{2}\big(\langle \underline{a},\nu\rangle\big)^2\mathrm{div}_{\partial\Sigma}\left(\frac{\underline{a}}{\langle \underline{a},\nu\rangle}-\nu\right)\quad H^{n-1}\text{-almost everywhere on }\partial\Sigma.$$

If $\operatorname{div}_{\partial\Sigma}\left(\frac{\underline{a}}{\langle \underline{a},\nu\rangle} - \nu\right) = 0$ H^{n-1}-almost everywhere on $\partial\Sigma$, we have for H^{n-1}-almost all $x \in \partial\Sigma$ the condition from the existence and uniqueness result for the classical solution. Furthermore for $\underline{a} = \nu$, i.e., the Robin problem, the condition reduces to $b > 0$ H^{n-1}-almost everywhere on $\partial\Sigma$. Finally, we are able to define for each bounded $C^{1,1}$-domain Σ, $\underline{a} \in H^{1,\infty}(\partial\Sigma;\mathbb{R}^n)$ and $b \in L^\infty(\partial\Sigma)$, fulfilling conditions (1) and (2), a continuous invertible linear solution operator $S^{\text{in}}_{\underline{a},b}$ by

$$\begin{aligned} S^{\text{in}}_{\underline{a},b} : \left(H^{1,2}(\Sigma)\right)' \times H^{-\frac{1}{2},2}(\partial\Sigma) &\to H^{1,2}(\Sigma), \\ (f,g) &\mapsto u, \end{aligned}$$

where u is the weak solution provided by Theorem 5. In addition this means that the inner weak problem is well posed.

3.3. A regularization result

In this section we will show that the weak solution from the previous section is even an element of $H^{2,2}(\Sigma)$ if we choose the inhomogeneities and coefficients smooth enough. The result for the oblique boundary problem is based a regularization result for the weak solution to the Neumann problem for the Poisson equation.

Theorem 6. *Let $\Sigma \subset \mathbb{R}^n$ be a bounded $C^{2,1}$-domain, $\underline{a} \in H^{2,\infty}(\partial\Sigma;\mathbb{R}^n)$ fulfilling condition* (1) *and $b \in H^{1,\infty}(\partial\Sigma)$. Then for all $f \in L^2(\Sigma)$ and $g \in H^{\frac{1}{2},2}(\partial\Sigma)$, the weak solution $u \in H^{1,2}(\Sigma)$ to the inner regular oblique boundary problem for the Poisson equation, provided in Theorem* 5*, is even in $H^{2,2}(\Sigma)$. Furthermore we have the a priori estimate*

$$\|u\|_{H^{2,2}(\Sigma)} \leq C_5\Big(\|f\|_{L^2(\Sigma)} + \|g\|_{H^{\frac{1}{2},2}(\partial\Sigma)}\Big).$$

for a constant $0 < C_5 < \infty$.

In order to prove the result it suffices to show that the normal derivative of the weak solution u of the oblique boundary problem is an element of $H^{\frac{1}{2},2}(\partial\Sigma)$. Therefore we use some results for Sobolev spaces defined on submanifolds. The weak solution in $H^{2,2}(\Sigma)$ is related to the classical solution in the following way. Let $u \in H^{2,2}(\Sigma)$ be the weak solution to the inner regular oblique boundary problem for the Poisson equation, provided by Theorem 6. Then we have

$$\begin{aligned} \Delta u = f \qquad & \lambda^n - \text{almost everywhere in } \Sigma, \\ \langle \underline{a}, \nabla u\rangle + bu = g \qquad & H^{n-1} - \text{almost everywhere on } \partial\Sigma. \end{aligned}$$

Such a solution we call strong solution to the inner regular oblique boundary problem for the Poisson equation.

3.4. Ritz–Galerkin approximation

In this section we provide a Ritz–Galerkin method which allows us to approximate the weak solution with help of a numerical computation. Let $a(\eta,u)$ and $F(\eta)$ be defined as above and the conditions of Theorem 9 be satisfied. Furthermore let $(U_n)_{n\in\mathbb{N}}$ be an increasing sequence of finite-dimensional subspaces of $H^{1,2}(\Sigma)$,

i.e.,$U_n \subset U_{n+1}$ such that $\overline{\bigcup_{n\in\mathbb{N}} U_n} = H^{1,2}(\Sigma)$. Because U_n is as a finite-dimensional subspace of the Hilbert space $H^{1,2}(\Sigma)$ itself a Hilbert space, we find for each $n \in \mathbb{N}$ a unique $u_n \in U_n$ with

$$a(\eta, u_n) = F(\eta) \quad \text{for all } \eta \in U_n.$$

Moreover, let $d := \dim(U_n)$ and $(\varphi_k)_{1\le k\le d}$ be a basis of U_n. Then $u_n \in U_n$ has the following unique representation

$$u_n = \sum_{i=1}^{d} h_i \varphi_i,$$

where $(h_i)_{1\le i\le d}$ is the solution of the linear system of equations given by

$$\sum_{i=1}^{d} a(\varphi_j, \varphi_i) h_i = F(\varphi_j) \quad 1 \le j \le d.$$

The following result from Céa proves that the sequence $(u_n)_{n\in\mathbb{N}}$ really approximates the weak solution u.

Theorem 7. *Let u be the weak solution provided by Theorem 5 and $(u_n)_{n\in\mathbb{N}}$ taken from above. Then:*

$$\|u - u_n\|_{H^{1,2}(\Sigma)} \le \frac{C_6}{C_7} \mathrm{dist}(u, U_n) \overset{n\to\infty}{\longrightarrow} 0,$$

where C_6 and C_7 are the continuity and the coercivity constants of a.

3.5. Stochastic extensions

First we define the spaces of stochastic functions. We are choosing a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, arbitrary but fixed, such that $L^2(\Omega, dP)$ is separable, and define

$$\begin{aligned}
\left(H^{1,2}(\Sigma)\right)'_\Omega &:= L^2(\Omega, P) \otimes \left(H^{1,2}(\Sigma)\right)' &&\cong L^2(\Omega, P; \left(H^{1,2}(\Sigma)\right)'),\\
H^{-\frac{1}{2},2}_\Omega(\partial\Sigma) &:= L^2(\Omega, P) \otimes H^{-\frac{1}{2},2}(\partial\Sigma) &&\cong L^2(\Omega, P; H^{-\frac{1}{2},2}(\partial\Sigma)),\\
H^{\frac{1}{2},2}_\Omega(\partial\Sigma) &:= L^2(\Omega, P) \otimes H^{\frac{1}{2},2}(\partial\Sigma) &&\cong L^2(\Omega, P; H^{\frac{1}{2},2}(\partial\Sigma)),\\
L^2_\Omega(\Sigma) &:= L^2(\Omega, P) \otimes L^2(\Sigma) &&\cong L^2(\Omega, P; L^2(\Sigma))\\
H^{1,2}_\Omega(\Sigma) &:= L^2(\Omega, P) \otimes H^{1,2}(\Sigma) &&\cong L^2(\Omega, P; H^{1,2}(\Sigma)),\\
H^{2,2}_\Omega(\Sigma) &:= L^2(\Omega, P) \otimes H^{2,2}(\Sigma) &&\cong L^2(\Omega, dP; H^{2,2}(\Sigma)),
\end{aligned}$$

with help of the tensor product. Now we can investigate the stochastic inner regular oblique boundary problem for the Poisson equation. We are searching for a solution $u \in H^{1,2}_\Omega(\Sigma)$ of:

$$\begin{aligned}
\Delta u(x,\omega) &= f(x,\omega) && \text{for all } x \in \Sigma, \text{ P-a.a. } \omega \in \Omega,\\
(\underline{a} \cdot \nabla u(x,\omega)) + b u(x,\omega) &= \tilde{g}(x,\omega) && \text{for all } x \in \partial\Sigma, \text{ P-a.a. } \omega \in \Omega,\\
|(\underline{a} \cdot \underline{\nu})| &\ge C_8 > 0 && \text{on } \partial\Sigma.
\end{aligned}$$

Using the argumentation from the first section we come immediately to the weak formulation of the stochastic boundary problem.

Definition 8. Find $u \in H_{\Omega}^{1,2}(\Sigma)$ with:

$$\int_{\Omega} \left({}_{H^{\frac{1}{2},2}(\partial\Sigma)}\langle \eta, \frac{g}{\langle \underline{a}, \nu \rangle} \rangle_{H^{-\frac{1}{2},2}(\partial\Sigma)} - \int_{\partial\Sigma} \eta \frac{b}{\langle \underline{a}, \nu \rangle} u \, dH^{n-1} \right) dP$$
$$- \int_{\Omega} \left(\sum_{i=1}^{n} {}_{H^{\frac{1}{2},2}(\partial\Sigma)}\langle \eta \frac{\underline{a}_i}{\langle \underline{a}, \nu \rangle}, (\nabla_{\partial\Sigma} u)_i \rangle_{H^{-\frac{1}{2},2}(\partial\Sigma)} \right) dP$$
$$- \int_{\Omega} \left(\int_{\Sigma} (\nabla \eta \cdot \nabla u) \, d\lambda^n - {}_{H^{1,2}(\Sigma)}\langle \eta, f \rangle_{(H^{1,2}(\Sigma))'} \right) dP = 0$$

for all $\eta \in H_{\Omega}^{1,2}(\Sigma)$. u is called *stochastic weak solution* of the stochastic inner regular oblique boundary problem for the Poisson equation.

Obviously, $u \in H_{\Omega}^{1,2}(\Sigma)$ is a stochastic weak solution of the stochastic regular oblique boundary problem for the Poisson equation if and only if for P-a.a. $\omega \in \Omega$, $u_\omega := u(\cdot, \omega)$ is a weak solution of the deterministic problem

$$\Delta u_\omega = f(\cdot, \omega) \quad \text{on } \Sigma,$$
$$\langle \underline{a} \cdot \nabla u_\omega \rangle + b u_\omega = g(\cdot, \omega) \quad \text{on } \partial\Sigma.$$

The solution operator of the deterministic problem extends to the stochastic setting in the following way

Theorem 9. *Let Σ be a bounded $C^{1,1}$-domain, $\underline{a} \in H^{1,\infty}(\partial\Sigma; \mathbb{R}^n)$, fulfilling condition (1), and $b \in L^{\infty}(\partial\Sigma)$ such that:*

$$\operatorname*{ess\,inf}_{\partial\Sigma} \left(\frac{b}{\langle \underline{a}, \nu \rangle} - \frac{1}{2} \operatorname{div}_{\partial\Sigma} \left(\frac{\underline{a}}{\langle \underline{a}, \nu \rangle} - \nu \right) \right) > 0.$$

Then for all $f \in \left(H^{1,2}(\Sigma)\right)'_{\Omega}$ and $g \in H_{\Omega}^{-\frac{1}{2},2}(\partial\Sigma)$ there exists one and only one stochastic weak solution $u \in H_{\Omega}^{1,2}(\Sigma)$ of the stochastic inner regular oblique boundary problem for the Poisson equation. Additionally we have for a constant $0 < C_9 < \infty$

$$\|u\|_{H_{\Omega}^{1,2}(\Sigma)} \leq C_9 \left(\|f\|_{(H^{1,2}(\Sigma))'_{\Omega}} + \|g\|_{H_{\Omega}^{-\frac{1}{2},2}(\partial\Sigma)} \right).$$

In the proof we use the results from the deterministic setting in order to prove the requirements of the Lax–Milgram Lemma to be fulfilled. Using the isomorphisms of the tensor product spaces to spaces of Hilbert space-valued random variables, also the regularization result translates to the stochastic setting.

Theorem 10. *Let $\Sigma \subset \mathbb{R}^n$ be a bounded $C^{2,1}$-domain, $\underline{a} \in H^{2,\infty}(\partial\Sigma; \mathbb{R}^n)$ fulfilling condition (1) and $b \in H^{1,\infty}(\partial\Sigma)$. Then for all $f \in L_{\Omega}^2(\Sigma)$ and $g \in H_{\Omega}^{\frac{1}{2},2}(\partial\Sigma)$, the weak solution $u \in H_{\Omega}^{1,2}(\Sigma)$ to the inner regular oblique boundary problem for*

the Poisson equation, provided in Theorem 5*, is even in* $H^{2,2}_{\Omega}(\Sigma)$*. Furthermore we have the a priori estimate*

$$\|u\|_{H^{2,2}_{\Omega}(\Sigma)} \leq C_{10} \left(\|f\|_{L^2_{\Omega}(\Sigma)} + \|g\|_{H^{\frac{1}{2},2}_{\Omega}(\partial\Sigma)} \right).$$

for a constant $0 < C_{10} < \infty$*.* u *is called stochastic strong solution and fulfills the classical problem almost everywhere.*

At the end of this section we want to mention that a Ritz–Galerkin approximation is available also for the stochastic weak solution, repeating the procedure from the deterministic problem. For details and proofs of the presented results we refer to [9].

4. Fundamental results for the outer problem

In this chapter we provide a solution operator for the outer oblique boundary problem for the Poisson equation. The results presented in this chapter are taken from [10] and as well the proofs as further details can be found in this reference. The outer problem is defined in an unbounded domain $\Sigma \subset \mathbb{R}^n$ which is representable as $\mathbb{R}^n \backslash \overline{D}$, where D is a bounded domain. Additionally we assume $0 \in D$ which is necessary for the Kelvin transformation. For unbounded Σ a Poincaré inequality is yet missing. Consequently we cannot use the technique used for the inner problem because we are unable to prove coercivity of the bilinear from a weak formulation corresponding to the outer problem. Thus we will not derive a weak formulation for the outer problem and thus we do not have to consider a regular outer problem. Our approach is to transform the outer problem to a corresponding inner problem for which a solution operator is available by the results of the previous chapter. In this way we will construct our weak solution and for this solution also a Ritz–Galerkin Method is available because of the continuity of the Kelvin transformation. Finally we again extend our results for stochastic inhomogeneities as well as stochastic solutions and present some examples from Geomathematics. The described procedure is presented in the following four sections.

4.1. Transformations to an inner setting

In this section we define the transformations which will be needed in order to transform the outer oblique boundary problem for the Poisson equation to a corresponding regular inner problem. Then we will apply the solution operator in order to get a weak solution in the inner domain. This solution will be transformed with help of the Kelvin transformation to a function defined in the outer domain. In the next section we will finally prove that this function solves the outer problem for sufficiently smooth data almost everywhere, which gives the connection to the original problem. The whole procedure is illustrated in the following Table 1.

We proceed in the following way. First we define the Kelvin transformation K_{Σ} of the outer domain Σ to a corresponding bounded domain Σ^K. At next the

Outer problem :	Σ	(f,g)	$\left(\overset{S^{\text{out}}_{\underline{a},b}}{\longrightarrow}\right)$	u
	$\downarrow K_\Sigma$	$T_1 \downarrow T_2$		$\uparrow K$
Inner problem:	Σ^K	$(T_1(f), T_2(g))$	$\overset{S^{\text{in}}_{T_3(\underline{a}),T_4(b)}}{\longrightarrow}$	v

TABLE 1. Transformation procedure

Kelvin transformation K of the solution for the inner problem will be presented. Finally we define the transformations T_1 and T_2 for the inhomogeneities as well as T_3 and T_4 for the coefficients. We will also show that the operators K, T_1 and T_2 are continuous. The consequence is that our solution operator

$$S^{\text{out}}_{\underline{a},b}(f,g) := K\left(S^{\text{in}}_{T_3(\underline{a}),T_4(b)}(T_1(f), T_2(f))\right),$$

forms a linear and continuous solution operator for the outer problem. Because all main results assume Σ to be at least an outer $C^{1,1}$-domain, we fix Σ for the rest of this chapter as such a domain, if not stated otherwise. At first we transform the outer domain Σ to a bounded domain Σ^K. The tool we use is the so called Kelvin transformation K_Σ for domains. We introduce the Kelvin transformation for outer $C^{1,1}$-domains in the following definition.

Definition 11. Let Σ be an outer $C^{1,1}$-domain and $x \in \Sigma$ be given. Then we define the Kelvin transformation $K_\Sigma(x)$ of x by

$$K_\Sigma(x) := \frac{x}{|x|^2}.$$

Furthermore, we define Σ^K as the Kelvin transformation of Σ via

$$\Sigma^K := K_\Sigma(\Sigma) \cup \{0\} = \left\{K_\Sigma(x)\middle| x \in \Sigma\right\} \cup \{0\}.$$

From this point on, we fix the notation in such a way that Σ^K always means the Kelvin transformation of Σ. Figure 3 illustrates the Kelvin transformation of Σ.

We have $K_\Sigma \in C^\infty(\mathbb{R}^n\backslash\{0\}; \mathbb{R}^n\backslash\{0\})$ with $K_\Sigma^2 = \text{Id}_{\mathbb{R}^n\backslash\{0\}}$. Furthermore we obtain by standard calculus, using the Leibnitz formula for the determinant, $|\text{Det}(D(K_\Sigma))(x)| \leq C_{11}|x|^{-2n}$ for all $x \in \mathbb{R}^n\backslash\{0\}$, $1 \leq i \leq n$. This is one of the reasons for the weighted measures of the Sobolev spaces introduced later on. Moreover the transformation leaves the regularity of the surface invariant. Let Σ be an outer $C^{2,1}$-domain. Then Σ^K is a bounded $C^{2,1}$-domain. Moreover we have that $\partial\Sigma^K = K_\Sigma(\partial\Sigma)$. Furthermore, if Σ is an outer $C^{1,1}$-domain, we have that Σ^K is a bounded $C^{1,1}$-domain. There are geometric situations in which $\partial\Sigma^K$ can be computed easily. For example if $\partial\Sigma$ is a sphere around the origin with radius R, then $\partial\Sigma^K$ is a sphere around the origin with radius R^{-1}. Furthermore, if $\partial\Sigma \subset \mathbb{R}^2$ is an ellipse with semi axes a and b around the origin, then $\partial\Sigma^K$ is also an ellipse around

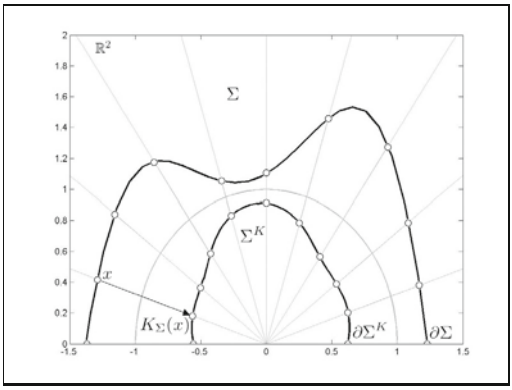


FIGURE 3. Kelvin transformation of Σ

the origin with semi axes b^{-1} and a^{-1}. At next we present the transformation for the weak solution of the inner problem back to the outer setting. Therefore we introduce the operator K. This is the so-called Kelvin transformation for functions. It transforms a given function u, defined in Σ^K, to a function $K(u)$, defined in Σ. In addition, it preserves some properties of the original function. We will state some of these properties. So, after the following considerations it will be clear why we choose exactly this transformation. It will also be clear how we have to choose the transformations $T_1, \ldots, T_4$ in the following. We start with a definition.

Definition 12. Let Σ be an outer $C^{1,1}$-domain and u be a function defined on Σ^K. Then we define the Kelvin transformation $K(u)$ of u, which is a function defined on Σ, via

$$K(u)(x) := \frac{1}{|x|^{n-2}} u\left(\frac{x}{|x|^2}\right),$$

for all $x \in \Sigma$.

Important is, that this transformation acts as a multiplier when applying the Laplace operator. Note that $-(n-2)$ is the only exponent for $|x|$ which has this property. We have for $u \in C^2(\Sigma^K)$ that $K(u) \in C^2(\Sigma)$ with

$$\Delta(K(u))(x) = \frac{1}{|x|^{n+2}} (\Delta u)\left(\frac{x}{|x|^2}\right),$$

for all $x \in \Sigma$. As already mentioned above we will apply K to functions from $H^{1,2}(\Sigma^K)$. So we want to find a normed function space $(V, \|\cdot\|_V)$ such that

$$K : H^{1,2}(\Sigma^K) \to V$$

defines a continuous operator.

It turns out that the weighted Sobolev space $H^{1,2}_{\frac{1}{|x|^2},\frac{1}{|x|}}(\Sigma)$ is a suitable choice. We have the following important result for K acting on $H^{1,2}(\Sigma^K)$.

Theorem 13. *Let Σ be an outer $C^{1,1}$-domain. For $u \in H^{1,2}(\Sigma^K)$ let $K(u)$ be defined as above for all $x \in \Sigma$. Then we have that*

$$K : H^{1,2}(\Sigma^K) \to H^{1,2}_{\frac{1}{|x|^2},\frac{1}{|x|}}(\Sigma)$$

is a continuous linear operator. Moreover K is injective.

It is left to provide the remaining transformations $T_1, \ldots, T_4$. In the first part we treat T_1, which transforms the inhomogeneity f of the outer problem in Σ to an inhomogeneity of the corresponding inner problem in Σ^K. Assume f to be a function defined on Σ. We want to define the function $T_1(f)$ on Σ^K, such that

$$\Delta u(x) = T_1(f)(x), \quad x \in \Sigma^K, \tag{3}$$

implies that

$$\Delta(K(u))(y) = f(y), \quad y \in \Sigma. \tag{4}$$

We are able to define T_1 for functions defined on Σ as follows.

Definition 14. Let Σ be an outer $C^{1,1}$-domain and f be a function defined in Σ. Then we define a function $T_1(f)$ on Σ^K by

$$T_1(f)(x) := \frac{1}{|x|^{n+2}} f\left(\frac{x}{|x|^2}\right),$$

for all $x \in \Sigma^K \backslash \{0\}$ and $T_1(f)(0) = 0$.

T_1 is well defined and fulfills the relation described by equations (3) and (4). Furthermore, T_1 defines a linear continuous isomorphism

$$T_1 : L^2_{|x|^2}(\Sigma) \to L^2(\Sigma^K),$$

with $(T_1)^{-1} = T_1$. We want to generalize our inhomogeneities in a way similar to the inner problem. This means we have to identify a normed vector space $(W, \|\cdot\|_W)$, such that

$$T_1 : W \to \left(H^{1,2}(\Sigma^K)\right)',$$

defines a linear continuous operator. Additionally, we want to end up with a Gelfand triple

$$U \subset L^2_{|x|^2}(\Sigma) \subset W.$$

Consequently $L^2_{|x|^2}(\Sigma)$ should be a dense subspace. It is possible to prove that the space $\left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'$ is a suitable choice. Recall the Gelfand triple, given by

$$H^{1,2}_{|x|^2,|x|^3}(\Sigma) \subset L^2_{|x|^2}(\Sigma) \subset \left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'.$$

Theorem 15. *We define a continuous linear operator*

$$T_1 : L^2_{|x|^2}(\Sigma) \to \left(H^{1,2}(\Sigma^K)\right)',$$

by

$$\left(T_1(f)\right)(h) := \int_{\Sigma^K} \left(T_1(f)\right)(y) h(y)\, d\lambda^n(y), \quad h \in H^{1,2}(\Sigma^K),$$

for $f \in L^2_{|x|^2}(\Sigma)$, *where* $L^2_{|x|^2}(\Sigma)$ *is equipped with the norm* $\|\cdot\|_{\left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'}$, *which, by the BLT Theorem, extends uniquely to a linear bounded operator*

$$T_1 : \left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)' \to \left(H^{1,2}(\Sigma^K)\right)'.$$

Next we define the transformations for the boundary inhomogeneity g and the coefficients $\underline{a}$ and b. This means we want to find transformations T_2, T_3 and T_4 such that

$$\langle \left(T_3(\underline{a})\right)(x), \nabla u(x)\rangle + \left(T_4(b)\right)(x) u(x) = \left(T_2(g)\right)(x), \tag{5}$$

for all $x \in \partial\Sigma^K$, yields that

$$\langle \underline{a}(y), \nabla\left(\left(K(u)\right)(y)\right)\rangle + b(y)u(x) = g(y), \tag{6}$$

for all $y \in \partial\Sigma^K$. We start with the transformation $T_2(g)$ of g.

Definition 16. Let Σ be an outer $C^{1,1}$-domain and g be a function defined on $\partial\Sigma$. Then we define a function $T_2(g)$ on $\partial\Sigma^K$ by

$$\left(T_2(g)\right)(x) := g\left(\frac{x}{|x|^2}\right), \quad x \in \partial\Sigma^K.$$

Again we use a Gelfand triple, namely

$$H^{\frac{1}{2},2}(\partial\Sigma) \subset L^2(\partial\Sigma) \subset H^{-\frac{1}{2},2}(\partial\Sigma).$$

We have that

$$\begin{aligned} T_2 &: L^2(\partial\Sigma) \to L^2(\partial\Sigma^K), \\ T_2 &: H^{\frac{1}{2},2}(\partial\Sigma) \to H^{\frac{1}{2},2}(\partial\Sigma^K), \end{aligned}$$

define linear, bounded isometries with $(T_2)^{-1} = T_2$. Moreover we define a continuous linear operator

$$T_2 : L^2(\partial\Sigma) \to H^{-\frac{1}{2},2}(\partial\Sigma^K),$$

by

$$\left(T_2(g)\right)(h) := \int_{\partial\Sigma^K} T_2(g)(y) h(y)\, dH^{n-1}(y), \quad h \in H^{-\frac{1}{2},2}(\partial\Sigma)$$

for $g \in L^2(\partial\Sigma)$, where $L^2(\partial\Sigma)$ is equipped with the norm $\|\cdot\|_{H^{-\frac{1}{2},2}(\partial\Sigma)}$. Hence again the BLT Theorem gives a unique continuous continuation

$$T_2 : H^{-\frac{1}{2},2}(\partial\Sigma) \to H^{-\frac{1}{2},2}(\partial\Sigma^K).$$

Closing this section, we give the definitions of the transformations T_3 and T_4.

Definition 17. Let Σ be an outer $C^{1,1}$-domain and $\underline{a}$ and b be defined on $\partial\Sigma$. We define the operators T_3 and T_4 via

$$(T_3(\underline{a}))(x) := |x|^n \cdot \left(\underline{a}\left(\frac{x}{|x|^2}\right) - 2\left\langle \underline{a}\left(\frac{x}{|x|^2}\right), \underline{e}_x \right\rangle \underline{e}_x \right),$$
$$(T_4(b))(x) := |x|^{n-2} \cdot \left(b\left(\frac{x}{|x|^2}\right) + (2-n)\left\langle \underline{a}\left(\frac{x}{|x|^2}\right), x \right\rangle \right),$$

for all $x \in \partial\Sigma^K$, where $\underline{e}_x$ denotes the unit vector in direction x. Furthermore we have

$$\begin{aligned} T_3 &: H^{1,\infty}(\partial\Sigma) \to H^{1,\infty}(\partial\Sigma^K), \\ T_4 &: L^{\infty}(\partial\Sigma) \to L^{\infty}(\partial\Sigma^K), \end{aligned}$$

if Σ is an outer $C^{1,1}$-domain and $\underline{a} \in H^{1,\infty}(\partial\Sigma)$ for T_4. All operators are well defined and give the relation formulated by equations (5) and (6).

These operators have the properties

$$\begin{aligned} T_3 &: H^{1,\infty}(\partial\Sigma) \to H^{1,\infty}(\partial\Sigma^K), \\ T_4 &: L^{\infty}(\partial\Sigma) \to L^{\infty}(\partial\Sigma^K), \end{aligned}$$

if Σ is an outer $C^{1,1}$-domain and $\underline{a} \in H^{1,\infty}(\partial\Sigma)$ for T_4 and

$$\begin{aligned} T_3 &: H^{2,\infty}(\partial\Sigma) \to H^{2,\infty}(\partial\Sigma^K), \\ T_4 &: H^{1,\infty}(\partial\Sigma) \to H^{1,\infty}(\partial\Sigma^K), \end{aligned}$$

if Σ is an outer $C^{2,1}$-domain and $\underline{a} \in H^{2,\infty}(\partial\Sigma)$ for T_4.

4.2. Solution operator for the outer problem

In this section we want apply the solution operator of the inner regular problem in order to get a weak solution of the outer problem. Therefore we will use a combination of all the operators defined in the previous section. In order to avoid confusion we denote the normal vector of $\partial\Sigma$ by ν and the normal vector of $\partial\Sigma^K$ by ν^K. We start with the classical formulation of the outer oblique boundary problem for the Poisson equation in the following definition.

Definition 18. Let Σ be an outer $C^{1,1}$-domain, $f \in C^0(\Sigma)$, $b, g \in C^0(\partial\Sigma)$ and $\underline{a} \in C^0(\partial\Sigma; \mathbb{R}^n)$ be given. A function $u \in C^2(\Sigma) \cap C^1(\overline{\Sigma})$ such that

$$\begin{aligned} \Delta u(x) &= f(x), && \text{for all } x \in \Sigma, \\ \langle \underline{a}(x) \cdot \nabla u(x) \rangle + b \cdot u(x) &= g(x), && \text{for all } x \in \partial\Sigma, \\ u(x) &\to 0, && \text{for } |x| \to \infty, \end{aligned}$$

is called classical solution of the outer oblique boundary problem for the Poisson equation.

Now we state the main result of this section which can be proved by the results about the transformations above.

Theorem 19. *Let Σ be an outer $C^{1,1}$-domain, $\underline{a} \in H^{1,\infty}(\partial\Sigma;\mathbb{R}^n)$, $b \in L^\infty(\partial\Sigma)$, $g \in H^{-\frac{1}{2},2}(\partial\Sigma)$ and $f \in \left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'$, such that*

$$\left|\langle (T_3(\underline{a}))(y), \nu^K(y)\rangle\right| > C > 0, \tag{7}$$

$$\operatorname*{ess\,inf}_{\partial\Sigma^K}\left\{\frac{T_4(b)}{\langle T_3(\underline{a}),\nu^K\rangle} - \frac{1}{2}\mathrm{div}_{\partial\Sigma^K}\left(\frac{T_3(\underline{a})}{\langle T_3(\underline{a}),\nu^K\rangle} - \nu^K\right)\right\} > 0, \tag{8}$$

for all $y \in \partial\Sigma^K$, where $0 < C < \infty$. Then we define

$$u := S^{\mathrm{out}}_{\underline{a},b}(f,g) := K\left(S^{\mathrm{in}}_{T_3(\underline{a}),T_4(b)}(T_1(f),T_2(g))\right),$$

as the weak solution to the outer oblique boundary problem for the Poisson equation from Definition 18. *$S^{\mathrm{out}}_{\underline{a},b}$ is injective and we have for a constant $0 < C_{12} < \infty$*

$$\|u\|_{H^{1,2}_{\frac{1}{|x|^2},\frac{1}{|x|}}(\Sigma)} \le C_{12}\left(\|f\|_{\left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'} + \|g\|_{H^{-\frac{1}{2},2}(\partial\Sigma)}\right).$$

We are able to prove that the Kelvin transformation for functions is also a continuous operator from $H^{2,2}(\Sigma^K)$ to $H^{2,2}_{\frac{1}{|x|^2},\frac{1}{|x|},1}(\Sigma)$. So we can prove the following regularization result, based on the regularization result for the inner problem, see Theorem 6. The following theorem shows, that the weak solution, defined by Theorem 19, is really related to the outer problem, given in Definition 18, although it is not derived by an own weak formulation.

Theorem 20. *Let Σ be an outer $C^{2,1}$-domain, $\underline{a} \in H^{2,\infty}(\partial\Sigma;\mathbb{R}^n)$, $b \in H^{1,\infty}(\partial\Sigma)$ such that* (7) *and* (8) *holds. If $f \in L^2_{|x|^2}(\Sigma)$ and $g \in H^{\frac{1}{2},2}(\Sigma)$ then we have that u provided by Theorem* 19 *is a strong solution, i.e., $u \in H^{2,2}_{\frac{1}{|x|^2},\frac{1}{|x|},1}(\Sigma)$, and*

$$\Delta u = f,$$
$$\langle \underline{a}, \nabla u\rangle + bu = g,$$

almost everywhere on Σ and $\partial\Sigma$, respectively. Furthermore we have an a priori estimate

$$\|u\|_{H^{2,2}_{\frac{1}{|x|^2},\frac{1}{|x|},1}(\Sigma)} \le C_{13}\Big(\|f\|_{L^2_{|x|^2}(\Sigma)} + \|g\|_{H^{\frac{1}{2},2}(\partial\Sigma)}\Big),$$

with a constant $0 < C_{13} < \infty$.

As a consequence we have that if the data in Theorem 20 fulfills the requirements of a classical solution, the weak solution u provided by Theorem 19 coincides with this classical solution. At the end of this section we investigate the conditions on the oblique vector field. Analogously to the regular inner problem, we have condition (8), which is a transformed version of (2) and gives a relation between $\underline{a}$ and b, depending on the geometry of the surface $\partial\Sigma$. Moreover condition (7) is a transformed version of (1) and gives the non-admissible direction for the oblique vector field $\underline{a}$. For the regular inner problem, (1) states the tangential directions as non-admissible for the oblique vector field. For the outer problem the direction

depends as well on the direction of the normal vector $\nu(y)$ at the point $y \in \partial\Sigma$ as on the direction of y itself. In this section we will investigate this dependency in detail. Using the definitions of T_3 and T_4, we can rewrite condition (7) into the equivalent form

$$\left|\cos\left(\angle_{\underline{a}(x),\nu^K\left(\frac{x}{|x|^2}\right)}\right) - 2\cdot\cos\left(\angle_{\underline{a}(x),\underline{e}_x}\right)\cdot\cos\left(\angle_{\underline{e}_x,\nu^K\left(\frac{x}{|x|^2}\right)}\right)\right| > C_{13} > 0, \tag{9}$$

for all $x \in \partial\Sigma$ and $0 < C_{13} < \infty$ independent of x. We use the formula

$$\frac{\langle y, z\rangle}{|y|\cdot|z|} =: \cos(\angle_{y,z}),$$

for vectors in $\mathbb{R}^n$, where $\angle_{y,z}$ denotes the angle $0 \le \angle_{y,z} \le \pi$ between y and z. Going to $\mathbb{R}^2$ and setting

$$C_{14}(x) := \cos(\angle_{\underline{e}_x,\nu^K(x)}),$$
$$C_{15}(x) := \sin(\angle_{\underline{e}_x,\nu^K(x)}).$$

we can explicitly characterize the non admissible direction as

$$\angle_{\underline{a}(x),\underline{e}_x} = \tan^{-1}\left|\frac{C_{14}(x)}{C_{15}(x)}\right|,$$

if $C_{15}(x) \neq 0$ and $\angle_{\underline{a}(x),\underline{e}_x} = \frac{\pi}{2}$ if $C_{15}(x) = 0$. Generally, transforming the problem to an inner setting transforms the conditions for the coefficients $\underline{a}$ and b. There are circumstances in which we have the same non-admissible direction as for the inner problem, i.e., the tangential directions are non-admissible. For example, this is the case if $\partial\Sigma$ is a sphere around the origin. In Figure 4 the situation for $\Sigma \subset \mathbb{R}^2$ is illustrated, the dashed line indicates the non-admissible direction, which occurs because of the transformed regularity condition $\langle T_3(\underline{a}), \nu^K\rangle > C_{14} > 0$, see (7).

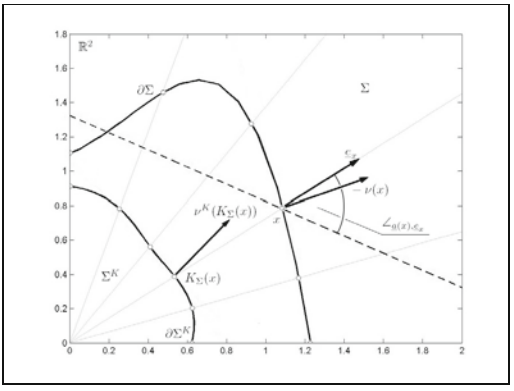


FIGURE 4. Non-admissible direction for the outer problem

4.3. Ritz–Galerkin method

In this subsection we provide a Ritz–Galerkin method for the weak solution to the outer problem. Therefore we use the approximation of the weak solution to the corresponding inner problem, provided in Chapter 3. Assume Σ to be an outer $C^{1,1}$-domain. Furthermore let $\underline{a} \in H^{1,\infty}(\partial\Sigma;\mathbb{R}^n)$, $b \in L^\infty(\partial\Sigma)$, $g \in H^{-\frac{1}{2},2}(\partial\Sigma)$ and $f \in \left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'$, such that condition (7) and condition (8) is fulfilled. We want to approximate the weak solution u to the outer oblique boundary problem, provided by Theorem 19. Let a and F be defined by

$$\begin{aligned} a(\eta,v) := &-\sum_{i=1}^{n} {}_{H^{\frac{1}{2},2}(\partial\Sigma)}\left\langle \eta\frac{T_3(\underline{a})_i}{\langle T_3(\underline{a}),\nu^K\rangle} - \nu_i^K, (\nabla_{\partial\Sigma}v)_i \right\rangle_{H^{-\frac{1}{2},2}(\partial\Sigma)} \\ &- \int_\Sigma (\nabla\eta,\nabla v)\,d\lambda^n - \int_{\partial\Sigma} \eta\frac{T_4(b)}{\langle T_3(\underline{a}),\nu^K\rangle} v\,dH^{n-1} \\ F(\eta) := &{}_{H^{\frac{1}{2},2}(\partial\Sigma)}\left\langle \eta, \frac{T_2(g)}{\langle T_3(\underline{a}),\nu^K\rangle}\right\rangle_{H^{-\frac{1}{2},2}(\partial\Sigma)} - {}_{H^{1,2}(\Sigma)}\langle \eta, T_1(f)\rangle_{(H^{1,2}(\Sigma))'} \end{aligned}$$

for $\eta, v \in H^{1,2}(\Sigma^K)$.

Furthermore, let $(V_n)_{n\in\mathbb{N}}$ be an increasing sequence of finite-dimensional subspaces of $H^{1,2}(\Sigma^K)$, i.e., $V_n \subset V_{n+1}$ such that $\overline{\bigcup_{n\in\mathbb{N}} V_n} = H^{1,2}(\Sigma^K)$. Then there exists for each $n \in \mathbb{N}$ a unique $v_n \in V_n$ with:

$$a(\eta, v_n) = F(\eta) \quad \text{for all } \eta \in V_n,$$

see Section 3.4. Moreover, v_n can be computed explicitly by solving a linear system of equations. In Section 3.4 we have also seen that

$$\|v - v_n\|_{H^{1,2}(\Sigma)} \le C_{16}\,\mathrm{dist}(v, V_n) \overset{n\to\infty}{\longrightarrow} 0.$$

So using the continuity of the operator K, see Theorem 13, we consequently get the following result

Theorem 21. *Let u be the weak solution provided by Theorem 19 to the outer problem and v, $(v_n)_{n\in\mathbb{N}}$ taken from Theorem 7 and Theorem 5, both corresponding to $\underline{a}$, b, g, f and Σ, given at the beginning of this section. Then:*

$$\|u - K(v_n)\|_{H^{1,2}(\Sigma)} \le C_{17}\,\mathrm{dist}(v, V_n) \overset{n\to\infty}{\longrightarrow} 0.$$

4.4. Stochastic extensions and examples

In this section we implement stochastic inhomogeneities as well as stochastic weak solutions for the outer setting. Again we start by defining the spaces of stochastic functions. So, let Σ be an outer $C^{1,1}$-domain and $(\Omega,\mathcal{F},\mathcal{P})$ a probability space,

arbitrary but fixed, such that $L^2(\Omega, P)$ is separable. We define

$$\begin{aligned}
\left(H^{2,2}_{\frac{1}{|x|^2},\frac{1}{|x|},1}(\Sigma)\right)_\Omega &:= L^2(\Omega,P)\otimes H^{2,2}_{\frac{1}{|x|^2},\frac{1}{|x|},1}(\Sigma) &&\cong L^2\left(\Omega,P;H^{2,2}_{\frac{1}{|x|^2},\frac{1}{|x|},1}(\Sigma)\right),\\
\left(H^{1,2}_{\frac{1}{|x|^2},\frac{1}{|x|}}(\Sigma)\right)_\Omega &:= L^2(\Omega,P)\otimes H^{1,2}_{\frac{1}{|x|^2},\frac{1}{|x|}}(\Sigma) &&\cong L^2\left(\Omega,P;H^{1,2}_{\frac{1}{|x|^2},\frac{1}{|x|}}(\Sigma)\right),\\
\left(L^2_{|x|^2}(\Sigma)\right)_\Omega &:= L^2(\Omega,P)\otimes L^2_{|x|^2}(\Sigma) &&\cong L^2\left(\Omega,P;L^2_{|x|^2}(\Sigma)\right),\\
\left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'_\Omega &:= L^2(\Omega,P)\otimes \left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)' &&\cong L^2\left(\Omega,P;\left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'\right),\\
H^{\frac{1}{2},2}_\Omega(\partial\Sigma) &:= L^2(\Omega,P)\otimes H^{\frac{1}{2},2}(\partial\Sigma) &&\cong L^2\left(\Omega,P;H^{\frac{1}{2},2}(\partial\Sigma)\right),\\
L^2_\Omega(\partial\Sigma) &:= L^2(\Omega,P)\otimes L^2(\partial\Sigma) &&\cong L^2\left(\Omega,P;L^2(\partial\Sigma)\right),\\
H^{-\frac{1}{2},2}_\Omega(\partial\Sigma) &:= L^2(\Omega,P)\otimes H^{-\frac{1}{2},2}(\partial\Sigma) &&\cong L^2\left(\Omega,P;H^{-\frac{1}{2},2}(\partial\Sigma)\right).
\end{aligned}$$

Because all spaces above are separable, we can again use the isomorphisms to Hilbert space-valued random variables. Thus we can prove the following main result of this section by defining the stochastic solution operator pointwisely.

Theorem 22. *Let Σ be an outer $C^{1,1}$-domain, $\underline{a}\in H^{1,\infty}(\partial\Sigma;\mathbb{R}^n)$, $b\in L^\infty(\partial\Sigma)$, $g\in H^{-\frac{1}{2},2}_\Omega(\partial\Sigma)$ and $f\in \left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'_\Omega$, such that (7) and (8) holds. Then we define*

$$u(\,\cdot\,,\omega) := S^{\mathrm{out}}_{\underline{a},b}(f(\,\cdot\,,\omega),g(\,\cdot\,,\omega)),$$

for dP-almost all $\omega\in\Omega$. u is called stochastic weak solution to the outer oblique boundary problem for the Poisson equation. Furthermore we have for a constant $0<C_{18}<\infty$

$$\|u\|_{\left(H^{1,2}_{\frac{1}{|x|^2},\frac{1}{|x|}}(\Sigma)\right)_\Omega}\le C_{18}\left(\|f\|_{\left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'_\Omega}+\|g\|_{H^{-\frac{1}{2},2}_\Omega(\partial\Sigma)}\right).$$

Moreover, we have the following result for a stochastic strong solution.

Theorem 23. *Let Σ be an outer $C^{2,1}$-domain, $\underline{a}\in H^{2,\infty}(\partial\Sigma;\mathbb{R}^n)$, $b\in H^{1,\infty}(\partial\Sigma)$ such that (7) and (8) holds. If $f\in\left(L^2_{|x|^2}(\Sigma)\right)_\Omega$ and $g\in H^{\frac{1}{2},2}_\Omega(\Sigma)$ then we have $u\in\left(H^{2,2}_{\frac{1}{|x|^2},\frac{1}{|x|},1}(\Sigma)\right)_\Omega$, for u provided by Theorem 22, and*

$$\begin{aligned}
\Delta u(x,\omega) &= f(x,\omega),\\
\langle\underline{a}(y),\nabla u(y,\omega)\rangle + b(y)u(y,\omega) &= g(y,\omega),
\end{aligned}$$

for λ^n-almost all $x\in\Sigma$, for H^{n-1}-almost all $y\in\partial\Sigma$ and for dP-almost all $\omega\in\Omega$. Furthermore, we have an a priori estimate

$$\|u\|_{\left(H^{2,2}_{\frac{1}{|x|^2},\frac{1}{|x|},1}(\Sigma)\right)_\Omega}\le C_{19}\Big(\|f\|_{\left(L^2_{|x|^2}(\Sigma)\right)_\Omega}+\|g\|_{H^{\frac{1}{2},2}_\Omega(\partial\Sigma)}\Big),$$

with a constant $0<C_{19}<\infty$.

Again a Ritz–Galerkin method is also available also for the stochastic weak solution. It is left to the reader to write down the details. As mentioned we close the section with examples for stochastic data. These are used in geomathematical applications in order to model noise on measured values. In the following we give the examples for the outer problem. They are also suitable for the inner problem.

4.4.1. Gaussian inhomogeneities. We choose the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ such that X_i, $1 \le i \le n_1$, are $P \otimes \lambda^n$-measurable and Y_j, $1 \le j \le n_2$, are $P \otimes H^{n-1}$-measurable with $X_i(\cdot, x)$, $x \in \Sigma$, and $Y_j(\cdot, x)$, $x \in \partial\Sigma$, Gaussian random variables with expectation value 0 and variance $f^2_{\sigma_i}(x)$ or variance $g^2_{\sigma_j}(x)$, respectively. Here $f_{\sigma_i} \in L^2_{|x|^2}(\Sigma)$ and $g_{\sigma_j} \in L^2(\partial\Sigma)$. We define:

$$f(\omega, x) := f_\mu(x) + \sum_{i=1}^{n_1} X_i(\omega, x), \qquad g(\omega, x) := g_\mu(x) + \sum_{j=1}^{n_2} Y_j(\omega, x),$$

where $f_\mu \in L^2_{|x|^2}(\Sigma)$ and $g_\mu \in L^2(\partial\Sigma)$. To use such kind of inhomogeneities we must show

$$f \in L^2(\Omega \times \Sigma, P \otimes |x|^4 \cdot \lambda^n) \qquad \text{and} \qquad g \in L^2(\Omega \times \partial\Sigma, P \otimes H^{n-1}).$$

It is easy to see that the inhomogeneities defined in this way fulfill these requirements and the main results are applicable. Such a Gaussian inhomogeneity is shown in Figure 5.

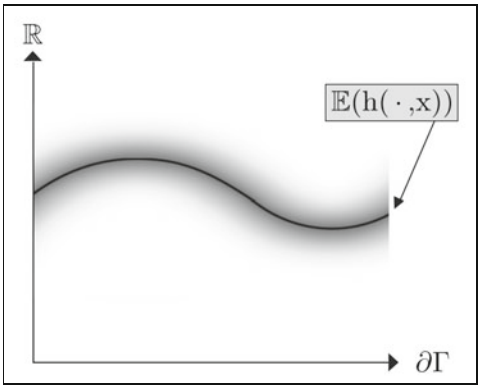


FIGURE 5. Data with Gaussian noise

4.4.2. Gauß–Markov model. Here we refer to [7], in which an application of the example from the previous paragraph can be found. The authors use a random field

$$h(\omega, x) := H(x) + Z(\omega, x)$$

to model an observation noise, where $x \in \partial B_1(0) \subset \mathbb{R}^3$ and $\omega \in \Omega$ with $(\Omega, \mathcal{F}, \mathcal{P})$ a probability space. Here we have that $Z(\cdot, x)$, $x \in \partial B_1(0)$, is a Gaussian random variable with expectation value 0 and variance $\sigma^2 > 0$. Additionally $H(x) \in L^2(\partial B_1(0))$ and the covariance is given by:

$$\operatorname{cov}(Z(\cdot, x_1), Z(\cdot, x_2)) = K(x_1, x_2),$$

where $K : \partial B_1(0) \times \partial B_1(0) \to \mathbb{R}$ is a suitable kernel.

Two geophysically relevant kernels are for example

$$K_1(x_1, x_2) := \frac{\sigma^2}{(M+1)^2} \sum_{n=1}^{M} \frac{2n+1}{4\pi} \mathrm{P}_n((x_1 \cdot x_2)) \quad 0 \leq M < \infty,$$

$$K_2(x_1, x_2) := \frac{\sigma^2}{\exp(-c)} \exp(-c(x_1 \cdot x_2)).$$

P_n, $1 \leq n \leq M$, are the Legendre polynomials defined on $\mathbb{R}$. The noise model corresponding to the second kernel is called first degree Gauß–Markov model. If one chooses a $P \otimes H^{n-1}$-measurable random field Z, then h fulfills the requirements. Existence of a corresponding probability measure P is provided in infinite-dimensional Gaussian Analysis, see, e.g., [4].

4.4.3. Noise model for satellite data. In this paragraph we give another precise application, which can be found in [3]. Here the authors are using stochastic inhomogeneities to implement a noise model for satellite data. Therefore random fields of the form

$$h(\omega, x) := \sum_{i=1}^{m} h_i(x) Z_i(\omega)$$

are used, where $x \in \partial\Sigma \subset \mathbb{R}^3$ and $\omega \in \Omega$ with $(\Omega, \mathcal{F}, \mathcal{P})$ a suitable probability space. Here $\partial\Sigma$ could be for example the Earth's surface and we are searching for harmonic functions in the space outside the Earth. Z_i are Gaussian random variables with expectation value 0 and variance $\sigma_i^2 > 0$ and h_i fulfilling the assumptions of Paragraph 4.4.1. If one chooses $(\Omega, \mathcal{F}, \mathcal{P})$ as $(\mathbb{R}^m, \mathbf{B}(\mathbb{R}), \gamma_{\mathrm{cov}_{ij}}^{0,\sigma_i})$, where:

$$\gamma_{\mathrm{cov}_{ij}}^{0,\sigma_i} := \frac{1}{\sqrt{(2\pi)^m \det(A)}} e^{-\frac{1}{2}(\underline{y}, A^{-1}\underline{y})} \, d\lambda^m,$$

$$a_{ij} := \operatorname{cov}(Z_i, Z_j), \quad 1 \leq i, j \leq m,$$

one has a realization of Z_i as the projection on the ith component in the separable space $L^2(\mathbb{R}^m, \gamma_{\mathrm{cov}_{ij}}^{0,\sigma_i})$.

5. Future directions

In this chapter we want to point out one direction of further investigations. We have seen how to provide the existence of a weak solution to the outer oblique boundary problem for the Poisson equation. Therefore we introduce several transformations.

In Theorem 15, we proved for the transformation of the space inhomogeneity f

$$T_1 : \left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)' \to \left(H^{1,2}(\Sigma^K)\right)'.$$

This transformation is not bijective, i.e., $T_1\left(\left(H^{1,2}_{|x|^2,|x|^3}(\Sigma)\right)'\right) \neq \left(H^{1,2}(\Sigma^K)\right)'$, Finding a Hilbert space V, such that $T_1 : V \to \left(H^{1,2}(\Sigma^K)\right)'$ is bijective would lead to the existence of a weak solution for a even larger class of inhomogeneities. Moreover we have for the transformation K of the weak solution to the inner problem

$$K : H^{1,2}(\Sigma^K) \to H^{1,2}_{\frac{1}{|x|^2},\frac{1}{|x|}}(\Sigma), \text{ where again } K\left(H^{1,2}(\Sigma^K)\right) \neq H^{1,2}_{\frac{1}{|x|^2},\frac{1}{|x|}}(\Sigma),$$

see Theorem 13. Finding a Hilbert space W such that $K : H^{1,2}(\Sigma^K) \to W$ is bijective, would give us uniqueness of the solution and more detailed information about the behavior of u and its weak derivatives, when x is tending to infinity. Additionally, we would be able to define a bijective solution operator for the outer problem. This could be used to find the right Hilbert spaces, such that a Poincaré inequality is available. Consequently the Lax–Milgram Lemma would be applicable directly to a weak formulation for the outer setting, which can be derived similar to the inner problem. Then we might have to consider a regular outer problem, because the tangential direction is forbidden for the oblique vector field, if we want to derive a weak formulation. In turn we get rid of the transformed regularity condition on $\underline{a}$. The results presented in this report are then still an alternative in order to get weak solutions for tangential $\underline{a}$. Moreover, the availability of a Poincaré inequality would lead to existence results for weak solutions to a broader class of second-order elliptic partial differential operators in outer domains. See, e.g., [2] for such second-order elliptic partial differential operators for inner domains.

Instead of using the Ritz–Galerkin approximation, it is also possible to approximate solutions to oblique boundary-value problems for harmonic functions with the help of geomathematical function systems, e.g., spherical harmonics. For such an approach, see, e.g., [8].

6. Conclusion

The analysis of inner oblique boundary value problems is rather good understood and we reached the limit when searching for weak solutions under as weak assumptions as possible. The outer problem causes still problems because of the unboundedness of the domain. As mentioned in Section 5, finding the right distribution spaces such that a Poincaré inequality holds, might lead to bijective solution operators for an even broader class of inhomogeneities. Nevertheless, we are already able to provide weak solutions to the outer problem as presented in the previous sections for very general inhomogeneities. Also stochastic weak solutions for stochastic inhomogeneities as used in geomathematical applications can be provided and approximation methods for the weak solutions are available.

References

[1] Adams, R.A. (1975) Sobolev Spaces. Academic Press, New York-San Francisco-London

[2] Alt, H.W. (2002) Lineare Funktionalanalysis. Springer, Berlin

[3] Bauer, F. (2004) An Alternative Approach to the Oblique Derivative Problem in Potential Theory. Shaker Verlag, Aachen

[4] Berezanskij Y.M. (1995) Spectral Methods in Infinite Dimensional Analysis. Kluwer Academic Publishers, Dordrecht

[5] Dautray, R., Lions, J.-L. (1988) Mathematical Analysis and Numerical Methods for Science and Technology, Volume 2: Functional and variational methods. Springer, Berlin

[6] Dobrowolski, M. (2006) Angewandte Funktionalanalysis. Springer, Berlin

[7] Freeden, W., Maier, T. (2002) On Multiscale Denoising of Spherical Functions: Basic Theory and Numerical Aspects. Electron. Trans. Numer. Anal. 14: 56–78

[8] Freeden, W., Michel, V. (2004) Multiscale Potential Theory (with Applications to Geoscience). Birkhäuser Verlag, Boston, Basel, Berlin

[9] Grothaus, M., Raskop, T. (2006) On the Oblique Boundary Problem with a Stochastic Inhomogeneity. Stochastics 78(4): 233–257

[10] Grothaus, M., Raskop, T. (2009) The Outer Oblique Boundary Problem of Potential Theory. Num. Funct. Anal. Optim. 30(7-8): 1–40

[11] Gilbarg, D., Trudinger, N.S. (1998) Elliptic Partial Differential Equations of Second Order. Springer, Berlin

[12] Gutting, M. (2008) Fast Multipole Methods for Oblique Derivative Problems. Shaker Verlag, Aachen

[13] Miranda, C. (1970) Partial Differential Equations of Elliptic Type. Springer, Berlin

[14] Reed. M., Simon, B. (1972) Methods of Modern Mathematical Physics, Volume 1: Functional Analysis. Academic Press, New York-San Francisco-London

[15] Rozanov, Y., Sanso, F. (2001) The Analysis of the Neumann and Oblique Derivative Problem: The Theory of Regularization and its Stochastic Version. J. Geod. 75(7-8): 391–398

[16] Rozanov, Y., Sanso, F. (2002a) On the Stochastic Versions of Neumann and Oblique Derivative Problems. Stochastics Rep. 74(1-2): 371–391

[17] Rozanov, Y., Sanso, F. (2002b) The Analysis of the Neumann and Oblique Derivative Problem: Weak Theory. In Geodesy: Challenge of the 3rd Millenium. Springer, Berlin

Martin Grothaus and Thomas Raskop
Functional Analysis and Stochastic Analysis Group
Fachbereich Mathematik
University of Kaiserslautern
D-67663 Kaiserslautern, Germany
e-mail: grothaus@mathematik.uni-kl.de

Handbook of Mathematical Geodesy
Geosystems Mathematics, 517–560

About the Importance of the Runge–Walsh Concept for Gravitational Field Determination

Matthias Augustin, Willi Freeden, and Helga Nutz

Abstract. On the one hand, the Runge–Walsh theorem plays a particular role in physical geodesy, because it allows to guarantee a uniform approximation of the Earth's gravitational potential within arbitrary accuracy by a harmonic function showing a larger analyticity domain. On the other hand, there are some less transparent manifestations of the Runge–Walsh context in the geodetic literature that must be clarified in more detail. Indeed, some authors make the attempt to apply the Runge–Walsh idea to the gravity potential of a rotating Earth instead of the gravitational potential in non-rotating status. Others doubt about the convergence of series expansions approximating the Earth's gravitational potential inside the *whole* outer space of the actual Earth.

The goal of this contribution is to provide the conceptual setup of the Runge–Walsh theorem such that geodetic expectation as well as mathematical justification become transparent and coincident. Even more, the Runge–Walsh concept in form of generalized Fourier expansions corresponding to certain harmonic trial functions (e.g., mono- and/or multi-poles) will be extended to the topology of Sobolev-like reproducing kernel Hilbert spaces thereby avoiding any need of (numerical) integration in the occurring spline solution process.

Keywords. Runge–Walsh theorem in physical geodesy, theoretical background, constructive solution concepts, (discrete) boundary value problems, generalized Fourier series, spline interpolation.

1. Introduction

In the theory of harmonic functions, a result first motivated by C. Runge [62] in one-dimensional complex analysis and later generalized, e.g., by J.L. Walsh [70], I.N. Vekua [69], and L. Hörmander [40] to potential theory in three-dimensional Euclidean space $\mathbb{R}^3$ is of basic interest. For geodetically relevant obligations (see,

e.g., [33, 44, 54, 56, 63], and the references therein) it may be formulated in accordance with [55]:

Geodetic version of the Runge–Walsh theorem. *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region, i.e., a bounded region $\mathcal{G} \subset \mathbb{R}^3$ dividing $\mathbb{R}^3$ uniquely into the inner space $\mathcal{G}$ and the outer space $\mathcal{G}^c = \mathbb{R}^3 \backslash \overline{\mathcal{G}}$, $\overline{\mathcal{G}} = \mathcal{G} \cup \partial\mathcal{G}$, such that $\mathcal{G}$ contains the origin and the boundary $\partial\mathcal{G}$ is an orientable smooth Lipschitzian manifold of dimension 2. Any harmonic function in $\mathcal{G}^c$ that is regular at infinity can be approximated by a function that is harmonic outside an arbitrarily given Runge (i.e., in geodetic nomenclature sometimes called Bjerhammar) ball $\mathcal{A} \Subset \mathcal{G}$, i.e., $\mathcal{A} \subset \mathcal{G}$ with $\operatorname{dist}(\overline{\mathcal{A}}, \partial\mathcal{G}) > 0$ (see Figure 1, right illustration) in the sense that, for any given $\varepsilon > 0$, the absolute error between the two functions is smaller than ε for all points outside and on any closed surface completely surrounding $\partial\mathcal{G}$ in its outer space. The value ε may be arbitrarily small, and the surrounding surface may be arbitrarily close to the surface $\partial\mathcal{G}$.*

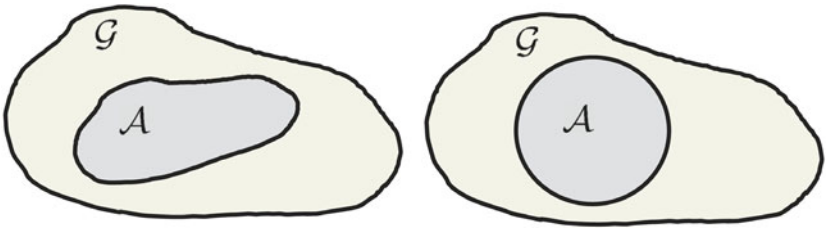


FIGURE 1. The geometric situation of the Runge–Walsh theorem (with $\mathcal{A}$ an arbitrary regular region such that $\mathcal{A} \Subset \mathcal{G}$ (left) and $\mathcal{A}$ an inner Runge (i.e., Bjerhammar) ball (right)).

Obviously, the *Runge–Walsh theorem* in the preceding formulation (with $\mathcal{G}$, e.g., chosen as the interior of the actual Earth) represents a pure existence theorem. It guarantees only the existence of an approximating function and does not provide a constructive method to find it. Nothing is said about the approximation procedure and the computational structure and methodology of the approximation. The theorem merely describes the theoretical background for the approximation of a potential by another potential defined on a larger harmonicity domain, i.e., the Runge region outside the sphere $\partial\mathcal{A}$.

The situation, however, is completely different if spherical geometries are exclusively involved in the Runge concept. Assuming that both $\mathcal{A}, \mathcal{G}$ are concentric balls around the origin with $\mathcal{A} \Subset \mathcal{G}$, a constructive approximation of a potential in the outer space $\mathcal{G}^c$ is available, e.g., by outer harmonic (orthogonal) expansions (see, e.g., [30, 41, 47, 48, 60, 71]). More concretely, within the classical context of a twofold spherical configuration, a constructive version of the Runge–Walsh theorem can be guaranteed by finite truncations of Fourier expansions in terms of outer harmonics, where the $L^2(\partial\mathcal{G})$-convergence of the Fourier series implies uni-

form convergence on any point set $\mathcal{K} \Subset \mathcal{G}^c$. The Fourier coefficients are obtained by integration over the sphere $\partial\mathcal{G}$. The gravitational potential is available (in spectral sense) by tables of the Fourier coefficients. Nowadays, in fact, outer harmonic expansions constitute the conventional geodetic tools in globally reflected approximation of the Earth's gravitational potential and its observables.

From a superficial point of view, one could suggest that the standard approximation by truncated series expansions in terms of outer harmonics is closely related to spherical geometries $\partial\mathcal{A}, \partial\mathcal{G}$. The purpose of our work, however, is to show that the essential steps to a constructive Fourier approach can be extended to any regular, i.e., *not-necessarily spherical* region $\mathcal{G}$ and to any regular, i.e., *not-necessarily spherical* Runge region $\mathcal{A} \Subset \mathcal{G}$ (see Figure 1, left illustration). As a matter of fact, the Runge–Walsh approach enables us to avoid any calamities with the convergence to the gravitational potential by the generalized Fourier series for arbitrary sets $\mathcal{K} \Subset \mathcal{G}^c$. In analogy to the spherical case, however, it likewise does not help to specify convergence inside $\mathcal{A}^c \backslash \overline{\mathcal{G}^c}$, so that any attempts (see [3]) to reduce gravitational information via infinite Fourier series downward from $\partial\mathcal{G}$ to the surface $\partial\mathcal{A}$ are not justifiable by the Runge–Walsh framework.

In summary, the Runge–Walsh concept as presented in this work reflects constructive approximation capabilities of the Earth's gravitational (and *not* gravity) potential even if geoscientifically realistic (i.e., not necessarily spherical) geometries come into play.

Mathematically, it should be pointed out that the main techniques for assuring the not-necessarily spherical results are the limit and jump relations and their formulations of potential theory in the Hilbert space nomenclature of $(L^2(\partial\mathcal{G}), \|\cdot\|_{L^2(\partial\mathcal{G})})$. The special function systems for use in constructive Runge–Walsh theorems are manifold. Moreover, all harmonic functions systems that are regular at infinity can be taken into account, whose restrictions to the boundary $\partial\mathcal{G}$ of a regular region $\mathcal{G}$ form an $L^2(\partial\mathcal{G})$-complete system. For numerical efficiency, however, we restrict ourselves to certain mono-pole and multi-pole configurations.

2. Special function systems

Spherical harmonics are the functions most commonly used in geosciences to represent scalar fields on the unit sphere $\Omega \subset \mathbb{R}^3$. They are used extensively in the gravitational and magnetic applications involving Laplace's equation. The introduction of (scalar) spherical harmonics and the derivation of some important properties can be found, e.g., in [16, 26, 59], and the references therein.

2.1. Spherical harmonics

Let $H_n : \mathbb{R}^3 \to \mathbb{R}$ be a homogeneous and harmonic polynomial of degree $n \in \mathbb{N}_0$, i.e., $H_n(\lambda x) = \lambda^n H_n(x)$, $\lambda \in \mathbb{R}$, and $\Delta H_n(x) = 0$, $x \in \mathbb{R}^3$. Then, the restriction $Y_n = H_n|_\Omega$ of H_n to the unit sphere Ω in $\mathbb{R}^3$ is called a (*scalar*) *spherical harmonic* (*of degree* n). The space of all spherical harmonics of degree n is denoted by

$Harm_n(\Omega)$. The spherical harmonics of degree n form a space of dimension $2n+1$, i.e., $\dim(Harm_n(\Omega)) = 2n+1$, $n \in \mathbb{N}_0$. Using the standard method of separation by spherical coordinates and observing the homogeneity, we have $H_n(x) = r^n Y_n(\xi)$, for $r = |x|$, $\xi = \frac{x}{|x|} \in \Omega$. From the identity $\left(\frac{1}{r^2}\frac{d}{dr}r^2\frac{d}{dr}\right) r^n = n(n+1)r^{n-2}$ it follows, in connection with the harmonicity of H_n, that $0 = \Delta H_n(x) = r^{n-2}n(n+1)Y_n(\xi)+r^{n-2}\Delta^*_\xi Y_n(\xi)$. As a consequence, we can state that any spherical harmonic $Y_n \in Harm_n(\Omega)$, $n \in \mathbb{N}_0$, is an infinitely often differentiable eigenfunction of the Beltrami operator corresponding to the eigenvalue $-n(n+1)$. More explicitly, $\Delta^* Y_n(\xi) = -n(n+1)Y_n(\xi)$, $\xi \in \Omega$. Conversely, every infinitely often differentiable eigenfunction of the Beltrami operator with respect to the eigenvalue $-n(n+1)$ constitutes a spherical harmonic of degree n. Using Green's formulas for the Beltrami operator, this implies that spherical harmonics of different degrees are orthogonal with respect to the $L^2(\Omega)$-inner product, i.e.,

$$(Y_n, Y_m)_{L^2(\Omega)} = \int_\Omega Y_n(\xi)Y_m(\xi)\, dS(\xi) = 0, \quad n \neq m, \tag{1}$$

where dS is the surface element in $\mathbb{R}^3$. The Gram–Schmidt method allows the orthonormalization of any set of linearly independent spherical harmonics of degree n with respect to the $L^2(\Omega)$-inner product. Throughout this work, a set $\{Y_{n,k}\}_{k=1,\dots,2n+1} \subset Harm_n(\Omega)$ always denotes an orthonormal basis of $Harm_n(\Omega)$.

The *Legendre polynomials* are one-dimensional orthogonal polynomials that are of great importance when treating spherical harmonics. A polynomial $P_n : [-1,1] \to \mathbb{R}$ of degree $n \in \mathbb{N}_0$ is called *Legendre polynomial* (*of degree* n) if

- $\int_{-1}^{1} P_n(t)P_m(t)dt = 0,\ n \neq m,$
- $P_n(1) = 1.$

The Legendre polynomials are uniquely determined by these properties. They have the explicit representation

$$P_n(t) = \sum_{s=0}^{[n/2]} (-1)^s \frac{(2n-2s)!}{2^n(n-2s)!(n-s)!s!} t^{n-2s}, \quad t \in [-1,+1], \tag{2}$$

where we use the abbreviation

$$\left[\frac{n}{2}\right] = \frac{1}{2}\left(n - \frac{1}{2}(1-(-1)^n)\right). \tag{3}$$

The zonal function $\eta \mapsto P_n(\xi \cdot \eta)$, $\eta \in \Omega$, is the only normalized (i.e., $P_n(\xi \cdot \xi) = 1$) spherical harmonic of degree n that is invariant with respect to orthogonal transformations which leave $\xi \in \Omega$ fixed. A consequence of this fact is the addition theorem (see, e.g., [58] and [7] for different approaches) which states the close relation of (univariate) Legendre polynomials to spherical harmonics: For $n \in \mathbb{N}_0$ and $\xi, \eta \in \Omega$, we have

$$\sum_{k=1}^{2n+1} Y_{n,k}(\xi)Y_{n,k}(\eta) = \frac{2n+1}{4\pi} P_n(\xi \cdot \eta). \tag{4}$$

As a direct consequence, estimates for Legendre polynomials and spherical harmonics are derivable by standard arguments. For $n \in \mathbb{N}_0$ and $k = 1, \ldots, 2n+1$, we have $|P_n(t)| \leq 1,\ t \in [-1,1]$. It is well-known that

$$\sum_{n=0}^{\infty} P_n(t) h^n = \frac{1}{\sqrt{1+h^2-2ht}}, \quad t \in [-1,1], \quad h \in (-1,1). \tag{5}$$

2.2. Mono- and multi-poles

Of special importance for our considerations is the so-called *fundamental solution for the Laplace operator*

$$G(\Delta; |x-y|) = \frac{1}{4\pi} \frac{1}{|x-y|}, \quad x, y \in \mathbb{R}^3, \quad x \neq y. \tag{6}$$

This function can be regarded as a *mono-pole at* y, it represents the gravitational potential between a mass point at y and a point x in the exterior (of the mass). An easy manipulation involving (5) yields

$$\frac{1}{|x-y|} = \frac{1}{|y|} \left(1 + \left(\frac{|x|}{|y|} \right)^2 - 2 \frac{|x|}{|y|} \xi \cdot \eta \right)^{-\frac{1}{2}}, \tag{7}$$

where $x, y \in \mathbb{R}^3$, $|x| < |y|$, and $\xi = \frac{x}{|x|}$, $\eta = \frac{y}{|y|}$. With $t = \xi \cdot \eta$ and $h = \frac{|x|}{|y|}$, this implies the series expansion

$$\frac{1}{|x-y|} = \frac{1}{|y|} \sum_{n=0}^{\infty} \left(\frac{|x|}{|y|} \right)^n P_n(\xi \cdot \eta). \tag{8}$$

Moreover, further calculations show that

$$\frac{1}{|x-y|} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} |x|^n (\xi \cdot \nabla_y)^n \frac{1}{|y|}, \tag{9}$$

where

$$\frac{(-1)^n}{n!} (\xi \cdot \nabla_y)^n \frac{1}{|y|} = \frac{P_n(\xi \cdot \eta)}{|y|^{n+1}}, \quad n \in \mathbb{N}_0. \tag{10}$$

The identity (10) is known as *Maxwell's representation formula*. As $y \mapsto |y|^{-1}$, $y \neq 0$, is (apart from a multiplicative constant) the fundamental solution for the Laplace operator, this representation tells us that the Legendre polynomials may be obtained by repeated differentiation of the fundamental solution in the radial direction of ξ. The potential on the right-hand side may be regarded as the potential of a *multi-pole of order* n *with the axis* ξ *at the origin.*

From any textbook about spherical harmonics (see, e.g., [16, 26]) we know that the system $\{Y_{n,k}\}_{\substack{n \in \mathbb{N}_0; \\ k=1,\ldots,2n+1}}$ is closed in $C^{(0)}(\Omega)$. That is, for any given $\varepsilon > 0$ and each $F \in C^{(0)}(\Omega)$, there exist coefficients $a_{n,k} \in \mathbb{R}$ such that

$$\left\| F - \sum_{n=0}^{N} \sum_{k=1}^{2n+1} a_{n,k} Y_{n,k} \right\|_{C^{(0)}(\Omega)} \leq \varepsilon. \tag{11}$$

This result also enables us to verify the closure in the Hilbert space $L^2(\Omega)$ with respect to the norm $\|\cdot\|_{L^2(\Omega)}$. That is, for any given $\varepsilon > 0$ and each $F \in L^2(\Omega)$, there exist coefficients $b_{n,k} \in \mathbb{R}$ such that

$$\left\| F - \sum_{n=0}^{N} \sum_{k=1}^{2n+1} b_{n,k} Y_{n,k} \right\|_{L^2(\Omega)} \leq \varepsilon. \tag{12}$$

Note that, in a Hilbert space, the closure property of a function system is equivalent to the *completeness property* (see, e.g., [6]). The completeness means: If F of class $L^2(\Omega)$ has vanishing *Fourier coefficients*

$$F^\wedge(n,k) = \int_\Omega F(\eta) Y_{n,k}(\eta) d\omega(\eta) = 0, \tag{13}$$

for all $n \in \mathbb{N}_0, k = 1, \ldots, 2n+1$, then $F = 0$ in $L^2(\Omega)$-sense. In other words, F is uniquely determined by its Fourier coefficients. Furthermore, we can state a constructive version of (12) in terms of the Fourier expansion, i.e., for all F of class $L^2(\Omega)$,

$$\lim_{N\to\infty} \left\| F - \sum_{n=0}^{N} \sum_{k=1}^{2n+1} F^\wedge(n,k) Y_{n,k} \right\|_{L^2(\Omega)} = 0. \tag{14}$$

Next we consider a sphere Ω_R around the origin with radius $R > 0$. By virtue of the isomorphism $\xi \mapsto R\xi$, $\xi \in \Omega$, we can assume a function $F : \Omega_R \to \mathbb{R}$ to be reduced to the unit sphere Ω. Obviously, an $L^2(\Omega)$-orthonormal system of spherical harmonics forms an $L^2(\Omega_R)$-orthogonal system. More explicitly,

$$(Y_{n,k}, Y_{p,q})_{L^2(\Omega_R)} = \int_{\Omega_R} Y_{n,k}\left(\frac{y}{|y|}\right) Y_{p,q}\left(\frac{y}{|y|}\right) dS(y) = R^2 \delta_{n,p} \delta_{k,q}. \tag{15}$$

Introducing the system $Y^R_{n,k}(x) = \frac{1}{R} Y_{n,k}\left(\frac{x}{|x|}\right)$, $x \in \Omega_R$, we get an orthonormal basis $\{Y^R_{n,k}\}_{n\in\mathbb{N}_0;\, k=1,\ldots,2n+1}$ of the space $L^2(\Omega_R)$.

(a) The functions

$$H^R_{n,k}(x) = \left(\frac{|x|}{R}\right)^n Y^R_{n,k}(x), \quad x \in \mathbb{R}^3, \tag{16}$$

for $n \in \mathbb{N}_0$, $k = 1, \ldots, 2n+1$, are called *inner harmonics* (*of degree n and order k*).

(b) The functions

$$H^R_{-n-1,k}(x) = \left(\frac{R}{|x|}\right)^{n+1} Y^R_{n,k}(x), \quad x \in \mathbb{R}^3 \backslash \{0\}, \tag{17}$$

for $n \in \mathbb{N}_0$, $k = 1, \ldots, 2n+1$, are called *outer harmonics* (*of degree n and order k*).

We let $\mathcal{B}_R(0) = \{x \in \mathbb{R}^3, |x| < R\}$ be the ball around the origin with radius R. Then it is not difficult to see that the inner harmonics are of class $Pot(\mathcal{B}_R(0))$, i.e.,

(i) $H_{n,k}^R$ is a member of $C^{(2)}(\mathcal{B}_R(0))$,
(ii) $H_{n,k}^R$ satisfies $\Delta H_{n,k}^R(x) = 0$, $x \in \mathcal{B}_R(0)$.

Furthermore, the inner harmonics show the following "boundary behavior" on $\Omega_R = \partial\mathcal{B}_R(0)$:

(iii) $H_{n,k}^R\big|_{\Omega_R} = Y_{n,k}^R$,
(iv) $\left(H_{n,k}^R, H_{p,q}^R\right)_{L^2(\Omega_R)} = \delta_{n,p}\delta_{k,q}$.

Analogously, the outer harmonics represent those functions that are harmonic in the exterior of Ω_R and regular at infinity, and which coincide with the spherical harmonics $Y_{n,k}^R$ on the boundary Ω_R. In shorthand nomenclature, the outer harmonics are of class $Pot(\mathbb{R}^3\backslash\overline{\mathcal{B}_R(0)})$, i.e.,

(i) $H_{-n-1,k}^R$ is a member of $C^{(2)}(\mathbb{R}^3\backslash\overline{\mathcal{B}_R(0)})$,
(ii) $H_{-n-1,k}^R$ satisfies $\Delta H_{-n-1,k}^R(x) = 0$, $x \in \mathbb{R}^3\backslash\overline{\mathcal{B}_R(0)}$,
(iii) H_{-n-1}^R is regular at infinity, i.e., $H_{-n-1,k}^R(x) = O(|x|^{-1})$, $|x| \to \infty$.

Furthermore, the outer harmonics show the following "boundary behavior" on $\Omega_R = \partial\mathcal{B}_R(0)$:

(iv) $H_{-n-1,k}^R\big|_{\Omega_R} = Y_{n,k}^R$,
(v) $\left(H_{-n-1,k}^R, H_{-p-1,q}^R\right)_{L^2(\Omega_R)} = \delta_{n,p}\ \delta_{k,q}$.

Moreover, it should be noted that an inner harmonic $H_{n,k}^R$ is related to its corresponding outer harmonic $H_{-n-1,k}^R$ in the following way:

$$H_{-n-1,k}^R(x) = \left(\frac{R}{|x|}\right)^{2n+1} H_{n,k}^R(x) = \frac{R}{|x|} H_{n,k}^R\left(\frac{R^2}{|x|^2}x\right), \quad x \in \mathbb{R}^3 \setminus \{0\}. \tag{18}$$

This observation leads us to the following interpretation: The mapping

$$x \mapsto \check{x}_R = \left(\frac{R}{|x|}\right)^2 x, \quad x \in \overline{\mathcal{B}_R(0)}\backslash\{0\}, \tag{19}$$

transforms $\mathcal{B}_R(0)\backslash\{0\}$ into $\mathbb{R}^3\backslash\overline{\mathcal{B}_R(0)}$ and $\Omega_R = \partial\mathcal{B}_R(0)$ onto itself. Referring to Figure 2, we observe that the two triangles with edges $(\check{x}_R, y, 0)$ and $(x, y, 0)$ are similar whenever $y \in \Omega_R$. Furthermore, the ratios $\frac{|x|}{|y|}$ and $\frac{|y|}{|\check{x}_R|}$ are equal, provided that $y \in \Omega_R$.

On the one hand, for $x = |x|\xi$, $\xi \in \Omega$, and $y = |y|\eta$, $R = |y|$, $\eta \in \Omega$, we have

$$|x - y|^2 = x^2 + y^2 - 2x \cdot y = |x|^2 + R^2 - 2|x|R\ \xi \cdot \eta. \tag{20}$$

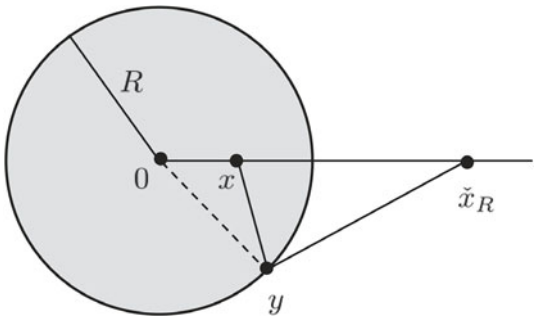


FIGURE 2. The inversion $x \mapsto \check{x}_R$ with respect to the sphere Ω_R.

On the other hand, we see that

$$\left(\frac{|x|}{R}\right)^2 \left|\frac{R^2}{|x|^2}x - y\right|^2 = \frac{|x|^2}{R^2}\left(\frac{R^4}{|x|^4}|x|^2 + R^2 - 2\frac{R^2}{|x|^2}x \cdot y\right)$$
$$= |x|^2 + R^2 - 2|x|R\,\xi \cdot \eta. \tag{21}$$

For all $y \in \Omega_R$ and $x \in \mathcal{B}_R(0)$, we have

$$|x - y| = \left(\frac{|x|}{R}\right)^2 |\check{x}_R - y|. \tag{22}$$

After these preparations about the inversion of points with respect to a sphere Ω_R, $R > 0$, we are able to discuss the *Kelvin transform* (see, e.g., [41]): Assume that U is of class $C^{(2)}(\mathcal{G}), \mathcal{G} \subset \mathbb{R}^3\backslash\{0\}$ open. Let $\check{\mathcal{G}}_R$ be the image of $\mathcal{G}$ under the inversion $x \mapsto \check{x}_R = R^2\ |x|^{-2}\ x$. Denote by $\check{U} = \mathcal{K}^R[U]\colon \check{\mathcal{G}}_R \to \mathbb{R}$, with

$$\check{U}(x) = \mathcal{K}^R[U](x) = \frac{R}{|x|}\ U\left(\left(\frac{R}{|x|}\right)^2 x\right), \tag{23}$$

the Kelvin transform of U with respect to $\Omega_R = \partial\mathcal{B}_R(0)$. Then

$$\Delta\check{U}(x) = \left(\frac{R}{|x|}\right)^5 \Delta U\left(\left(\frac{R}{|x|}\right)^2 x\right). \tag{24}$$

As a consequence, the outer harmonics are obtainable by the Kelvin transform $\mathcal{K}^R$ from their inner counterparts, and vice versa. More precisely, we have

$$H^R_{-n-1,k}(x) = \mathcal{K}^R\left[H^R_{n,k}\right](x), \tag{25}$$
$$H^R_{n,k}(x) = \mathcal{K}^R\left[H^R_{-n-1,k}\right](x), \tag{26}$$

for $x \in \mathbb{R}^3 \setminus \{0\}$.

3. Runge–Walsh closure theorems

We begin our considerations with a result on the special function system of outer harmonics (see [8, 15, 23]).

Lemma 3.1 (Linear Independence). *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region such that $\underline{R} < \inf_{x\in\partial\mathcal{G}} |x|$, i.e., $\mathcal{B}_{\underline{R}}(0) \Subset \mathcal{G}$. Then the sequence $\left\{H^{\underline{R}}_{-n-1,j}\right\}_{n\in\mathbb{N}_0;\, j=1,\ldots,2n+1}$ is a subsystem of Pot$(\mathbb{R}^3\backslash\overline{\mathcal{B}_{\underline{R}}(0)})$, and its restriction to $\partial\mathcal{G}$*

$$\left\{H^{\underline{R}}_{-n-1,j}\Big|_{\partial\mathcal{G}}\right\}_{\substack{n\in\mathbb{N}_0;\\ j=1,\ldots,2n+1}} \tag{27}$$

forms a linearly independent system.

Proof. In order to verify the statement under the assumption $\mathcal{B}_{\underline{R}}(0) \Subset \mathcal{G}$ we have to derive that, for any linear combination H of the form

$$H = \sum_{n=0}^{N}\sum_{j=1}^{2n+1} a_{n,j} H^{\underline{R}}_{-n-1,j}, \tag{28}$$

the condition $H|_{\partial\mathcal{G}} = 0$ implies $a_{0,1} = \cdots = a_{N,1} = \cdots = a_{N,2N+1} = 0$. Indeed, from the uniqueness theorem of the exterior Dirichlet problem (see, e.g., [15]) we know that $H|_{\partial\mathcal{G}} = 0$ yields $H|_{\overline{\mathcal{G}^c}} = 0$. Therefore, for every sphere with radius $\overline{R} > \sup_{x\in\partial\mathcal{G}} |x|$ around the origin 0, it follows that

$$\int_{\Omega_{\overline{R}}} H^{\underline{R}}_{-n-1,j}(x) H(x)\, dS(x) = 0 \tag{29}$$

for $n = 0, \ldots, N, j = 1, \ldots, 2n+1$. Inserting (28) into (29) yields, in connection with the well-known completeness property of the spherical harmonics (see, e.g., [16]), that $a_{n,j} = 0$ for all $n = 0, \ldots, N$, $j = 1, \ldots, 2n+1$, as required. □

3.1. L^2-closure and truncated Fourier series expansions

Next, our purpose is to prove completeness and closure theorems (see [8]).

Theorem 3.2 (Completeness). *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region such that $\mathcal{B}_{\underline{R}}(0) \Subset \mathcal{G}$. Then the restriction of $\left\{H^{\underline{R}}_{-n-1,j}\right\}_{n\in\mathbb{N}_0, j=1,\ldots,2n+1}$ to the surface $\partial\mathcal{G}$*

$$\left\{H^{\underline{R}}_{-n-1,j}\Big|_{\partial\mathcal{G}}\right\}_{\substack{n\in\mathbb{N}_0;\\ j=1,\ldots,2n+1}} \tag{30}$$

is complete in $L^2(\partial\mathcal{G})$.

Proof. Suppose that $F \in L^2(\partial\mathcal{G})$ satisfies

$$\begin{aligned}(F, H^{\underline{R}}_{-n-1,j}\Big|_{\partial\mathcal{G}})_{L^2(\partial\mathcal{G})} &= \int_{\partial\mathcal{G}} F(y) H^{\underline{R}}_{-n-1,j}(y)\, dS(y)\\ &= 0,\end{aligned} \tag{31}$$

$n \in \mathbb{N}_0$, $j = 1, \ldots, 2n+1$. We have to show that $F = 0$ in $L^2(\partial\mathcal{G})$.

We know that the fundamental solution $G(\Delta; |\cdot - y|) = (4\pi|\cdot - y|)^{-1}$ of the Laplace operator given in terms of its spherical harmonic series expansion (see, e.g., [15])

$$G(\Delta; |x-y|) = \frac{1}{4\pi}\frac{1}{|x-y|} = \sum_{n=0}^{\infty} \frac{1}{2n+1}\,\frac{|x|^n}{|y|^{n+1}} \sum_{j=1}^{2n+1} Y_{n,j}(\xi) Y_{n,j}(\eta), \tag{32}$$

$x = |x|\xi$, $y = |y|\eta$, is analytic in the variable x on the ball $\mathcal{B}_{\underline{R}}(0)$ around the origin 0 with radius $\underline{R}$, if y is a member of $\mathbb{R}^3 \setminus \mathcal{B}_{\underline{R}}(0)$. For all $x \in \mathcal{B}_{\underline{R}}(0)$, we thus find by virtue of (31)

$$\begin{aligned} P(x) &= \int_{\partial\mathcal{G}} F(y) G(\Delta; |x-y|)\; dS(y) \\ &= \sum_{n=0}^{\infty} \frac{\underline{R}}{2n+1} \sum_{j=1}^{2n+1} H_{n,j}^{\underline{R}}(x) \int_{\partial\mathcal{G}} F(y) H_{-n-1,j}^{\underline{R}}(y)\; dS(y) = 0. \end{aligned} \tag{33}$$

Analytic continuation shows that the single-layer potential P vanishes in $\mathcal{G}$. In other words, the equations

$$P(x - \tau\nu(x)) = 0, \tag{34}$$

$$\frac{\partial P}{\partial \nu}(x - \tau\nu(x)) = 0 \tag{35}$$

hold true for all $x \in \partial\mathcal{G}$ and all sufficiently small $\tau > 0$, where $\nu(x)$ is the outer unit normal at the point x. Therefore, using the L^2-limit and jump relations (see [8]), we obtain

$$\lim_{\tau\to 0+} \int_{\partial\mathcal{G}} \Big| P\left(x + \tau\nu(x)\right) \Big|^2\; dS(x) = 0, \tag{36}$$

$$\lim_{\tau\to 0+} \int_{\partial\mathcal{G}} \left| \frac{\partial P}{\partial \nu}(x + \tau\nu(x)) + F(x) \right|^2\; dS(x) = 0, \tag{37}$$

and

$$\lim_{\tau\to 0+} \int_{\partial\mathcal{G}} \left| \frac{\partial P}{\partial \nu}(x) + \frac{1}{2}F(x) \right|^2\; dS(x) = 0. \tag{38}$$

The limit in the last equation can be omitted. Hence, the identity (38) can also be understood as

$$\int_{\partial\mathcal{G}} F(y) \frac{\partial}{\partial \nu(x)} G(\Delta; |x-y|)\; dS(y) = -\frac{1}{2}F(x), \tag{39}$$

in the sense of $L^2(\partial\mathcal{G})$. The left-hand side of (39) constitutes a continuous function (see, e.g., [15]). Thus, the function F is continuous itself. For continuous functions, however, the classical limit and jump relations are valid:

$$\lim_{\tau\to 0+} P(x + \tau\nu(x)) = 0, \qquad x \in \partial\mathcal{G}, \tag{40}$$

$$\lim_{\tau\to 0+} \frac{\partial P}{\partial \nu}(x + \tau\nu(x)) = -F(x), \quad x \in \partial\mathcal{G}. \tag{41}$$

Consequently, the uniqueness theorem of the exterior Dirichlet problem (see, e.g., [41]) shows us that $P(x) = 0$ for all $x \in \mathbb{R}^3 \backslash \overline{\mathcal{G}^c}$. But this means that $F = 0$ on the surface $\partial\mathcal{G}$, as required. □

From approximation theory (see, e.g., [6]) we know that the properties of completeness and closure are equivalent in a Hilbert space such as $L^2(\partial\mathcal{G})$. The equivalence leads us to the following statement

Corollary 3.3 (Closure). *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region such that $\mathcal{B}_{\underline{R}}(0) \Subset \mathcal{G}$. Then the system $\left\{ H^{\underline{R}}_{-n-1,j}\big|_{\partial\mathcal{G}} \right\}_{n\in\mathbb{N}_0, j=1,\dots,2n+1}$ is closed in $L^2(\partial\mathcal{G})$, i.e., for any given $F \in L^2(\partial\mathcal{G})$ and arbitrary $\varepsilon > 0$ there exist coefficients $a_{n,j}$, $n = 0, \dots, N$, $j = 1, \dots, 2n+1$, constituting the linear combination*

$$F_N = \sum_{n=0}^{N} \sum_{j=1}^{2n+1} a_{n,j} H^{\underline{R}}_{-n-1,j}\Big|_{\partial\mathcal{G}} \tag{42}$$

such that

$$\|F - F_N\|_{L^2(\partial\mathcal{G})} \leq \varepsilon \,. \tag{43}$$

Based on our results on outer harmonics, i.e., multi-pole expansions, a large variety of countable systems of potentials can be shown to possess the L^2-closure property on $\partial\mathcal{G}$. Probably best known are *mono-poles,* (i.e., fundamental solutions of the Laplace operator). Their $L^2(\partial\mathcal{G})$-closure can be adequately described by using the concept of *fundamental systems*, which should be recapitulated briefly (see, e.g., [8, 15, 23]).

Definition 3.4 (Fundamental System). Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be regular regions satisfying the "Runge condition" $\mathcal{A} \Subset \mathcal{G}$ (cf. Figure 3). A point set $\mathcal{Y} = \{y_n\}_{n=0,1,\dots} \subset \mathcal{A}$ (with $y_n \neq y_l$ for $n \neq l$) is called a *fundamental system in* $\mathcal{G}$, if for each $Q \in Pot(\mathcal{G})$, i.e., for each $Q \in C^{(2)}(\mathcal{G})$ with $\Delta Q = 0$ in $\mathcal{G}$, the condition $Q(y_n) = 0$ for all $n = 0, 1, \dots$ implies $Q = 0$ in $\mathcal{G}$.

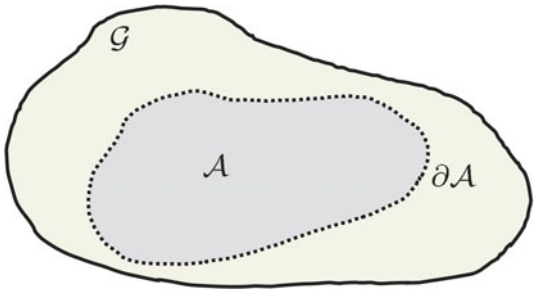


FIGURE 3. Illustration of the positioning of a fundamental system on $\partial\mathcal{A}$ in $\mathcal{G}$.

Two examples of fundamental systems in $\mathcal{G}$ should be listed: $\mathcal{Y} = \{y_n\}_{n=0,1,\dots} \subset \mathcal{A}$ is a fundamental system in $\mathcal{G}$ if

(1) $\mathcal{Y}$ is a dense subset of points in $\mathcal{A} \Subset \mathcal{G}$,
(2) $\mathcal{Y}$ is a dense subset of points on the boundary $\partial\mathcal{A}$ with $\mathcal{A} \Subset \mathcal{G}$ (cf. Figure 3).

Theorem 3.5. *Let $\mathcal{G}$ be a regular region. Then the following statement is valid: For every fundamental system $\mathcal{Y} = \{y_n\}_{n=0,1,\ldots} \subset \mathcal{A}$ in $\mathcal{G}$, $\mathcal{A} \Subset \mathcal{G}$, the system*

$$\left\{ G(\Delta; |\cdot - y_n|)\Big|_{\partial\mathcal{G}} \right\}_{n=0,1,\ldots} \tag{44}$$

is closed in $L^2(\partial\mathcal{G})$.

Proof. Since $y_n \neq y_m$ for all $n \neq m$, it immediately follows that the system $\{G(\Delta; |\cdot - y_n|)\}_{n=0,1,\ldots}$ is linearly independent.

Our purpose is to verify the completeness of the system (44) in $L^2(\partial\mathcal{G})$. To this end, we consider a function $F \in L^2(\partial\mathcal{G})$ with

$$\int_{\partial\mathcal{G}} F(x) G(\Delta; |x - y_n|)\, dS(x) = 0, \quad n \in \mathbb{N}_0. \tag{45}$$

We have to prove that $F = 0$ in $L^2(\partial\mathcal{G})$. We consider the single-layer potential Q given by

$$Q(y) = \int_{\partial\mathcal{G}} F(x) G(\Delta; |x - y|)\, dS(x). \tag{46}$$

Since Q is harmonic in $\mathcal{G}$, the properties of the fundamental system $\{y_n\}_{n=0,1,\ldots}$ in $\mathcal{G}$ imply that $Q(y) = 0$ for all $y \in \mathcal{G}$. Then, the same arguments as given in the proof of Theorem 3.2 guarantee that $F = 0$ in the sense of $L^2(\partial\mathcal{G})$, as desired. □

Besides the outer harmonics, i.e., multi-poles (see Corollary 3.3) and the mass (single-)poles (see Theorem 3.5), there exist a variety of countable systems of potentials showing the properties of completeness and closure in $L^2(\partial\mathcal{G})$. Many systems, however, are much more difficult to handle numerically (for instance, the ellipsoidal systems of Lamé or Mathieu functions). Although they are originally meant for particular use in series expansions corresponding to ellipsoidal boundaries (see, e.g., [34]), they can be likewise taken in *any* regular (i.e., also in spherical) Runge–Walsh framework. Nonetheless, our particular aim here is to show that all constructive approximation in physical geodesy can be provided by mono-pole and multi-pole (i.e., outer harmonics) conglomerates even if the Runge region inside is an ellipsoid $\mathcal{G}$ such that convergence problems do not occur in $\mathcal{G}$ if the approximation method is suitably organized.

Seen from numerical point of view it is preferable to study some further kernel systems generated by superposition (i.e., infinite sums) of outer harmonics (as described, e.g., in [15]). Indeed, if they are explicitly available as elementary functions like kernel representations known from minimum norm interpolation, i.e., in the jargon of physical geodesy "least squares collocation" (see, e.g., [55]), these systems turn out to be particularly suitable for numerical purposes because of their simple implementation.

Theorem 3.6. *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region such that $\mathcal{B}_{\underline{R}}(0) \Subset \mathcal{G}$. Suppose that the kernel function $K(\cdot,\cdot) : \mathbb{R}^3 \backslash \mathcal{B}_{\underline{R}}(0) \times \mathcal{B}_{\underline{R}}(0) \to \mathbb{R}$ is given by*

$$\begin{aligned} K(x,y) &= \sum_{k=0}^{\infty} \sum_{l=1}^{2k+1} K^{\wedge}(k) H^{\underline{R}}_{-k-1,l}(x) H^{\underline{R}}_{k,l}(y) \\ &= \frac{\underline{R}}{|x|} \sum_{k=0}^{\infty} \frac{2k+1}{4\pi \underline{R}^2} K^{\wedge}(k) \left(\frac{|y|}{|x|} \right)^k P_k \left(\frac{x}{|x|} \cdot \frac{y}{|y|} \right) \end{aligned} \tag{47}$$

for $x \in \mathbb{R}^3 \backslash \mathcal{B}_{\underline{R}}(0), y \in \mathcal{B}_{\underline{R}}(0)$. Let $\mathcal{Y} = \{y_n\}_{n=0,1,\ldots}$ be a fundamental system in $\mathcal{B}_{\underline{R}}(0)$. Suppose that

$$\sum_{k=0}^{\infty} (2k+1) \left| K^{\wedge}(k) \right| \left(\frac{\underline{R}}{\inf\limits_{x \in \partial \mathcal{G}} |x|} \right)^k < \infty \tag{48}$$

with $K^{\wedge}(k) \neq 0$ for $k \in \mathbb{N}_0$. Then the system

$$\left\{ K(\cdot, y_n) \Big|_{\partial \mathcal{G}} \right\}_{n=0,1,\ldots}$$

is closed in $L^2(\partial \mathcal{G})$.

Proof. Let F be of class $L^2(\partial \mathcal{G})$. The function P given by

$$P(y) = \int_{\partial \mathcal{G}} K(x,y) F(x) \, dS(x), \tag{49}$$

is analytic in $\mathcal{B}_{\underline{R}}(0)$. Indeed, for all $y \in \mathbb{R}^3$ with $|y| < \underline{R}$, it follows from (49) that

$$P(y) = \sum_{k=0}^{\infty} K^{\wedge}(k) \sum_{j=1}^{2k+1} H^{\underline{R}}_{k,j}(y) \int_{\partial \mathcal{G}} F(x) H^{\underline{R}}_{-k-1,j}(x) \, dS(x). \tag{50}$$

Assume that $P(y_n) = 0$ for $n = 0, 1, \ldots$. Since $\mathcal{Y} = \{y_n\}_{n=0,1,\ldots}$ is a fundamental system in $\mathcal{B}_{\underline{R}}(0)$, the function P vanishes in $\mathcal{B}_{\underline{R}_0}(0)$ for all $\underline{R}_0 \leq \underline{R}$. This implies that

$$(F, H^{\underline{R}}_{-k-1,j})_{L^2(\partial \mathcal{G})} = \int_{\partial \mathcal{G}} F(x) H^{\underline{R}}_{-k-1,j}(x) \, dS(x) = 0, \tag{51}$$

for $k \in \mathbb{N}_0$, $j = 1, \ldots, 2k+1$. Hence, by virtue of the completeness of the system of outer harmonics (Theorem 3.2), we obtain $F = 0$ in the topology of $L^2(\partial \mathcal{G})$, as required. □

Examples of kernel representations (47) are easily obtainable from known series expansions in terms of Legendre polynomials (see, e.g., elementary representations in [5, 12, 16, 23, 55] based on identities as presented in, e.g., [50]).

Applying the Kelvin transform with respect to the sphere $\Omega_{\underline{R}}(0)$ around the origin with radius $\underline{R}$, we are led to systems $\left\{\check{K}(\cdot,\check{y}_n)\right\}_{n=0,1,\ldots}$ with

$$\begin{aligned}\check{K}(x,\check{y}_n) &= \sum_{k=0}^{\infty}\sum_{l=1}^{2k+1} K^{\wedge}(k) H^{\underline{R}}_{-k-1,l}(x) H^{\underline{R}}_{-k-1,l}(\check{y}_n) \\ &= \sum_{k=0}^{\infty} \frac{2k+1}{4\pi\underline{R}^2} K^{\wedge}(k) \left(\frac{\underline{R}^2}{|x||\check{y}_n|}\right)^{k+1} P_k\left(\frac{x}{|x|}\cdot\frac{\check{y}_n}{|\check{y}_n|}\right), \end{aligned} \tag{52}$$

where $\check{\mathcal{Y}} = \{\check{y}_n\}_{n=0,1,\ldots}$ is the point system generated by application of the Kelvin transform to $\mathcal{Y}$, i.e., by letting

$$\check{y}_n = \frac{\underline{R}^2}{|y_n|^2} y_n, \quad n = 0,1,\ldots \tag{53}$$

(assuming that $0 \notin \mathcal{Y}$).

Theorem 3.7. *Suppose that $\check{\mathcal{Y}} = \{\check{y}_n\}_{n=0,1,\ldots}$ is given as described above. Then the system*

$$\left\{\check{K}(\cdot,\check{y}_n)\Big|_{\partial\mathcal{G}}\right\}_{n=0,1,\ldots} \tag{54}$$

is closed in $L^2(\partial\mathcal{G})$.

Particularly helpful in geosciences is a fundamental system $\mathcal{Y} = \{y_n\}_{n=0,1,\ldots}$ in $\mathcal{B}_{\underline{R}_0}(0)$ that yields $\check{\mathcal{Y}} = \{\check{y}_n\}_{n=0,1,\ldots} \subset \partial\mathcal{G}$ (cf. Figure 4). In other words, the closure property is related to points lying on the actual (Earth's) surface $\partial\mathcal{G}$ (note that the Kelvin transform is easily invertible, so that $\mathcal{Y}$ and $\check{\mathcal{Y}}$ can be easily constructed from each other).

The (Kelvin modified) kernels $\check{K}(\cdot,\check{y}_n), n \in \mathbb{N}_0$, as given by (52) define reproducing kernel spaces in Sobolev-like Hilbert spaces. They play a central role in

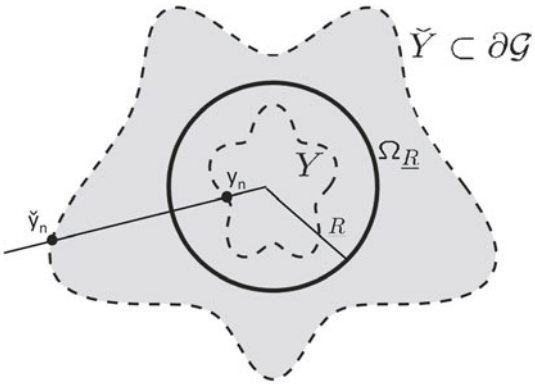


FIGURE 4. Fundamental system $\mathcal{Y}$ and the Kelvin transformed system $\check{\mathcal{Y}} \subset \partial\mathcal{G}$ (see also [15]).

the theory of (Runge-type) harmonic splines (see [9, 12, 13, 15, 64]). Of particular significance as examples are the spline kernels corresponding to the following "symbols" $\{K^\wedge(k)\}_{k\in\mathbb{N}_0}$:

(a) *Abel–Poisson kernel:*

$$K^\wedge(k) = C_0,\ k \in \mathbb{N}_0, \tag{55}$$

(b) *singularity kernel:*

$$K^\wedge(k) = \frac{2C_0}{2k+1},\ k \in \mathbb{N}_0, \tag{56}$$

(c) *logarithmic kernel:*

$$K^\wedge(k) = \frac{C_0}{(k+1)(2k+1)},\ k \in \mathbb{N}_0, \tag{57}$$

where $C_0 \neq 0$ is a real constant (see Examples 4.4–4.6).

From potential theory (see, e.g., [15]) we know the following *regularity theorem*: Suppose that V is of class $Pot^{(0)}(\overline{\mathcal{G}^c})$, i.e.,

(i) $V \in C^{(2)}(\mathcal{G}^c) \cap C^{(0)}(\overline{\mathcal{G}^c})$,
(ii) $\Delta V = 0$ in $\mathcal{G}^c$,
(iii) V is regular at infinity, i.e., $V(x) = O(|x|^{-1})$, $|x| \to \infty$.

Then, for all $k \in \mathbb{N}_0$ and all $\mathcal{K} \Subset \mathcal{G}^c$ there exist a constant C (dependent on $\mathcal{K}$, and $\partial\mathcal{G}$) (note that $\nabla^{(1)}V = \nabla V$ and $\nabla^{(2)}V = (\nabla \otimes \nabla)V$, etc.) such that

$$\sup_{x\in\overline{\mathcal{K}}} |(\nabla^{(k)}V)(x)| \leq C \left(\int_{\partial\mathcal{G}} |V(x)|^2\ dS(x) \right)^{1/2} \tag{58}$$

(note that the norm of the matrix $(\nabla \otimes \nabla)V$ is given by $\sum_{i,j=1}^{3}(((\nabla \otimes \nabla)V)_{i,j})^2$).

Combining the $L^2(\partial\mathcal{G})$-closure (Theorem 3.2) and the regularity theorems (Theorem 3.7), we obtain the following result for the system of outer harmonics.

Theorem 3.8. *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region such that $\mathcal{B}_{\underline{R}}(0) \Subset \mathcal{G}$. For given $F \in C^{(0)}(\partial\mathcal{G})$, let V be the potential of class $Pot^{(0)}(\overline{\mathcal{G}^c})$ with $V|_{\partial\mathcal{G}} = F$. Then, for any given $\varepsilon > 0$ and any given $\mathcal{K} \Subset \mathcal{G}^c$, there exist an integer N (dependent on ε) and a set of coefficients $a_{0,1}, \ldots, a_{N,1}, \ldots, a_{N,2N+1}$ such that*

$$\left(\int_{\partial\mathcal{G}} \left| F(x) - \sum_{n=0}^{N} \sum_{j=1}^{2n+1} a_{n,j} H^{\underline{R}}_{-n-1,j}(x) \right|^2 dS(x) \right)^{1/2} \leq \varepsilon \tag{59}$$

and

$$\sup_{x\in\overline{\mathcal{K}}} \left| \left(\nabla^{(k)}V\right)(x) - \sum_{n=0}^{N} \sum_{j=1}^{2n+1} a_{n,j} \left(\nabla^{(k)} H^{\underline{R}}_{-n-1,j}\right)(x) \right| \leq C\varepsilon \tag{60}$$

hold for all $k \in \mathbb{N}_0$.

In other words, the L^2-approximation in terms of outer harmonics on $\partial\mathcal{G}$ implies the uniform approximation (in the ordinary sense) on each subset $\overline{\mathcal{K}}$ with positive distance to $\partial\mathcal{G}$.

Unfortunately, although the Runge–Walsh framework is formulated for special trial systems (such as mono- and/or multi-pole systems), the theorems developed until now are non-constructive since further information about the choice of the truncation order N and the coefficients of the approximating linear combination is needed. In order to derive a constructive approximation theorem, the system of potential values and normal derivatives, respectively, can be orthonormalized on $\partial\mathcal{G}$. As a result, we obtain a (*generalized*) *Fourier series expansion* (orthogonal Fourier series expansion) that shows locally uniform approximation.

Theorem 3.9. *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region such that $\mathcal{B}_{\underline{R}}(0) \Subset \mathcal{G}$. For given $F \in C^{(0)}(\partial\mathcal{G})$, let V be the solution of the Dirichlet problem $V \in Pot^{(0)}(\overline{\mathcal{G}^c})$, $V|_{\partial\mathcal{G}} = F$. Corresponding to $\{H^{\underline{R}}_{-n-1,j}\}_{n\in\mathbb{N}_0,j=1,\ldots,2n+1}$ there exists a system $\{H_{-n-1,j}(\partial\mathcal{G};\cdot)\}_{n\in\mathbb{N}_0,j=1,\ldots,2n+1} \subset Pot^{(0)}(\mathbb{R}^3\backslash\mathcal{B}_{\underline{R}}(0))$ such that*

$$\{H_{-n-1,j}(\partial\mathcal{G};\cdot)|_{\partial\mathcal{G}}\}_{\substack{n\in\mathbb{N}_0;\\ j=1,\ldots,2n+1}}$$

is orthonormal in the sense that

$$\int_{\partial\mathcal{G}} H_{-n-1,j}(\partial\mathcal{G};y)H_{-l-1,k}(\partial\mathcal{G};y)\ dS(y) = \delta_{n,l}\ \delta_{j,k}. \tag{61}$$

Consequently, V is representable in the form

$$V(x) = \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\left(\int_{\partial\mathcal{G}} F(y)H_{-n-1,j}(\partial\mathcal{G};y)\ dS(y)\right)H_{-n-1,j}(\partial\mathcal{G};x) \tag{62}$$

for all points $x \in \mathcal{K} \Subset \mathcal{G}^c$. Moreover, for each $V^{(N)}$ given by

$$V^{(N)}(x) = \sum_{n=0}^{N}\sum_{j=1}^{2n+1}\left(\int_{\partial\mathcal{G}} F(y)H_{-n-1,j}(\partial\mathcal{G};y)\ dS(y)\right)H_{-n-1,j}(\partial\mathcal{G};x) \tag{63}$$

we have the estimate

$$\sup_{x\in\overline{\mathcal{K}}}\left|\left(\nabla^{(k)}V\right)(x) - \left(\nabla^{(k)}V^{(N)}\right)(x)\right| \tag{64}$$

$$\leq C\left(\int_{\partial\mathcal{G}}|F(y)|^2\ dS(y) - \sum_{n=0}^{N}\sum_{j=1}^{2n+1}\left|\int_{\partial\mathcal{G}} F(y)H_{-n-1,j}(\partial\mathcal{G};y)\ dS(y)\right|^2\right)^{1/2}.$$

Note that the orthonormalization procedure can be performed (e.g., by the well-known Gram–Schmidt orthonormalization process) once and for all in the case that the boundary surface $\partial\mathcal{G}$ of a regular region $\mathcal{G}$ is specified.

Next, we rewrite our generalized Fourier approach in a more abstract form. For that purpose we introduce the concept of so-called Dirichlet bases.

Definition 3.10 (Dirichlet Runge Basis). Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be arbitrary regular regions such that $\mathcal{A} \Subset \mathcal{G}$ (cf. Figure 1, left illustration). A linearly independent system $\{D_n\}_{n=0,1,\ldots} \subset Pot(\mathcal{A}^c)$ is called a ($Pot(\mathcal{A}^c)$-generated) $L^2(\partial\mathcal{G})$-*Dirichlet Runge basis* if

$$\overline{\operatorname{span}_{n\in\mathbb{N}_0}\{D_n|_{\partial\mathcal{G}}\}}^{\|\cdot\|_{L^2(\partial\mathcal{G})}} = L^2(\partial\mathcal{G}). \tag{65}$$

Dirichlet Runge bases are constituted by all special function system discussed earlier (i.e., mono- and muli-pole system, certain kernel function systems, Lamé and Mathieu systems, etc.)

Corollary 3.11. *Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be regular regions such that $\mathcal{A} \Subset \mathcal{G}$ holds true. Let $\{D_n^*\}_{n=0,1,\ldots} \subset Pot(\mathcal{A}^c)$, be a function system generated by (Gram–Schmidt) orthonormalization of an $L^2(\partial\mathcal{G})$-Dirichlet Runge basis $\{D_n\}_{n=0,1,\ldots} \subset Pot(\mathcal{A}^c)$, such that*

$$(D_n^*, D_m^*)_{L^2(\partial\mathcal{G})} = \int_{\partial\mathcal{G}} D_n^*(x) D_m^*(x)\, dS(x) \;=\; \delta_{n,m}. \tag{66}$$

If $F \in C^{(0)}(\partial\mathcal{G})$, then

$$\lim_{N\to\infty} \left(\int_{\partial\mathcal{G}} \left| F(x) - F^{(N)}(x) \right|^2 \, dS(x) \right)^{\frac{1}{2}} = 0, \tag{67}$$

where $F^{(N)}$ denotes the truncated Fourier series expansion

$$F^{(N)} = \sum_{n=0}^{N} (F, D_n^*)_{L^2(\partial\mathcal{G})}\, D_n^*\big|_{\partial\mathcal{G}}. \tag{68}$$

The potential $V \in Pot^{(0)}(\overline{\mathcal{G}^c})$ satisfying $V\big|_{\partial\mathcal{G}} = F$ can be represented in the form

$$\lim_{N\to\infty} \sup_{x\in\overline{\mathcal{K}}} \left| V(x) - V^{(N)}(x) \right| = 0, \tag{69}$$

for every $\mathcal{K} \Subset \mathcal{G}^c$, where

$$V^{(N)} = \sum_{n=0}^{N} (F, D_n^*)_{L^2(\partial\mathcal{G})} \;\; D_n^*. \tag{70}$$

The concrete versions of the Runge–Walsh theorem, i.e., the (*generalized*) *Fourier expansions* (68), are indeed constructed to have the *permanence property*: The transition from $F^{(N)}$ to $F^{(N+1)}$, and therefore from $V^{(N)}$ to $V^{(N+1)}$, merely necessitates the addition of one more term; all the other terms obtained formerly remain unchanged. This is characteristic of orthogonal expansions.

In connection with the $L^2(\partial\mathcal{G})$-regularity theorems, we additionally find the following estimate: For given $F \in C^{(0)}(\partial\mathcal{G})$, let V satisfy $V \in Pot^{(0)}(\overline{\mathcal{G}^c})$,

$V|_{\partial\mathcal{G}} = F$. Then

$$\begin{aligned}\sup_{x\in\overline{\mathcal{K}}}\left|\left(\nabla^{(k)}V\right)(x) - \sum_{n=0}^{N}(F,D_n^*)_{L^2(\partial\mathcal{G})}\left(\nabla^{(k)}D_n^*\right)(x)\right| \\ \leq C\left(\|F\|^2_{L^2(\partial\mathcal{G})} - \sum_{n=0}^{N}(F,D_n^*)^2_{L^2(\partial\mathcal{G})}\right)^{1/2}\end{aligned} \tag{71}$$

holds for all $k \in \mathbb{N}_0$ and all subsets $\mathcal{K} \Subset \mathcal{G}^c$. In addition, Corollary 3.11 indicates that $F - F^{(N)}$ is $L^2(\partial\mathcal{G})$-orthogonal to all members of the $L^2(\partial\mathcal{G})$-orthonormal Runge basis up to the index N. This observation is valid for the Dirichlet problem.

Corollary 3.12. *Let $\mathcal{A},\mathcal{G} \subset \mathbb{R}^3$ be regular regions such that the "Runge property" $\mathcal{A} \Subset \mathcal{G}$ holds true. Let $\{D_n\}_{n=0,1,\ldots} \subset Pot(\mathcal{A}^c)$ be an $L^2(\partial\mathcal{G})$-Dirichlet Runge basis. If $F \in C^{(0)}(\partial\mathcal{G})$, then*

$$\lim_{N\to\infty}\left(\int_{\partial\mathcal{G}}\left|F(x) - F^{(N)}(x)\right|^2\, dS(x)\right)^{\frac{1}{2}} = 0, \tag{72}$$

where the coefficients $a_0^N,\ldots,a_N^N$ of the function

$$F^{(N)} = \sum_{n=0}^{N} a_n^N D_n\big|_{\partial\mathcal{G}} \tag{73}$$

satisfy the "normal equations"

$$\sum_{n=0}^{N} a_n^N\,(D_k,D_n)_{L^2(\partial\mathcal{G})} = (D_k,F)_{L^2(\partial\mathcal{G})}\,, \quad k=0,\ldots,N. \tag{74}$$

The potential $V \in Pot^{(0)}(\overline{\mathcal{G}^c})$ satisfying $V\Big|_{\partial\mathcal{G}} = F$ can be represented in the form

$$\lim_{N\to\infty}\sup_{x\in\overline{\mathcal{K}}}\left|V(x) - V^{(N)}(x)\right| = 0, \tag{75}$$

where

$$V^{(N)} = \sum_{n=0}^{N} a_n^N D_n \tag{76}$$

for every $\mathcal{K} \Subset \mathcal{G}^c$.

Remark. Later on, the construction principle based on Corollary 3.12 formulated in an appropriately defined Sobolev space structure will lead to harmonic spline interpolation involving (Sobolev-like) reproducing kernel Hilbert space (RKHS) features.

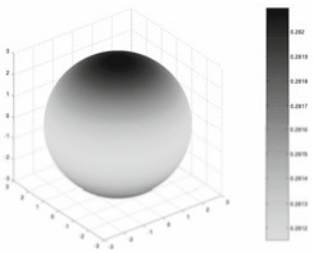
The approximation of boundary values and the gravitational potential of, e.g., the Earth by the method of generalized Fourier expansion in terms of outer

harmonics is achieved by superposition of functions with *oscillating character*. The oscillations grow in number, but they decrease in size with increasing truncation order. The oscillating character of the generalized Fourier expansions remains true (cf. [11]) if other trial bases are used, for example, mono-poles and certain kernel function representations such as Abel–Poisson (cf. Example 4.4) and singularity kernels (cf. Example 4.5). Thus, generalized Fourier expansions provide least squares approximation by successive oscillations, which become larger and larger in number, but smaller and smaller in amplitude. It is therefore not (as [65] has pointed out) a technique of *osculating character* (as, e.g., interpolation in reproducing Hilbert spaces by harmonic splines (as proposed by W. Freeden [9] and L. Shure et al. [64])).

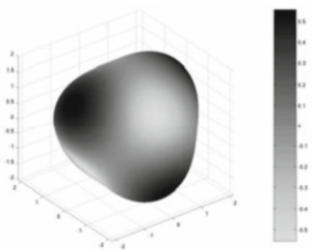
Since the time of [30], there is evidence – at least in the spherical context using multi-poles, i.e., outer harmonics – that a Fourier series expansion provides an excellent (spherically reflected) *trend approximation* of a harmonic function such as the Earth's gravitational and magnetic potential. The ideal *frequency localization,* more accurately momentum localization of outer harmonics – each of them referring to a certain degree of oscillation – has proved to be extraordinarily advantageous due to the physical interpretability and the immediate comparability of the Fourier coefficients for observables. From a numerical point of view, however, trial functions would be desirable that show ideal frequency as well as space localization on the reference sphere. The uncertainty principle (see, e.g., [14, 26] and the references therein) teaches us that both properties are mutually exclusive (except in the trivial case). This explains some problems in the Fourier technique of approximation, at least by means of outer harmonics. Fourier expansions in terms of outer harmonics are well suited to resolve low-frequency ingredients in an observable, while their application is critical to obtain high-resolution phenomena. The kernel functions provided by Example 4.4, Example 4.5, and Example 4.6 (among others which are similarly constructed) show a reduced frequency but increased space localization (on a Runge reference surface) as the series conglomerates of outer harmonics are constructed to cover various spectral bands (see, e.g., [26]).

As graphical examples (taken from the Ph.D.-thesis [49]) we illustrate the members of the Runge–Walsh orthonormalized low-degree outer harmonics on the International Reference Ellipsoid (IRE) $\partial\mathcal{G}$ thereby choosing the standard parameters as specified in physical geodesy (for more detailed IRE-parameter specifications see, e.g., [39]) (see figure 5).
The figures have been obtained via the well-known Gram–Schmidt orthonormalizing process in canonical way. Since the ellipsoid is quite close to a sphere, the phenotype of the Runge–Walsh orthonormal functions does not differ so much from outer harmonics on a sphere.

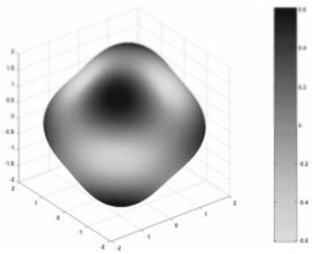
It should also be mentioned that the Runge–Walsh concept also leads to $\|\cdot\|_{C^{(0)}(\partial\mathcal{G})}$-approximation. To be more concrete, from our considerations leading to locally uniform approximation, we know for a given regular region $\mathcal{G}$ with



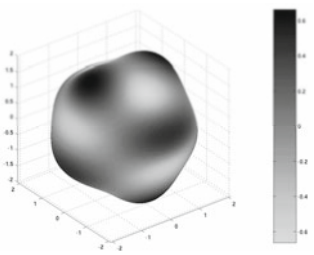
(a) Degree 1 and order 1.



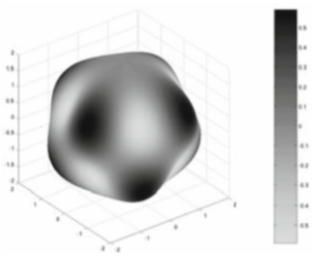
(b) Degree 4 and order 2.



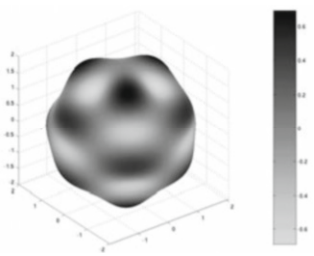
(c) Degree 5 and order 7.



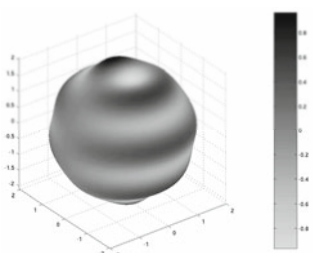
(d) Degree 6 and order 4.



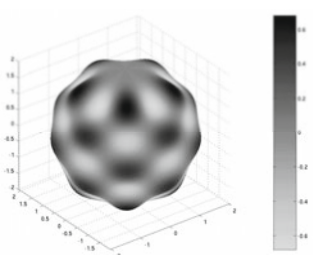
(e) Degree 6 and order 10.



(f) Degree 8 and order 5.



(g) Degree 9 and order 10.



(h) Degree 10 and order 15.

FIGURE 5. Outer harmonics of different degree and order on the International Reference Ellipsoid (IRE).

$\mathcal{B}_{\underline{R}}(0) \Subset \mathcal{G}$ that

$$\overline{\operatorname{span}_{\substack{n=0,1,\ldots;\\ j=1,\ldots,2n+1}} \{H^{\underline{R}}_{-n-1,j}|_{\partial\mathcal{G}}\}}^{\|\cdot\|_{L^2(\partial\mathcal{G})}} = L^2(\partial\mathcal{G}). \tag{77}$$

The same results remain valid when the regular surface $\partial\mathcal{G}$ is replaced by any inner parallel surface $\partial\mathcal{G}(-\tau)$ of distance $|\tau|$ to $\partial\mathcal{G}$ (where $|\tau|$ is chosen sufficiently small). This fact can be exploited to verify the following closure properties (see [8] for a detailed proof).

Theorem 3.13. *Let $\partial\mathcal{G}$ be the boundary of a regular region such that $\mathcal{B}_{\underline{R}}(0) \Subset \mathcal{G}$. Then the system $\{H^{\underline{R}}_{-n-1,j}|_{\partial\mathcal{G}}\}_{n\in\mathbb{N}_0, j=1,\ldots,2n+1}$ is closed in $C^{(0)}(\partial\mathcal{G})$:*

$$\overline{\operatorname{span}_{\substack{n=0,1,\ldots;\\ j=1,\ldots,2n+1}} \{H^{\underline{R}}_{-n-1,j}|_{\partial\mathcal{G}}\}}^{\|\cdot\|_{C^{(0)}(\partial\mathcal{G})}} = C^{(0)}(\partial\mathcal{G}). \tag{78}$$

Remark. The same arguments leading to the $C^{(0)}(\partial\mathcal{G})$-closure of outer harmonics on $\partial\mathcal{G}$ apply to all other systems for which the $L^2(\partial\mathcal{G})$-closure is known, e.g., ellipsoidal harmonics, Lamé functions, etc.

Combining our results obtained by Theorem 3.13, we easily arrive at the following statement.

Theorem 3.14. *Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region such that $\mathcal{B}_{\underline{R}}(0) \Subset \mathcal{G}$. For a given (boundary) function $F \in C^{(0)}(\partial\mathcal{G})$, let $V \in Pot^{(0)}(\overline{\mathcal{G}^c})$ satisfy $V|_{\partial\mathcal{G}} = F$. Then, for every $\varepsilon > 0$, there exist an integer N (depending on ε) and a finite set of real numbers $a_{n,j}$ such that*

$$\begin{aligned} &\sup_{x\in\overline{\mathcal{G}^c}} \left| V(x) - \sum_{n=0}^{N} \sum_{j=1}^{2n+1} a_{n,j} H^{\underline{R}}_{-n-1,j}(x) \right| \\ &\quad \le \sup_{x\in\partial\mathcal{G}} \left| F(x) - \sum_{n=0}^{N} \sum_{j=1}^{2n+1} a_{n,j} H^{\underline{R}}_{-n-1,j}(x) \right| \le \varepsilon. \end{aligned} \tag{79}$$

Unfortunately, a constructive procedure of determining best approximate coefficients $a_{n,j}$ in the $C^{(0)}(\partial\mathcal{G})$-topology seems to be unknown. Therefore, harmonic splines (see, e.g., [9, 12, 64]) can be introduced in reproducing kernel Hilbert subspaces of $Pot(\overline{\mathcal{A}^c})$ (characterized by variational principles), so that the spline method can be regarded as an immediate extension of the method of generalized Fourier series expansions to reproducing kernel subspaces of $Pot(\overline{\mathcal{A}^c})$, hence, providing coefficients that are optimal in a different (Sobolev like) norm. Moreover, bandlimited as well as non-bandlimited spline wavelets (see [13, 15, 20], and the references therein) can be introduced to guarantee constructive multiscale variants of the Runge–Walsh concept (for more details about the role of non-bandlimited (Runge-type) spline solutions in geodetic boundary value problems the reader is referred to the next section).

3.2. RKHS-closure and spline interpolation

Next, our purpose is to formulate closure theorems for a Sobolev-like reproducing kernel Hilbert subspace of $Pot(\overline{\mathcal{A}^c})$ (see [10, 12]): Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be regular regions so that $\mathcal{A} \Subset \mathcal{G}$ (see Figure 1, left illustration). Suppose that $\{D_n^*\}_{n=0,1,...}$ is an $L^2(\partial\mathcal{A})$-Dirichlet Runge basis (in the sense of Definition 3.10) obeying the orthonormality condition

$$(D_n^*|_{\partial\mathcal{A}}, D_m^*|_{\partial\mathcal{A}})_{L^2(\partial\mathcal{A})} = \int_{\partial\mathcal{A}} D_n^*(x) D_m^*(x)\, dS(x) \;=\; \delta_{n,m}. \tag{80}$$

Assume that the system $\{E_n\}_{n=0,1,...}$ is given by

$$E_n = \sigma_n D_n^*, \quad \sigma_n \in \mathbb{R}\backslash\{0\}, \;\; n = 0, 1, \ldots, \tag{81}$$

and that the sequence $\{\sigma_n\}_{n\in\mathbb{N}_0}$ is summable in the sense

$$\sum_{n=0}^{\infty} \sigma_n^2 < \infty. \tag{82}$$

Then, for every $k \in \mathbb{N}_0$ and $\tilde{F} \in L^2(\partial\mathcal{A})$, the sum

$$\left(\nabla^{(k)} F\right)(x) = \sum_{n=0}^{\infty} (\tilde{F}, D_n^*)_{L^2(\partial\mathcal{A})} \left(\nabla^{(k)} E_n(x)\right), \quad x \in \mathcal{A}^c, \tag{83}$$

satisfies the estimate

$$\sup_{x\in\overline{\mathcal{K}}} \left|\nabla^{(k)} F\right| \le \left(\sum_{n=0}^{\infty} \left(\tilde{F}, D_n^*\right)_{L^2(\partial\mathcal{A})}\right)^{\frac{1}{2}} \sup_{x\in\overline{\mathcal{K}}} \left(\sum_{n=0}^{\infty} \left|\nabla^{(k)} E_n(x)\right|^2\right)^{\frac{1}{2}}, \tag{84}$$

where $\mathcal{K} \Subset \mathcal{A}^c$. The L^2-regularity condition implies with a constant $\tilde{C}$ (dependent on $\mathcal{A}, \mathcal{K}$) that

$$\sup_{x\in\overline{\mathcal{K}}} \left|\nabla^{(k)} D_n^*(x)\right|^2 \le \tilde{C}^2 \underbrace{\int_{\partial\mathcal{A}} |D_n^*(x)|^2\, dS(x)}_{=1}. \tag{85}$$

Even more, the mean value theorem of multi-variate analysis shows (cf. [15]) that there exist a positive constant $\tilde{\tilde{C}}$ such that

$$\left|\nabla^{(k)} D_n^*(x) - \nabla^{(k)} D_n^*(y)\right| \le \tilde{\tilde{C}}^2 |x - y|^2 \tag{86}$$

is valid for $x, y \in \overline{\mathcal{K}}$. From (85) it follows that

$$\sup_{x\in\overline{\mathcal{K}}} \left(\sum_{n=0}^{\infty} \left|\nabla^{(k)} E_n(x)\right|^2\right)^{\frac{1}{2}} \le C \left(\sum_{n=0}^{\infty} \sigma_n^2\right)^{\frac{1}{2}}. \tag{87}$$

In other words, the expansion on the right of (83) exists such that F is harmonic in $\mathcal{A}^c$ and regular at infinity. In particular,

$$\sup_{x\in\overline{\mathcal{G}^c}} \left|\nabla^{(k)} F(x)\right| \le C \left(\sum_{n=0}^{\infty} \sigma_n^2\right)^{\frac{1}{2}} \left(\sum_{n=0}^{\infty} \left(\tilde{F}, D_n^*\right)_{L^2(\partial\mathcal{A})}\right)^{\frac{1}{2}} \tag{88}$$

(note that $\mathcal{G}$ is assumed to be a regular region satisfying $\mathcal{A} \Subset \mathcal{G}$). All functions F which can be expressed by such series as given in (83) form a linear space $H(\mathcal{A}^c)$ on which we are able to impose the structure of a separable Hilbert space (cf. [10])

$$H(\mathcal{A}^c) = \left\{ F = \sum_{n=0}^{\infty} (\tilde{F}, D_n^*)_{L^2(\partial\mathcal{A})} \; E_n, \;\; \tilde{F} \in L^2(\partial\mathcal{A}) \right\} \tag{89}$$

by taking as an inner product, for members $F, G \in H(\mathcal{A}^c)$ associated to $\tilde{F}$, $\tilde{G} \in L^2(\partial\mathcal{A})$, respectively, the expression

$$\begin{aligned} (F, G)_{H(\mathcal{A}^c)} &= (\tilde{F}, \tilde{G})_{L^2(\partial\mathcal{A})} \\ &= \sum_{n=0}^{\infty} \left(\tilde{F}, D_n^* \right)_{L^2(\partial\mathcal{A})} \left(\tilde{G}, D_n^* \right)_{L^2(\partial\mathcal{A})} \\ &= \sum_{n=0}^{\infty} \frac{1}{\sigma_n^2} \, (F, D_n^*)_{L^2(\partial\mathcal{A})} \, (G, D_n^*)_{L^2(\partial\mathcal{A})} \, . \end{aligned} \tag{90}$$

Theorem 3.15. *$(H(\mathcal{A}^c), (\cdot,\cdot)_{H(\mathcal{A}^c)})$ is a separable Hilbert space possessing the (uniquely determined) reproducing kernel*

$$K_{H(\mathcal{A}^c)}(x, y) = \sum_{n=0}^{\infty} \; E_n(x) \; E_n(y) = \sum_{n=0}^{\infty} \sigma_n^2 \; D_n^*(x) \; D_n^*(y) \tag{91}$$

for all $x, y \in \mathcal{A}^c$.

Suppose that $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ are regular regions such that $\mathcal{A} \Subset \mathcal{G}$. Assume that

$$X \;\; = \;\; \bigcup_{N=1}^{\infty} X_N, \tag{92}$$

where

$$X_N = \{x_1, \ldots, x_N\} \subset \partial\mathcal{G}, \quad x_i \neq x_j, \quad i \neq j \tag{93}$$

is a countable dense set of points on $\partial\mathcal{G}$. Then we are able to show the following *$H(\mathcal{A}^c)$-closure result:*

Lemma 3.16.

$$H(\mathcal{A}^c) = \overline{\operatorname*{span}_{x \in X} K_{H(\mathcal{A}^c)}(x, \cdot)}^{\|\cdot\|_{H(\mathcal{A}^c)}}. \tag{94}$$

Proof. Our aim is to prove that $F \in H(\mathcal{A}^c)$ and $\left(F, K_{H(\mathcal{A}^c)}(x, \cdot)\right)_{H(\mathcal{A}^c)} = 0$ for all $x \in X$ implies that $F = 0$. The reproducing kernel Hilbert space structure tells us that $\left(F, K_{H(\mathcal{A}^c)}(x, \cdot)\right)_{H(\mathcal{A}^c)} = 0$ is equivalent to $F(x) = 0$. According to our construction, F is continuous on $\partial\mathcal{G}$. Hence, if $F(x) \neq 0$ for some $x \in \partial\mathcal{G}$, then F is different from zero for a whole neighborhood of x on $\partial\mathcal{G}$. But this contradicts the density of X. Hence, $F = 0$ on $\partial\mathcal{G}$. The analyticity of F finally yields $F = 0$ in $H(\mathcal{A}^c)$, as desired. □

Lemma 3.16 motivates to discuss discrete boundary value problems within the framework of $H(\mathcal{A}^c)$, in other words, contrary to the classical boundary value problems, where the solution process is based on the continuous knowledge of the "boundary function" as a whole (note that, in the case of a discrete boundary value problem, the boundary information is given only in a set of discrete points).

For simplicity, we start with the *exterior Dirichlet boundary value problem* (EDP) in its classical formulation: Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be regular regions so that $\mathcal{A} \Subset \mathcal{G}$ (see Figure 1, left illustration). Given a function F of class $C^{(0)}(\partial\mathcal{G})$, find a function V of class $Pot^{(0)}(\overline{\mathcal{G}^c})$ satisfying the boundary condition $V|_{\partial\mathcal{G}}(x) = F(x),\ x \in \partial\mathcal{G}$. Our goal is to construct an $H(\mathcal{A}^c)$-spline (interpolation) solution of the *discrete exterior Dirichlet problem* (DEDP). To this end, we are interested in a (Runge-type) spline potential $P \in H(\mathcal{A}^c)$ with observed values $F(x_i) = \mathcal{E}_{x_i}[P] = P(x_i) = \beta_i, i = 1, \ldots, N$, where the points $x_1, \ldots, x_N \in \partial\mathcal{G}$ are assumed to be associated with linearly independent bounded evaluation functionals $\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}$ (with respect to the $H(\mathcal{A}^c)$-topology) applied to the "(Runge) restriction" $P|_{\overline{\mathcal{G}^c}} \in H(\mathcal{A}^c)|_{\overline{\mathcal{G}^c}}$ of the potential $P \in H(\mathcal{A}^c)$. In doing so, we are able to find a minimum norm solution $S_N^P \in H(\mathcal{A}^c)$ as a linear combination of the representers $\mathcal{E}_{x_i}[K_{H(\mathcal{A}^c)}(\cdot,\cdot)]$ to the functionals $\mathcal{E}_{x_i}$, i.e., S_N^P is exactly the projection of P to the N-dimensional linear subspace spanned by the linearly independent representers $\mathcal{E}_{x_i}[K_{H(\mathcal{A}^c)}(\cdot,\cdot)],\ i = 1, \ldots, N$ (see, e.g., [6]). In other words, the solution of (DEDP) is sought in the reproducing kernel Hilbert space $H(\mathcal{A}^c)$ under the assumption that $\{\beta_1, \ldots, \beta_N\}$ with $\beta_i = \mathcal{E}_{x_i}[P] = P(x_i)$, $i = 1, \ldots, N$, is the (observed) given data set for the unknown potential P corresponding to the discrete set $X_N = \{x_1, \ldots, x_N\}$ of points on $\partial\mathcal{G}$. All in all, the aim of minimum norm interpolation in $H(\mathcal{A}^c)$ as proposed here is to find the "smoothest" $S_N^P \in H(\mathcal{A}^c)$ within the set of all $H(\mathcal{A}^c)$-interpolants, where the norm is minimized in the metric of $H(\mathcal{A}^c)$. Equivalently, the problem is to find a function S_N^P within the interpolatory set

$$\mathcal{I}^P_{\mathcal{E}_{x_1},\ldots,\mathcal{E}_{x_N}} = \big\{Q \in H(\mathcal{A}^c) : \mathcal{E}_{x_i}[Q] = \mathcal{E}_{x_i}[P] = P(x_i) = \beta_i,\ x_i \in \partial\mathcal{G},\ i = 1, \ldots, N\big\}, \tag{95}$$

such that

$$\big\|S_N^P\big\|_{H(\mathcal{A}^c)} = \inf_{Q \in \mathcal{I}^P_{\mathcal{E}_{x_1},\ldots,\mathcal{E}_{x_N}}} \|Q\|_{H(\mathcal{A}^c)}\,. \tag{96}$$

For any *$\mathcal{E}$-unisolvent system* $X_N = \{x_1, \ldots, x_N\} \subset \partial\mathcal{G}$, i.e., for any system $X_N = \{x_1, \ldots, x_N\}$ such that $\{\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}\}$ forms a set of N linearly independent bounded linear functionals on $H(\mathcal{A}^c)$ we introduce *$H(\mathcal{A}^c)$-splines relative to* $\{\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}\}$ in the following way:

Definition 3.17 (Harmonic Splines). Let $X_N = \{x_1, \ldots, x_N\} \subset \partial\mathcal{G}$ be an $\mathcal{E}$-unisolvent system on $\partial\mathcal{G}$. Then, any function $S \in H(\mathcal{A}^c)$ given by

$$S(x) = \sum_{i=1}^{N} a_i\ \mathcal{E}_{x_i}[K_{H(\mathcal{A}^c)}(\cdot, x)] = \sum_{i=1}^{N} a_i\ K_{H(\mathcal{A}^c)}(x_i, x) \tag{97}$$

with arbitrarily given (real) coefficients $a_1, \ldots, a_N$ is called an $H(\mathcal{A}^c)$-spline relative to $\{\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}\}$.

The space of all $H(\mathcal{A}^c)$-splines relative to $\{\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}\}$ is denoted by

$$Spline_{H(\mathcal{A}^c)}(\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}).$$

Clearly, $Spline_{H(\mathcal{A}^c)}(\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N})$ is an N-dimensional subspace of $H(\mathcal{A}^c)$. Moreover, by virtue of the reproducing property in $H(\mathcal{A}^c)$, we immediately obtain the so-called $H(\mathcal{A}^c)$*-spline formula*.

Lemma 3.18. *Let S be a function of class $Spline_{H(\mathcal{A}^c)}(\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N})$. Then, for each $F \in H(\mathcal{A}^c)$,*

$$(S, F)_{H(\mathcal{A}^c)} = \sum_{i=1}^{N} a_i \mathcal{E}_{x_i}[F] = \sum_{i=1}^{N} a_i F(x_i). \tag{98}$$

By virtue of the $\mathcal{E}$-unisolvence of the system $X_N = \{x_1, \ldots, x_N\} \subset \partial\mathcal{G}$ it is not difficult to verify the *uniqueness of interpolation.*

Lemma 3.19. *For a given potential $P \in H(\mathcal{A}^c)$, there exist a unique element S_N^P characterized by the property $S_N^P \in Spline_{H(\mathcal{A}^c)}(\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}) \cap \mathcal{I}^P_{\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}}$.*

Proof. The application of the N bounded linear functionals $\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}$ on $H(\mathcal{A}^c)$ to the $H(\mathcal{A}^c)$-spline of the form (97) yields N linear equations in the unknowns $a_1^N, \ldots, a_N^N$, i.e.,

$$\sum_{j=1}^{N} a_j^N \mathcal{E}_{x_i} \mathcal{E}_{x_j}[K_{H(\mathcal{A}^c)}(\cdot, \cdot)] = \mathcal{E}_{x_i}[P] \;=\; \beta_i, \quad i = 1, \ldots, N, \tag{99}$$

where the coefficient matrix is given by

$$\big(\mathcal{E}_{x_i} \mathcal{E}_{x_j}[K_{H(\mathcal{A}^c)}(\cdot, \cdot)]\big)_{i,j=1,\ldots,N} = \big(K_{H(\mathcal{A}^c)}(x_i, x_j)\big)_{i,j=1,\ldots,N}. \tag{100}$$

From multi-variate interpolation theory (see, e.g., [6]) we know that (100) constitutes a Gram matrix of N linearly independent functions

$$\mathcal{E}_{x_1}[K_{H(\mathcal{A}^c)}(\cdot, \cdot)], \ldots, \mathcal{E}_{x_N}[K_{H(\mathcal{A}^c)}(\cdot, \cdot)],$$

hence, it is non-singular such that the linear system (99) is uniquely solvable. The coefficients $a_1^N, \ldots, a_N^N$ determine the unique interpolating spline S_N^P. □

The following minimum norm properties for the interpolating spline S_N^P are easily derivable (see, e.g., [10] for comparable conclusions).

Lemma 3.20 (First Minimum Property). *If $F \in \mathcal{I}^P_{\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}}$, then*

$$||F||^2_{H(\mathcal{A}^c)} = ||S_N^P||^2_{H(\mathcal{A}^c)} + ||S_N^P - F||^2_{H(\mathcal{A}^c)}. \tag{101}$$

Lemma 3.21 (Second Minimum Property). *Suppose that*

$$S \in Spline_{H(\mathcal{A}^c)}(\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}) \quad and \quad F \in \mathcal{I}^P_{\mathcal{E}_{x_1},\ldots,\mathcal{E}_{x_N}}.$$

Then

$$||S - F||^2_{H(\mathcal{A}^c)} = ||S^P_N - F||^2_{H(\mathcal{A}^c)} + ||S - S^P_N||^2_{H(\mathcal{A}^c)}. \tag{102}$$

Summarizing our results on $H(\mathcal{A}^c)$-spline interpolation of a finite set of evaluation (Dirichlet) functionals we obtain

Theorem 3.22 (Spline Interpolation). *The minimum norm interpolation problem for solving* DEDP *from N given data* $P(x_i) = \beta_i,\ i = 1, \ldots, N,$

$$||S^P_N||_{H(\mathcal{A}^c)} \quad = \quad \inf_{Q \in \mathcal{I}^P_{\mathcal{E}_{x_1},\ldots,\mathcal{E}_{x_N}}} ||Q||_{H(\mathcal{A}^c)} \tag{103}$$

is well posed in the sense that its solution exists, is unique, and depends continuously on the data $\beta_1, \ldots, \beta_N$. *The uniquely determined solution* S^P_N *is given in the explicit form*

$$S^P_N(x) = \sum_{i=1}^N a_i^N \mathcal{E}_{x_i}[K_{H(\mathcal{A}^c)}(\cdot, x)] = \sum_{i=1}^N a_i^N K_{H(\mathcal{A}^c)}(x_i, x), \quad x \in \mathcal{A}^c, \tag{104}$$

where the coefficients $a_1^N, \ldots, a_N^N$ *satisfy the linear equations*

$$\sum_{j=1}^N a_j^N \mathcal{E}_{x_i}\mathcal{E}_{x_j}[K_{H(\mathcal{A}^c)}(\cdot,\cdot)] = \mathcal{E}_{x_i}[P] \ = \ \beta_i, \quad i = 1, \ldots, N. \tag{105}$$

Let ϑ_{X_N} denote the X_N*-width on* $\partial\mathcal{G}$, i.e., the maximal distance for any point of $\partial\mathcal{G}$ to the system X_N:

$$\vartheta_{X_N} \ = \ \max_{x \in \partial\mathcal{G}} \left(\min_{y \in X_N} |x - y| \right). \tag{106}$$

Our interest is the *stability of the solution* obtained by spline interpolation by letting $\vartheta_{X_N} \to 0$ as $N \to \infty$. As already known, for every $\mathcal{E}$-unisolvent system $X_N = \{x_1, \ldots, x_N\} \subset \partial\mathcal{G}$ and for every function $P \in H(\mathcal{A}^c)$ there exists a unique element $S^P_N \in H(\mathcal{A}^c)$ satisfying the conditions $\mathcal{E}_{x_i}[P] \ = \ \mathcal{E}_{x_i}[S^P_N],\ i = 1, \ldots, N.$

Lemma 3.23. *Let* P *be a member of class* $H(\mathcal{A}^c)$. *Suppose that* $X_N \subset \partial\mathcal{G}$ *is an* $\mathcal{E}$*-unisolvent system. Then there exists a constant* $C > 0$ (*dependent on* $\partial\mathcal{G}$ *and* $\mathcal{A}$) *such that*

$$\sup_{x \in \partial\mathcal{G}} \left| \mathcal{E}_x[P] - \mathcal{E}_x[S^P_N] \right| \ \le \ C \ \vartheta_{X_N} \ \|P\|_{H(\mathcal{A}^c)}. \tag{107}$$

Proof. For $x \in \partial\mathcal{G}$, there exists a point $y \in X_N$ with $|x - y| \le \vartheta_{X_N}$. Observing the interpolation property $\mathcal{E}_y[P] = \mathcal{E}_y[S^P_N]$, $y \in X_N$, we see that

$$\mathcal{E}_x[P] - \mathcal{E}_x[S^P_N] \ = \ (\mathcal{E}_x[P] - \mathcal{E}_y[P]) \ - \ \left(\mathcal{E}_x[S^P_N] - \mathcal{E}_y[S^P_N]\right). \tag{108}$$

The reproducing kernel structure of $H(\mathcal{A}^c)$ enables us to derive the estimates

$$|\mathcal{E}_x[P] - \mathcal{E}_y[P]| \leq (\kappa(x,y))^{\frac{1}{2}} \ \|P\|_{H(\mathcal{A}^c)}, \tag{109}$$

$$\left|\mathcal{E}_x[S_N^P] - \mathcal{E}_y[S_N^P]\right| \leq (\kappa(x,y))^{\frac{1}{2}} \ \|S_N^P\|_{H(\mathcal{A}^c)}, \tag{110}$$

where

$$\kappa(x,y) = \left(\mathcal{E}_x\mathcal{E}_x - 2\mathcal{E}_x\mathcal{E}_y + \mathcal{E}_y\mathcal{E}_y\right)[K_{H(\mathcal{A}^c)}(\cdot,\cdot)]. \tag{111}$$

S_N^P is the smoothest $H(\mathcal{A}^c)$-interpolant, i.e., $\|S_N^P\|_{H(\mathcal{A}^c)} \leq \|P\|_{H(\mathcal{A}^c)}$. From (108), (109), and (110) we therefore obtain

$$\sup_{x\in\partial\mathcal{G}} |\mathcal{E}_x[S_N^P] - \mathcal{E}_x[P]| \leq 2 \ (\kappa(x,y))^{\frac{1}{2}} \ \|P\|_{H(\mathcal{A}^c)}, \tag{112}$$

where $x \in \partial\mathcal{G}$ and $y \in X_N = \{x_1, \ldots, x_N\} \subset \partial\mathcal{G}$. More explicitly, we have

$$\kappa(x,y) = \sum_{n=0}^{\infty} \sigma_n^2 \left(\mathcal{E}_x D_n^* - \mathcal{E}_y D_n^*\right)^2. \tag{113}$$

By use of (86) we find

$$|\mathcal{E}_x D_n^* - \mathcal{E}_y D_n^*| \leq C \ |x-y|, \tag{114}$$

where $C > 0$ is a constant (depending on $\mathcal{A}^c$ and $\partial\mathcal{G}$). This proves Lemma 3.23. □

Summarizing our results we obtain

Theorem 3.24. *Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be regular regions so that $\mathcal{A} \Subset \mathcal{G}$. Suppose that P is of class $H(\mathcal{A}^c)$. Let $X_N = \{x_1, \ldots, x_N\}$ be an $\mathcal{E}$-unisolvent system on $\partial\mathcal{G}$. Let S_N^P denote the uniquely determined solution of the spline interpolation problem* (103). *Then there exists a constant B (dependent on $\mathcal{A}$ and $\mathcal{G}$) such that*

$$\sup_{x\in\overline{\mathcal{G}^c}} |S_N^P(x) - P(x)| \leq \ B \ \vartheta_{X_N} \ \|P\|_{H(\mathcal{A}^c)}. \tag{115}$$

As consequences we are able to deduce the following $C^{(0)}(\partial\mathcal{G})$*-closure theorem*

$$C^{(0)}(\partial\mathcal{G}) = \overline{\operatorname*{span}_{x\in X} \mathcal{E}_{|\partial\mathcal{G}}[K_{H(\mathcal{A}^c)}(\cdot,\cdot)]}^{\|\cdot\|_{C^{(0)}(\partial\mathcal{G})}} = \overline{\operatorname*{span}_{x\in X} K_{H(\mathcal{A}^c)}(x,\cdot)}^{\|\cdot\|_{C^{(0)}(\partial\mathcal{G})}}, \tag{116}$$

as well as the $L^2(\partial\mathcal{G})$*-closure theorem*

$$L^2(\partial\mathcal{G}) = \overline{\operatorname*{span}_{x\in X} \mathcal{E}_{|\partial\mathcal{G}}[K_{H(\mathcal{A}^c)}(\cdot,\cdot)]}^{\|\cdot\|_{L^2(\partial\mathcal{G})}} = \overline{\operatorname*{span}_{x\in X} K_{H(\mathcal{A}^c)}(x,\cdot)}^{\|\cdot\|_{L^2(\partial\mathcal{G})}}, \tag{117}$$

provided that X is the union of a sequence $\{X_N\}$ of $\mathcal{E}$-unisolvent systems X_N on $\partial\mathcal{G}$ with $\vartheta_{X_N} \to 0$ as $N \to \infty$.

4. Runge–Walsh solution of geodetic boundary value problems

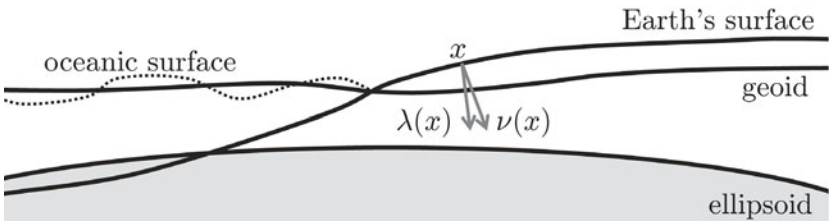


FIGURE 6. Earth's surface, geoid, ellipsoid (λ = oblique gravity vector normal to the geoid, ν = normal vector to the actual Earth's surface).

Terrestrial observations of the gravity field on the real (known) Earth's surface do not generally provide normal derivatives (cf. Figure 6). Instead, oblique derivatives are measured, since the actual Earth's surface does not coincide with the equipotential surface of the geoid (at least not for large parts over continents). In the following we are interested in discussing a locally uniform approximation implied by generalized L^2-Fourier series expansions with respect to certain trial functions (such as outer harmonics (multi-poles) or mono-pole configurations).

Classically, a solution procedure for the oblique derivative problem is undertaken by virtue of integral equations using the potential of a single-layer. These results were essentially worked out by A.V. Bitzadse [4] and C. Miranda [53] (see also the references therein). In accordance with this work, K.R. Koch, A.J. Pope [43] applied the integral equation procedure to the so-called *geodetic boundary value problem using the known surface of the Earth*. However, the strong nature of the singularities demanding Cauchy's principal integral value understanding turned out to be a serious obstacle. For numerical computation, alternative techniques have to be taken into account. The integral equation method also represents the point of departure for some subsequent work by W. Freden and H. Kersten [17–19]. They provide a new concept of approximation, viz. generalized Runge–Walsh Fourier expansions, thereby transferring strongly singular integrals into regular ones. As for the classical Dirichlet and Neumann boundary value problems (see [8, 11, 12]), the generalized Fourier series approach yields Fourier coefficients of the boundary values within the $L^2(\partial\mathcal{G})$-framework, and it simultaneously implies locally uniform approximation of the solution for subsets totally contained in the outer space. Even more, in a series of papers, [17–19, 23], and [15] successfully provide the basis for closure theorems in oblique derivative problems in different topologies such as uniform as well as Hölder norms. Additionally, [35] deal with Sobolev norms. However, to the knowledge of the authors, up to now only if the approximation of the boundary values is implemented as a generalized Fourier expansion in the L^2-context or the method of generalized Fourier series expansions is transferred as spline procedure to a reproducing kernel Hilbert $C^{(0)}$-substructure,

constructive realizations of the oblique derivative problem have been implemented successfully and efficiently to (noisy) oblique data sets (see, e.g., [36–38]).

Our work concerning the exterior Dirichlet problem leads to a remarkable conclusion: The L^2-method of generalized Fourier series expansions involving kernel functions of type (52) in $Pot(\mathcal{A}^c)$ can be seen in parallel to the minimum norm (spline) interpolation in the Sobolev-like $H(\mathcal{A}^c)$-reproducing kernel Hilbert space. Methodologically, indeed, the generalized Fourier series expansion as well as minimum norm (spline) interpolation constitute the same "Runge manifestations", however, corresponding to different topologies.

4.1. Oblique boundary value problem corresponding to the actual Earth's surface

The *oblique boundary value problem* (OBVP) can be formulated briefly as follows: Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be regular regions so that $\mathcal{A} \Subset \mathcal{G}$. Given a function F of class $C^{(0,\alpha)}(\partial\mathcal{G})$, $0 < \alpha < 1$ (i.e., Hölder-continuous with Hölder coefficient α), find a function V of class $Pot^{(1,\alpha)}(\overline{\mathcal{G}^c}) = Pot(\mathcal{G}^c) \cap C^{(1,\alpha)}(\overline{\mathcal{G}^c})$ satisfying the boundary condition

$$\frac{\partial V}{\partial \lambda}(x) = F(x), \quad x \in \partial\mathcal{G}, \tag{118}$$

where λ is a $c^{(1,\alpha)}(\partial\mathcal{G})$-(unit) vector field (i.e., a Hölder-continuous vector field with Hölder coefficient α) satisfying the physically plausible condition

$$\inf_{x \in \partial\mathcal{G}} (\lambda(x) \cdot \nu(x)) > 0 \tag{119}$$

with ν being the (unit) normal field on $\partial\mathcal{G}$ directed outward into $\mathcal{G}^c$ (cf. Figure 6).

Remark. If the field λ coincides with the normal field ν on $\partial\mathcal{G}$, Eq. (118) becomes the boundary condition of the ordinary exterior Neumann boundary value problem. In this case, we know from [17, 18] that the smoothness conditions imposed on the boundary values F may be weakened from Hölder continuity to just continuity.

In order to discuss the well-posedness of the *exterior oblique derivative problem* (*EODP*), we can follow the standard solution theory by use of the potential of a single-layer. Existence and uniqueness are recapitulated briefly in accordance with the work of A.V. Bitzadse [4] and C. Miranda [53]. Moreover, as in classical theory, we are interested in a regularity theorem providing the solution in locally uniform topology in the outer space from the L^2-approximation of the boundary values on $\partial\mathcal{G}$ (cf. [17–19]).

The point of departure for our considerations concerning $L^2(\partial\mathcal{G})$-approximation is the following result (for the proof see [18]).

Theorem 4.1. *Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be regular regions such that $\mathcal{A} \Subset \mathcal{G}$ holds true. If $\{D_n\}_{n=0,1,\ldots}$ is an $L^2(\partial\mathcal{A})$-Dirichlet Runge basis (in the sense of Definition* 3.10*),*

then

$$\mathrm{span}_{n=0,1,\ldots}\left\{\frac{\partial D_n}{\partial\lambda}\Big|_{\partial\mathcal{G}}\right\} \tag{120}$$

is dense in $(C^{(0,\alpha)}(\partial\mathcal{G}),\|\cdot\|_{L^2(\partial\mathcal{G})})$.

For numerical purposes we orthonormalize the members of an $L^2(\partial\mathcal{A})$-Dirichlet Runge basis $\{D_n\}_{n=0,1,\ldots}$ (e.g., certain systems of mono-poles (fundamental solutions), outer harmonics (multi-poles), and/or appropriate kernel functions such as Abel–Poisson kernel (55), singularity kernel (56), logarithmic kernel (57), etc.). We obtain a system $\{D_n^*\}_{n=0,1,\ldots}$, $D_n^*\in Pot(\mathcal{A}^c)$, $D_n^*\in\mathrm{span}\,(D_0,\ldots,D_n)$ satisfying the orthonormality condition

$$\left(\frac{\partial D_n^*}{\partial\lambda},\frac{\partial D_m^*}{\partial\lambda}\right)_{L^2(\partial\mathcal{G})}=\int_{\partial\mathcal{G}}\frac{\partial D_n^*}{\partial\lambda}(x)\frac{\partial D_m^*}{\partial\lambda}(x)\,dS(x)=\delta_{n,m}\,. \tag{121}$$

We are able to derive the following limit relation (see [15, 17, 18]): If $F\in C^{(0,\alpha)}(\partial\mathcal{G})$, then

$$\lim_{N\to\infty}\left(\int_{\partial\mathcal{G}}\left|F(x)-\sum_{n=0}^{N}\left(F,\frac{\partial D_n^*}{\partial\lambda}\right)_{L^2(\partial\mathcal{G})}\frac{\partial D_n^*}{\partial\lambda}(x)\right|^2\,dS(x)\right)^{\frac{1}{2}}=0. \tag{122}$$

Consequently, the uniquely determined $V\in Pot^{(1,\alpha)}(\overline{\mathcal{G}^c})$, $\frac{\partial V}{\partial\lambda}\big|_{\partial\mathcal{G}}=F$, can be approximated in the form

$$V^{(N)}=\sum_{n=0}^{N}\left(F,\frac{\partial D_n^*}{\partial\lambda}\right)_{L^2(\partial\mathcal{G})}D_n^* \tag{123}$$

with

$$\lim_{N\to\infty}\sup_{x\in\overline{\mathcal{K}}}\left|V(x)-V^{(N)}(x)\right|=0, \tag{124}$$

for every $\mathcal{K}\Subset\mathcal{G}^c$. Equivalently, $V^{(N)}$ can be obtained by use of the $L^2(\partial\mathcal{A})$-Dirichlet Runge basis $\{D_n\}_{n=0,1,\ldots}$, $D_n\in Pot(\mathcal{A}^c)$, in the form

$$V^{(N)}=\sum_{n=0}^{N}a_n^N\;D_n, \tag{125}$$

where the coefficients $a_0^N,\ldots,a_N^N$ satisfy the "normal equations"

$$\sum_{n=0}^{N}a_n^N\left(\frac{\partial D_n}{\partial\lambda},\frac{\partial D_k}{\partial\lambda}\right)_{L^2(\partial\mathcal{G})}=\left(F,\frac{\partial D_k}{\partial\lambda}\right)_{L^2(\partial\mathcal{G})},\;k=0,\ldots,N. \tag{126}$$

Next we come to the $H(\mathcal{A}^c)$-realization of the *discrete exterior oblique derivative problem* (DEODP) (see also [12, 15] and the references therein). DEODP demands to study the boundedness of the functional $\mathcal{D}_x=\partial/\partial\lambda(x)$ of the oblique

derivative $\mathcal{D} = \partial/\partial\lambda$ at a point $x \in \partial\mathcal{G}$ with respect to the $H(\mathcal{A}^c)$-topology (as introduced by Lemma 3.16). For $x \in \partial\mathcal{G}$ and $\varepsilon > 0$ we are able to conclude that

$$\frac{1}{\varepsilon}|F(x) - F(x + \varepsilon\lambda(x))| \leq \frac{1}{\varepsilon}\|F\|_{H(\mathcal{A}^c)}\|K_{H(\mathcal{A}^c)}(x,\cdot) - K_{H(\mathcal{A}^c)}(x + \varepsilon\lambda(x),\cdot)\|_{H(\mathcal{A}^c)}$$
$$\leq C\ \|F\|_{H(\mathcal{A}^c)} \tag{127}$$

for some constant $C > 0$ (depending on $\mathcal{A}^c$ and $\partial\mathcal{G}$) provided that F is of class $H(\mathcal{A}^c)$ (cf. [12]). Consequently, by the same minimum norm procedure as for the evaluation functional we obtain (thereby formally replacing $\mathcal{E}$ by $\mathcal{D}$)

Theorem 4.2. *Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be regular regions so that $\mathcal{A} \Subset \mathcal{G}$. Let P be a member of $H(\mathcal{A}^c)$ satisfying $\frac{\partial V}{\partial\lambda}(x_i) = \frac{\partial P}{\partial\lambda}(x_i) = \mathcal{D}_{x_i}[P] = \beta_i,\ i = 1,\dots,N$. Then the minimum norm interpolation problem*

$$\|S_N^P\|_{H(\mathcal{A}^c)} = \inf_{F \in \mathcal{I}^P_{\mathcal{D}_{x_1},\dots,\mathcal{D}_{x_N}}} \|F\|_{H(\mathcal{A}^c)} \tag{128}$$

with

$$\mathcal{I}^P_{\mathcal{D}_{x_1},\dots,\mathcal{D}_{x_N}} = \big\{Q \in H(\mathcal{A}^c) : \mathcal{D}_{x_i}[Q] = \mathcal{D}_{x_i}[P] = \beta_i,\ x_i \in \partial\mathcal{G},\ i = 1,\dots,N\big\}, \tag{129}$$

is well posed in the sense that its solution exists, is unique, and depends continuously on the data $\frac{\partial V}{\partial\lambda}(x_i) = \frac{\partial P}{\partial\lambda}(x_i) = \beta_i,\ i = 1,\dots,N$. The uniquely determined solution is given in the form

$$S_N^P(x) = \sum_{i=1}^{N} a_i^N \mathcal{D}_{x_i}[K_{H(\mathcal{A}^c)}(x,\cdot)], \quad x \in \mathcal{A}^c, \tag{130}$$

where the coefficients $a_1^N,\dots,a_N^N$ satisfy the linear equations

$$\sum_{i=1}^{N} a_i^N \mathcal{D}_{x_i}\mathcal{D}_{x_j}[K_{H(\mathcal{A}^c)}(\cdot,\cdot)] = \beta_j, \quad j = 1,\dots,N. \tag{131}$$

As in the Dirichlet case the *stability* should be investigated. Analogously to Lemma 3.23 we get

$$\sup_{x\in\partial\mathcal{G}} |\mathcal{D}_x[S_N^P] - \mathcal{D}_x[P]| \leq 2\ (\kappa(x,y))^{\frac{1}{2}}\ \|P\|_{H(\mathcal{A}^c)} \tag{132}$$

where

$$\kappa(x,y) = (\mathcal{D}_x\mathcal{D}_x - 2\mathcal{D}_x\mathcal{D}_y + \mathcal{D}_y\mathcal{D}_y)[K_{H(\mathcal{A}^c)}(\cdot,\cdot)], \tag{133}$$

$x \in \partial\mathcal{G}$, and $y \in X_N = \{x_1,\dots,x_N\} \subset \partial\mathcal{G}$. More explicitly, we have

$$\kappa(x,y) = \sum_{n=0}^{\infty} \sigma_n^2 \left(\frac{\partial D_n^*}{\partial\lambda}(x) - \frac{\partial D_n^*}{\partial\lambda}(y)\right)^2. \tag{134}$$

By use of (86) we find (cf. [15]) that

$$\left|\frac{\partial D_n^*}{\partial\lambda}(x) - \frac{\partial D_n^*}{\partial\lambda}(y)\right| \leq C\ |x - y|, \tag{135}$$

where $C > 0$ is some constant (depending on $\mathcal{A}^c$ and $\partial\mathcal{G}$).

Theorem 4.3. *Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be regular regions so that $\mathcal{A} \Subset \mathcal{G}$. Suppose that P is of class $H(\mathcal{A}^c)$. Let $X_N = \{x_1, \dots, x_N\}$ be a $\mathcal{D}$-unisolvent system on $\partial\mathcal{G}$. Let S_N^P denote the uniquely determined solution of the spline interpolation problem* (128). *Then there exists a constant B (dependent on $\mathcal{A}$ and $\mathcal{G}$) such that*

$$\sup_{x \in \overline{\mathcal{G}^c}} |S_N^P(x) - P(x)| \leq \; B \; \vartheta_{X_N} \; \|P\|_{H(\mathcal{A}^c)}. \tag{136}$$

Obviously,

$$C^{(0)}(\partial\mathcal{G}) = \overline{\operatorname*{span}_{x \in X} \mathcal{D}|_{\partial\mathcal{G}}[K_{H(\mathcal{A}^c)}(\cdot,\cdot)]}^{\|\cdot\|_{C^{(0)}(\partial\mathcal{G})}} \tag{137}$$

and

$$L^2(\partial\mathcal{G}) = \overline{\operatorname*{span}_{x \in X} \mathcal{D}|_{\partial\mathcal{G}}[K_{H(\mathcal{A}^c)}(\cdot,\cdot)]}^{\|\cdot\|_{L^2(\partial\mathcal{G})}}, \tag{138}$$

where X is the union of a sequence $\{X_N\}$ of $\mathcal{D}$-unisolvent systems X_N on $\partial\mathcal{G}$ with $\vartheta_{X_N} \to 0$ as $N \to \infty$.

For computational reasons, reproducing Hilbert space kernels with closed expressions in terms of elementary functions are welcome (see, e.g., [10, 12, 68] for more details). For that purpose, we specialize to an inner Runge-ball $\mathbb{B}_{\underline{R}_0}$ around the origin, so that $\mathbb{B}_{\underline{R}_0} \Subset \mathcal{A} = \mathbb{B}_{\underline{R}}(0)$, $\underline{R} < \inf_{x \in \partial\mathcal{G}} |x|$. Moreover, in practical applications, an outer harmonics $L^2(\partial\mathcal{A})$-Dirichlet Runge basis is of frequent use at least in physical geodesy (see, e.g., [10, 12, 44, 55, 57]). In fact, via the addition theorem of spherical harmonics, a large number of representations can be derived from series expansions in terms of Legendre polynomials:

$$K_{H\left(\mathbb{R}^3 \setminus \overline{\mathbb{B}_{\underline{R}}(0)}\right)}(x,y) = \sum_{n=0}^{\infty} \sigma_n^2 \, \frac{2n+1}{4\pi \underline{R}^2} \left(\frac{\underline{R}^2}{|x||y|}\right)^{n+1} P_n\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right), \; x, y \in \mathbb{R}^3 \setminus \overline{\mathbb{B}_{\underline{R}}(0)}. \tag{139}$$

In this approach we restrict ourselves to three important cases.

Example 4.4.

$$\sigma_n^2 = \left(\frac{\underline{R}_0^2}{\underline{R}^2}\right)^n, \quad \underline{R}_0 < \underline{R}, \quad n = 0, 1, \dots \tag{140}$$

Abel–Poisson kernel:

$$K_{H\left(\mathbb{R}^3 \setminus \overline{\mathbb{B}_{\underline{R}}(0)}\right)}(x,y) = \frac{|x|\,|y|}{4\pi \underline{R}_0^2} \frac{|x|^2|y|^2 - \underline{R}_0^4}{(L(x,y))^{\frac{3}{2}}} \tag{141}$$

with

$$L(x,y) = |x|^2|y|^2 - 2\underline{R}_0^2 \, x \cdot y + \underline{R}_0^4 \tag{142}$$

and

$$\begin{aligned}
&\frac{\partial}{\partial\lambda_x}\frac{\partial}{\partial\lambda_y}K_{H(\mathbb{R}^3\setminus\overline{\mathcal{B}_{\underline{R}}(0)})}(x,y)\\
&\quad=\frac{1}{4\pi\underline{R}_0^2}\left(\frac{9|x|^2|y|^2-\underline{R}_0^4}{(L(x,y))^{\frac{3}{2}}}\frac{(\lambda(x)\cdot x)(\lambda(y)\cdot y)}{|x|\ |y|}\right)\\
&\qquad+\frac{3}{4\pi\underline{R}_0{}^2}\frac{3|x|^2|y|^2-\underline{R}_0{}^4}{(L(x,y))^{\frac{5}{2}}}\left(\frac{(\lambda(x)\cdot x)|y|}{|x|}(\underline{R}_0{}^2(\lambda(y)\cdot x)-(\lambda(y)\cdot y)|x|^2)\right)\\
&\qquad+\frac{3}{4\pi\underline{R}_0{}^2}\frac{3|x|^2|y|^2-\underline{R}_0{}^4}{(L(x,y))^{\frac{5}{2}}}\left(\frac{(\lambda(y)\cdot y)|x|}{|y|}(\underline{R}_0{}^2(\lambda(x)\cdot y)-(\lambda(x)\cdot x)|y|^2)\right)\\
&\qquad+\frac{3}{4\pi\underline{R}_0{}^2}\frac{|x|^2|y|^2-\underline{R}_0{}^4}{(L(x,y))^{\frac{5}{2}}}|x|\,|y|\left(\underline{R}_0{}^2(\lambda(x)\cdot\lambda(y))-2(\lambda(y)\cdot y)(\lambda(x)\cdot x)\right)\\
&\qquad+\frac{15}{4\pi\underline{R}_0{}^2}\frac{|x|^2|y|^2-\underline{R}_0{}^4}{(L(x,y))^{7/2}}|x|\,|y|(\underline{R}_0{}^2(\lambda(x)\cdot y)-(\lambda(x)\cdot x)|y|^2)\\
&\qquad\qquad\times(\underline{R}_0{}^2(\lambda(y)\cdot x)-(\lambda(y)\cdot y)|x|^2).
\end{aligned}\tag{143}$$

Example 4.5.

$$\sigma_n^2=\frac{2}{2n+1}\left(\frac{\underline{R}_0^2}{\underline{R}^2}\right)^n,\quad \underline{R}_0<\underline{R},\quad n=0,1,\ldots \tag{144}$$

Singularity kernel:

$$K_{H(\mathbb{R}^3\setminus\overline{\mathcal{B}_{\underline{R}}(0)})}(x,y)=\frac{1}{2\pi}\frac{1}{(L(x,y))^{\frac{1}{2}}} \tag{145}$$

and

$$\begin{aligned}
&\frac{\partial}{\partial\lambda_x}\frac{\partial}{\partial\lambda_y}K_{H(\mathbb{R}^3\setminus\overline{\mathcal{B}_{\underline{R}}(0)})}(x,y)\\
&\quad=\frac{1}{2\pi}\frac{1}{(L(x,y))^{\frac{3}{2}}}(\underline{R}_0^2(\lambda(x)\cdot\lambda(y))-2(\lambda(x)\cdot x)(\lambda(y)\cdot y))\\
&\qquad+\frac{3}{2\pi}\frac{1}{(L(x,y))^{\frac{3}{2}}}(\underline{R}_0^2(\lambda(x)\cdot y)-(\lambda(x)\cdot x)|y|^2)(\underline{R}_0^2(\lambda(y)\cdot x)-(\lambda(y)\cdot y)|x|^2).
\end{aligned}\tag{146}$$

Example 4.6.

$$\sigma_n^2=\frac{1}{(2n+1)(n+1)}\left(\frac{\underline{R}_0^2}{\underline{R}^2}\right)^n,\quad \underline{R}_0<\underline{R},\quad n=0,1\ldots \tag{147}$$

Logarithmic kernel:

$$K_{H(\mathbb{R}^3\setminus\overline{\mathcal{B}_{\underline{R}}(0)})}(x,y)=\frac{1}{4\pi\underline{R}_0^2}\ln\left(1+\frac{2\underline{R}_0^2}{M(x,y)}\right) \tag{148}$$

with

$$M(x,y)=(L(x,y))^{\frac{1}{2}}+|x|\,|y|-\underline{R}_0^2 \tag{149}$$

and

$$\begin{aligned}
&\frac{\partial}{\partial\lambda_x}\frac{\partial}{\partial\lambda_y}K_{H(\mathbb{R}^3\setminus\overline{\mathcal{B}_{\underline{R}}(0)})}(x,y)\\
&=\frac{1}{2\pi}\frac{1}{(M(x,y))^2+2\underline{R}_0^2M(x,y)}\\
&\quad\times\left[(L(x,y))^{-\frac{3}{2}}(\underline{R}_0^2(\lambda(y)\cdot x)-|x|^2(\lambda(y)\cdot y))(\underline{R}_0^2(n(x)\cdot y)-|y|^2(\lambda(x)\cdot x))\right]\\
&\quad+\frac{1}{2\pi}\frac{1}{(M(x,y))^2+2\underline{R}_0^2M(x,y)}\\
&\quad\times\left[(L(x,y)^{-\frac{1}{2}}(\underline{R}_0^2(\lambda(x)\cdot\lambda(y))-2(\lambda(x)\cdot x)(\lambda(y)\cdot y)))-\frac{(\lambda(x)\cdot x)(\lambda(y)\cdot y)}{|x||y|}\right]\\
&\quad+\frac{1}{\pi}\frac{M(x,y)+\underline{R}_0^2}{((M(x,y))^2+2\underline{R}_0^2M(x,y))^2}\\
&\quad\times\left[(L(x,y))^{-\frac{1}{2}}(|x|^2(\lambda(y)\cdot y)-\underline{R}_0^2(\lambda(y)\cdot x))\frac{|x|}{|y|}(\lambda(y)\cdot y)\right]\\
&\quad\times\left[(L(x,y))^{-\frac{1}{2}}(|y|^2(\lambda(x)\cdot x)-\underline{R}_0^2(\lambda(x)\cdot y))+\frac{|y|}{|x|}(\lambda(x)\cdot x)\right].
\end{aligned}\tag{150}$$

Example 4.7.

$$\sigma_n^2=\frac{1}{\underline{R}_0^3}\frac{1}{(2n+1)^2(2n+3)}\left(\frac{\underline{R}_0^2}{\underline{R}^2}\right)^n,\quad \underline{R}_0<\underline{R},\quad n=0,1,\ldots\tag{151}$$

Newton kernel:

$$K_{H(\mathbb{R}^3\setminus\overline{\mathcal{B}_{\underline{R}}(0)})}(x,y)=\left(\frac{1}{4\pi}\right)^2\int_{\mathcal{B}_{\underline{R}_0}(0)}\frac{1}{|x-z||y-z|}\,dV(z)\tag{152}$$

and

$$\frac{\partial}{\partial\lambda_x}\frac{\partial}{\partial\lambda_y}K_{H(\mathbb{R}^3\setminus\overline{\mathcal{B}_{\underline{R}}(0)})}(x,y)=\left(\frac{1}{4\pi}\right)^2\int_{\mathcal{B}_{\underline{R}_0}(0)}\frac{\lambda(x)\cdot(x-z)\;\;\lambda(y)\cdot(y-z)}{|x-z|^3|y-z|^3}\,dV(z).\tag{153}$$

In other words, the iterated Newton kernel leads back to the volume-based reproducing kernel Hilbert space structure recently developed by W. Freeden, C. Gerhards [15].

Finally it should be noted that the advantage of a sphere-based reproducing kernel Hilbert space $(H(\mathcal{A}^c),(\cdot,\cdot)_{H(\mathcal{A}^c)})$ is twofold:

(1) The reproducing kernel contains outer harmonic contributions of any degree like the Earth's gravitational potential itself.
(2) The geometry of the regular region $\mathcal{G}$ may be arbitrary so that especially the actual Earth's surface $\partial\mathcal{G}$ can be easily handled in numerical computations thereby taking advantage from the fact that there is no need for numerical integration. The coefficient matrix of the occurring linear (spline) systems is

symmetric and positive definite, hence, they are solvable by standard methods of linear algebra.

Even better, multi-pole (far and near field) methods in combination with suitable domain decomposition procedures (see [36–38] and the references therein) make spline interpolation (and/or smoothing in the case of error affected data) an efficient as well as economical technique for numerical application.

Nevertheless, it should be mentioned that the particular choice of the reproducing kernel, i.e., the appropriate topology of $H(\mathcal{A}^c)$ is a problem in minimum norm (spline) interpolation. In principle, seen from a theoretical point of view, all topologies are equivalent. In practice, however, the reproducing kernel structure should be in adaptation to the characteristics of the available dataset (if possible).

Altogether, Runge-type spline interpolation is a constructive method for solving the oblique derivative problem. The difficulties are the suitable choice of an inner Runge-sphere and the Sobolev structure, the positioning of the point systems on $\partial\mathcal{G}$, and the efforts to solve the occurring linear systems.

4.2. Molodensky boundary value problem in physical geodesy

The gravimetric determination of the geoid is a current research area in physical geodesy. It has become even more important, since the GPS techniques deliver accurate measurements with dense data coverage. In particular, for geodetic purposes, locally reflected approximation methods resulting in high wavelength geoidal reconstructions are of future significance.

The original problem of Molodensky can briefly be formulated as follows: Given, at all points on the geoid $\partial\mathcal{G}$, the gravity potential W and the gravity vector $w = \nabla W$, then the aim is to determine the geoidal surface $\partial\mathcal{G}$. It is clear by the definition of the geoid, that W is constant on $\partial\mathcal{G}$, such that only a gauge value W_0 has to be given. Furthermore we will not discuss in detail here, how the gravity vector w is obtained on $\partial\mathcal{G}$ from measurements on the real Earth's surface. For a detailed discussion of determining w from the Earth's surface to the (a priori not known) geoid, the reader is referred to the literature (an important approach is given, e.g., by L. Hörmander [40], see also the references therein). Our description of the linearized Molodensky problem essentially follows the conventional concept of, e.g., [33, 34, 45, 54].

The geoidal height determination is based on the fact that the geoid $\partial\mathcal{G}$ is approximated by the boundary $\partial\mathcal{T}$ of a regular region $\mathcal{T}$ called the *telluroid* with known gravitational potential U in $\overline{\mathcal{T}^c}$ (cf. Figure 7). We assume that there exists a one-to-one correspondence between $\partial\mathcal{G}$ and $\partial\mathcal{T}$. W is the actual potential and U is an approximation of W called the *normal potential*. We define $u = \nabla U$ which is called the *normal gravity* and $w = \nabla W$ called the *actual gravity* which is given on $\partial\mathcal{G}$. Assume that, for given $x \in \partial\mathcal{T}$, the point $y \in \partial\mathcal{G}$ is the one associated to x by the one-to-one correspondence between $\partial\mathcal{G}$ and $\partial\mathcal{T}$ (cf. Figure 7). The two points are connected by the vector $d = y - x$. A substitute formulation of the classical Molodensky problem is now to determine the length of d, i.e., the distance of the geoid and the approximating telluroid along the one-to-one correspondence

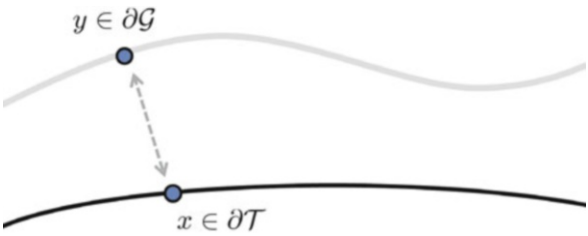


FIGURE 7. Geoid $\partial\mathcal{G}$, telluroid $\partial\mathcal{T}$, and their one-to-one correspondence.

between $\partial\mathcal{G}$ and $\partial\mathcal{T}$. To this end we introduce

$$\delta W = W|_{\partial\mathcal{G}} - U|_{\partial\mathcal{T}}, \tag{154}$$

$$\delta w = w|_{\partial\mathcal{G}} - u|_{\partial\mathcal{T}}, \tag{155}$$

where δW is called the *potential anomaly* and δw is called *gravity anomaly* (see [44, 45, 54]). Furthermore, we define the *disturbing potential* T by

$$T = W - U \tag{156}$$

in $\overline{\mathcal{G}}$, so that we have

$$\delta W = T|_{\partial\mathcal{G}} + U|_{\partial\mathcal{G}} - U|_{\partial\mathcal{T}}, \tag{157}$$

$$\delta w = w|_{\partial\mathcal{G}} - u|_{\partial\mathcal{T}}\,. \tag{158}$$

Using the Taylor expansion of u and U in terms of d and neglecting terms of higher order in d (which represents no substantial loss of accuracy if a sufficiently close telluroid is chosen) we finally arrive at the approximations

$$\delta W(x) = T(x) + u(x) \cdot d, \tag{159}$$

$$\delta w(x) = w(y) - u(y) + \mathbf{m}(x)\, d, \tag{160}$$

$x \in \partial\mathcal{T}$, $y \in \partial\mathcal{G}$, where we set

$$\mathbf{m} = \nabla u = \left(\frac{\partial^2 U}{\partial x_i \partial x_j} \right)_{i,j=1,2,3}. \tag{161}$$

Observing the relations

$$w(y) - u(y) = (\nabla W)(y) - (\nabla U)(y) = (\nabla T)(y) = (\nabla T)(x) \tag{162}$$

we arrive at

$$\delta W(x) = T(x) + u(x) \cdot d, \tag{163}$$

$$\delta w(x) = (\nabla T)(x) + \mathbf{m}(x)\, d. \tag{164}$$

Equation (163) is called *Bruns formula*. Actually it connects the disturbing potential T on the telluroid $\partial\mathcal{T}$ with the *geoid anomalies* d, i.e., the anomalies between

the geoid $\partial\mathcal{G}$ and the telluroid $\partial\mathcal{T}$. If we assume that $\mathbf{m}(x)$ is invertible for all $x \in \partial\mathcal{T}$, we get by virtue of (162)

$$d = \mathbf{m}(x)^{-1}(\delta w(x) - (\nabla T)(x)). \tag{165}$$

Inserting the identity (165) into equation (163) we end up with

$$T(x) - u(x) \cdot (\mathbf{m}(x))^{-1}(\nabla T)(x) = \delta W(x) - u(x) \cdot \mathbf{m}(x)^{-1}\delta w(x). \tag{166}$$

This is the so-called *fundamental boundary condition of physical geodesy.*

Following [45] (see also the report [54]) the vector $u(x)(\mathbf{m}(x))^{-1}$ can be seen in first order of d to be oriented in the direction of the exterior unit normal field ν on the telluroid $\partial\mathcal{T}$. More specifically,

$$u(x)(\mathbf{m}(x))^{-1} = -\frac{|x|}{2}\,\nu(x). \tag{167}$$

Inserting expression (167) into equation (166) therefore results in the identity

$$\nu(x) \cdot (\nabla T)(x) + \frac{2}{|x|}\,T(x) = F(x), \tag{168}$$

where we have used the abbreviation

$$F(x) = \nu(x) \cdot \delta w(x) + \frac{2}{|x|}\,\delta W(x) \tag{169}$$

(note that the boundary condition (168) can be seen to be equivalent to (166) transformed in an appropriate coordinate system).

Summarizing all the steps of the linearization procedure we are led to discuss the following type of a boundary value problem in potential theory. In fact, our goal is to solve this problem by a constructive Runge approach as presented in this work (cf. [22]).

Exterior Molodensky Problem (*EMP*)*:* Find $T \in Pot^{(1)}(\overline{\mathcal{T}^c})$, i.e., $T \in C^{(2)}(\mathcal{T}^c) \cap C^{(1)}(\overline{\mathcal{T}^c})$ with $\Delta T = 0$ in $\mathcal{T}^c$ and $|T(x)| = O(|x|^{-1})$, $|x| \to \infty$, such that

$$\frac{\partial T}{\partial \nu}(x) + \mu(x)T(x) = F(x), \quad x \in \partial\mathcal{T}, \tag{170}$$

where μ, $F \in C^{(0)}(\partial\mathcal{T})$ are known functions on the boundary surface $\partial\mathcal{T}$ of the regular region $\mathcal{T}$ (it should be noted that in modern mathematical nomenclature, the exterior Molodensky problem (EDP) forms a special Robin problem).

Remark. In the case that $\partial\mathcal{T}$ is a sphere, the problem becomes the well-known Stokes problem (see [39] or [54]) and in the case of an ellipsoid it is called ellipsoidal Stokes problem (see, e.g., [33, 52, 54]). Locally reflected multi-scale solutions of Stokes' problem are due to [27, 29] (see also the references in [15]).

Next, we discuss the well-posedness of the Molodensky boundary value problem corresponding to a regular telluroidal surface $\partial\mathcal{T}$. First, we will reformulate the problem in our notation.

Exterior Molodensky Problem (EMP): Given $F, \mu \in C^{(0)}(\partial\mathcal{T})$, find $T \in Pot^{(1)}(\overline{\mathcal{T}^c})$ such that

$$\left(\frac{\partial T}{\partial \nu} + \mu T\right)(x) = F(x), \quad x \in \partial\mathcal{T}. \tag{171}$$

From [22] we borrow the *regularity theorem* in the $L^2(\partial\mathcal{T})$-context for the Molodensky problem.

Theorem 4.8. *Let T be of class $Pot^{(1)}(\overline{\mathcal{T}^c})$. Then there exists a constant $C(= C(k; \mathcal{K}, \partial\mathcal{T}))$ such that*

$$\sup_{x\in\mathcal{K}} \left|\left(\nabla^{(k)} T\right)\right|(x) \leq C \left(\left\|\frac{\partial T}{\partial \nu}\right\|_{L^2(\partial\mathcal{T})} + \|\mu\|_{L^2(\partial\mathcal{T})}\ \|T\|_{L^2(\partial\mathcal{T})}\right) \tag{172}$$

for all $\mathcal{K} \Subset \mathcal{T}^c$ and all $k \in \mathbb{N}_0$.

For numerical purposes in the sense of the Runge–Walsh approach we again orthonormalize the members of a Dirichlet Runge basis $(D_n)_{n=0,1,\ldots}$ obtaining a system $\{D_n^*\}_{n=0,1,\ldots}$, $D_n^* \in Pot(\mathcal{A}^c)$, $D_n^* \in \text{span}\,(D_0, \ldots, D_n)$, satisfying the orthonormality condition

$$\left(\frac{\partial D_n^*}{\partial \nu} + \mu D_n^*, \frac{\partial D_m^*}{\partial \nu} + \mu D_m^*\right)_{L^2(\partial\mathcal{T})} = \delta_{n,m}. \tag{173}$$

In connection with the regularity result we are then able to derive the following conclusion in the framework of the Molodensky problem: If $F \in C^{(0)}(\partial\mathcal{T})$, then

$$\lim_{N\to\infty}\left(\int_{\partial\mathcal{T}}\left|F(x) - \sum_{n=0}^{N}\left(F, \left(\frac{\partial}{\partial \nu} + \mu\right) D_n^*\right)_{L^2(\partial\mathcal{T})}\right.\right.$$
$$\left.\left.\times \left(\frac{\partial}{\partial \nu} + \mu(x)\right) D_n^*(x)\right|^2 dS(x)\right)^{1/2} = 0. \tag{174}$$

Consequently, the potential $T \in Pot^{(1)}(\overline{\mathcal{T}^c})$, $\left(\frac{\partial}{\partial \nu} + \mu\right) T = F$ on $\partial\mathcal{T}$, can be represented in the form

$$\lim_{N\to\infty} \sup_{x\in\mathcal{K}} \left|T(x) - T^{(N)}(x)\right| = 0, \tag{175}$$

where

$$T^{(N)} = \sum_{n=0}^{N}\left(F, \left(\frac{\partial}{\partial \nu} + \mu\right) D_n^*\right)_{L^2(\partial\mathcal{T})} D_n^*. \tag{176}$$

Finally it should be noted that the whole solution context developed for the discrete exterior oblique derivative problem (DEOP) also remains valid for the *discrete exterior Molodensky problem* (DEMP) in an obvious way by using the linear functional $\mathcal{M} = \frac{\partial}{\partial \nu} + \mu$ instead of $\mathcal{D} = \frac{\partial}{\partial \lambda}$. We summarize the results.

Theorem 4.9. *Let $\mathcal{A}, \mathcal{G} \subset \mathbb{R}^3$ be regular regions so that $\mathcal{A} \Subset \mathcal{G}$. Let P be a member of $H(\mathcal{A}^c)$ satisfying $\mathcal{M}_{x_i}[P] = \beta_i,\ i = 1, \ldots, N$. Then the minimum norm interpolation problem*

$$\|S_N^P\|_{H(\mathcal{A}^c)} = \inf_{F \in \mathcal{I}^P_{\mathcal{M}_{x_1}, \ldots, \mathcal{M}_{x_N}}} \|F\|_{H(\mathcal{A}^c)} \tag{177}$$

with

$$\mathcal{I}^P_{\mathcal{M}_{x_1}, \ldots, \mathcal{M}_{x_N}} = \big\{Q \in H(\mathcal{A}^c) : \mathcal{M}_{x_i}[Q] = \mathcal{M}_{x_i}[P] = \beta_i,\ x_i \in \partial\mathcal{G}, \quad i = 1, \ldots, N\big\}, \tag{178}$$

is well posed in the sense that its solution exists, is unique, and depends continuously on the data $\mathcal{M}(V)(x_i) = \mathcal{M}(P)(x_i) = \beta_i,\ i = 1, \ldots, N$. The uniquely determined solution is given in the form

$$S_N^P(x) = \sum_{i=1}^{N} a_i^N \mathcal{M}_{x_i}[K_{H(\mathcal{A}^c)}(x, \cdot)], \quad x \in \mathcal{A}^c, \tag{179}$$

where the coefficients $a_1^N, \ldots, a_N^N$ satisfy the linear equations

$$\sum_{i=1}^{N} a_i^N \mathcal{M}_{x_i} \mathcal{M}_{x_j}[K_{H(\mathcal{A}^c)}(\cdot, \cdot)] = \beta_j, \quad j = 1, \ldots, N. \tag{180}$$

A multi-scale (Runge-type) method for solving (DEMP) obtained by regularization of layer potentials can be found in [21].

Once again, we notice that the exterior Molodensky problem leads to the remarkable conclusion that the L^2-method of generalized Fourier series expansions involving kernel functions of type (52) in $Pot(\mathcal{A}^c)$ can be recognized in parallel to the minimum norm (spline) interpolation in the Sobolev-like $H(\mathcal{A}^c)$-reproducing kernel Hilbert space. More explicitly, the generalized Fourier series expansion as well as minimum norm (spline) interpolation constitute the same "Runge manifestations" of solutions for the Molodensky problem, however, corresponding to different topologies. Nevertheless, the numerical realization is based on different data assumptions. In case of the L^2-method of generalized Fourier series expansions the data set has to be equidistributed all over the boundary, since integration and equidistribution are mutually dependent according to the famous Weyl law (cf. [72]). Spline interpolation may be performed to discrete points, but the stability of the solution of the occurring linear system and the prevention of oscillations of the spline approximations imply a reasonable data structure avoiding larger gaps of the data distribution.

5. Conclusion

This contribution provides constructive realizations of the Runge–Walsh theorem in order to solve geodetic boundary value problems such as the exterior oblique derivative problem from discrete data sets. The numerical methods proposed here, respectively, are $Pot(\mathcal{A}^c)$- and $H(\mathcal{A}^c)$-generalized Fourier series expansions. Particular kernels serving as trial functions for use in our numerics are mono- and multi-poles and their Kelvin transforms relative to a "Runge sphere", i.e., the boundary of a Runge ball. A mono-pole interrelates the length of its spectral bands to the distance of the mono-pole from the Runge sphere. The mono-pole, i.e., the fundamental solution (as well as its Kelvin transformed singularity kernel) is more and more space localized and simultaneously in accordance with the uncertainty principle (cf. [14]) less frequency localized, the closer and closer the mono-pole is positioned to the Runge sphere. As a matter of fact, seen from a methodological point of view, $Pot(\mathcal{A}^c)$-Fourier approaches using a sequence of kernel functions corresponding to an inner fundamental system can be realized in a manner equivalent to outer harmonic expansions for completely recovering the gravitational potential within the framework of boundary value problems. A sequence of kernel functions (such as the Abel–Poisson, singularity, and logarithmic kernel) is even conceptually easier to implement than outer harmonic expansions, as long as the kernels are available in closed form as elementary functions.

Mono-pole, i.e., fundamental solution approximations have a long history. Early attempts in potential theory to make the so-called *method of fundamental solutions* reality date back to the middle of the 19th century (cf. [66]). Related studies are due to [61, 67]. Further ideas are, e.g., due to [31, 42, 46, 69, 70]. The line to the Fourier approach as presented here follows [8, 11, 15, 20, 21, 23, 32, 69, 70]. All these approaches take advantage of the Kelvin transform in potential theory that is not transferable for more general elliptic partial differential equations.

In the meantime, however, generalized Fourier expansions are theoretically established and practically applied not only to the Laplace equation, but also to more general elliptic partial differential equations, e.g., the reduced (Helmholtz) wave equation (see [25, 28]), the Cauchy–Navier equation (see [1, 24]), (reduced) Maxwell equations (see [22]), the (linear) Stokes equations (see [51] and the references therein). [2] used the method of fundamental solutions in poroelasticity to model stress fields. The drawback of the numerical realization is the need for an adequate selection of a finite number of points out of the infinite inner fundamental system. An optimal strategy for positioning a finite system in a computationally efficient and physically relevant way remains a great challenge for future work.

References

[1] Abeyratne, M.K., Freeden, W., Mayer, C.: Multiscale Deformation Analysis by Cauchy–Navier Wavelets. *J. Appl. Math.*, **12**, 605–645, 2003.

[2] Augustin, M.A.: *A Method of Fundamental Solutions in Poroelasticity to Model the Stress Field in Geothermal Reservoirs*. Lecture Notes in Geosystems Mathematics and Computing, Birkhäuser, Basel, 2015.

[3] Bjerhammar, A.: *Gravity Reduction to an Internal Sphere*. Division of Geodesy, Stockholm, Sweden, 1962.

[4] Bitzadse, A.V.: *Boundary-Value Problems for Second-Order Elliptic Equations*. North-Holland, Amsterdam, 1968.

[5] Colombo, O.L.: The Convergence of the External Spherical Harmonic Expansion of the Gravitational Potential. *Boll. Geod. Sci. Affin.*, **42**, 221–239, 1983.

[6] Davis, P.J.: *Interpolation and Approximation*. Blaisdell Publishing Company, Waltham, MA, 1963.

[7] Freeden W.: *Über eine Klasse von Integralformeln der Mathematischen Geodäsie*. Veröff. *Geod. Inst. RWTH Aachen*, Report **27**, 1979.

[8] Freeden, W.: On the Approximation of External Gravitational Potential with Closed Systems of (Trial) Functions. *Bull. Géod.*, **54**, 1–20, 1980.

[9] Freeden, W.: On Spherical Spline Interpolation and Approximation. *Math. Meth. in the Appl. Sci.*, **3**, 551–575, 1981.

[10] Freeden, W.: On Approximation by Harmonic Splines. *Manuscr. Geod.*, **6**, 193-244, 1981.

[11] Freeden, W.: *Least Squares Approximation by Linear Combinations of (Multi)-Poles*. Scientific Report, No. **344**, The Ohio State University, Department of Geodetic Science, Columbus, 1983.

[12] Freeden, W.: A Spline Interpolation Method for Solving Boundary Value Problems of Potential Theory from Discretely Given Data. *Math. Part. Diff. Equations*, **3**, 375–398, 1987.

[13] Freeden, W.: *Multiscale Modelling of Spaceborne Geodata*. B.G. Teubner, Leipzig, 1999.

[14] Freeden, W.: The Uncertainty Principle and Its Role in Physical Geodesy. Progress in Geodetic Science, 225–237, Shaker, Aachen, 1999.

[15] Freeden, W., Gerhards, C.: *Geomathematically Oriented Potential Theory*. Chapman and Hall/CRC Press, Boca Raton, 2013.

[16] Freeden, W., Gutting, M.: *Special Functions of Mathematical (Geo-)Physics*. Birkhäuser, Basel, 2013.

[17] Freeden, W., Kersten, H.: *The Geodetic Boundary-Value Problem Using the Known Surface of the Earth*. Veröff. Geod. Inst. RWTH Aachen, No. **29**, 1980.

[18] Freeden, W., Kersten, H.: A Constructive Approximation Theorem for the Oblique Derivative Problem in Potential Theory. *Math. Meth. Appl. Sci.* **3**, 104–114, 1981.

[19] Freeden, W., Kersten, H.: An Extended Version of Runge's Theorem. *Manuscr. Geod.*, **7**, 267–278, 1982.

[20] Freeden, W., Mayer, C.: Wavelets Generated by Layer Potentials. *Appl. Comput. Harm. Anal.* (*ACHA*), **14**, 195–237, 2003.

[21] Freeden, W., Mayer, C.: Multiscale Solution for the Molodenski Problem on Regular Telluroidal Surfaces. *Acta Geod. Geophys. Hung.*, **41**, 55–86, 2006.

[22] Freeden, W., Mayer, C.: Modeling Tangential Vector Fields on Regular Surfaces by Means of Mie Potentials. *Int. J. Wavel. Multires. Inf. Process.* **5**, 417–449, 2007.

[23] Freeden, W., Michel, V.: *Multiscale Potential Theory* (*with Applications to Geoscience*). Birkhäuser, Boston, 2004.

[24] Freeden, W., Reuter, R.: A Constructive Method for Solving the Displacement Boundary-Value Problem of Elastostatics by Use of Global Basis Systems. *Math. Meth. Appl. Sci.*, **12**, 105–128, 2011.

[25] Freeden, W., Schneider, F.: Runge–Walsh Wavelet Approximation for the Helmholtz Equation. *J. Math. Anal. Appl.*, **235**, 533-566, 1999.

[26] Freeden, W., Schreiner, M.: *Spherical Functions of Mathematical Geosciences* (*a Scalar, Vectorial, and Tensorial Setup*). Springer, Heidelberg, 2009.

[27] Freeden, W., Wolf, K.: Klassische Erdschwerefeldbestimmung aus der Sicht moderner Geomathematik. *Math. Semesterb.*, **56**, 53–77, 2008.

[28] Freeden, W., Mayer, C., Schreiner, M.: Tree Algorithms in Wavelet Approximation by Helmholtz Potential Operators. *Num. Funct. Anal. Optim.*, **24**, 747–782, 2003.

[29] Freeden, W., Fehlinger, T., Klug, M., Mathar, D., Wolf, K.: Classical Globally Reflected Gravity Field Determination in Modern Locally Oriented Multiscale Framework. *J. of Geodesy*, **83**, 1171–1191, 2009.

[30] Gauss, C.F.: *Allgemeine Theorie des Erdmagnetismus*. Resultate aus den Beobachtungen des magnetischen Vereins, Göttingen, 1838.

[31] Golberg, M.: The Method of Fundamental Solutions for Poisson's Equation. *Eng. Anal. Bound. Elem.* **16**, 205–213, 1995.

[32] Golberg, M., Chen, C.: The Method of Fundamental Solutions for Potential, Helmholtz, and Diffusion Problems. In: *Boundary Integral Methods: Numerical and Mathematical Aspects* (Goldberg, M., ed.), 103–176, WIT Press, Computational Mechanics Publications, Southampton, 1998.

[33] Grafarend, E.W.: The Reference Figure of the Rotating Earth in Geometry and Gravity Space and an Attempt to Generalize the Celebrated Runge–Walsh Approximation Theorem for Irregular Surfaces. *GEM Int. J. Geomath.*, **6**, 101–140, 2015

[34] Grafarend, E.W., Klapp, M., Martinec. Z.: Spacetime Modelling of the Earth's Gravity Field by Ellipsoidal Harmonics. In: *Handbook of Geomathematics, Vol.* 1 (Freeden, W., Nashed, M.Z., Sonar, T., Eds.), 2nd edition, 381–496, Springer, New York, 2015.

[35] Grothaus, M., Raskop, T.: Oblique Stochastic Boundary-Value Problem. In: *Handbook of Geomathematics, Vol.* 3 (Freeden, W., Nashed, M.Z., Sonar, T., Eds.), 2nd edition, 2285–2316, Springer, New York, 2015.

[36] Gutting, M., *Fast Multipole Methods for Oblique Derivative Problems*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen, 2007.

[37] Gutting, M.: Fast Multipole Accelerated Solution of the Oblique Derivative Boundary-Value Problem. *GEM Int. J. Geomath.*, **3**, 223–252, 2012.

[38] Gutting, M.: Fast Spherical/Harmonic Spline Modeling. In: *Handbook of Geomathematics, Vol.* 3 (Freeden, W., Nashed, M.Z., Sonar, T., Eds.), 2nd edition, 2711–2746, Springer, New York, 2015.

[39] Heiskanen, W.A., Moritz, H.: *Physical Geodesy*. W.H. Freeman and Company (1967).

[40] Hörmander, L.: *The Boundary Problems of Physical Geodesy*. The Royal Institute of Technology, Division of Geodesy, Stockholm, Report **9**, 1975 (also: *Arch. Rat. Mech. Anal.*, **62**, 1–52, 1976).

[41] Kellogg, O.D.: *Foundations of Potential Theory*. Frederick Ungar Publishing Company, New York, 1929.

[42] Kita, E., Kamiya, N.: Trefftz Method: An Overview. *Adv. Eng. Softw.*, **24**, 3–13, 1995.

[43] Koch, K.R., Pope, A.J.: Uniqueness and Existence for the Geodetic Boundary Value Problem Using the Known Surface of the Earth. *Bull. Géod.*, **106**, 467–476, 1972.

[44] Krarup, T.: *A Contribution to the Mathematical Foundation of Physical Geodesy*. Danish Geodetic Institute, Report No. **44**, Copenhagen, 1969.

[45] Krarup, T.: *Letters on Molodensky's Problem I–IV*. Communication to the members of the IAG Special Study Group 4.31, 1973. In: Mathematical Foundation of Geodesy: Selected Papers of Torben Krarup (K. Borre, ed.), Springer, Berlin, Heidelberg, 2006

[46] Kupradze, V., Aleksidze M.: The Method of Functional Equations for the Approximate Solution of Certain Boundary Value Problems. *USSR Comp. Math. Math. Phys.*, **4**, 82–126, 1964.

[47] Laplace, P.S. de: *Théorie des attractions des sphéroides et de la figure des planètes*, Mèm. de l'Acad., Paris, 1785.

[48] Legendre, A.M.: *Sur l'attraction des Sphéroides*. Mém. Math. et Phys. présentés á l'Ac. r. des. sc. par divers savants, **10**, 1785.

[49] Litzenberger, R.: *Pyramid Schemes for Harmonic Wavelets in Boundary-Value Problems*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen 2001.

[50] Magnus, W., Oberhettinger, F., Soni, R.P.: Formulas and Theorems for the Special Functions of Mathematical Physics. In: *Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen*, Band **52**, Springer, Berlin, 3. Auflage, 1966.

[51] Mayer, C., Freeden, W.: Stokes Problem, Layer Potentials and Regularizations, and Multiscale Applications. In: *Handbook of Geomathematics, Vol.* 2 (Freeden, W., Nashed, M.Z., Sonar, T., Eds.), 2nd edition, 1155–1254, Springer, New York, 2015.

[52] Martinec, Z.: *Boundary-Value Problems for Gravimetric Determination of a Precise Geoid*. Springer, Berlin, Heidelberg, New York, 1999.

[53] Miranda, C.: *Partial Differential Equations of Elliptic Type*. Springer, Berlin, 1970.

[54] Moritz, H.: *Recent Developments in the Geodetic Boundary Value Problem*. Scientif. Report, No. **266**, The Ohio State University, Department of Geodetic Science, Columbus, Ohio, 1977.

[55] Moritz, H.: *Advanced Physical Geodesy*. Wichmann Verlag, Karlsruhe, 1980.

[56] Moritz, H.: Classical Physical Geodesy. In: *Handbook of Geomathematics, Vol.* 1 (Freeden, W., Nashed, M.Z., Sonar, T., Eds.), 2nd edition, 253–289, Springer, New York, 2015.

[57] Moritz, H., Sünkel, H., Eds.: *Approximation Methods in Geodesy*: Lectures delivered at 2nd International Summer School in the Mountains on Mathematical Methods in Physical Geodesy, Wichmann, Karlsruhe, 1978.

[58] Müller, C.: *Spherical Harmonics*, Lecture Notes in Mathematics. Vol. **17**, Springer, Berlin, 1966.

[59] Müller C.: *Foundations of the Mathematical Theory of Electromagnetic Waves*. Springer, Berlin, Heidelberg, New York, 1969.

[60] Neumann, F.: *Vorlesungen über die Theorie des Potentials und der Kugelfunktionen*. B.G. Teubner, Leipzig, 1887.

[61] Ritz, W.: Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. *J. Reine Angew. Math.*, **135**, 1–6, 1909.

[62] Runge, C.: Zur Theorie der eindeutigen analytischen Funktionen. *Acta Math.*, **6**, 229–234, 1885.

[63] Sanso, F.: A Note on Density Problems and the Runge Krarup's Theorem. *Boll. Geod. Sci. Affin.*, **41**, 422–477, 1982.

[64] Shure, L., Parker, R.L., Backus, G.E.: Harmonic Splines for Geomagnetic Modelling. *Phys. Earth Planet. Inter.*, **28**, 215–229, 1982.

[65] Sommerfeld, A.: *Vorlesungen der Theoretischen Physik II*. 6th edition, Verlag Harri Deutsch, Frankfurt, 1978.

[66] Thomson, W.: *On Electrical Images, Report of the* 17*th Meeting of the British Association for the Advancement of Science*. Oxford, June 1847, John Murray, London, 1848. Notices and Abstracts of Miscellaneous Communications to the Sections, 6, John Murray, London, 1848.

[67] Trefftz, E.: Ein Gegenstück zum Ritzschen Verfahren. In: *Verh. d.* 2. *Intern. Kongr. f. Techn. Mech.*, 131–137, Zürich, 1926.

[68] Tscherning, C.C., Rapp, R.H.: *Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations, and Deflections of the Vertical Implied by Anomaly Degree Variance Models*, Scientific Report, No. **208**, The Ohio State University, Department of Geodetic Science, Columbus, 1974.

[69] Vekua, I.N.: Über die Vollständigkeit des Systems harmonischer Polynome im Raum. *Dokl. Akad. Nauk*, **90**, 495–498, 1953.

[70] Walsh, J.L.: The Approximation of Harmonic Functions by Harmonic Polynomials and by Harmonic Rational Functions. *Bull. Amer. Math. Soc.*, **35**, 499–544, 1929.

[71] Wangerin, A.: *Theorie des Potentials und der Kugelfunktionen*. 1884, published by G.J. Göschen, Leipzig, 1921.

[72] Weyl, H.: Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.,* **77**, 313–352, 1916.

Matthias Augustin
Mathematical Image Analysis Group
Saarland University
Fakultät 6
D-66123 Saarbrücken, Germany

Willi Freeden and Helga Nutz
Geomathematics Group
University of Kaiserslautern
MPI-Gebäude, Paul-Ehrlich-Str. 26
D-67663 Kaiserslautern, Germany

Handbook of Mathematical Geodesy
Geosystems Mathematics, 561–604

Geomathematical Advances in Satellite Gravity Gradiometry (SGG)

Willi Freeden, Helga Nutz, and Michael Schreiner

Abstract. A promising technique of globally establishing the fine structure and the characteristics of the external Earth's gravitational field is satellite gravity gradiometry (SGG). Satellites such as ESA's gradiometer satellite GOCE are able to provide sufficiently large data material of homogeneous quality and accuracy. In geodesy, traditionally the external Earth's gravitational potential and its Hesse matrix are described using orthogonal (Fourier) expansions in terms of (outer) spherical harmonics. Spherical and outer harmonics are introduced for the global modeling of (scalar / tensor) fields. We briefly recapitulate the results interconnecting spherically the potential coefficients with respect to tensor spherical harmonics at Low Earth Orbiter's (LEO) altitude to the corresponding coefficients with respect to scalar spherical harmonics at the Earth's surface. The relation between the known tensorial measurements $\mathbf{g}$ (i.e., gradiometer data) and the gravitational potential F on the Earth's surface is expressed by a linear integral equation of the first kind. This operator equation is discussed in the framework of pseudodifferential operators as an invertible mapping between Sobolev spaces under the assumption that the data are not erroneous. In reality, however, the data $\mathbf{g}$ are noisy such that the Sobolev reference space for the (noisy) tensorial data $\mathbf{g}$ must be embedded in a larger Sobolev space. Under these conditions, we base our inversion process on the fact that the reference Sobolev subspace is dense in the larger Sobolev space and that, e.g., a smoothing spline process or a signal-to-noise procedure in multiscale framework open appropriate perspectives to approximate F (in suitable accuracy) from noisy data $\mathbf{g}$.

Keywords. Tensorial spherical framework, satellite gravity gradiometry problem, multiscale spline and wavelet modeling.

1. Introduction

Due to the non-spherical shape, the irregularities of the interior mass distribution, and the movement of the lithospheric plates as well as volcanic and magmatic activities, the external gravitational field of the Earth shows significant local vari-

ations. The recognition of the microstructure of the Earth's external gravitational potential is of tremendous importance for a large palette of geoscientific questions, for example, studies of the processes in the Earth's interior, models of sea surface topography and circulations of the ocean, and investigations of the (global) climate change. An important measurement technique to determine locally the fine structure of the external gravitational field is terrestrial gravimetry, i.e., the reconstruction of the density variations inside and on the Earth's surface from the gravitational potential and its functionals. A promising technique of globally establishing the fine structure and the characteristics of the external Earth's gravitational field is satellite gravity gradiometry (SGG) . Its principle can be described roughly as follows: Several test masses in a low Earth's orbiting (LEO)-satellite feel – due to their distinct positions and the local changes of the gravitational field – different forces, thus yielding different accelerations. The measurements of the relative accelerations between two test masses provide information about the second-order partial derivatives of the gravitational potential at flight position. Assuming an ideal situation, the full Hesse matrix can be observed globally above the Earth by an array of test masses.

The traditional way to describe the external Earth's gravitational potential and its Hesse matrix of second derivatives is to use orthogonal (Fourier) expansions with (outer) spherical harmonics as reference trial system, a procedure dating back to C.F. Gauss in the nineteenth century. During the last decades many gravitational models have been enhanced based on the (until now unrealistic) assumption of continuously improvable terrestrial gravitational observations (with respect to both in quality and quantity). But the actual non-uniform distribution and the strong heterogeneity of the data material have set a limit for an intensive improvement of such spherically based (outer) harmonic models down to local scale. A tremendous step forward in the measurement of data for global modeling was provided by modern satellites such as ESA's gradiometer satellite GOCE (mission duration: 17 March 2009 till 11 November 2013). This satellite was able to provide sufficiently large data material of homogeneous quality and accuracy, at least for the diagonal components of the Hesse matrix. However, the great drawback of acquiring gravitational data at LEO's altitude is that the upward continuation of the gravitation amounts to an exponential spectral smoothing of the potential coefficients in terms of outer harmonics with increasing height. In other words, satellite measurements do not contain the same signal information at LEO's height (i.e., 200–250 km) as on the Earth's surface. This is the reason why the gravitational potential is obtainable from satellite data only in an attenuated form when continued to the Earth's surface. Even more, it may happen in downward continuation that the noise in the measurements is amplified. Nevertheless, for satellite gravity gradiometry (SGG), as provided by GOCE, an advantage can be taken from the fact that second derivatives instead of the potential itself are used as observations on LEO's orbit. Mathematically, this means that the exponential decay of the outer harmonic coefficients is reduced polynomially by two degrees. In other words, SGG takes advantage of the fact that second derivatives produce a rougher data set than

the potential itself such that the resolution of the gravitational structure is much finer. In addition, in the frequency context of outer harmonics, the Meissl scheme (see, e.g., [10, 20, 26, 31]) enables us in spectral nomenclature to relate the orthogonal coefficients at LEO's height to the orthogonal coefficients at the surface of the Earth, at least in the context of a spherical model and under the restrictive assumption of bandlimited outer harmonic modeling without observational errors.

In this contribution the relation between the known tensorial measurements $\mathbf{g}$ (i.e., gradiometer data) and the gravitational potential F on the Earth's surface is expressed by the linear integral equation of the first kind

$$\mathbf{\Lambda}_{SGG}^{R;r} F(x) = \int_{\Omega_R} \nabla_x \otimes \nabla_x K_{\Lambda^{up}}(x,y)\, F(y)\, d\omega_R(y) = \mathbf{g}(x), \quad x \in \Omega_r,$$

where Ω_R and Ω_r are the spheres with radii R and r, respectively, $\nabla_x \otimes \nabla_x$ denotes the Hesse tensor and $K_{\Lambda^{up}}$ is the Abel–Poisson kernel for the upward continuation (more details are explained in Section 4). This operator equation is discussed in the framework of pseudodifferential operators in, e.g., [9] as an invertible mapping between Sobolev spaces under the assumption that the data are not erroneous. In reality, however, the data $\mathbf{g}$ are noisy such that the Sobolev reference space for the (noisy) tensorial data $\mathbf{g}$ must be embedded in a larger Sobolev space. Under these conditions, we base our inversion process on the fact that the reference Sobolev subspace is dense in the larger Sobolev space and that, e.g., a smoothing spline process or a signal-to-noise procedure in multiscale framework (see [7, 8]) open perspectives to approximate F (in suitable accuracy) from noisy data $\mathbf{g}$.

Our work yields a modified wavelet approach for regularization of the inverse gradiometer problem based on ideas presented in [9]. Moreover a tree algorithm for multiscale decorrelation of the Earth's external gravitational potential is introduced. In adequate consistency with the reality the spaceborne gradiometer data are assumed to be of tensorial nature. As an essential tool tensorial radial basis functions (see, e.g., [12, 14, 26]) are used for multiscale regularization of the exponentially ill-posed downward continuation of satellite gradiometer data.

2. Potential theoretic aspects

Gravity as observed on the Earth's surface is the combined effect of the gravitational mass attraction and the centrifugal force due to the Earth's rotation. The force of gravity provides a directional structure to the space above the Earth's surface. It is tangential to the vertical plumb lines and perpendicular to all level surfaces. Any water surface at rest is part of a level surface. As if the Earth were a homogeneous, spherical body gravity turns out to be always constant all over the Earth's surface. The plumb lines are directed towards the Earth's center of mass, and this implies that all level surfaces are nearly spherical, too.

However, the gravity shows local (temporal and spatial) variations due to mass density inhomogeneities and temporal mass rearrangement in the Earth's

interior which enable the investigation of geological structures as fault zone, geological faults, salt domes and volcanic formations as well as mineral deposits. The strongest variations in the gravity are caused by the periodically changing positions of moon and sun relative to the Earth and the associated deformations of the body of the Earth (ocean and Earth tide and the induced loading effects).

The geoid is a virtual surface shaped by the gravity field of the Earth in the absence of external influences such as winds and tides. The level surfaces are ideal reference surfaces, for example, for heights. In more detail, the *gravity acceleration* (*gravity*) w is the resultant of gravitation v and centrifugal acceleration c, i.e., $w = v + c$. The centrifugal force c arises as a result of the rotation of the Earth around its axis. We assume here a rotation of constant angular velocity around the rotational axis x_3, which is further assumed to be fixed with respect to the Earth.

The direction of the gravity w is known as the direction of the *plumb line*, the quantity $|w|$ is called the *gravity intensity* (often just *gravity*). The *gravity potential of the Earth* can be expressed in the form: $W = V + C$. The gravity acceleration w is given by $w = \nabla W = \nabla V + \nabla C$. The surfaces of constant gravity potential $W(x) = \text{const}$, $x \in \mathbb{R}^3$, are designated as *equipotential* (*level,* or *geopotential*) *surfaces of gravity.* The *gravity potential* W of the Earth is the sum of the *gravitational potential* V and the *centrifugal potential* C, i.e., $W = V + C$. In an Earth's fixed coordinate system the centrifugal potential C is explicitly known. Hence, the determination of equipotential surfaces of the potential W is strongly related to the knowledge of the potential V. The gravity vector w given by $w(x) = \nabla_x W(x)$, where the point $x \in \mathbb{R}^3$ is located outside or on a sphere around the origin with Earth's radius R, is normal to the equipotential surface passing through the same point. Thus, equipotential surfaces intuitively express the notion of tangential surfaces, as they are normal to the plumb lines given by the direction of the gravity vector (for more details see, e.g., [17]).

A tremendous step forward in the measurement of data for global modeling was provided with modern satellites such as CHAMP (2000–2010), GRACE (launch 2002; designed for a mission lifetime of five years, GRACE is currently operating in an extended mission phase suffering from aging of the components, e.g., health of the batteries), and GOCE (2009–2013). These satellites yield sufficiently large data material of homogeneous quality and accuracy. However, as already pointed out, the great drawback of acquiring gravitational data at LEO's altitude is that the gravitational potential at ground level is obtainable from satellite data only in a "rough" form if continued downward to the Earth's surface. Even more, it is unavoidable for ill-posed problems such as downward continuation that the noise in the measurements is amplified. For SGG, however, a certain compensation effect can be taken from the fact that second derivatives instead of the potential itself are used. As already mentioned, this mathematically means that the exponential decay of the outer harmonic coefficients is reduced polynomially by two degrees. In other words, SGG takes advantage of the fact that second derivatives produce a rougher data set than the potential itself such that the resolution of the gravitational structure is much finer. A mathematical model relating observables

in geodesy to each other in the frequency context of outer harmonics is the Meissl scheme (see, e.g., [20, 21, 26, 28–31]). By specifying the spectral properties of the Fourier coefficients, this scheme enables us to relate the orthogonal coefficients at LEO's height to the orthogonal coefficients at the Earth's surface, at least in the context of a spherical model and under the restrictive assumption of bandlimited outer harmonic modeling.

Spherical notation

We begin by introducing some basic notation that will be used throughout our work: Let $x, y, \ldots$ represent the elements of the Euclidean space $\mathbb{R}^3$. For all $x \in \mathbb{R}^3$, $x = (x_1, x_2, x_3)^T$, different from the origin, we have $x = r\xi$, $r = |x| = \sqrt{x_1^2 + x_2^2 + x_3^2}$, where $\xi = (\xi_1, \xi_2, \xi_3)^T$ is the uniquely determined directional unit vector of $x \in \mathbb{R}^3$. The unit sphere in $\mathbb{R}^3$ will be denoted by Ω, whereas Ω_α designates the sphere around the origin with radius α. If the vectors $\varepsilon^1, \varepsilon^2, \varepsilon^3$ form the canonical orthonormal basis in $\mathbb{R}^3$, we may represent $\xi \in \Omega$ in polar coordinates by

$$\xi = t\varepsilon^3 + \sqrt{1-t^2}\left(\cos\varphi\varepsilon^1 + \sin\varphi\varepsilon^2\right), \quad -1 \le t \le 1,\ 0 \le \varphi < 2\pi,\ t = \cos\theta. \quad (1)$$

Inner, vector, and dyadic (tensor) product of two vectors $x, y \in \mathbb{R}^3$, respectively, are denoted by $x \cdot y$, $x \wedge y$ and $x \otimes y$. As usual, a *second-order tensor* $\mathbf{f} \in \mathbb{R}^3 \otimes \mathbb{R}^3$ is understood to be a linear mapping that assigns to each $x \in \mathbb{R}^3$ a vector $y \in \mathbb{R}^3$. The (cartesian) components F_{ij} of $\mathbf{f}$ are defined by $F_{ij} = \varepsilon^i \cdot (\mathbf{f}\varepsilon^j) = (\varepsilon^i)^T(\mathbf{f}\varepsilon^j)$, so that $y = \mathbf{f}x$ is equivalent to $y \cdot \varepsilon^i = \sum_{j=1}^3 F_{ij}(x \cdot \varepsilon^j)$. We write $\mathbf{f}^T$ for the transpose of $\mathbf{f}$; it is the unique tensor satisfying $(\mathbf{f}y) \cdot x = y \cdot (\mathbf{f}^T x)$ for all $x, y \in \mathbb{R}^3$. The dyadic (tensor) product $x \otimes y$ of two elements $x, y \in \mathbb{R}^3$ is the tensor that assigns to each $u \in \mathbb{R}^3$ the vector $(y \cdot u)x$. More explicitly, $(x \otimes y)u = (y \cdot u)x$ for every $u \in \mathbb{R}^3$. The inner product $\mathbf{f} \cdot \mathbf{g}$ of two second-order tensors $\mathbf{f}, \mathbf{g} \in \mathbb{R}^3 \otimes \mathbb{R}^3$ is defined by $\mathbf{f} \cdot \mathbf{g} = \sum_{i,j=1}^3 F_{ij}G_{ij}$, while $|\mathbf{f}| = (\mathbf{f} \cdot \mathbf{f})^{1/2}$ is called the norm of $\mathbf{f}$.

Next we come to some differential operators, which are of particular importance in the tensorial context. In terms of polar coordinates (1) the gradient ∇ in $\mathbb{R}^3$ allows the representation $\nabla_x = \xi\partial/\partial r + (1/r)\nabla^*_\xi$, where ∇^* is the *surface gradient* of the unit sphere $\Omega \subset \mathbb{R}^3$. The operator $\Delta^* = \nabla^* \cdot \nabla^*$ is called the *Beltrami operator* of the unit sphere Ω. Obviously, it can be understood as the angular part of the Laplace operator.

Note that, throughout this paper, scalar-valued (resp. vector-valued, tensor-valued) functions are denoted by capital (resp. small, small bold) letters. A function $F : \Omega \to \mathbb{R}$ (resp. $f : \Omega \to \mathbb{R}^3$, $\mathbf{f} : \Omega \to \mathbb{R}^3 \otimes \mathbb{R}^3$) possessing k continuous derivatives on the unit sphere Ω is said to be of class $C^{(k)}(\Omega)$ (resp. $c^{(k)}(\Omega)$, $\mathbf{c}^{(k)}(\Omega)$). $C^{(0)}(\Omega)$ (resp. $c^{(0)}(\Omega)$, $\mathbf{c}^{(0)}(\Omega)$) is the class of real continuous scalar-valued (resp. vector-valued, tensor-valued) functions on Ω. For $F \in C^{(1)}(\Omega)$ we introduce the *surface curl gradient* L^*_ξ by $L^*_\xi F(\xi) = \xi \wedge \nabla^*_\xi F(\xi)$, $\xi \in \Omega$, while $\nabla^*_\xi \cdot f(\xi)$, $\xi \in \Omega$, and $L^*_\xi \cdot f(\xi)$, $\xi \in \Omega$, respectively, denote the *surface divergence* and *surface curl* of the vector field f at $\xi \in \Omega$. For more details the reader is referred to [5].

Scalar spherical harmonics

Scalar spherical harmonics are defined as restrictions of homogeneous harmonic polynomials to the unit sphere Ω. In all geosciences interested in global modeling, spherical harmonics are the functions which are usually taken to represent scalar fields on a spherical surface such as the Earth's (mean) sphere.

Definition 2.1. Let H_n be a homogeneous harmonic polynomial of degree n in $\mathbb{R}^3$, $n \in \mathbb{N}_0$, i.e., $H_n \in Harm_n(\mathbb{R}^3)$. The restriction $Y_n = H_n|_\Omega$ is called (scalar) spherical harmonic of degree n. The space $\{Y_n \mid Y_n = H_n|_\Omega,\ H_n \in Harm_n(\mathbb{R}^3)\}$ of all (scalar) spherical harmonics of degree n is denoted by $Harm_n(\Omega)$.

$Harm_n(\Omega)$ is known to be of dimension $2n+1$. Spherical harmonics of different degrees are orthogonal in $L^2(\Omega)$-sense, that is

$$(Y_n, Y_{\tilde{n}})_{L^2(\Omega)} = \int_\Omega Y_n(\xi) Y_{\tilde{n}}(\xi)\, d\omega(\xi) = 0, \quad n \neq \tilde{n},$$

where $d\omega$ is the surface element on Ω. Throughout this text a capital letter Y followed by one or two indices always denotes a spherical harmonic of the degree given by the first index and order given by the second index. Two indices mean that the function, for example $Y_{n,m}$, is a member of an $L^2(\Omega)$-orthonormal system of functions $\{Y_{n,1}, \ldots, Y_{n,2n+1}\}_{n\in\mathbb{N}_0}$. By use of the scalar spherical harmonics every function $F \in L^2(\Omega)$ can be written as a Fourier series

$$F = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} F^{\wedge_{L^2(\Omega)}}(n,m) Y_{n,m},$$

(in $L^2(\Omega)$-sense) with Fourier coefficients

$$F^{\wedge_{L^2(\Omega)}}(n,m) = (F, Y_{n,m})_{L^2(\Omega)} = \int_\Omega F(\eta) Y_{n,m}(\eta) d\omega(\eta).$$

The system $\{Y_{n,m}\}_{n=0,1,\ldots;m=1,\ldots,2n+1}$ is closed in $C(\Omega)$ with respect to the norm $\|\cdot\|_{C(\Omega)}$, i.e., for any number $\varepsilon > 0$ and any function $F \in C(\Omega)$, there exists a linear combination $F_N = \sum_{n=0}^{N} \sum_{m=1}^{2n+1} d_{n,m} Y_{n,m}$ such that $\|F - F_N\|_{C(\Omega)} \leq \varepsilon$. The system $\{Y_{n,m}\}_{n=0,1,\ldots;m=1,\ldots,2n+1}$ is, furthermore, complete in $L^2(\Omega)$ with respect to $(\cdot,\cdot)_{L^2(\Omega)}$, i.e., $F \in L^2(\Omega)$ with $F^{\wedge_{L^2(\Omega)}}(n,m) = 0$ for all $n = 0, 1, \ldots$; $m = 1, \ldots, 2n+1$ implies $F = 0$ (see, e.g., [14]).

Theorem 2.2 (Addition Theorem for Scalar Spherical Harmonics). *Let the system $\{Y_{n,m}\}_{m=1,\ldots,2n+1}$ be an $L^2(\Omega)$-orthonormal one in $Harm_n(\Omega)$. Then, for any pair $(\xi,\eta) \in \Omega^2$, the addition theorem reads*

$$\sum_{m=1}^{2n+1} Y_{n,m}(\xi) Y_{n,m}(\eta) = \frac{2n+1}{4\pi} P_n(\xi \cdot \eta),$$

where $P_n : [-1,1] \to [-1,1]$ is the Legendre polynomial of degree n.

Tensor spherical harmonics

By $\mathbf{l}^2(\Omega)$ we denote the Hilbert space of square-integrable tensor fields $\mathbf{f} : \Omega \to \mathbb{R}^3 \otimes \mathbb{R}^3$ with the inner product

$$(\mathbf{f}, \mathbf{g})_{\mathbf{l}^2(\Omega)} = \int_\Omega \mathbf{f}(\xi) \cdot \mathbf{g}(\xi)\, d\omega(\xi), \quad \mathbf{f}, \mathbf{g} \in \mathbf{l}^2(\Omega),$$

and the associated norm $\|\cdot\|_{\mathbf{l}^2(\Omega)}$. Note that the space $\mathbf{l}^2(\Omega)$ is the completion of $\mathbf{c}(\Omega)$ with respect to the norm $\|\cdot\|_{\mathbf{l}^2(\Omega)}$. The operators $\mathbf{o}^{(i,k)} : C^{(\infty)}(\Omega) \to \mathbf{c}^{(\infty)}(\Omega)$, $i, k = 1, 2, 3$, transform scalar functions into tensor fields (cf. [12]):

$$\begin{aligned}
\mathbf{o}^{(1,1)}F(\xi) &= \xi \otimes \xi F(\xi),\\
\mathbf{o}^{(1,2)}F(\xi) &= \xi \otimes \nabla_\xi^* F(\xi),\\
\mathbf{o}^{(1,3)}F(\xi) &= \xi \otimes L_\xi^* F(\xi),\\
\mathbf{o}^{(2,1)}F(\xi) &= \nabla_\xi^* F(\xi) \otimes \xi,\\
\mathbf{o}^{(3,1)}F(\xi) &= L_\xi^* F(\xi) \otimes \xi,\\
\mathbf{o}^{(2,2)}F(\xi) &= \mathbf{i}_{\tan}(\xi) F(\xi),\\
\mathbf{o}^{(2,3)}F(\xi) &= \left(\nabla_\xi^* \otimes \nabla_\xi^* - L_\xi^* \otimes L_\xi^*\right) F(\xi) + 2\nabla_\xi^* F(\xi) \otimes \xi,\\
\mathbf{o}^{(3,2)}F(\xi) &= \left(\nabla_\xi^* \otimes L_\xi^* + L_\xi^* \otimes \nabla_\xi^*\right) F(\xi) + 2L_\xi^* F(\xi) \otimes \xi,\\
\mathbf{o}^{(3,3)}F(\xi) &= \mathbf{j}_{\tan}(\xi) F(\xi),
\end{aligned}$$

$F \in \mathrm{C}^{(2)}(\Omega)$, $\xi \in \Omega$. Note that the tensors $\mathbf{i}_{\tan} = \mathbf{i} - \xi \otimes \xi$ and $\mathbf{j}_{\tan} = \xi \wedge \mathbf{i}$ are the surface identity tensor and the surface rotation tensor, respectively. The adjoint operators $O^{(i,k)}$ to $\mathbf{o}^{(i,k)}$ satisfying

$$\left(\mathbf{o}^{(i,k)}F, \mathbf{f}\right)_{\mathbf{l}^2(\Omega)} = \left(F, O^{(i,k)}\mathbf{f}\right)_{L^2(\Omega)}$$

for $F \in C^{(2)}(\Omega)$ and $\mathbf{f} \in \mathbf{c}^{(2)}(\Omega)$ are given by

$$\begin{aligned}
O^{(1,1)}\mathbf{f}(\xi) &= \xi^T \mathbf{f}(\xi) \xi,\\
O^{(1,2)}\mathbf{f}(\xi) &= -\nabla_\xi^* \cdot p_{\tan}\left(\xi^T \mathbf{f}(\xi)\right),\\
O^{(1,3)}\mathbf{f}(\xi) &= -L_\xi^* \cdot p_{\tan}\left(\xi^T \mathbf{f}(\xi)\right),\\
O^{(2,1)}\mathbf{f}(\xi) &= -\nabla_\xi^* \cdot p_{\tan}\left(\mathbf{f}(\xi)\xi\right),\\
O^{(3,1)}\mathbf{f}(\xi) &= -L_\xi^* \cdot p_{\tan}\left(\mathbf{f}(\xi)\xi\right),\\
O^{(2,2)}\mathbf{f}(\xi) &= \mathbf{i}_{\tan}(\xi) \cdot \mathbf{f}(\xi),\\
O^{(2,3)}\mathbf{f}(\xi) &= \nabla_\xi^* \cdot p_{\tan}\left(\nabla_\xi^* \cdot \mathbf{p}_{\tan,*}\mathbf{f}(\xi)\right) - L_\xi^* \cdot p_{\tan}\left(L_\xi^* \cdot \mathbf{p}_{\tan,*}\mathbf{f}(\xi)\right)\\
&\quad - 2\nabla_\xi^* \cdot p_{\tan}\left(\mathbf{f}(\xi)\xi\right),\\
O^{(3,2)}\mathbf{f}(\xi) &= L_\xi^* \cdot p_{\tan}\left(\nabla_\xi^* \cdot \mathbf{p}_{\tan}\mathbf{f}(\xi)\right) + \nabla_\xi^* \cdot p_{\tan}\left(L_\xi^* \cdot \mathbf{p}_{\tan}\mathbf{f}(\xi)\right)\\
&\quad - 2L_\xi^* \cdot p_{\tan}\left(\mathbf{f}(\xi)\xi\right),\\
O^{(3,3)}\mathbf{f}(\xi) &= \mathbf{j}_{\tan}(\xi) \cdot \mathbf{f}(\xi),
\end{aligned}$$

where

$$p_{\tan} f(\xi) = f(\xi) - (\xi \cdot f(\xi))\xi,$$
$$\mathbf{p}_{\tan,*}\mathbf{f}(\xi) = \mathbf{f}(\xi) - \xi \otimes ((\mathbf{f}(\xi))^T \xi),$$

$\xi \in \Omega$. With the help of the operators $\mathbf{o}^{(i,k)}$ we are able to define a set of tensor spherical harmonics $\{\mathbf{y}_{n,m}^{(i,k)}\}_{i,k=1,2,3;\ n=0_{ik},\ldots;\ m=1,\ldots,2n+1}$ by setting

$$\mathbf{y}_{n,m}^{(i,k)} = \left(\mu_n^{(i,k)}\right)^{-1/2} \mathbf{o}^{(i,k)} Y_{n,m}, \tag{2}$$

where the normalization constants $\mu_n^{(i,k)}$ are given by

$$\mu_n^{(i,k)} = \begin{cases} 1, & (i,k) = (1,1), \\ 2, & (i,k) \in \{(2,2),(3,3)\}, \\ n(n+1), & (i,k) \in \{(1,2),(1,3),(2,1),(3,1)\}, \\ 2n(n+1)(n(n+1)-2), & (i,k) \in \{(2,3),(3,2)\}. \end{cases}$$

For simplicity, we use the abbreviation

$$0_{ik} = \begin{cases} 0, & (i,k) \in \{(1,1),(2,2),(3,3)\}, \\ 1, & (i,k) \in \{(1,2),(1,3),(2,1),(3,1)\}, \\ 2, & (i,k) \in \{(2,3),(3,2)\}. \end{cases}$$

By $\mathbf{harm}_n^{(i,k)}(\Omega)$ we denote the space of all tensor spherical harmonics of degree n and kind (i,k). If $\{\mathbf{y}_{n,m}^{(i,k)}\}_{m=1,\ldots,2n+1}$ is an $\mathbf{l}^2(\Omega)$-orthonormal basis of $\mathbf{harm}_n^{(i,k)}(\Omega)$, then the *tensorial addition theorem* reads

$$\sum_{m=1}^{2n+1} \mathbf{y}_{n,m}^{(i,k)}(\xi) \otimes \mathbf{y}_{n,m}^{(p,q)}(\eta) = \frac{2n+1}{4\pi} \mathbf{P}_n^{(i,k,p,q)}(\xi,\eta),$$

$i,k,p,q \in \{1,2,3\}$, where $\mathbf{P}_n^{(i,k,p,q)} : \Omega \times \Omega \to \mathbb{R}^3 \otimes \mathbb{R}^3 \otimes \mathbb{R}^3 \otimes \mathbb{R}^3$ denote the *Legendre tensors of degree n* defined by

$$\mathbf{P}_n^{(i,k,p,q)} = \left(\mu_n^{(i,k)}\right)^{-1/2} \left(\mu_n^{(p,q)}\right)^{-1/2} \mathbf{o}_\xi^{(i,k)} \mathbf{o}_\eta^{(p,q)} P_n(\xi \cdot \eta), \quad \xi,\eta \in \Omega,$$

(for explicit representations see [12]). Note that, for sufficiently smooth tensor fields $\mathbf{f} : \Omega \to \mathbb{R}^3 \otimes \mathbb{R}^3$ of the form

$$\mathbf{f}(\xi) = \sum_{i,k=1}^{3} F_{i,k}(\xi) \varepsilon^i \otimes \varepsilon^k, \quad x \in \Omega,$$

we set

$$\mathbf{o}_\xi^{(p,q)} \mathbf{f}(\xi) = \sum_{i,k=1}^{3} \left(\mathbf{o}_\xi^{(p,q)} F_{i,k}(\xi)\right) \otimes \varepsilon^i \otimes \varepsilon^k.$$

By $\mathbf{harm}_n(\Omega)$ we denote the space of all tensor spherical harmonics of degree n.

We have defined the system $\{\mathbf{y}_{n,m}^{(i,k)}\}$ of tensor spherical harmonics concentrating on the fact that the decomposition into normal and tangential tensor fields is fulfilled (cf. [26]). But one disadvantage of this set of tensor spherical harmonics

is that these functions are no eigenfunctions of the (scalar) Beltrami operator. As turns out, this property enables us to define so-called outer harmonics in such a way that they fulfill the Laplace equation in the outer space. To this end we introduce the operators $\tilde{\mathbf{o}}^{(i,k)} : C^{(\infty)}(\Omega) \to \mathbf{c}^{(\infty)}(\Omega)$, $i,k = 1,2,3$, based on the operators $\mathbf{o}^{(i,k)}$ and $O^{(i,k)}$ by

$$\begin{pmatrix} \tilde{\mathbf{o}}^{(1,1)}Y_n \\ \tilde{\mathbf{o}}^{(1,2)}Y_n \\ \tilde{\mathbf{o}}^{(2,1)}Y_n \\ \tilde{\mathbf{o}}^{(2,2)}Y_n \\ \tilde{\mathbf{o}}^{(3,3)}Y_n \end{pmatrix} = \mathbf{a}_D \begin{pmatrix} Y_n \\ Y_n \\ Y_n \\ Y_n \\ Y_n \end{pmatrix},$$

and

$$\begin{pmatrix} \tilde{\mathbf{o}}^{(1,3)}Y_n \\ \tilde{\mathbf{o}}^{(2,3)}Y_n \\ \tilde{\mathbf{o}}^{(3,1)}Y_n \\ \tilde{\mathbf{o}}^{(3,2)}Y_n \end{pmatrix} = \mathbf{b}_D \begin{pmatrix} Y_n \\ Y_n \\ Y_n \\ Y_n \end{pmatrix},$$

where the matrix operators $\mathbf{a}_D$ and $\mathbf{b}_D$ are defined by (see [12])

$$\mathbf{a}_D = \begin{pmatrix} \mathbf{o}^{(1,1)}(D+1)(D+2) & -\mathbf{o}^{(1,2)}(D+2) & -\mathbf{o}^{(2,1)}(D+2) & -\frac{1}{2}o^{(2,2)}(D+2)(D+1) & \frac{1}{2}\mathbf{o}^{(2,3)} \\ \mathbf{o}^{(1,1)}D^2 & \mathbf{o}^{(1,2)}D & -\mathbf{o}^{(2,1)}(D-1) & -\frac{1}{2}\mathbf{o}^{(2,2)}D(D-1) & -\frac{1}{2}\mathbf{o}^{(2,3)} \\ \mathbf{o}^{(1,1)}(D+1)^2 & -\mathbf{o}^{(1,2)}(D+1) & \mathbf{o}^{(2,1)}(D+2) & \frac{1}{2}\mathbf{o}^{(2,2)}(D+2)(D+1) & -\frac{1}{2}\mathbf{o}^{(2,3)} \\ \mathbf{o}^{(1,1)}D(D-1) & \mathbf{o}^{(1,2)}(D-1) & \mathbf{o}^{(2,1)}(D-1) & -\frac{1}{2}\mathbf{o}^{(2,2)}D(D-1) & \frac{1}{2}\mathbf{o}^{(2,3)} \\ 0 & 0 & \mathbf{o}^{(2,1)} & -\frac{1}{2}\mathbf{o}^{(2,2)}D(D+1) & -\frac{1}{2}\mathbf{o}^{(2,3)} \end{pmatrix}$$

and

$$\mathbf{b_D} = \begin{pmatrix} \mathbf{o}^{(1,3)}(D+1) & \mathbf{o}^{(3,1)} & -\frac{1}{2}\mathbf{o}^{(3,2)} & -\frac{1}{2}\mathbf{o}^{(3,3)}D(D+1) \\ \mathbf{o}^{(1,3)}D & -\mathbf{o}^{(3,1)} & \frac{1}{2}\mathbf{o}^{(3,2)} & \frac{1}{2}\mathbf{o}^{(3,3)}D(D+1) \\ 0 & \mathbf{o}^{(3,1)}(D+2) & -\frac{1}{2}\mathbf{o}^{(3,2)} & \frac{1}{2}\mathbf{o}^{(3,3)}(D+2)(D+1) \\ 0 & \mathbf{o}^{(3,1)}(D-1) & \frac{1}{2}\mathbf{o}^{(3,2)} & -\frac{1}{2}\mathbf{o}^{(3,3)}D(D-1) \end{pmatrix},$$

and D is the pseudodifferential operator $D = (-\Delta + \frac{1}{4})^{1/2} - \frac{1}{2}$ of order 1 satisfying $DY_n = D^{\wedge}(n)Y_n = nY_n$ for all $Y_n \in Harm_n(\Omega)$.

The adjoint operators $\tilde{O}^{(i,k)} : \mathbf{c}^{(\infty)}(\Omega) \to C^{(\infty)}(\Omega)$, $i,k = 1,2,3$, to the operators $\tilde{\mathbf{o}}^{(i,k)}$ satisfying the equation $(\tilde{\mathbf{o}}^{(i,k)}G, \mathbf{f})_{\mathbf{l}^2(\Omega)} = (G, \tilde{O}^{(i,k)}\mathbf{f})_{L^2(\Omega)}$, $\mathbf{f} \in \mathbf{c}^{(\infty)}(\Omega)$, $G \in \mathrm{C}^{(\infty)}(\Omega)$, are easily obtainable as follows

$$\begin{pmatrix} \tilde{O}^{(1,1)}Y_n \\ \tilde{O}^{(1,2)}Y_n \\ \tilde{O}^{(2,1)}Y_n \\ \tilde{O}^{(2,2)}Y_n \\ \tilde{O}^{(3,3)}Y_n \end{pmatrix} = \mathbf{a}_D \begin{pmatrix} Y_n \\ Y_n \\ Y_n \\ Y_n \\ Y_n \end{pmatrix},$$

and

$$\begin{pmatrix} \tilde{O}^{(1,3)}Y_n \\ \tilde{O}^{(2,3)}Y_n \\ \tilde{O}^{(3,1)}Y_n \\ \tilde{O}^{(3,2)}Y_n \end{pmatrix} = \mathbf{b}_D \begin{pmatrix} Y_n \\ Y_n \\ Y_n \\ Y_n \end{pmatrix}.$$

After these preliminaries we are now in the position to introduce the tensor spherical harmonics

$$\tilde{\mathbf{y}}_{n,m}^{(i,k)} = \left(\tilde{\mu}_n^{(i,k)}\right)^{-1/2} \tilde{\mathbf{o}}^{(i,k)} Y_{n,m}, \tag{3}$$

$n = \tilde{0}_{ik}, \ldots; m = 1, \ldots, 2n+1$, where we use the abbreviation

$$\tilde{0}_{ik} = \begin{cases} 0, & (i,k) \in \{(1,1),(2,1),(3,1)\}, \\ 1, & (i,k) \in \{(1,2),(1,3),(2,3),(3,3)\}, \\ 2, & (i,k) \in \{(2,2),(3,2)\}, \end{cases}$$

and the normalization constants

$$\begin{aligned}
\tilde{\mu}_n^{(1,1)} &= (n+2)(n+1)(2n+3)(2n+1), \\
\tilde{\mu}_n^{(1,2)} &= 3n^4, \\
\tilde{\mu}_n^{(1,3)} &= n(n+1)^2(2n+1), \\
\tilde{\mu}_n^{(2,1)} &= (n+1)^2(2n+3)(2n+1), \\
\tilde{\mu}_n^{(2,2)} &= n(n-1)(2n+1)(2n-1), \\
\tilde{\mu}_n^{(2,3)} &= n^2(n+1)^2, \\
\tilde{\mu}_n^{(3,1)} &= n^2(n+1)(2n+1), \\
\tilde{\mu}_n^{(3,2)} &= n(n+1)^2(2n+1), \\
\tilde{\mu}_n^{(3,3)} &= n^2(n-1)(2n+1).
\end{aligned}$$

According to this construction, in contrary to the system (2), each member of the system $\{\tilde{\mathbf{y}}_{n,m}^{(i,k)}\}$ is an eigenfunction of the Beltrami operator. More explicitly, we have

Theorem 2.3. *Let $\{Y_{n,m}\}_{n=0,1,\ldots;\ m=1,\ldots,2n+1}$ be an $L^2(\Omega)$-orthonormal set of scalar spherical harmonics. Then, the set*

$$\left\{\tilde{\mathbf{y}}_{n,m}^{(i,k)}\right\}_{\substack{i,k=1,2,3; \\ n=\tilde{0}_{ik},\ldots; \\ m=1,\ldots,2n+1}},$$

as defined by (3), forms an $\mathbf{l}^2(\Omega)$-orthonormal set of tensor spherical harmonics which is closed in $\mathbf{c}(\Omega)$ and $\mathbf{l}^2(\Omega)$ with respect to $\|\cdot\|_{\mathbf{c}(\Omega)}$ and $\|\cdot\|_{\mathbf{l}^2(\Omega)}$, respectively, and complete in $\mathbf{l}^2(\Omega)$ with respect to $(\cdot,\cdot)_{\mathbf{l}^2(\Omega)}$. Furthermore, we are able to verify that

$$\begin{aligned}
\Delta_\xi^* \tilde{\mathbf{y}}_{n,m}^{(1,1)} &= -(n+2)(n+3)\tilde{\mathbf{y}}_{n,m}^{(1,1)}, \\
\Delta_\xi^* \tilde{\mathbf{y}}_{n,m}^{(1,2)} &= -n(n+1)\tilde{\mathbf{y}}_{n,m}^{(1,2)}, \\
\Delta_\xi^* \tilde{\mathbf{y}}_{n,m}^{(2,1)} &= -n(n+1)\tilde{\mathbf{y}}_{n,m}^{(2,1)}, \\
\Delta_\xi^* \tilde{\mathbf{y}}_{n,m}^{(2,2)} &= -(n-1)(n-2)\tilde{\mathbf{y}}_{n,m}^{(2,2)}, \\
\Delta_\xi^* \tilde{\mathbf{y}}_{n,m}^{(3,3)} &= -n(n+1)\tilde{\mathbf{y}}_{n,m}^{(3,3)},
\end{aligned}$$

$$\begin{aligned}
\Delta_\xi^* \tilde{\mathbf{y}}_{n,m}^{(1,3)} &= -(n+1)(n+2)\tilde{\mathbf{y}}_{n,m}^{(1,3)}, \\
\Delta_\xi^* \tilde{\mathbf{y}}_{n,m}^{(2,3)} &= -n(n-1)\tilde{\mathbf{y}}_{n,m}^{(2,3)}, \\
\Delta_\xi^* \tilde{\mathbf{y}}_{n,m}^{(3,1)} &= -(n+1)(n+2)\tilde{\mathbf{y}}_{n,m}^{(3,1)}, \\
\Delta_\xi^* \tilde{\mathbf{y}}_{n,m}^{(3,2)} &= -n(n-1)\tilde{\mathbf{y}}_{n,m}^{(3,2)},
\end{aligned}$$

where the application of the Beltrami operator is understood component-by-component.

Because of the completeness of the tensor spherical harmonics every tensor field $\mathbf{f} \in \mathbf{l}^2(\Omega)$ can be written as a Fourier series

$$\mathbf{f} = \sum_{i,k=1}^{3} \sum_{n=\bar{0}_{ik}}^{\infty} \sum_{m=1}^{2n+1} \mathbf{f}^{(i,k)\wedge_{\mathbf{l}^2(\Omega)}}(n,m)\tilde{\mathbf{y}}_{n,m}^{(i,k)}$$

(in $\|\cdot\|_{\mathbf{l}^2(\Omega)}$-sense), where the Fourier coefficients are given by

$$\mathbf{f}^{(i,k)\wedge_{\mathbf{l}^2(\Omega)}}(n,m) = (\mathbf{f}, \tilde{\mathbf{y}}_{n,m}^{(i,k)})_{\mathbf{l}^2(\Omega)} = \int_\Omega \mathbf{f}(\xi) \cdot \tilde{\mathbf{y}}_{n,m}^{(i,k)}(\xi) d\omega(\xi).$$

For a more detailed introduction to the theory of scalar and tensor spherical harmonics including the development of associated addition theorems the reader is referred to [12].

Outer harmonics

Up to now, we assumed spherical geometry, i.e., we presented spherical harmonics which are adequate for the conventional approach, where the reference surface of the Earth is supposed to be a sphere and the data are assumed to be given on a spherical satellite orbit. Next, we make the first steps to a Runge concept by specifying two spheres as illustrated in Figure 1, thereby using the specific properties of outer harmonics.

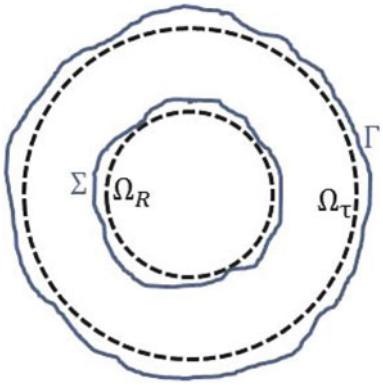


FIGURE 1. *The geometric Runge concept underlying this paper.*

Our idea can be explained as follows: Starting from the data given on the real satellite orbit Γ (which is not necessarily required to build a closed surface) we pull down the tensorial information to a sphere Ω_τ of radius τ such that $\mathrm{dist}(\Omega_\tau, \Gamma) > 0$ by use of tensor outer harmonics. By virtue of "downward continuation" from Ω_τ to the sphere Ω_R such that $\mathrm{dist}(\Omega_R, \Sigma) > 0$ corresponding to the real Earth's surface Σ we are able to calculate the desired solution in terms of scalar outer harmonics, i.e., the gravitational potential on the real Earth's surface Σ from data on the real orbit Γ. In consequence, we have to base our considerations on scalar as well as tensor outer harmonics that are consistently related to each other.

Scalar outer harmonics. We begin our considerations with the introduction of the scalar outer harmonics

$$H_{n,m}^R(x) = \frac{1}{R}\left(\frac{R}{|x|}\right)^{n+1} Y_{n,m}\left(\frac{x}{|x|}\right),\ x \in \overline{\Omega_R^{\mathrm{ext}}},\ n = 0, 1, \ldots;\ m = 1, \ldots, 2n+1.$$

They obey the following properties:

- $H_{n,m}^R$ is of class $C^{(\infty)}(\Omega_R^{\mathrm{ext}})$,
- $H_{n,m}^R$ is harmonic in Ω_R^{ext}: $\Delta_x H_{n,m}^R(x) = 0$ for $x \in \Omega_R^{\mathrm{ext}}$,
- $H_{n,m}^R|_{\Omega_R} = (1/R)Y_{n,m}$,
- $(H_{n,m}^R, H_{l,s}^R)_{L^2(\Omega_R)} = \int_{\Omega_R} H_{n,m}^R(x) H_{l,s}^R(x) d\omega(x) = \delta_{n,l}\delta_{m,s}$,
- $|H_{n,m}^R(x)| = \mathcal{O}\left(|x|^{-1}\right)$, $|x| \to \infty$.

Accordingly, the space $Harm_n(\overline{\Omega_R^{\mathrm{ext}}})$ is defined by

$$Harm_n(\overline{\Omega_R^{\mathrm{ext}}}) = \mathrm{span}_{m=1,\ldots,2n+1}(H_{n,m}),$$

while $Harm_{0,\ldots,n}(\overline{\Omega_R^{\mathrm{ext}}})$ denotes the space

$$Harm_{0,\ldots,n}(\overline{\Omega_R^{\mathrm{ext}}}) = \bigoplus_{k=0}^{n} Harm_k(\overline{\Omega_R^{\mathrm{ext}}}).$$

Tensor outer harmonics. Next we introduce an associated class of *tensor outer harmonics* by using tensor spherical harmonics as defined in (3)

$$\mathbf{h}_{n,m}^{R;(1,1)}(x) = \frac{1}{R}\left(\frac{R}{|x|}\right)^{n+3} \tilde{\mathbf{y}}_{n,m}^{(1,1)}\left(\frac{x}{|x|}\right), \tag{4}$$

$$\mathbf{h}_{n,m}^{R;(1,2)}(x) = \frac{1}{R}\left(\frac{R}{|x|}\right)^{n+1} \tilde{\mathbf{y}}_{n,m}^{(1,2)}\left(\frac{x}{|x|}\right), \tag{5}$$

$$\mathbf{h}_{n,m}^{R;(1,3)}(x) = \frac{1}{R}\left(\frac{R}{|x|}\right)^{n+2} \tilde{\mathbf{y}}_{n,m}^{(1,3)}\left(\frac{x}{|x|}\right), \tag{6}$$

$$\mathbf{h}_{n,m}^{R;(2,1)}(x) = \frac{1}{R}\left(\frac{R}{|x|}\right)^{n+1} \tilde{\mathbf{y}}_{n,m}^{(2,1)}\left(\frac{x}{|x|}\right), \tag{7}$$

$$\mathbf{h}_{n,m}^{R;(2,2)}(x) = \frac{1}{R}\left(\frac{R}{|x|}\right)^{n-1} \tilde{\mathbf{y}}_{n,m}^{(2,2)}\left(\frac{x}{|x|}\right), \tag{8}$$

$$\mathbf{h}_{n,m}^{R;(2,3)}(x) = \frac{1}{R}\left(\frac{R}{|x|}\right)^{n} \tilde{\mathbf{y}}_{n,m}^{(2,3)}\left(\frac{x}{|x|}\right), \tag{9}$$

$$\mathbf{h}_{n,m}^{R;(3,1)}(x) = \frac{1}{R}\left(\frac{R}{|x|}\right)^{n+2} \tilde{\mathbf{y}}_{n,m}^{(3,1)}\left(\frac{x}{|x|}\right), \tag{10}$$

$$\mathbf{h}_{n,m}^{R;(3,2)}(x) = \frac{1}{R}\left(\frac{R}{|x|}\right)^{n} \tilde{\mathbf{y}}_{n,m}^{(3,2)}\left(\frac{x}{|x|}\right), \tag{11}$$

$$\mathbf{h}_{n,m}^{R;(3,3)}(x) = \frac{1}{R}\left(\frac{R}{|x|}\right)^{n+1} \tilde{\mathbf{y}}_{n,m}^{(3,3)}\left(\frac{x}{|x|}\right), \tag{12}$$

$x \in \overline{\Omega_R^{\text{ext}}}$, $n = \tilde{0}_{ik}, \ldots;\ m = 1, \ldots, 2n+1$. It is not difficult to show that the following properties are satisfied:

- $\mathbf{h}_{n,m}^{R;(i,k)}$ is of class $\mathbf{c}^{(\infty)}(\Omega_R^{\text{ext}})$,
- $\Delta_x \mathbf{h}_{n,m}^{R;(i,k)}(x) = 0$ for $x \in \Omega_R^{\text{ext}}$, i.e., the component functions of $\mathbf{h}_{n,m}^{R;(i,k)}$ fulfill the Laplace equation,
- $\mathbf{h}_{n,m}^{R;(i,k)}|_{\Omega_R} = (1/R)\tilde{\mathbf{y}}_{n,m}^{(i,k)}$,
- $(\mathbf{h}_{n,m}^{R;(i,k)}, \mathbf{h}_{l,s}^{R;(p,q)})_{\mathbf{l}^2(\Omega_R)} = \int_{\Omega_R} \mathbf{h}_{n,m}^{R;(i,k)}(x) \cdot \mathbf{h}_{l,s}^{R;(p,q)}(x) d\omega_R(x) = \delta_{i,p}\delta_{k,q}\delta_{n,l}\delta_{m,s}$, where $\mathbf{l}^2(\Omega_R)$ is the space of square-integrable tensor fields on Ω_R,
- $|\mathbf{h}_{n,m}^{R;(i,k)}(x)| = \mathcal{O}\left(|x|^{-1}\right)$, $|x| \to \infty$.

It must be emphasized that the spherically reflected formulation of the tensorial SGG problem exclusively uses the tensor outer harmonics of kind $(1,1)$ specified by (4). However, for reasons of completeness of our tensor spherical approach, we have listed all kinds of outer harmonics.

Sphere to sphere interconnection between SGG-data and the gravitational potential

Next we are interested in characterizing the essential players involved in the SGG-match as members of infinite-dimensional potential spaces. We begin with the scalar space

$$Pot^{(0)}(\Omega_R^{\text{ext}}) = \{F \in C^{(2)}(\Omega_R^{\text{ext}}) :\ \Delta F = 0 \text{ in } \Omega_R^{\text{ext}},\ F(x) = \mathcal{O}\left(|x|^{-1}\right),\ |x| \to \infty\}.$$

In addition, $Pot^{(0)}(\overline{\Omega_R^{\text{ext}}})$ is the space of continuous functions $F : \overline{\Omega_R^{\text{ext}}} \to \mathbb{R}$ whose restrictions $F|_{\Omega_R^{\text{ext}}}$ are members of $Pot^{(0)}(\Omega_R^{\text{ext}})$. In brief,

$$Pot^{(0)}(\overline{\Omega_R^{\text{ext}}}) = Pot^{(0)}(\Omega_R^{\text{ext}}) \cap C(\overline{\Omega_R^{\text{ext}}}).$$

Clearly, in accordance with the well-posedness of the classical Dirichlet problem (see, e.g., [6]) on $Pot^{(0)}(\overline{\Omega_R^{\text{ext}}})$ we are able to impose an inner product by

$$(F,G)_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})} = (F,G)_{L^2(\Omega_R)} = \int_{\Omega_R} F(x)G(x)d\omega_R(x).$$

Furthermore,

$$Pot^{(\infty)}(\overline{\Omega_R^{\text{ext}}}) = Pot(\Omega_R^{\text{ext}}) \cap C^{(\infty)}(\overline{\Omega_R^{\text{ext}}}).$$

Now, let $\mathcal{A}$ be the linear space consisting of all sequences $\{A_n\}_{n\in\mathbb{N}}$ of real numbers $A_n \neq 0$, $n \in \mathbb{N}_0$:

$$\mathcal{A} = \{\{A_n\} : A_n \in \mathbb{R},\, A_n \neq 0,\, n \in \mathbb{N}_0\}.$$

Following [5] we consider the set $\mathcal{E} = \mathcal{E}(\{A_n\}; \overline{\Omega_R^{\text{ext}}})$ defined by

$$\mathcal{E} = \left\{ F \in Pot^{(\infty)}(\overline{\Omega_R^{\text{ext}}}) : \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} A_n^2 \left(F^{\wedge_{L^2(\Omega_R)}}(n,m) \right)^2 < \infty \right\},$$

where

$$F^{\wedge_{L^2(\Omega_R)}}(n,m) = \int_{\Omega_R} F(y) H_{n,m}^R(y)\, d\omega_R(y).$$

On $\mathcal{E}$ we define an inner product $(\cdot,\cdot)_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}$ by

$$(F,G)_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})} = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} A_n^2 F^{\wedge_{L^2(\Omega_R)}}(n,m) G^{\wedge_{L^2(\Omega_R)}}(n,m), \quad F,G \in \mathcal{E}.$$

As usual, the associated norm is given by

$$\|F\|_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})} = \left((F,F)_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})} \right)^{1/2}.$$

Definition 2.4. The (*scalar*) *Sobolev space* $\mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ $(= \mathcal{H}(\{A_n\}; \overline{\Omega_R^{\text{ext}}}))$ is the completion of $\mathcal{E}$ under the norm $\|\cdot\|_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}$:

$$\mathcal{H}(\{A_n\}; \overline{\Omega_R^{\text{ext}}}) = \overline{\mathcal{E}}^{\|\cdot\|_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}}.$$

$\mathcal{H}(\{A_n\}; \overline{\Omega_R^{\text{ext}}})$ equipped with the inner product $(\cdot,\cdot)_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}$ is a Hilbert space. The system $\{H_{n,m}^{*\{A_n\}}(R;\cdot)\}$ given by

$$H_{n,m}^{*\{A_n\}}(R;x) = A_n^{-1} H_{n,m}^R(x), \quad x \in \overline{\Omega_R^{\text{ext}}},$$

is a Hilbert basis. We simply write $H_{n,m}^*(R;\cdot)$ instead of $H_{n,m}^{*\{A_n\}}(R;\cdot)$ if no confusion is likely to arise. As is well known, any function $F \in \mathcal{H}(\{A_n\}; \overline{\Omega_R^{\text{ext}}})$ can be expanded as a Fourier series in terms of the basis $H_{n,m}^{*\{A_n\}}(R;\cdot)$:

$$F = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} F^{\wedge_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}}(n,m) H_{n,m}^{*\{A_n\}}(R;\cdot),$$

where

$$F^{\wedge_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}}(n,m) = (F, H_{n,m}^{*\{A_n\}}(R;\cdot))_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}.$$

If no confusion is likely to arise we will also use the notation

$$F^{\wedge_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}}(n,m) = F^{\wedge_{\mathcal{H}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}}(n,m).$$

Our next goal is the introduction of a class of scalar Sobolev spaces based on the spherical symbol $\{(\Delta^{*;R})^\wedge(n)\}_{n\in\mathbb{N}_0}$ of the Beltrami operator $\Delta^{*;R}$ on the sphere Ω_R. We know that

$$\Delta^{*;R}Y_{n,m} = \frac{1}{R^2}\Delta^* Y_{n,m} = -\frac{1}{R^2}n(n+1)Y_{n,m},\ n\in\mathbb{N}_0;\ m=1,\dots,2n+1.$$

In particular, we have $\Delta^{*;R}Y_{0,1} = 0$, which requires a shift by a constant, for example $\frac{1}{4R^2}$, to obtain invertibility. We formally have

$$\left(-\Delta^{*;R}+\frac{1}{4R^2}\right)^{s/2} Y_{n,m} = \left(\frac{n+1/2}{R}\right)^s Y_{n,m}$$

and

$$\left(\left(-\Delta^{*;R}+\frac{1}{4R^2}\right)^{s/2} F\right)^\wedge (n,m) = \left(\frac{n+1/2}{R}\right)^s F^\wedge(n,m),$$

$n\in\mathbb{N}_0$, $m=1,\dots,2n+1$. For $s\in\mathbb{R}$ we let

$$\mathcal{H}_s(\overline{\Omega_R^{\text{ext}}}) = \mathcal{H}\left(\left\{\left(\frac{n+1/2}{R}\right)^s\right\};\overline{\Omega_R^{\text{ext}}}\right)$$

and the norm in $\mathcal{H}_s(\overline{\Omega_R^{\text{ext}}})$ fulfills

$$\|F\|_{\mathcal{H}_s(\overline{\Omega_R^{\text{ext}}})} = \|(-\Delta_x^{*;R}+\frac{1}{4R^2})^{s/2}F\|_{L^2(\Omega_R)}.$$

Remark. For the space $\mathcal{H}_0(\overline{\Omega_R^{\text{ext}}})$ we identify the norm $\|\cdot\|_{\mathcal{H}_0(\overline{\Omega_R^{\text{ext}}})}$ with the $\|\cdot\|_{L^2(\Omega_R)}$-norm. The space $\mathcal{H}_0(\overline{\Omega_R^{\text{ext}}})$ may be understood to be the space of all solutions of the Dirichlet boundary value problem in $\overline{\Omega_R^{\text{ext}}}$ corresponding to $L^2(\Omega_R)$-boundary values on Ω_R. Note that the potential in $\mathcal{H}_0(\overline{\Omega_R^{\text{ext}}})$ corresponding to the $L^2(\Omega_R)$-(Dirichlet) boundary conditions on Ω_R is uniquely determined. Furthermore, if $t<s$, then $\|F\|_{\mathcal{H}_t(\overline{\Omega_R^{\text{ext}}})} \le \|F\|_{\mathcal{H}_s(\overline{\Omega_R^{\text{ext}}})}$ and $\mathcal{H}_s(\overline{\Omega_R^{\text{ext}}}) \subset \mathcal{H}_t(\overline{\Omega_R^{\text{ext}}})$.

In order to formulate some results about the convergence of the expansion in terms of outer harmonics to a function in ordinary sense (Sobolev Lemma) we need the concept of summable sequences $\{A_n\}_{n\in\mathbb{N}_0}\in\mathcal{A}$ satisfying

$$\sum_{n=0}^{\infty}\frac{2n+1}{A_n^2}<\infty.$$

Theorem 2.5 (Sobolev Lemma). *Assume that the sequences $\{A_n\}_{n\in\mathbb{N}_0},\{B_n\}_{n\in\mathbb{N}_0}\in\mathcal{A}$ are given in such a way that $\{B_n^{-1}A_n\}_{n\in\mathbb{N}_0}$ is summable. Then each $F\in\mathcal{H}\{B_n^{-1}A_n\};\overline{\Omega_R^{\text{ext}}}$ corresponds to a potential of class $Pot^{(0)}(\overline{\Omega_R^{\text{ext}}})$.*

The Sobolev Lemma (see [5] for its proof) states that in the case of summability of the sequence $\{B_n^{-1}A_n\}_{n\in\mathbb{N}_0}$, the Fourier series in terms of the basis functions $H_{n,m}^*\in\mathcal{H}\{B_n^{-1}A_n\};\overline{\Omega_R^{\text{ext}}}$ is continuous on the boundary Ω_R. In particular, we have the following statement.

Theorem 2.6. *If $F \in \mathcal{H}_s(\overline{\Omega_R^{\text{ext}}})$ with $s > 1$, then F corresponds to a function of class $Pot^{(0)}(\overline{\Omega_R^{\text{ext}}})$.*

For any F in $L^2(\Omega_R)$, there exists one and only one "harmonic continuation" $U \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ which is given by

$$U(x) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} A_n^2 F^{\wedge_{L^2(\Omega_R)}}(n,m) H_{n,m}^*(R;x), \quad x \in \overline{\Omega_R^{\text{ext}}}, \tag{13}$$

where

$$F^{\wedge_{L^2(\Omega_R)}}(n,m) = \int_{\Omega_R} F(y) H_{n,m}^R(y) d\omega_R(y) = \frac{1}{A_n^2} U^{\wedge_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}}(n,m),$$

$n = 1, 2, \ldots$; $m = 1, \ldots, 2n+1$.

The scalar Sobolev space $\mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ is a separable Hilbert space and the system $\{H_{n,m}^*(R;\cdot)\}$ is a Hilbert basis. The space $\mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ has the reproducing kernel function $K_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}(\cdot,\cdot) : \overline{\Omega_R^{\text{ext}}} \times \overline{\Omega_R^{\text{ext}}} \to \mathbb{R}$ given by

$$K_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}(x,y) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} H_{n,m}^*(R;x) H_{n,m}^*(R;y), \quad x, y \in \overline{\Omega_R^{\text{ext}}}.$$

In analogy to the scalar case we now introduce its tensorial counterpart $\mathbf{pot}(\Omega_R^{\text{ext}})$ as follows:

$$\mathbf{pot}(\Omega_R^{\text{ext}}) = \{\mathbf{f} \in \mathbf{c}^{(1)}(\Omega_R^{\text{ext}}) : \nabla \cdot \mathbf{f} = 0, \quad \nabla \wedge \mathbf{f} = 0 \text{ in } \Omega_R^{\text{ext}}, \\ |\mathbf{f}(x)| = \mathcal{O}\left(|x|^{-3}\right), |x| \to \infty\}.$$

Similarly, we let

$$\mathbf{pot}^{(0)}(\overline{\Omega_R^{\text{ext}}}) = \mathbf{pot}(\Omega_R^{\text{ext}}) \cap \mathbf{c}^{(0)}(\overline{\Omega_R^{\text{ext}}}),$$

and

$$\mathbf{pot}^{(\infty)}(\overline{\Omega_R^{\text{ext}}}) = \mathbf{pot}(\Omega_R^{\text{ext}}) \cap \mathbf{c}^{(\infty)}(\overline{\Omega_R^{\text{ext}}}).$$

In order to introduce Sobolev spaces for tensor fields we remember the sequences $\{A_n\}_{n\in\mathbb{N}_0} \in \mathcal{A}$. Then we define

$$\mathbf{e} = \left\{\mathbf{f} \in \mathbf{pot}^{(\infty)}(\overline{\Omega_R^{\text{ext}}}) : \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} A_n^2 (\mathbf{f}^{\wedge_{\mathbf{l}^2(\Omega_R)}}(n,m))^2 < \infty\right\},$$

where

$$\mathbf{f}^{\wedge_{\mathbf{l}^2(\Omega_R)}}(n,m) = \int_{\Omega_R} \mathbf{f}(y) \mathbf{h}_{n,m}^{R;(1,1)}(y) d\omega_R(y).$$

Equipped with the inner product

$$(\mathbf{f},\mathbf{g})_{\mathbf{h}(\{A_n\};\overline{\Omega_R^{\text{ext}}})} = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} A_n^2 \mathbf{f}^{\wedge_{\mathbf{l}^2(\Omega_R)}}(n,m) \mathbf{g}^{\wedge_{\mathbf{l}^2(\Omega_R)}}(n,m),$$

$\mathbf{f}, \mathbf{g} \in \mathbf{e}$, the space $\mathbf{e}$ becomes a pre-Hilbert space. We define the Sobolev space $\mathbf{h}(\overline{\Omega_R^{\text{ext}}}) = \mathbf{h}(\{A_n\}; \overline{\Omega_R^{\text{ext}}})$ to be the completion of $\mathbf{e}$ under the norm $\|\cdot\|_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})} = \|\cdot\|_{\mathbf{h}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}$, which denotes the norm associated to $(\cdot,\cdot)_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})} = (\cdot,\cdot)_{\mathbf{h}(\{A_n\};\Omega_R^{\text{ext}})}$:

$$\mathbf{h}(\{A_n\}; \overline{\Omega_R^{\text{ext}}}) = \overline{\mathbf{e}}^{\|\cdot\|_{\mathbf{h}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}}.$$

The space $\mathbf{h}(\overline{\Omega_R^{\text{ext}}})$ $(= \mathbf{h}(\{A_n\}; \overline{\Omega_R^{\text{ext}}}))$ equipped with the inner product $(\cdot,\cdot)_{\mathbf{h}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}$ is a Hilbert space. The system $\{\mathbf{h}_{n,m}^{*\{A_n\}}(R;\cdot)\}_{n\in\mathbb{N}_0;m=1,\ldots,2n+1}$, given by

$$\mathbf{h}_{n,m}^{*\{A_n\}}(R;x) = A_n^{-1}\mathbf{h}_{n,m}^{R;(1,1)}(x), \quad x \in \overline{\Omega_R^{\text{ext}}},$$

represents an $\mathbf{h}(\overline{\Omega_R^{\text{ext}}})$-orthonormal Hilbert basis in $\mathbf{h}(\overline{\Omega_R^{\text{ext}}})$. We simply write $\mathbf{h}_{n,m}^*(R;\cdot)$ instead of $\mathbf{h}_{n,m}^{*\{A_n\}}(R;\cdot)$ if no confusion is likely to arise. As a consequence we can expand a function $\mathbf{f} \in \mathbf{h}(\overline{\Omega_R^{\text{ext}}})$ as a Fourier series in terms of the basis $\mathbf{h}_{n,m}^{*\{A_n\}}(R;\cdot)$:

$$\mathbf{f} = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1} \mathbf{f}^{\wedge_{\mathbf{h}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}}(n,m)\mathbf{h}_{n,m}^{*\{A_n\}}(R;\cdot),$$

where

$$\mathbf{f}^{\wedge_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})}}(n,m) = \mathbf{f}^{\wedge_{\mathbf{h}(\{A_n\};\overline{\Omega_R^{\text{ext}}})}}(n,m) = (\mathbf{f}, \mathbf{h}_{n,m}^{*\{A_n\}}(R;\cdot))_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})}.$$

Finally, we are led to define

$$\mathbf{h}_s(\overline{\Omega_R^{\text{ext}}}) = \mathbf{h}\left(\left\{\left(\frac{n+1/2}{R}\right)^s\right\}; \overline{\Omega_R^{\text{ext}}}\right).$$

The Sobolev Lemma (Theorem 2.5) can be extended in the same way to tensor fields.

Theorem 2.7 (Tensorial Sobolev Lemma). *Suppose that the sequences* $\{A_n\}_{n\in\mathbb{N}_0}$, $\{B_n\}_{n\in\mathbb{N}_0} \in \mathcal{A}$ *are given such that* $\{B_n^{-1}A_n\}_{n\in\mathbb{N}_0} \in \mathcal{A}$ *is summable. Then each* $\mathbf{f} \in \mathbf{h}\big(\{B_n^{-1}A_n\}; \overline{\Omega_R^{\text{ext}}}\big)$ *corresponds to a function of class* $\mathbf{pot}^{(0)}\big(\overline{\Omega_R^{\text{ext}}}\big)$.

For any $\mathbf{f} \in \mathbf{l}^2(\Omega_R)$, there exists one and only one tensorial "harmonic upward continuation" $\mathbf{u} \in \mathbf{h}(\overline{\Omega_R^{\text{ext}}})$ of the form

$$\mathbf{u}(x) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1} A_n^2 \mathbf{f}^{\wedge_{\mathbf{l}^2(\Omega_R)}}(n,m)\mathbf{h}_{n,m}^*(R;x), \quad x \in \overline{\Omega_R^{\text{ext}}}, \tag{14}$$

where

$$\mathbf{f}^{\wedge_{\mathbf{l}^2(\Omega_R)}}(n,m) = \int_{\Omega_R} \mathbf{f}(y)\cdot\mathbf{h}_{n,m}^{R;(1,1)}(y)d\omega_R(y) = \frac{1}{A_n^2}\mathbf{u}^{\wedge_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})}}(n,m).$$

The tensorial Sobolev space $\mathbf{h}(\overline{\Omega_R^{\text{ext}}})$ is a separable Hilbert space, and the system $\{\mathbf{h}_{n,m}^{*;\{A_n\}}(R;\cdot)\}$ is a Hilbert basis. The space $\mathbf{h}(\overline{\Omega_R^{\text{ext}}})$ has the reproducing kernel

function

$$\mathbf{K}_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})}(\cdot,\cdot) : \overline{\Omega_R^{\text{ext}}} \otimes \overline{\Omega_R^{\text{ext}}} \to \mathbb{R}^3 \otimes \mathbb{R}^3 \otimes \mathbb{R}^3 \otimes \mathbb{R}^3$$

given by

$$\mathbf{K}_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})}(x,y) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \mathbf{h}_{n,m}^{*\{A_n\}}(x) \otimes \mathbf{h}_{n,m}^{*\{A_n\}}(y),$$

$x, y \in \overline{\Omega_R^{\text{ext}}}$. This means that

- for all $x \in \overline{\Omega_R^{\text{ext}}}$, $\tilde{O}_R^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})}(\cdot, x) \in \mathbf{h}(\overline{\Omega_R^{\text{ext}}})$, where the operator $\tilde{O}_R^{(1,1)}$ is the extension of the adjoint operator of $\tilde{\mathbf{o}}_R^{(1,1)}$ to tensor fields of rank four,
- $\tilde{O}_R^{(1,1)} \mathbf{f}(x) = \left(\tilde{O}_R^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})}(\cdot, x), \mathbf{f}\right)_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})}$ for every $\mathbf{f} \in \mathbf{h}(\overline{\Omega_R^{\text{ext}}})$ and all $x \in \overline{\Omega_R^{\text{ext}}}$.

More detailed information about tensorial Sobolev spaces can be found in the Ph.D.-thesis [26]. The interrelation between scalar outer harmonics and their Hesse tensor is known from [9]

$$(\nabla_x \otimes \nabla_x) H_{n,m}^R(x) = \frac{1}{R^2} \sqrt{\tilde{\mu}_n^{(1,1)}} \mathbf{h}_{n,m}^{R;(1,1)}(x), \quad n = 0, 1, \ldots; m = 1, \ldots, 2n+1.$$

The last identity enables us to deduce that, for all $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$, the "Meissl relation"

$$\left(\nabla \otimes \nabla F, \mathbf{h}_{n,m}^*(\tau;\cdot)\right)_{\mathbf{h}(\{A_n\};\overline{\Omega_\tau^{\text{ext}}})} = \left(\frac{R}{\tau}\right)^n \frac{\sqrt{\tilde{\mu}_{n,m}^{(1,1)}}}{\tau^2} \left(F, H_{n,m}^*(R;\cdot)\right)_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}$$

holds true for all $\tau > R$ and all n, m. This immediately leads us to the scalar outer harmonic expansion in terms of tensorial Hesse data

$$F = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \left(\frac{\tau}{R}\right)^n \frac{\tau^2}{\sqrt{\tilde{\mu}_{n,m}^{(1,1)}}} \left(\nabla \otimes \nabla F\right)^{\wedge_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}}(n,m) H_{n,m}^*(R;\cdot). \tag{15}$$

The correlations between the potential F and the full Hesse tensor of F on the Earth's surface and the satellite orbit can be presented in a so-called Meissl scheme (cf. [31]) as shown in Figure 2. Detailed information about Meissl schemes both in the framework of outer harmonics and multiscale analysis can be found elsewhere in this handbook.

Clearly, this formula is extremely suitable in the determination of the scalar gravitational potential on a spherical Earth Ω_R from tensorial SGG-data on a spherical orbit Ω_τ. It expresses the gravitational potential $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ in terms of the gravitational tensor $\nabla \otimes \nabla F$ on the spherical satellite orbit Ω_τ in terms of a spherical harmonic expansion, where the convergence of the series (15) is understood in uniform sense on every subset $S \subset \Omega_R^{\text{ext}}$ with $\text{dist}(S, \Omega_R) > 0$. Even more, the convergence on Ω_R can also be understood in the $L^2(\Omega_R)$-topology.

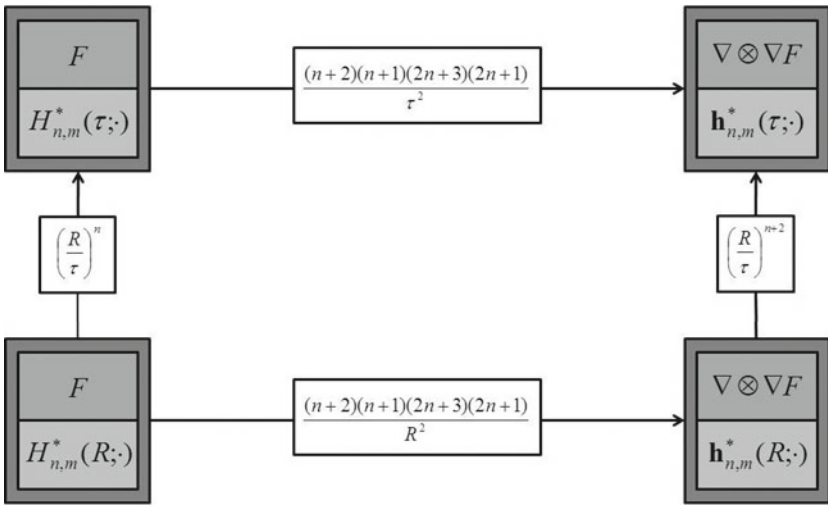


FIGURE 2. The full Meissl scheme for the Hesse tensor on the Earth's surface and on the satellite orbit (see the contribution [10] in this handbook).

The outer harmonic expansion (15) actually represents the basic setting for the multiresolution approach by means of (outer) harmonic wavelets as proposed later on in this work.

Scalar Runge–Walsh approximation property

From [3] we know the following constructive version of the Runge theorem in terms of outer harmonics

$$Pot^{(0)}(\overline{\Sigma^{\text{ext}}}) = \overline{\text{span}\,\underset{\substack{m=1,\ldots,2n+1}}{n=0,1,\ldots;}\,(H^R_{n,m})|_{\overline{\Sigma^{\text{ext}}}}}^{\|\cdot\|_{C(\overline{\Sigma^{\text{ext}}})}}, \tag{16}$$

where Σ is any regular surface (for example, a sphere, an ellipsoid, a telluroid, the geoid, or the real Earth's surface) and Ω_R is a sphere inside Σ, such that the "Runge condition" $R < \sigma = \inf_{x\in\Sigma}|x|$ is valid.

The Runge–Walsh approximation property(16) justifies the approximation of the Earth's gravitational potential on and outside the regular surface by a linear combination of scalar outer harmonics, i.e., by harmonic "trial functions of polynomial nature" showing a "harmonicity domain" $\Omega^{\text{ext}}_R \supset \overline{\Sigma^{\text{ext}}}$. It should be remarked that the same property holds true, for example, for outer ellipsoidal harmonics. However, for reasons of numerical economy and efficiency, we restrict ourselves to outer spherical harmonics.

Tensorial Runge–Walsh approximation property

In an analogous way we are able to deduce from [26] that

$$\mathbf{pot}^{(0)}(\overline{\Sigma^{\text{ext}}}) = \overline{\text{span}\underset{\substack{n=0,1,\ldots;\\ m=1,\ldots,2n+1}}{} (\mathbf{h}_{n,m}^{R;(1,1)})|_{\overline{\Sigma^{\text{ext}}}}}^{\|\cdot\|_{\mathbf{c}(\overline{\Sigma^{\text{ext}}})}}.$$

The calamity of evaluating the gravitational potential in terms of outer harmonics is that these basis functions are globally supported. This is the reason why they do not show any space localization but ideal frequency (momentum) localization (for a more detailed description see [12]). This property makes outer harmonics difficult to use for high resolution modeling at local scale. As a matter of fact, the uncertainty principle leads us to the conclusion that outer harmonics are well suited to resolve low-frequency problems at global scale, i.e., to provide trend approximations.

Roughly spoken, seen from a numerical point of view, suitable ansatz spaces should consist of harmonic functions possessing ideal space as well as frequency localization. But this is mutually exclusive for several reasons (see [4]). Nevertheless, as well promising compromise, we are able to handle "sum conglomerates" of outer harmonics, i.e., so-called (outer) harmonic kernel functions, offering a limited but appropriately balanced range of frequency as well as space localization. Even better, we can construct families of kernels which control the increase of space localization at the cost of the decrease of frequency localization by specifying a scale parameter. Before we come to the definition of such families, called *scaling functions*, within regularization procedures of the exponentially ill-posed SGG-problem, however, it is advisable to describe the SGG-problem as a pseudodifferential equation between the Sobolev spaces $\mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ and $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$.

3. Functional analytic background

For the convenience of the reader, we present here a brief course of basic facts on regularization in a Hilbert space setting, which is useful to understand the spline and multiscale solution strategies in the framework of pseudodifferential equations. The explanations are based on functional analytic tools as described in [1, 2, 18] where much more additional material can be found even for more general reference spaces, too.

Let $\mathcal{H}$ and $\mathcal{K}$ be two Hilbert spaces with inner products $(\cdot,\cdot)_{\mathcal{H}}$ and $(\cdot,\cdot)_{\mathcal{K}}$, respectively. Let

$$\Lambda : \mathcal{H} \longrightarrow \mathcal{K}$$

be a linear bounded operator. Given $y \in \mathcal{K}$, we are looking for a solution of

$$\Lambda x = y. \tag{17}$$

In accordance to Hadamard (cf. [16]) we call such a problem *well-posed*, if the following properties are valid:

- For all admissible data, a solution exists.
- For all admissible data, the solution is unique.
- The solution depends continuously on the data.

In our setting, these requirements can be translated into

- Λ is injective, i.e., $\mathcal{R}(\Lambda) = \mathcal{K}$.
- Λ is surjective, i.e., $\mathcal{N}(\Lambda) = \{0\}$.
- Λ^{-1} is bounded and continuous.

If one of the three conditions is not fulfilled, the problem (17) is called *ill-posed*. It will turn out, that the satellite problems we are concerned with are ill posed, the most critical problem being the unboundedness of the inverse operator Λ^{-1}.

Let us discuss the consequences of the violations of the above requirements for the well-posedness of equation (17). The lack of injectivity of Λ is perhaps the easiest problem. From theoretical point of view, the space $\mathcal{H}$ can be replaced by the orthogonal complement $\mathcal{N}(\Lambda)^{\perp}$, and the restriction of the operator Λ to $\mathcal{N}(\Lambda)^{\perp}$ yields an injective problem. But in practice, one is normally confronted with the problem that $\mathcal{R}(\Lambda) \neq \mathcal{K}$, since the right-hand side is given by measurements and is, therefore, disturbed by errors. Now, we assume that $y \in \mathcal{R}(\Lambda)$, but only a perturbed right-hand side y^{δ} is known. We suppose that

$$\|y - y^{\delta}\|_{\mathcal{K}} < \delta.$$

Our aim is to solve

$$\Lambda x^{\delta} = y^{\delta}. \tag{18}$$

Since y^{δ} might not be in $\mathcal{R}(\Lambda)$, the solution of this equation might not exist, and we have to generalize what is meant by a solution. x^{δ} is called *least squares solution* of (18), if

$$\|\Lambda x^{\delta} - y^{\delta}\|_{\mathcal{K}} = \inf\{\|\Lambda z - y^{\delta}\|_{\mathcal{K}} \ : \ z \in \mathcal{H}\}. \tag{19}$$

The solution of (19) might not be unique, and therefore one looks for the solution of (19) with minimal norm. x^{δ} is called *best approximate solution* of $\Lambda x^{\delta} = y^{\delta}$, if x^{δ} is a least squares solution and

$$\|x^{\delta}\|_{\mathcal{H}} = \inf\{\|z\|_{\mathcal{H}} \ : \ z \text{ is a least squares solution of } \Lambda z = y^{\delta}\} \tag{20}$$

holds.

The notion of a best-approximate solution is closely related to the Moore–Penrose (generalized) inverse of Λ (see [22, 24], and a large amount of subsequent contributions). We let

$$\tilde{\Lambda} : \mathcal{N}(\Lambda)^{\perp} \longrightarrow \mathcal{R}(\Lambda) \quad \text{with} \quad \tilde{\Lambda} = \Lambda|_{\mathcal{N}(\Lambda)^{\perp}}$$

and define the *Moore–Penrose* (*generalized*) *inverse* Λ^{+} to be the unique linear extension of $\tilde{\Lambda}^{-1}$ to

$$\mathcal{D}(\Lambda^{+}) := \mathcal{R}(\Lambda) + \mathcal{R}(\Lambda)^{\perp}$$

with

$$\mathcal{N}(\Lambda^{+}) = \mathcal{R}(\Lambda)^{\perp}.$$

A standard result is provided by

Theorem 3.1. *If $y \in \mathcal{D}(\Lambda^+)$, then $\Lambda x = y$ has a unique best-approximate solution which is given by*

$$x^+ = \Lambda^+ y.$$

Note that the best-approximate solution is defined for all perturbed data $y^\delta \in \mathcal{K}$, whereas the last theorem requires that the right-hand side is an element of $\mathcal{D}(\Lambda^+)$.

A serious problem for ill-posed problems occurs when Λ^{-1} or Λ^+ are not continuous. That means that small errors in the data or even small numerical noise can cause large errors in the solution. In fact, in most cases the application of an unbounded Λ^{-1} or Λ^+ does not make any sense. The usual strategy to overcome this difficulty is to substitute the unbounded inverse operator

$$\Lambda^{-1} : \mathcal{R}(\Lambda) \longrightarrow \mathcal{H}$$

by a suitable bounded approximation

$$R : \mathcal{K} \longrightarrow \mathcal{H}\,.$$

The operator R is not chosen to be fix, but dependent on a *regularization parameter* α. According to the conventional approach in the theory of ill-posed problems we are led to introduce the following definition:

Definition 3.2. A *regularization strategy* is a family of linear bounded operators

$$R_\alpha : \mathcal{K} \longrightarrow \mathcal{H},\ \alpha > 0,$$

so that

$$\lim_{\alpha \to 0} R_\alpha \Lambda x = x \quad \text{for all } x \in \mathcal{H},$$

i.e., the operators $\mathcal{R}_\alpha \Lambda$ converge pointwise to the identity.

From the theory of inverse problems (see, e.g., [24, 25]) it is also clear that if $\Lambda : \mathcal{H} \to \mathcal{K}$ is compact and $\mathcal{H}$ has infinite dimension (as it is the case for the application we have in mind), then the operators $\mathcal{R}_\alpha$ are not uniformly bounded, i.e., there exists a sequence (α_j) with $\lim_{j\to\infty} \alpha_j = 0$ and

$$\|R_{\alpha_j}\|_{\mathcal{L}(\mathcal{K},\mathcal{H})} \to \infty \quad \text{for } j \to \infty.$$

Note that the convergence of $\mathcal{R}_\alpha \Lambda x$ in Definition 3.2 is based on $y = \Lambda x$, i.e., on unperturbed data. In practice, the right-hand side is affected by errors and then no convergence is achieved. Instead, one is (or has to be) satisfied with an approximate solution based on a certain choice of the regularization parameter.

Let us discuss the error of the solution. For that purpose, we let $y \in \mathcal{R}(\Lambda)$ be the (unknown) exact right-hand side and $y^\delta \in \mathcal{K}$ be the measured data with

$$\|y - y^\delta\|_{\mathcal{K}} < \delta.$$

For a fixed $\alpha > 0$, we let

$$x^{\alpha,\delta} = R_\alpha y^\delta,$$

and look at $x^{\alpha,\delta}$ as an approximation of the solution x of $\Lambda x = y$. Then the error can be split as follows:

$$\begin{aligned}\|x^{\alpha,\delta} - x\|_{\mathcal{H}} &= \|R_\alpha y^\delta - x\|_{\mathcal{H}} \\ &\le \|R_\alpha y^\delta - R_\alpha y\|_{\mathcal{H}} + \|R_\alpha y - x\|_{\mathcal{H}} \\ &\le \|R_\alpha\|_{\mathcal{L}(\mathcal{K},\mathcal{H})}\ \|y^\delta - y\|_{\mathcal{K}} + \|R_\alpha y - x\|_{\mathcal{H}},\end{aligned}$$

such that

$$\|x^{\alpha,\delta} - x\|_{\mathcal{H}} \le \delta\|R_\alpha\|_{\mathcal{L}(\mathcal{K},\mathcal{H})} + \|R_\alpha \Lambda x - x\|_{\mathcal{H}}.$$

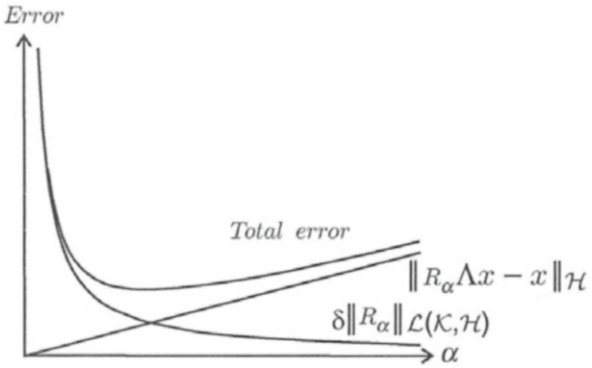


FIGURE 3. Typical behavior of the total error in a regularization process.

We see that the error between the exact and the approximate solution consists of two parts: The first term is the product of the bound for the error in the data and the norm of the regularization parameter R_α. This term will usually tend to infinity for $\alpha \to 0$ if the inverse Λ^{-1} is unbounded and Λ is compact (cf. (3)). The second term denotes the approximation error $\|(R_\alpha - \Lambda^{-1})y\|_{\mathcal{H}}$ for the exact right-hand side $y = \Lambda x$. This error tends to zero as $\alpha \to 0$ by the definition of a regularization strategy. Thus, both parts of the error show a diametrically oriented behavior. A typical picture of the errors in dependence on the regularization parameter α is sketched in Figure 3. Thus, a strategy is needed to choose α dependent on δ in order to keep the error as small as possible, i.e., we would like to minimize

$$\delta\|R_\alpha\|_{\mathcal{L}(\mathcal{K},\mathcal{H})} + \|R_\alpha \Lambda x - x\|_{\mathcal{H}}.$$

In principle, we distinguish two classes of parameter choice rules: If $\alpha = \alpha(\delta)$ only depends on δ, we call $\alpha = \alpha(\delta)$ an *a priori* parameter choice rule. Otherwise α depends also on y^δ and we call $\alpha = \alpha(\delta, y^\delta)$ an *a posteriori* parameter choice rule. It is usual to say that a parameter choice rule is convergent, if for $\delta \to 0$ the rule fulfills the limit relations

$$\lim_{\delta\to 0} \sup\{\|R_{\alpha(\delta,y^\delta)} y^\delta - \Lambda^+ y\|_{\mathcal{H}} :\ y^\delta \in \mathcal{K},\ \|y^\delta - y\|_{\mathcal{K}} \le \delta\} = 0 \tag{21}$$

and

$$\lim_{\delta\to 0}\sup\{\alpha(\delta,y^\delta):\ y^\delta\in\mathcal{K},\ \|y-y^\delta\|_{\mathcal{K}}\le\delta\}=0. \tag{22}$$

We stop here the discussion of parameter choice rules. For more material the interested reader is referred to any textbook on inverse problems, e.g., [2, 18, 19, 27].

The remaining part of this section is devoted to the case that Λ is compact, since then we gain benefits from the spectral representations of the operators. If $\Lambda:\mathcal{H}\to\mathcal{K}$ is compact, a singular system $(\sigma_n;v_n,u_n)$ is defined as follows: $\{\sigma_n^2\}_{n\in\mathbb{N}}$ are the nonzero eigenvalues of the self-adjoint operator $\Lambda^*\Lambda$ (Λ^* is the adjoint operator of Λ), written down in decreasing order with corresponding multiplicity. The family $\{v_n\}_{n\in\mathbb{N}}$ constitutes a corresponding complete orthonormal system of eigenvectors of $\Lambda^*\Lambda$. We let $\sigma_n>0$ and define the family $\{u_n\}_{n\in\mathbb{N}}$ via $u_n=\Lambda v_n/\|\Lambda v_n\|_{\mathcal{K}}$. The sequence $\{u_n\}_{n\in\mathbb{N}}$ forms a complete orthonormal system of eigenvectors of $\Lambda\Lambda^*$, and the following formulas are valid:

$$\Lambda v_n=\sigma_n u_n, \tag{23}$$

$$\Lambda^* u_n=\sigma_n v_n, \tag{24}$$

$$\Lambda x=\sum_{n=1}^{\infty}\sigma_n(x,v_n)_{\mathcal{H}}u_n,\quad x\in\mathcal{H}, \tag{25}$$

$$\Lambda^* y=\sum_{n=1}^{\infty}\sigma_n(y,u_n)_{\mathcal{K}}v_n,\quad y\in\mathcal{K}. \tag{26}$$

The convergence of the infinite series is understood with respect to the Hilbert space norms under consideration. The identities (25) and (26) are called the *singular value expansions* of the corresponding operators. If there are infinitely many singular values, they tend to 0, i.e., $\lim_{n\to\infty}\sigma_n=0$.

Theorem 3.3. *Let $(\sigma_n;v_n,u_n)$ be a singular system for the compact linear operator Λ, $y\in\mathcal{K}$. Then we have*

$$y\in\mathcal{D}(\Lambda^+)\ \textit{if and only if}\ \sum_{n=1}^{\infty}\frac{|(y,u_n)_{\mathcal{K}}|^2}{\sigma_n^2}<\infty, \tag{27}$$

and for $y\in\mathcal{D}(\Lambda^+)$ it holds

$$\Lambda^+ y=\sum_{n=1}^{\infty}\frac{(y,u_n)_{\mathcal{K}}}{\sigma_n}v_n. \tag{28}$$

The condition (27) is the *Picard criterion*. It says that a best-approximate solution of $\Lambda x=y$ exists only if the Fourier coefficients of y decrease fast enough relative to the singular values.

The representation (28) of the best-approximate solution motivates a method for the construction of regularization operators, namely by damping the factors

$1/\sigma_n$ in such a way that the series converges for all $y \in \mathcal{K}$. We are looking for filters

$$q : (0, \infty) \times (0, \|\Lambda\|_{\mathcal{L}(\mathcal{H},\mathcal{K})}) \longrightarrow \mathbb{R} \tag{29}$$

such that

$$R_\alpha y := \sum_{n=1}^{\infty} \frac{q(\alpha, \sigma_n)}{\sigma_n} (y, u_n)_{\mathcal{K}} \, v_n, \quad y \in \mathcal{K},$$

is a regularization strategy. The following statement is known from [18].

Theorem 3.4. *Let $\Lambda : \mathcal{H} \to \mathcal{K}$ be compact with singular system $(\sigma_n; v_n, u_n)$. Assume that q from* (29) *has the following properties:*

(i) $|q(\alpha, \sigma)| \leq 1$ *for all* $\alpha > 0$ *and* $0 < \sigma \leq \|\Lambda\|_{\mathcal{L}(\mathcal{H},\mathcal{K})}$.
(ii) *For every* $\alpha > 0$ *there exists a* $c(\alpha)$ *so that* $|q(\alpha, \sigma)| \leq c(\alpha)\sigma$ *for all* $0 < \sigma \leq \|\Lambda\|_{\mathcal{L}(\mathcal{H},\mathcal{K})}$.
(iii) $\lim_{\alpha \to 0} q(\alpha, \sigma) = 1$ *for every* $0 \leq \sigma \leq \|\Lambda\|_{\mathcal{L}(\mathcal{H},\mathcal{K})}$.

Then the operator $R_\alpha : \mathcal{K} \to \mathcal{H}$, $\alpha > 0$, defined by

$$R_\alpha y := \sum_{n=1}^{\infty} \frac{q(\alpha, \sigma_n)}{\sigma_n} (y, u_n)_{\mathcal{K}} \, v_n, \quad y \in \mathcal{K},$$

is a regularization strategy with $\|R_\alpha\|_{\mathcal{L}(\mathcal{K},\mathcal{H})} \leq c(\alpha)$.

The function q is called a *regularizing filter* for Λ. Two important examples should be mentioned:

$$q(\alpha, \sigma) = \frac{\sigma^2}{\alpha + \sigma^2}$$

defines the *Tikhonov regularization*, whereas

$$q(\alpha, \sigma) = \begin{cases} 1, & \sigma^2 \geq \alpha, \\ 0, & \sigma^2 < \alpha \end{cases}$$

leads to the regularization by *truncated singular value decomposition.*

4. SGG as exponentially ill-posed problem

After the discussion of the spherical settings and the outer harmonic based nomenclature leading to orthogonal expansions we are interested in formulating the SGG problem by use of the concept of pseudodifferential operators. To this end we shortly introduce the framework of tensorial spherical pseudodifferential operators (PDO) (for a profound definition the reader is referred to [12]): Let $\{\mathbf{\Lambda}^\wedge(n)\}_{n \in \mathbb{N}_0}$ be a sequence of real numbers. The operator $\mathbf{\Lambda} : \mathcal{H}(\overline{\Omega_R^{\text{ext}}}) \to \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$, $\tau \geq R > 0$, defined by

$$\mathbf{\Lambda} F = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \mathbf{\Lambda}^\wedge(n) F^{\wedge_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}}(n, m) \mathbf{h}^*_{n,m}(\tau; \cdot),$$

is called a *tensorial pseudodifferential operator of order t* if

$$\lim_{n\to\infty} \frac{|\mathbf{\Lambda}^\wedge(n)|}{\left(n+\frac{1}{2}\right)^t} = \text{const} \neq 0$$

holds true for some $t \in \mathbb{R}$. If the limit

$$\lim_{n\to\infty} \frac{|\mathbf{\Lambda}^\wedge(n)|}{\left(n+\frac{1}{2}\right)^t} = 0$$

holds true for all $t \in \mathbb{R}$, the operator Λ is called a *tensorial pseudodifferential operator of exponential order*. The sequence $\{\mathbf{\Lambda}^\wedge(n)\}_{n\in\mathbb{N}_0}$ is called the *symbol of the tensorial PDO* $\mathbf{\Lambda}$.

In the following we define scalar and tensorial kernel functions which are of basic importance for the consideration of the SGG problem in terms of pseudodifferential operators.

Definition 4.1. Suppose that $R, \tau \in \mathbb{R}$, $\tau \geq R \geq 0$ as usual. Then any kernel $K^{R,R}(\cdot,\cdot) : \overline{\Omega_R^{\text{ext}}} \times \overline{\Omega_R^{\text{ext}}} \to \mathbb{R}$ of the form

$$K^{R,R}(x,y) = \sum_{n=0}^{\infty} K^\wedge(n) \sum_{m=1}^{2n+1} H_{n,m}^*(R;x) H_{n,m}^*(R;y), \quad x,y \in \overline{\Omega_R^{\text{ext}}}$$

is called an $\mathcal{H}_{R,R}$-kernel. Any kernel $\mathbf{k}^{R,\tau}(\cdot,\cdot) : \overline{\Omega_R^{\text{ext}}} \times \overline{\Omega_\tau^{\text{ext}}} \to \mathbb{R}^3 \otimes \mathbb{R}^3$ of the form

$$\mathbf{k}^{R,\tau}(x,y) = \sum_{n=0}^{\infty} \mathbf{k}^\wedge(n) \sum_{m=1}^{2n+1} H_{n,m}^*(R;x) \mathbf{h}_{n,m}^*(\tau;y)$$

$(x,y) \in \overline{\Omega_R^{\text{ext}}} \times \overline{\Omega_\tau^{\text{ext}}}$ is called an $\mathbf{h}_{R,\tau}$-kernel.

The sequence $\{K^\wedge(n)\}_{n\in\mathbb{N}_0}$ is called the *symbol of the* $\mathcal{H}_{R,R}$*-kernel*, whereas $\mathbf{k}^\wedge(n)$ is called the *symbol of the* $\mathbf{h}_{R,\tau}$*-kernel*.

Definition 4.2. An $\mathcal{H}_{R,R}$-kernel $K^{R,R}(\cdot,\cdot)$ with the symbol $\{K^\wedge(n)\}_{n=0,\ldots}$ is called admissible, if the following conditions are satisfied:

1. $\sum_{n=0}^{\infty} (K^\wedge(n))^2 < \infty$,
2. $\sum_{n=0}^{\infty} (2n+1)^2 \left(\frac{K^\wedge(n)}{A_n}\right)^2 < \infty$.

In analogy, an $\mathbf{h}_{R,\tau}$-kernel $\mathbf{k}^{R,\tau}(\cdot,\cdot)$ with the symbol $\{\mathbf{k}^\wedge(n)\}_{n=0,\ldots}$ is called admissible, if these two conditions are satisfied for the symbol $\mathbf{k}^\wedge(n)$.

We define the convolution of an admissible $\mathcal{H}_{R,R}$-kernel against a function $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ via the Parseval identity as follows

$$(K^{R,R} \star_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})} F)(x) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} K^\wedge(n) F^{\wedge_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}}(n,m) H_{n,m}^*(R;x),$$

$x \in \overline{\Omega_R^{\text{ext}}}$. In analogy, we introduce the convolution of an admissible $\mathbf{h}_{R,\tau}$-kernel against $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ and $\mathbf{f} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$, respectively, as follows

$$(\mathbf{k}^{R,\tau} \star_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})} F)(x) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \mathbf{k}^{\wedge}(n) F^{\wedge_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}}(n,m) \mathbf{h}_{n,m}^*(\tau; x), \quad x \in \overline{\Omega_\tau^{\text{ext}}},$$

$$(\mathbf{k}^{R,\tau} *_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})} \mathbf{f})(x) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \mathbf{k}^{\wedge}(n) \mathbf{f}^{\wedge_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})}}(n,m) H_{n,m}^*(R; x), \quad x \in \overline{\Omega_R^{\text{ext}}}.$$

Within the context of pseudodifferential operators we are able to formulate the (tensorial) interrelation of SGG presuming tensorial data at the orbital height and requiring the potential at the Earth's surface, however, starting first from a spherical configuration (for more details concerning the tensorial SGG PDO see also [9, 13]).

The pseudodifferential operator of the Hesse tensor

The operator of the second derivative $(\nabla \otimes \nabla) : \mathcal{H}(\overline{\Omega_R^{\text{ext}}}) \to \mathbf{h}(\overline{\Omega_R^{\text{ext}}})$ is defined by

$$(\nabla \otimes \nabla) H_{n,m}^*(R; x) = \frac{\sqrt{(n+2)(n+1)(2n+3)(2n+1)}}{R^2} \mathbf{h}_{n,m}^*(R; x), \quad x \in \overline{\Omega_R^{\text{ext}}},$$

such that the symbol of this operator can be written as

$$(\nabla \otimes \nabla)^{\wedge}(n,m) = \frac{\sqrt{(n+2)(n+1)(2n+3)(2n+1)}}{R^2},$$

$n = 0, 1, \ldots; m = 1, \ldots, 2n+1$. For $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$, the operator of the Hesse tensor is given by

$$(\nabla \otimes \nabla) F(x) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} (\nabla \otimes \nabla)^{\wedge}(n,m) F^{\wedge_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}}(n,m) \mathbf{h}_{n,m}^*(R; x).$$

The associated tensorial kernel can be expressed in the form

$$\mathbf{k}_{\nabla\otimes\nabla}(x,y) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \frac{\sqrt{(n+2)(n+1)(2n+3)(2n+1)}}{R^2} H_{n,m}^*(R; x) \mathbf{h}_{n,m}^*(R; y),$$

$x, y, \in \overline{\Omega_R^{\text{ext}}}$, which leads to

$$(\nabla \otimes \nabla) F(x) = (\mathbf{k}_{\nabla\otimes\nabla} \star_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})} F)(x).$$

The tensorial SGG pseudodifferential operator

The SGG operator $\mathbf{\Lambda}_{SGG}^{R,\tau} : \mathcal{H}(\overline{\Omega_R^{\text{ext}}}) \to \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$ can be formulated in terms of outer harmonics by

$$\mathbf{\Lambda}_{SGG}^{R,\tau} H_{n,m}^*(R; x) = \frac{\sqrt{(n+2)(n+1)(2n+3)(2n+1)}}{\tau^2} \left(\frac{R}{\tau}\right)^n \mathbf{h}_{n,m}^*(\tau; x), \; x \in \overline{\Omega_\tau^{\text{ext}}}.$$

The symbol of this operator is given by

$$\left(\mathbf{\Lambda}_{SGG}^{R,\tau}\right)^{\wedge}(n,m) = \frac{\sqrt{(n+2)(n+1)(2n+3)(2n+1)}}{\tau^2}\left(\frac{R}{\tau}\right)^n, \tag{30}$$

$n = 0, 1, \ldots; m = 1, \ldots, 2n+1$. The SGG operator applied to $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ is expressible as a series as follows

$$\mathbf{\Lambda}_{SGG}^{R,\tau}F(x) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\left(\mathbf{\Lambda}_{SGG}^{R,\tau}\right)^{\wedge}(n,m)F^{\wedge_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}}(n,m)\mathbf{h}_{n,m}^*(\tau;x).$$

The associated kernel is given by

$$\mathbf{k}_{\mathbf{\Lambda}_{SGG}^{R,\tau}}(x,y) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1}\frac{\sqrt{(n+2)(n+1)(2n+3)(2n+1)}}{\tau^2}\left(\frac{R}{\tau}\right)^n H_{n,m}^*(R;x)\mathbf{h}_{n,m}^*(\tau;y)$$

and

$$\mathbf{\Lambda}_{SGG}^{R,\tau}F(x) = (\mathbf{k}_{\mathbf{\Lambda}_{SGG}^{R,\tau}} \star_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})} F)(x). \tag{31}$$

All in all, the inverse problem

$$\mathbf{\Lambda}_{SGG}^{R,\tau}F = \mathbf{g} \tag{32}$$

of determining $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ from a given $\mathbf{g} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$ is exponentially ill-posed (see, e.g., [19, 23, 24] for a more detailed classification in inverse theory and [15, 32] for more information about SGG as inverse problem).

5. Spline inversion

We are now interested in calculating the gravitational potential $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ from the observable $\mathbf{g} = \mathbf{\Lambda}_{SGG}^{R,\tau}F$ via the tensorial SGG operator $\mathbf{\Lambda}_{SGG}^{R,\tau}$:

$$\mathbf{\Lambda}_{SGG}^{R,\tau}F = \mathbf{k}_{\mathbf{\Lambda}_{SGG}^{R,\tau}} \star_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})} F = \mathbf{g}. \tag{33}$$

As already mentioned, the inverse problem $\mathbf{\Lambda}_{SGG}^{R,\tau}F = \mathbf{g}$ of determining $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ from $\mathbf{g} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$ is exponentially ill posed.

In the following we propose a spline approximation technique for the regularization of the tensorial SGG problem (33).

Definition 5.1. Suppose that $Y_N = \{y_1, \ldots, y_N\} \subset \Sigma$ (with Σ representing the real Earth's surface as shown in Fig. 1). Any function $U_N \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ of the form

$$U_N(x) = \sum_{i=1}^{N} K_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}(y_i,x)a_i, \quad x \in \overline{\Omega_R^{\text{ext}}},$$

with arbitrarily given coefficients $a_1, \ldots, a_N \in \mathbb{R}$ is called a scalar harmonic spline in the space $\mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ relative to the system $Y_N \subset \Sigma$, provided that the functions

$K_{\mathcal{H}(\overline{\Omega_R^{\rm ext}})}(y_1,\cdot)$, …, $K_{\mathcal{H}(\overline{\Omega_R^{\rm ext}})}(y_N,\cdot)$ are linearly independent. The class of all scalar harmonic splines in $\mathcal{H}(\overline{\Omega_R^{\rm ext}})$ relative to Y_N is denoted by $\mathcal{S}^{Y_N}_{\mathcal{H}(\overline{\Omega_R^{\rm ext}})}$.

Definition 5.2. Suppose that $X_N = \{x_1,\dots,x_N\} \subset \Gamma$ (with Γ representing the real LEO's orbit). Any function $\mathbf{u}_N \in \mathbf{h}(\overline{\Omega_\tau^{\rm ext}})$ of the form

$$\mathbf{u}_N(x) = \sum_{i=1}^{N} \tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})}(x_i,x)a_i, \quad x \in \overline{\Omega_\tau^{\rm ext}},$$

with arbitrarily given coefficients $a_1,\dots,a_N \in \mathbb{R}$ is called a tensorial harmonic spline in $\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})$ relative to the system $X_N \subset \Gamma$, provided that $\tilde{O}_\tau^{(1,1)}\mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})}(x_1,\cdot)$, …, $\tilde{O}_\tau^{(1,1)}\mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})}(x_N,\cdot)$ are linearly independent. The class of all tensorial harmonic splines in $\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})$ relative to X_N is denoted by $\mathbf{s}^{x_N}_{\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})}$.

Following the usual constituents of harmonic spline theory (see [5]) it is not difficult to verify the following minimum norm interpolation result.

Theorem 5.3. *Let there be known from a function* $\mathbf{g} \in \mathbf{h}(\overline{\Omega_\tau^{\rm ext}})$ *the data points* $(x_i, \mathbf{g}(x_i)) \subset \Gamma \times (\mathbb{R}^3 \otimes \mathbb{R}^3)$, $i = 1,\dots,N$. *Then the spline interpolation problem*

$$\|\mathbf{u}_N^{\mathbf{g}}\|_{\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})} = \inf_{\mathbf{v}\in \mathbf{i}_N^{\mathbf{g}}} \|\mathbf{v}\|_{\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})}$$

with

$$\mathbf{i}_N^{\mathbf{g}} = \{\mathbf{v} \in \mathbf{h}(\overline{\Omega_\tau^{\rm ext}}) \ : \ \mathbf{v}(x_i) = \mathbf{g}(x_i), \quad i = 1,\dots,N\}$$

is well posed in the sense that its solution exists, is unique, and depends continuously on the data $\mathbf{g}(x_1),\dots,\mathbf{g}(x_N)$. *The uniquely determined solution* $\mathbf{u}_N^{\mathbf{g}}$ *is given in the explicit form*

$$\mathbf{u}_N^{\mathbf{g}}(x) = \sum_{i=1}^{N} \tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})}(x_i,x)a_i,$$

where the coefficients $a_1,\dots,a_N$ *solve the linear equations*

$$\sum_{i=1}^{N} \tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})}(x_i,x_j)a_i = \mathbf{g}(x_j). \tag{34}$$

The unique solvability of (34) easily follows from the fact that

$$\tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\rm ext}})}(\cdot,y) = \sum_{n=0}^{\infty}\sum_{m=1}^{2n+1} \sqrt{\tilde{\mu}_{n,m}^{(1,1)}} H_{n,m}^*(R;\cdot)\mathbf{h}_{n,m}^*(\tau;y) \left(\frac{R}{\tau}\right)^{n+2}$$

such that

$$\sum_{i=1}^{N} \tilde{K}(x_i,x_j)a_i = G(x_j)$$

with

$$\tilde{K}(x,y) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \tilde{\mu}_{n,m}^{(1,1)} H_{n,m}^*(R;x) H_{n,m}^*(R;y),$$

and

$$G(x) = \tilde{O}_R^{(1,1)} \mathbf{g}(x).$$

Using the same coefficients a_i, $i = 1, \ldots, N$, we are led to an interpolating scalar spline of the gravitational potential F satisfying $\mathbf{\Lambda}_{SGG}^{R,\tau} F = \mathbf{g}$ in Ω_R^{ext} in the following way:

$$U_N^F(x) = \sum_{i=1}^{N} K_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})}(x_i, x) a_i$$

such that

$$\mathbf{\Lambda}_{SGG}^{R,\tau} U_N^F(x_j) = \mathbf{g}(x_j)$$

holds true for all $x_j \in \Gamma$.

We finally mention the tensorial counterpart of the *Shannon sampling theorem*.

Theorem 5.4. *Let* $\mathbf{f}$ *be in* $\mathbf{s}^{X_N}_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}$. *Then* $\mathbf{f}$ *can be reconstructed from its samples by the following interpolation formula*

$$\mathbf{f}(x) = \sum_{k=1}^{N} \tilde{O}_\tau^{(1,1)} \mathbf{f}(x_k) \mathbf{p}_k^N(x), \quad x \in \overline{\Omega_\tau^{\text{ext}}}, \tag{35}$$

where the "Lagrangians" $\mathbf{p}_k^N \in \mathbf{s}^{Y_N}_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}$, $k = 1, \ldots, N$, *are given by*

$$\mathbf{p}_k^N = \sum_{l=1}^{N} w_{l,k}^N \tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}(x_l, x), \tag{36}$$

and the coefficients $w_{l,k}^N$ *have to satisfy the linear equations*

$$\sum_{l=1}^{N} w_{l,k}^N \tilde{O}_\tau^{(1,1)} \left(\tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}(x_i, x_l) \right) = \delta_{i,k}, \quad i,k = 1, \ldots, N. \tag{37}$$

Proof. As a member of $\mathbf{s}^{X_N}_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}$ the function $\mathbf{f}$ is uniquely determined by

$$\mathbf{f} = \sum_{k=1}^{N} a_k \tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}(x_k, x)$$

with

$$\sum_{k=1}^{N} a_k \tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}(x_k, x_i) = \mathbf{f}(x_i).$$

If we define the Lagrangians as in Equation (36) with coefficients given by (37) we get

$$\tilde{O}_\tau^{(1,1)} \sum_{k=1}^{N} w_{k,l}^N \mathbf{f}(x_k) = \sum_{k=1}^{N} w_{k,l}^N \sum_{i=1}^{N} a_i \tilde{O}_\tau^{(1,1)} \left(\tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}(x_i, x_k) \right)$$
$$= \sum_{i=1}^{N} a_i \delta_{il} = a_l.$$

Using $w_{l,k}^N = w_{k,l}^N$ we get

$$\mathbf{f}(x) = \sum_{l=1}^{N} a_l \tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}(x_l, x)$$
$$= \sum_{l=1}^{N} \sum_{k=1}^{N} w_{k,l}^N \tilde{O}_\tau^{(1,1)} \mathbf{f}(x) \tilde{O}_\tau^{(1,1)} \mathbf{K}_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}(x_l, x)$$
$$= \sum_{k=1}^{N} \tilde{O}_\tau^{(1,1)} \mathbf{f}(x_k) \mathbf{p}_k^N,$$

which is the desired result. □

Remark. In a realistic situation we have to deal with noisy tensorial measurements $\mathbf{g}$ and thus we have to turn over to an operator equation $\mathbf{\Lambda}_{SGG}^{R,\tau} : \mathcal{H} \to \tilde{\mathbf{h}}$ involving a Sobolev space $\tilde{\mathbf{h}} \supset \mathbf{h}$ (e.g., $\tilde{\mathbf{h}} = \mathbf{h}_0$) such that the invertibility of $\mathbf{\Lambda}_{SGG}^{R,\tau}$ cannot be assumed anymore and regularization has to come into play. In [9] we presented a multiscale regularization, whereas in case of the spline approximation presented here we take advantage of the fact that the space $\mathbf{h}$ is dense in $\tilde{\mathbf{h}}$ such that the interpolation procedure is replaceable by a smoothing procedure in the framework of $\mathcal{H}$ and $\mathbf{h}$, thereby taking advantage of the denseness of $\mathbf{h}$ in $\tilde{\mathbf{h}}$.

6. Multiscale inversion

Next we introduce tensorial wavelets for the approximation of tensor fields (a more detailed presentation can be found in [5] for the scalar case and in [26] for the tensorial case).

Definition 6.1. A family $\{\{\varphi_j(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ of sequences $\{\varphi_j(n)\}_{n\in\mathbb{N}_0}$ is called a generator of a scaling function if it satisfies the following requirements (see [11]):

(i) For all $j \in \mathbb{N}_0$

$$(\varphi_j(0))^2 = 1,$$

(ii) for all $j, j' \in \mathbb{N}_0$ with $j \le j'$ and all $n = 1, 2, \ldots$

$$(\varphi_j(n))^2 \le (\varphi_{j'}(n))^2,$$

(iii) for all $n = 1, 2, \ldots$

$$\lim_{j\to\infty} (\varphi_j(n))^2 = 1.$$

Based on the definition of a generator of a scaling function, we now introduce $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$-scaling functions.

Definition 6.2. A family $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ of $\mathbf{h}_{\tau,\tau}$-kernels defined by $\mathbf{\Phi}_j^\wedge(n) = \varphi_j(n)$, $j \in \mathbb{N}_0$, $n \in \mathbb{N}_0$, i.e.,

$$\mathbf{\Phi}_j(x,y) = \sum_{n=0}^{\infty} \varphi_j(n) \sum_{m=1}^{2n+1} H_{n,m}^*(\tau;x)\mathbf{h}_{n,m}^*(\tau;y), \quad x, y \in \overline{\Omega_\tau^{\text{ext}}},$$

is called an $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$-scaling function, if it satisfies the following properties:

(i) $\mathbf{\Phi}_j(\cdot,\cdot)$ is an admissible $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$-kernel for every $j \in \mathbb{N}_0$,

(ii) $\{\{\mathbf{\Phi}_j^\wedge(n)\}_{n\in\mathbb{N}_0}\}_{j\in\mathbb{N}_0}$ constitutes a generator of a scaling function.

We now come to the definition of the multiresolution analysis.

Definition 6.3. Let $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be a family of admissible $\mathbf{h}_{\tau,\tau}$-kernels. Then the family $\{\mathbf{v}_j(\overline{\Omega_\tau^{\text{ext}}})\}_{j\in\mathbb{N}_0}$ of scale spaces $\mathbf{v}_j(\overline{\Omega_\tau^{\text{ext}}})$ defined by

$$\mathbf{v}_j(\overline{\Omega_\tau^{\text{ext}}}) = \{\mathbf{\Phi}_j \star_{\mathcal{H}(\overline{\Omega_\tau^{\text{ext}}})} \mathbf{\Phi}_j *_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})} \mathbf{f} : \mathbf{f} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})\},$$

is called an $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$-*multiresolution analysis*, if the following properties are satisfied:

(i) $\mathbf{v}_0(\overline{\Omega_\tau^{\text{ext}}}) \subset \cdots \subset \mathbf{v}_j(\overline{\Omega_\tau^{\text{ext}}}) \subset \mathbf{v}_{j+1}(\overline{\Omega_\tau^{\text{ext}}}) \subset \cdots \subset \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$,

(ii) $\overline{\bigcup_{j\in\mathbb{N}_0} \mathbf{v}_j(\overline{\Omega_\tau^{\text{ext}}})}^{\|\cdot\|_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}} = \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$.

It is not hard to show that the following multiscale approximation theorem is valid.

Theorem 6.4. *Let $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$-scaling function. Then*

$$\lim_{j\to\infty} \|\mathbf{f} - \mathbf{\Phi}_j \star_{\mathcal{H}(\overline{\Omega_\tau^{\text{ext}}})} \mathbf{\Phi}_j *_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})} \mathbf{f}\|_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})} = 0$$

holds for all $\mathbf{f} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$. In conclusion, $\{\mathbf{v}_j(\overline{\Omega_\tau^{\text{ext}}})\}_{j\in\mathbb{N}_0}$ forms an $\mathbf{h}_{\tau,\tau}$-multiresolution analysis.

The next purpose is to introduce wavelets via a tensorial refinement equation.

Definition 6.5. Let $\{\mathbf{\Phi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ be an $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$-scaling function. Then the family of $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$-kernels $\{\mathbf{\Psi}_j(\cdot,\cdot)\}_{j\in\mathbb{N}_0}$ given by

$$(\mathbf{\Psi}_j)^\wedge(n) = \psi_j(n), \quad j \in \mathbb{N}_0; n \in \mathbb{N}_0,$$

is called $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$-wavelet, if the $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$-kernels $\mathbf{\Psi}_j(\cdot,\cdot)$, $j \in \mathbb{N}_0$, are admissible and the symbols $\{\psi_j(n)\}$, in addition, satisfy the refinement equation

$$(\psi_j(n))^2 = (\varphi_{j+1}(n))^2 - (\varphi_j(n))^2$$

for all $j \in \mathbb{N}_0$, $n \in \mathbb{N}_0$.

As usual, we define the detail space $\mathbf{w}_j(\overline{\Omega_\tau^{\text{ext}}})$ at scale j by

$$\mathbf{w}_j(\overline{\Omega_\tau^{\text{ext}}}) = \left\{ \mathbf{\Psi}_j \star_{\mathcal{H}(\overline{\Omega_\tau^{\text{ext}}})} \mathbf{\Psi}_j *_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})} \mathbf{f} : \mathbf{f} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}}) \right\}.$$

In case of low-to-medium wavelength approximation bandlimited wavelets (based on the Shannon kernel and its modifications) are used for the evaluation of F, whereas the non-bandlimited variant of the wavelet evaluation (using non-bandlimited kernels such as Tikhonov, rational, exponential, and "locally supported" kernels) is applied to short wavelength approximation.

Examples for SGG-regularization strategies

We discuss three examples of multiresolution analyses by specifying the sequence $\{\varphi_j(n)\}$ in more detail. We let $\sigma_n = \sigma_{n,m} = \left(\mathbf{\Lambda}_{SGG}^{R,\tau}\right)^\wedge(n,m)$ (see Eq. (30)).

Tikhonov regularization. The symbol of the scaling function in case of Tikhonov regularization is given by

$$\varphi_j(n) = \frac{\sigma_n}{\sigma_n^2 + \gamma_j}$$

with $\lim_{j\to\infty} \gamma_j = 0$ and $\lim_{j\to-\infty} \gamma_j = \infty$ (see Figure 4).

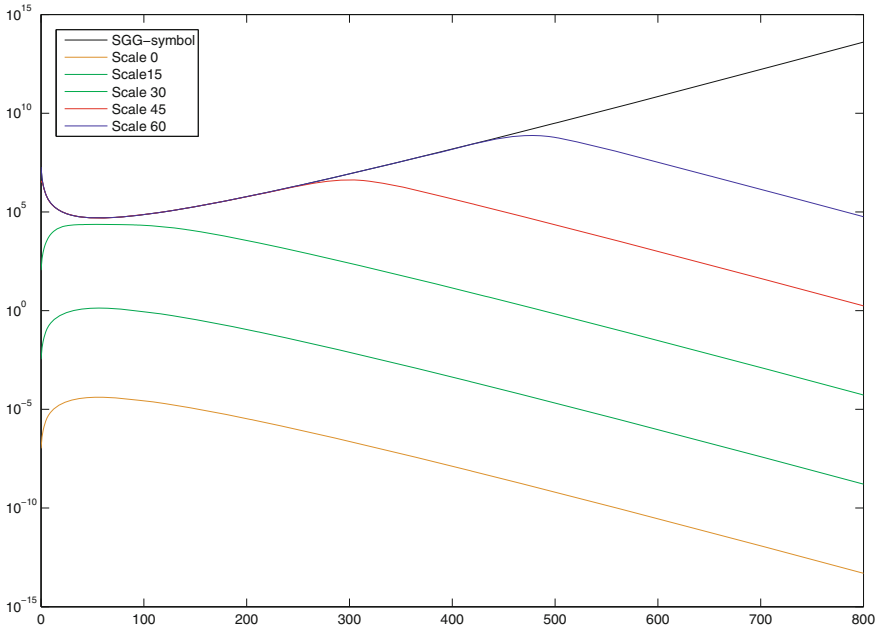


FIGURE 4. Symbol of the SGG operator with $R = 6378.127$, $\tau = R + 225.0$ (black) and of the scaling function in case of Tikhonov regularization for different dyadic scales $\gamma_j = 2^j$.

Tikhonov–Phillips regularization. The symbol of the scaling function in case of Tikhonov–Phillips regularization is given by

$$\varphi_j(n) = \frac{\sigma_n}{\sigma_n^2 + \gamma_j \frac{1}{R^2}(n(n+1) + \frac{1}{4})}$$

with $\lim_{j\to\infty} \gamma_j = 0$ and $\lim_{j\to-\infty} \gamma_j = \infty$ (see Figure 5).

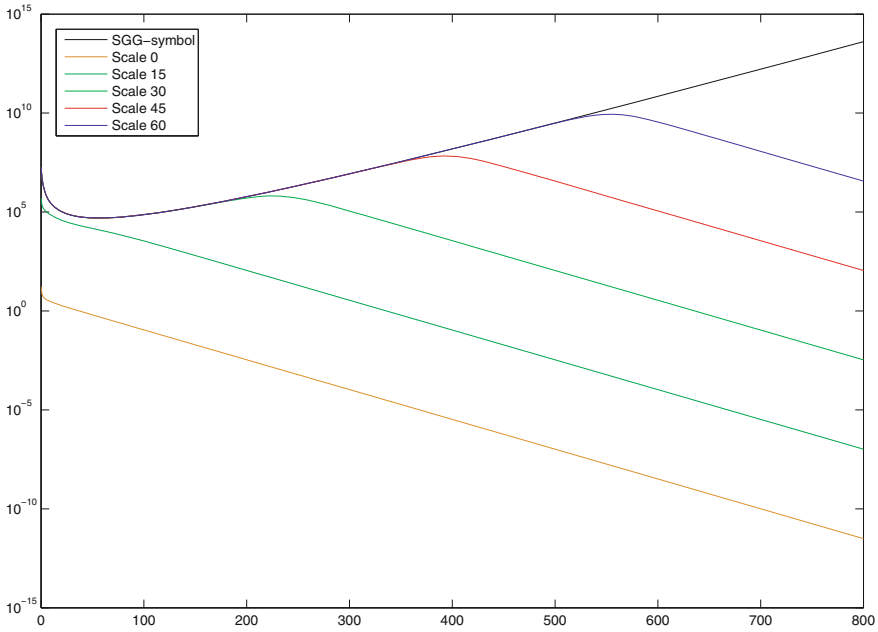


FIGURE 5. Symbol of the SGG operator with $R = 6378.127$, $\tau = R + 225.0$ (black) and of the scaling function in case of Tikhonov–Phillips regularization for different dyadic scales $\gamma_j = 2^j$.

Orthogonal truncated singular value decomposition. The symbol of the scaling function in case of the orthogonal truncated singular value decomposition is given by

$$\varphi_j(n) = \begin{cases} \sigma_n^{-1}, & \text{for} \quad n = 0, \dots, N_j, \\ 0\,, & \text{for} \quad n \geq N_j + 1, \end{cases}$$

with

$$N_j = \begin{cases} 0\,, & \text{for} \quad j < 0, \\ 2^j - 1\,, & \text{for} \quad j \geq 0\,. \end{cases}$$

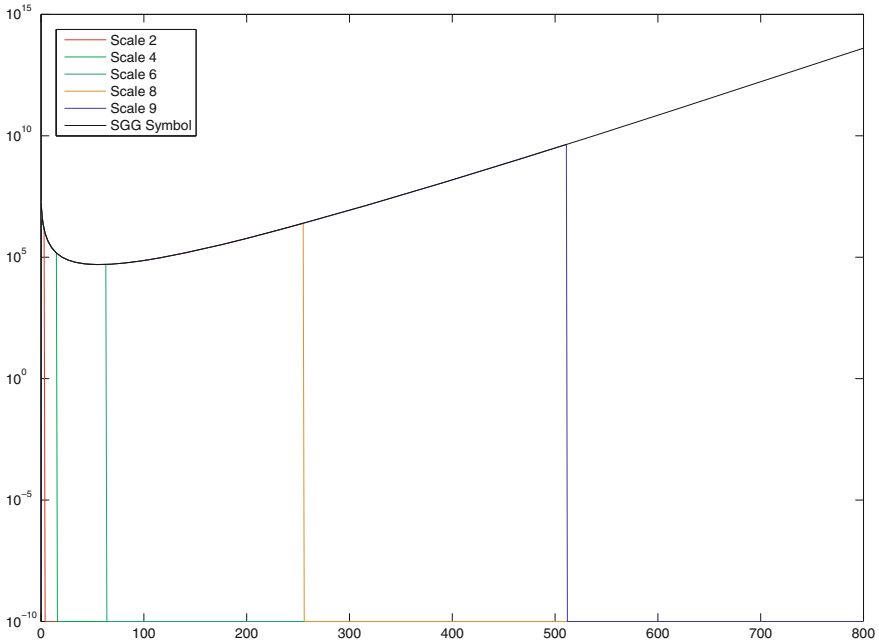


FIGURE 6. Symbol of the SGG operator with $R = 6378.127$, $\tau = R + 225.0$ (black) and of the scaling function in case of truncated singular value decomposition regularization for different scales.

Runge regularization by bandlimited outer harmonic wavelet integration

The idea now is to use a two step strategy for observing real geometries in the SGG-inverse problem: For that purpose we formally relate a scalar potential F of class $\mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ to a tensorial potential $\mathbf{g} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$ by multiscale regularization of the spherically based SGG-equation $\mathbf{\Lambda}_{SGG}^{R,\tau} F = \mathbf{g}$. In doing so, the scalar harmonic wavelet expansion of F outside Ω_R by means of bandlimited wavelets can be constructed in such a way that its scalar wavelet (potential) coefficients on Ω_R become expressible as integrals over Ω_τ convolving $\mathbf{g}$ against tensorial bandlimited harmonic wavelet kernels. Turning over to Step 1 we are led to understand the convolution integrals over Ω_τ involving the tensorial bandlimited harmonic wavelet kernels as linear functionals that can be written as linear combinations in terms of the actual SGG-data on Γ, hence, in combination with the spherical approach, the scalar wavelet (potential) coefficients on Ω_R become available as linear combinations of the actual tensorial SGG-data on Γ. Step 2 enables us to establish the multiscale regularization of F on and outside Σ such that an approximation of the gravitational potential from SGG-data becomes available obeying the real geometries.

In order to derive the integration rules for Runge regularization involving the actual geometry we remember $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ to possess a restriction $F|_{\Omega_R} \in L^2(\Omega_R)$ (see (13)). In the same way $\mathbf{g} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$ implies $\mathbf{g}|_{\Omega_\tau} \in \mathbf{l}^2(\Omega_\tau)$ (see (14)). Note that $\mathbf{l}^2(\Omega_\tau)$ is the Hilbert space of square integrable tensor fields $\mathbf{f} : \Omega_\tau \to \mathbb{R}^3 \otimes \mathbb{R}^3$. We now formulate a two step method for the Runge realization of multiscale regularization by integration in more detail:

Step 1. We suppose the scale discrete scaling functions and the corresponding wavelets to be bandlimited, i.e., there exists a sequence

$$0 \le m_0 < m_1 < m_2 < \cdots, \quad \lim_{j\to\infty} m_j = \infty,$$

with

$$\mathbf{\Phi}_0^{R,\tau}(x,\cdot) = \sum_{n=0}^{m_0} (\Phi_0)^\wedge(n) \sum_{m=1}^{2n+1} H_{n,m}^R(x) \mathbf{h}_{n,m}^{\tau;(1,1)}(\cdot),$$

and

$$\mathbf{\Psi}_j^{R,\tau}(x,\cdot) = \sum_{n=0}^{m_j} (\Psi_j)^\wedge(n) \sum_{m=1}^{2n+1} H_{n,m}^R(x) \mathbf{h}_{n,m}^{\tau;(1,1)}(\cdot), \quad j \ge 0.$$

This implies

$$\int_{\Omega_\tau} \mathbf{g}(y) \cdot \mathbf{\Phi}_0^{R,\tau}(\cdot, y) d\omega_\tau(y) \in Harm_{0,\ldots,m_0}(\Omega_R^{\text{ext}})$$

as well as

$$\int_{\Omega_\tau} \mathbf{g}(y) \cdot \mathbf{\Psi}_j^{R,\tau}(\cdot, y) d\omega_\tau(y) \in Harm_{0,\ldots,m_j}(\Omega_R^{\text{ext}}), \quad j \ge 0.$$

In consequence we are able to formulate

Theorem 6.6. *Let $X_{M_j}^R = \{x_1^{M_j}, \ldots, x_M^{M_j}\} \subset \Omega_R$, $M_j = (2m_j+1)^2$, denote fundamental systems with respect to $Harm_{0,\ldots,2m_j}(\Omega_R^{\text{ext}})$, $j \ge 0$, i.e., the matrix*

$$\begin{pmatrix} H_{1,1}^R(x_1^{M_j}) & \cdots & H_{1,1}^R(x_{M_j}^{M_j}) \\ \vdots & & \vdots \\ H_{m,2m+1}^R(x_1^{M_j}) & \cdots & H_{m,2m+1}^R(x_{M_j}^{M_j}) \end{pmatrix}$$

is assumed to be regular. Then, for $\mathbf{g} \in \text{im}(\mathbf{\Lambda}_{SGG}^{R,\tau})$ the potential $F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}})$ given by

$$\begin{aligned} F(z) = &\sum_{k=1}^{M_0} a_k^0 \int_{\Omega_\tau} \mathbf{g}(y) \cdot \mathbf{\Phi}_0^{R,\tau}(x_k^{M_0}, y) d\omega_\tau(y) \Phi_0^{R,R}(x_k^{M_0}, z) \\ &+ \sum_{j=0}^{\infty} \sum_{k=1}^{M_j} a_k^j \int_{\Omega_\tau} \mathbf{g}(y) \cdot \mathbf{\Psi}_j^{R,\tau}(x_k^{M_j}, y) d\omega_\tau(y) \Psi_j^{R,R}(x_k^{M_j}, z), \end{aligned}$$

$z \in \Omega_R^{\text{ext}}$, is the solution of the inverse problem

$$\mathbf{\Lambda}_{SGG}^{R,\tau} F = \mathbf{g}, \quad F \in \mathcal{H}(\overline{\Omega_R^{\text{ext}}}), \quad \mathbf{g} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}}).$$

In case of arbitrary $\mathbf{g} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$ we have the regularized solution

$$\begin{aligned} F_J(z) = &\sum_{k=1}^{M_0} a_k^0 \int_{\Omega_\tau} \mathbf{g}(y) \cdot \mathbf{\Phi}_0^{R,\tau}(x_k^{M_0}, y) d\omega_\tau(y) \Phi_0^{R,R}(x_k^{M_0}, z) \\ &+ \sum_{j=0}^{J-1} \sum_{k=1}^{M_j} a_k^j \int_{\Omega_\tau} \mathbf{g}(y) \cdot \mathbf{\Psi}_j^{R,\tau}(x_k^{M_j}, y) d\omega_\tau(y) \Psi_j^{R,R}(x_k^{M_j}, z), \end{aligned}$$

$z \in \Omega_R^{\text{ext}}$, where the integration weights $a_1^j, \ldots, a_M^j$, $j = 0, \ldots, J-1$, satisfy the linear systems

$$\begin{aligned} \sum_{k=1}^{M_j} a_k^j H_{n,m}(x_k^{M_j}) &= \int_{\Omega_R} H_{n,m}(x) d\omega_R(x), \\ n &= 0, \ldots, 2m_j, \quad i = 1, \ldots, 2n+1. \end{aligned} \tag{38}$$

Step 2. The "orbit" Γ is assumed to be totally contained in the exterior of the sphere Ω_τ. It is helpful to introduce the space

$$\tilde{\mathbf{h}}(\overline{\Omega_\tau^{\text{ext}}}) = \left\{ \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \mathbf{g}^{\wedge_{\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})}}(n,m) \left(\frac{\tau}{|x|} \right)^2 \mathbf{h}_{n,m}^{\tau;(1,1)}(x) \,:\, \mathbf{g} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}}),\, x \in \overline{\Omega_\tau^{\text{ext}}} \right\}.$$

Indeed, $(\tilde{\mathbf{h}}(\overline{\Omega_\tau^{\text{ext}}}), (\cdot,\cdot)_{\mathrm{l}^2(\Omega_\tau)})$ is a Hilbert space. Any $\mathbf{g} \in \tilde{\mathbf{h}}(\overline{\Omega_\tau^{\text{ext}}})$ can be uniquely represented in the form

$$\mathbf{g}(x) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \mathbf{g}^{\wedge_{\mathrm{l}^2(\Omega_\tau)}}(n,m) \left(\frac{\tau}{|x|} \right)^2 \mathbf{h}_{n,m}^{\tau;(1,1)}(x), \quad x \in \overline{\Omega_\tau^{\text{ext}}}.$$

Using the functions

$$\tilde{\mathbf{h}}_{n,m}^{\tau;(1,1)}(x) = \left(\frac{\tau}{|x|} \right)^2 \mathbf{h}_{n,m}^{\tau;(1,1)}(x), \quad x \in \overline{\Omega_\tau^{\text{ext}}},$$

we are led to the space

$$\tilde{\mathbf{harm}}_{0,\ldots,m}^{(1,1)}(\overline{\Omega_\tau^{\text{ext}}}) = \operatorname{span}_{\substack{n=0,\ldots,m;\\ m=1,\ldots,2n+1}} (\tilde{\mathbf{h}}_{n,m}^{\tau;(1,1)}).$$

The set $X_M^\Sigma = \{y_1^M, \ldots, y_M^M\} \subset \Sigma$, $M = (m+1)^2$ is called a fundamental system with respect to $\tilde{\mathbf{harm}}_{0,\ldots,m}^{(1,1)}(\overline{\Omega_\tau^{\text{ext}}})$, if the matrix

$$\begin{pmatrix} \tilde{\mathbf{h}}_{1,1}^{\tau;(1,1)}(y_1^M) & \ldots & \tilde{\mathbf{h}}_{1,1}^{\tau;(1,1)}(y_M^M) \\ \vdots & & \vdots \\ \tilde{\mathbf{h}}_{m,2m+1}^{\tau;(1,1)}(y_1^M) & \ldots & \tilde{\mathbf{h}}_{m,2m+1}^{\tau;(1,1)}(y_M^M) \end{pmatrix}$$

is regular. These settings allow us to formulate

Theorem 6.7. *Let $X_M^\Sigma = \{y_1^M, \dots, y_M^M\} \subset \Sigma$, $M = (m+1)^2$ be a fundamental system with respect to $\tilde{\mathbf{harm}}_{0,\dots,m}^{(1,1)}(\overline{\Omega_\tau^{\text{ext}}})$. Furthermore, suppose that*

$$\mathbf{g} \in \tilde{\mathbf{harm}}_{0,\dots,m}^{(1,1)}(\overline{\Omega_\tau^{\text{ext}}}) \quad \text{and} \quad \mathbf{\Psi} \in \mathbf{h}(\overline{\Omega_\tau^{\text{ext}}}).$$

Then

$$\int_{\Omega_\tau} \mathbf{g}(y) \cdot \mathbf{\Psi}(y) d\omega_\tau(y) = \sum_{p=1}^{M} \mathbf{a}_p \cdot \mathbf{g}(y_p^M)$$

holds true, if $\mathbf{a}_1, \dots, \mathbf{a}_M$ satisfy

$$\sum_{p=1}^{M} \mathbf{a}_p \cdot \tilde{\mathbf{h}}_{n,m}^{\tau;(1,1)}(y_p^M) = \int_{\Omega_\tau} \mathbf{\Psi}(y) \cdot \tilde{\mathbf{h}}_{n,m}^{\tau;(1,1)}(y) d\omega_\tau(y),$$

$n = 0, \dots, m; j = 1, \dots, 2n+1$.

Altogether, we end up with the following theorem about the resulting gravitational potential F on the Earth's surface Σ computed from SGG data given on the real orbit Γ:

Theorem 6.8. *Let $\mathbf{g}$ be a bandlimited function of class $\tilde{\mathbf{h}}(\overline{\Omega_\tau^{\text{ext}}})$ that is given at $M = (m+1)^2$ points $\{y_1^M, \dots, y_M^M\} \subset \Sigma$ which form a fundamental system with respect to $\tilde{\mathbf{harm}}_{0,\dots,m}^{(1,1)}(\overline{\Omega_\tau^{\text{ext}}})$. Furthermore, let $\{\mathbf{\Psi}_j^{R,\tau}\}$, $j \geq 0$, denote a scale discrete bandlimited spherical regularization decomposition wavelet of order 0 with respect to (32) and corresponding decomposition scaling function $\{\mathbf{\Phi}_j^{R,\tau}\}$, $j \geq 0$. Let $\{\Psi_j^{R,R}\}$ and $\{\Phi_j^{R,R}\}$ be the corresponding reconstruction wavelet and scaling function, respectively. Then the regularized solution F_J of (32) is given by*

$$\begin{aligned} F_J(x) = &\sum_{k=1}^{M_0} \mathbf{a}_k^0 \cdot \sum_{s=1}^{M} \hat{b}_s^{0,k} \mathbf{g}(y_s^M) \Phi_0^{R,R}(x, x_k^{M_0}) \\ &+ \sum_{j=0}^{J-1} \sum_{k=1}^{M_j} \mathbf{a}_k^j \cdot \sum_{s=1}^{M} b_s^{j,k} \mathbf{g}(y_s^M) \Psi_j^{R,R}(x, x_k^{M_j}), \end{aligned}$$

where $\hat{b}_1^{0,k}, \dots, \hat{b}_M^{0,k}$, $k = 1, \dots, M_0$, satisfy

$$\sum_{p=1}^{M} \mathbf{a}_p^j \cdot \mathbf{h}_{n,m}^{\tau_i(1,1)}(y_p^{M_j}) = (\Phi^{R,R})^\wedge(n) H_{n,j}^{R;(1,1)}(x_p^{M_j}), \quad n = 0, \dots, m;\ i = 1, \dots, 2n+1,$$

$b_1^{j,k}, \dots, b_M^{j,k}$, $j = 0, \dots, J-1$, $k = 1, \dots, M_j$, *satisfy*

$$\sum_{s=1}^{M} b_s^{j,k} \tilde{\mathbf{h}}_{n,i}^{\tau;(1,1)}(y_s^M) = (\mathbf{\Psi}_j^{R,\tau})^\wedge(n) \mathbf{h}_{n,i}^{R;(1,1)}(x_k^{M_j}), \quad n = 0, \dots, m; i = 1, \dots, 2n+1,$$

$\mathbf{a}_1^j, \dots, \mathbf{a}_{M_j}^j$, $j = 0, \dots, J-1$, *satisfy* (38), *and* $X_{M_j}^R = \{x_1^{M_j}, \dots, x_{M_j}^{M_j}\} \subset \Omega_R$, $M_j = (2m_j+1)^2$ *denote fundamental systems with respect to* $Harm_{0,\dots,2m_j}(\overline{\Omega_R^{\text{ext}}})$.

At a first glance, the regularization procedure requires the solution of lots of linear systems. But if we have a careful look we recognize that always the same matrix has to be inverted, and all weights for the numerical integration are obtained by a matrix-vector multiplication. In principle, the assumption that $\mathbf{g}$ is bandlimited can be omitted and the calculated weights can be used for the numerical integration formulas (which, however, are then no longer exact).

Note that our wavelet based regularization technique is presented for the tensorial SGG problem (32) thereby applying new types of appropriate integration formula involving outer harmonic spaces. The basic concept is to express convolution integrals on spheres by linear combinations of data on real geometries taking strong advantage of the harmonicity of the wavelet kernels. As the SGG problem demands a global approach we assume the data to be given on discrete points uniformly distributed over the satellite orbit Γ. Nevertheless, because of the space localizing properties of scaling and wavelet functions, our theory can be extended canonically to local areas for which locally given data sets can be used. An extension to locally oriented multiscale regularization can be found in [9].

7. A tree algorithm based on harmonic spline exact approximation

In what follows, the SGG problem will be solved within a multiscale procedure involving interpolating splines. Suppose that a SGG data set of values corresponding to N_J points on Γ is known. Let the family $\{\mathbf{\Phi}_J(\cdot,\cdot)\}_{J\in\mathbb{N}_0}$ be an $\mathbf{h}(\overline{\Omega_\tau^{\text{ext}}})$-scaling function which is bandlimited. We formally set

$$\mathbf{\Phi}_J^{(2)} = \sum_{n=0}^{N_J}\sum_{m=1}^{2n+1} (\varphi_J(n))^2 \mathbf{h}_{n,m}^*(\tau;\cdot)\otimes\mathbf{h}_{n,m}^*(\tau;\cdot)$$

(with $\varphi_J(n) = \Phi_J^\wedge(n)$) for the iterated kernel function. Note that $\mathbf{\Phi}_J^{(2)}$ is a tensorial kernel function of rank four. In the framework of convolutions we obviously have

$$\mathbf{\Phi}_J^{(2)} *_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})} \mathbf{g} = \mathbf{\Phi}_J \star_{\mathcal{H}(\overline{\Omega_R^{\text{ext}}})} \mathbf{\Phi}_J *_{\mathbf{h}(\overline{\Omega_R^{\text{ext}}})} \mathbf{g}.$$

In accordance with the presentation in [5] for the scalar case it follows that, in the terminology of Sobolev spaces, $\mathbf{\Phi}_J^{(2)}$ is the unique reproducing kernel of $\mathbf{h}(\{A_n/\varphi_J^\wedge(n)\};\overline{\Omega_\tau^{\text{ext}}})$. For sufficiently large J, corresponding to the tensor data $\mathbf{g}$, there is in arbitrarily close accuracy to $\mathbf{\Phi}_J^{(2)} * \mathbf{g}$ a spline $\mathbf{u}_{N_J}^{\mathbf{\Phi}_J^{(2)}*\mathbf{g}}$ consistent with a set of N_J original data $\mathbf{g}(x_1^{N_J}),\ldots,\mathbf{g}(x_{N_J}^{N_J})$:

$$\mathbf{u}_{N_J}^{\mathbf{\Phi}_J^{(2)}*\mathbf{g}}(x) \approx \sum_{i=1}^{N_J} \tilde{O}_\tau^{(1,1)} \mathbf{\Phi}_J^{(2)}(x_i^{N_J},x) a_i^{N_J}, \quad x\in\Omega_\tau^{\text{ext}}.$$

For $j = 0, \ldots, J$, we assume that the coefficients $w_{l,k}^{N_j}$ have been determined by solving the linear system (see Theorem 5.4)

$$\sum_{i=1}^{N_j} \left(\tilde{O}_\tau^{(1,1)} \tilde{O}_\tau^{(1,1)} \mathbf{\Phi}_j^{(2)}(x_l^{N_j}, x_i^{N_j}) \right) w_{i,k}^{N_j} = \delta_{lk}, \quad l,k = 1, \ldots, N_j. \tag{39}$$

Remark. In its tensorial generality as formulated here the condition (39) certainly is a bottleneck of the presented method seen from numerical point of view. Nonetheless, our approach is mathematically interesting. Even more, if we restrict ourselves to constituting ingredients of the Hesse tensor (such as second radial derivatives), the tree algorithm can be established in the same way requiring much less numerical effort. Observe that the solution of the linear system (39) has to be calculated once and can be stored elsewhere, as far as the same nodal system is in use.

The tree algorithm consists of the following ingredients:

The initial step (*or sampling step*). The point of departure is the observation that for sufficiently large J

$$\begin{aligned} \mathbf{g}(x) &= \mathbf{\Phi}_J \star_{\mathcal{H}(\overline{\Omega_\tau^{\mathrm{ext}}})} \mathbf{\Phi}_J *_{\mathbf{h}(\overline{\Omega_\tau^{\mathrm{ext}}})} \mathbf{g} \\ &\approx \mathbf{u}_{N_J}^{\mathbf{\Phi}_J^{(2)}} * \mathbf{g} \\ &= \sum_{i=1}^{N_J} \tilde{O}_\tau^{(1,1)} \mathbf{\Phi}_J^{(2)}(x_i^{N_J}, x) a_i^{N_J}, \quad x \in \Omega_\tau^{\mathrm{ext}}, \end{aligned}$$

where $a_i^{N_J}$ are determined by

$$a_i^{N_J} = \sum_{k=1}^{N_J} w_{i,k}^{N_J} \tilde{O}_\tau^{(1,1)} (\mathbf{\Phi}_J^{(2)} *_{\mathbf{h}(\overline{\Omega_\tau^{\mathrm{ext}}})} \mathbf{g})(x_k^{N_J}) \approx \sum_{k=1}^{N_J} w_{i,k}^{N_J} \tilde{O}_\tau^{(1,1)} \mathbf{g}(x_k^{N_J}).$$

The pyramid step. We start from

$$(\mathbf{\Phi}_j^{(2)} * \mathbf{g})(x) \approx \mathbf{u}_{N_j}^{\mathbf{\Phi}_j^{(2)} * \mathbf{g}}(x) = \sum_{i=1}^{N_j} \tilde{O}_\tau^{(1,1)} \mathbf{\Phi}_j^{(2)}(x_i^{N_j}, x) a_i^{N_j}, \quad x \in \Omega_\tau^{\mathrm{ext}},$$

with

$$a_i^{N_j} = \sum_{k=1}^{N_j} w_{i,k}^{N_j} \tilde{O}_\tau^{(1,1)} (\mathbf{\Phi}_j^{(2)} *_{\mathbf{h}(\overline{\Omega_\tau^{\mathrm{ext}}})} \mathbf{g})(x_k^{N_j}), \quad \text{for} \quad j = J_0, \ldots, J.$$

From

$$\mathbf{\Phi}_j^{(2)} * \mathbf{g} \approx \sum_{l=1}^{N_{j+1}} a_l^{N_{j+1}} \tilde{O}_\tau^{(1,1)} \mathbf{\Phi}_j^{(2)}(\cdot, x_l^{N_{j+1}})$$

we get the following *recursion relation*

$$a_i^{N_j} \approx \sum_{k=1}^{N_j} \sum_{l=1}^{N_{j+1}} \tilde{O}_\tau^{(1,1)} \tilde{O}_\tau^{(1,1)} \mathbf{\Phi}_j^{(2)}(x_k^{N_j}, x_l^{N_{j+1}}) w_{i,k}^{N_j} a_l^{N_{j+1}},$$

$i = 1, \ldots, N_j$, which leads us to the following *decomposition scheme*:

$$\begin{array}{ccccccc} F & \to & a^{N_J} & \to & a^{N_{J-1}} & \to \cdots & a^{N_{J_0}} \\ & & \downarrow & & \downarrow & & \downarrow \\ & & \mathbf{\Phi}_J^{(2)} * \mathbf{g} & & \mathbf{\Phi}_{J-1}^{(2)} * \mathbf{g} & & \mathbf{\Phi}_{J_0}^{(2)} * \mathbf{g}. \end{array}$$

Note that by use of the corresponding iterated scalar scaling function (see, e.g., [5]) given by

$$\Phi_J^{(2)} = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} (\varphi_J(n))^2 H_{n,m}^R H_{n,m}^R$$

we arrive at a representation of the spline approximation of F satisfying $\mathbf{\Lambda}_{SGG}^{R,\tau} F = \mathbf{g}$ on Ω_R^{ext} in the following way

$$U_{N_j}^{\Phi_j^{(2)} * F}(x) = \sum_{i=1}^{N_j} \Phi_j^{(2)}(x_i^{N_j}, x) a_i^{N_j}, \quad x \in \Omega_R^{\text{ext}}.$$

Note that, in case of bandlimited scaling functions, "$\approx$" may be replaced by "$=$".

Remark. Accordingly, the tree algorithm can be realized if only scalar data within the Hesse tensor (for example, second radial derivatives) are used for approximation (cf. [5]). The trace of the Hesse tensor (which is equal to zero) offers the possibility to validate the method. Moreover, observational errors can be handled by smoothing or filtering techniques within the tree algorithm (see [5, 7, 8]). Since these procedures are well documented in the literature, they will be not discussed in this approach.

8. Conclusion

A gradiometer mission ideally produces a coverage of the entire Earth with tensorial measurements at a certain altitude. Our multiscale method yields decorrelations of the scalar internal gravitational potential of the Earth by a tree algorithm within the framework of inverse problems under real geometric situations. Even more, from the SGG-approach presented here, it can be expected that the knowledge about the representation of the Earth's gravitational potential will be improved in considerable way, providing the transition form low- and meso-based to high(er) reflected resolution at global scale, thereby offering significant local features.

Although an impressive rate of the Earth's gravitational potential can be detected globally at the orbit of a satellite (like GOCE), the computational drawback of satellite techniques in geoscientific research is the fact that measurements must be performed at a certain altitude. Consequently, a "downward continuation" process must be applied to handle the potential at the Earth's surface, hence, a loss of information for the signal is unavoidable. Indeed, "downward continuation" causes severe problems, since the amount of amplification for the potential is not known suitably (as an a priori amount) and even small errors in the measurements may produce huge errors in the potential at the Earth's surface.

However it is of great advantage that satellite data are globally available, at least in principle. Nevertheless, from mathematical point of view, we are not confronted with a boundary value of potential theory. Satellite techniques such as SGG require the solution of an inverse problem to produce gravitational information at the Earth's surface, where it is needed actually. SGG can be formulated adequately as (Fredholm) pseudodifferential equation of the first kind, which is exponentially ill posed, and this fact makes the development of suitable mathematical methods with strong relation to the nature and structure of the data indispensable.

In this respect it should be mentioned that each method in approximation theory has its own aim and character. Even more, it is the essence of any numerical realization that it becomes optimal only with respect to certain specified features. For example, Fourier expansion methods with polynomial trial functions (spherical harmonics) offer the canonical "trend-approximation" of low-frequency phenomena (for global modeling), they offer an excellent control and comparison of spectral properties of the signal, since any spherical harmonic relates to one frequency. This is of tremendous advantage for relating data types under spectral aspects. But it is at the price that the polynomials are globally supported such that local modeling results into serious problems of economy and efficiency. Bandlimited kernels can be used for the transition from long-wavelength to short-wavelength phenomena (global to local modeling) in the signal. Because of their excellent localization properties in the space domain, the non-bandlimited kernels can be used for the modeling of short-wavelength phenomena. Local modeling is effective and economic. But the information obtained by kernel approximations is clustered in frequency bands so that spectral investigations are laborious and time consuming. In other words, for numerical work to be done, we have to make an a priori choice. We have to reflect the different stages of space/frequency localization so that the modeling process can be adapted to the localization requirements necessary and sufficient for our geophysical or geodetic interpretation.

In conclusion, an algorithm establishing an approximate solution for the inverse SGG-problem has to reflect the intention of the applicant. Different techniques for regularization are at the disposal of the numerical analyst for global as well as local purposes. Each effort only gives certain progress in the particular field of pre-defined interest. If a broad field of optimality should be covered, only a combined approach is the strategic instrument to make an essential step forward. Thus, for computational aspects of determining the Earth's gravitational potential, at least a twofold combination is demanded, viz. combining globally available satellite data (including the SGG-contribution) with local airborne and/or terrestrial data and combining tools and means of constructive approximation such as polynomials, splines, wavelets, etc. Altogether, in numerical modeling of the Earth's gravitational potential, there is no best universal method, there exist only optimized procedures with respect to certain features and the option and the feasibility for their suitable combination.

Acknowledgment

The first two authors thank the "Federal Ministry for Economic Affairs and Energy, Berlin" and the "Project Management Jülich" for funding the project "SPE" (funding reference number 0324016, CBM – Gesellschaft für Consulting, Business und Management mbH, Bexbach, Germany).

References

[1] Eggermont, P.N., LaRiccia, V., Nashed, M.Z. (2014) Noise Models for Ill-Posed Problems. In: Freeden,W., Nashed, M.Z., Sonar, T. (Eds.) Handbook of Geomathematics, 2nd edition, Springer, New York, 1633–1658

[2] Engl, H., Hanke, M., Neubauer, A. (1996) Regularization of Inverse Problems, Kluwer, Dordrecht, Boston, London

[3] Freeden, W. (1980a) On the Approximation of External Gravitational Potential With Closed Systems of (Trial) Functions, Bull. Géod., **54**: 1–20

[4] Freeden, W. (1998) The Uncertainty Principle and Its Role in Physical Geodesy. In: Freeden, W. (ed.) Progress in Geodetic Science, 225–236. Shaker, Aachen

[5] Freeden, W. (1999) Multiscale Modelling of Spaceborne Geodata, B.G. Teubner, Leipzig

[6] Freeden, W., Gerhards, C. (2013) Geomathematically Oriented Potential Theory, Chapman & Hall/CRC Press, Boca Raton

[7] Freeden, W., Maier, T. (2002) On Multiscale Denoising of Spherical Functions: Basic Theory and Numerical Aspects. Electronic Transactions on Numerical Analysis (ETNA), **14**: 40–62

[8] Freeden, W., Maier, T. (2003) Spectral and Multiscale Signal-to-Noise Thresholding of Spherical Vector Fields. Computational Geosciences, **7**: 215–250

[9] Freeden, W., Nutz, H. (2011) Satellite Gravity Gradiometry as Tensorial Inverse Problem. Int. J. Geomath. **2**: 177–218

[10] Freeden, W., Nutz, H. (2018) Geodetic Observables and Their Mathematical Treatement in Multiscale Framework. In: Freeden, W., Nashed , M.Z. (Eds.), Handbook of Mathematical Geodesy, this volume, Springer International Publishing, 315–458

[11] Freeden, W., Schneider, F. (1998) Regularization Wavelets and Multiresolution, Inverse Problems, **14**: 493–515

[12] Freeden, W., Schreiner, M. (2009) Spherical Functions of Mathematical Geosciences. Springer, Heidelberg

[13] Freeden, W., Schreiner, M. (2014) Satellite Gravity Gradiometry (SGG): From Scalar to Tensorial Solution. In: Handbook of Geomathematics, Freeden, W., Nashed, M.Z., and Sonar, T. (Eds.), Springer, New York, Heidelberg, 2nd edition, 339–380

[14] Freeden W., Gervens, T., Schreiner, M. (1998) Constructive Approximation on the Sphere (with Applications to Geomathematics), Oxford Science Publications, Clarendon, Oxford

[15] Freeden, W., Schneider, F., Schreiner, M. (1997) Gradiometry – An Inverse Problem in Modern Satellite Geodesy, in: Engl, H.W., Louis, A., Rundell, W. (Eds.), GAMM–SIAM Symposium on Inverse Problems: Geophysical Applications, 179–239

[16] Hadamard, J. (1923) Lectures on the Cauchy Problem in Linear Partial Differential Equations, Yale University Press, New Haven

[17] Heiskanen, W.A., Moritz, H. (1967) Physical Geodesy, Freeman, San Francisco, CA

[18] Kirsch, A. (1996) Introduction to the Mathematical Theory of Inverse Problems, Springer, New York

[19] Louis, A.K. (1989) Inverse und schlecht gestellte Probleme, Teubner, Leipzig

[20] Meissl, P. (1971) A Study of Covariance Functions Related to the Earth's Disturbing Potential. Department of Geodetic Science, No. 151, The Ohio State University, Columbus, OH

[21] Meissl, P. (1971) On the Linearization of the Geodetic Boundary Value Problem. Department of Geodetic Science, No. 152, The Ohio State University, Columbus, OH

[22] Nashed, M.Z. (1976) Generalized Inverses and Applications, Proceedings of an Advanced Seminar Conducted by the Mathematics Research Center, University of Wisconsin, Madison

[23] Nashed, M.Z. (1981) Operator-Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory. IEEE Transactions on Antennas and Propagation, 29(**2**): 220–231

[24] Nashed, M.Z. (1987) A New Approach to Classification and Regularization of Ill-Posed Operator Equations. In: Inverse and Ill-Posed Problems, Engl, H.W., Groetsch, C.W. (Eds.), Notes and Reports in Mathematics in Science and Engineering **4**: 53–75, Academic Press, New York

[25] Nashed, M.Z. (1987) Inner, Outer, and Generalized Inverses in Banach and Hilbert Spaces. Numer. Funct. Anal. and Optimiz., **9**: 261–325

[26] Nutz, H. (2002) A Unified Setup of Gravitational Field Observables. Ph.D.-thesis, University of Kaiserslautern, Geomathematics Group, Shaker, Aachen

[27] Rieder, A (2003) Keine Probleme mit Inversen Problemen. Vieweg, Wiesbaden

[28] Rummel, R. (1997) Spherical Spectral Properties of the Earth's Gravitational Potential and Its First and Second Derivatives. In Lecture Notes in Earth Science, Springer, Berlin, **65**: 359–404

[29] Rummel, R. (2014) GOCE: Gravitational Gradiometry in a Satellite. In: Handbook of Geomathematics, Freeden, W., Nashed, M.Z., and Sonar, T. (eds)., Springer, New York, Heidelberg, 2nd edition, 211–226

[30] Rummel, R., van Gelderen, M. (1992) Spectral Analysis of the Full Gravity Tensor. Geophys. J. Int., **111**: 159–169

[31] Rummel, R., van Gelderen, M. (1995) Meissl Scheme – Spectral Characteristics of Physical Geodesy. Manuscr. Geod. **20**: 379–385

[32] Schreiner, M. (1994) Tensor Spherical Harmonics and Their Application in Satellite Gradiometry. Ph.D.-thesis, University of Kaiserslautern, Geomathematics Group

Willi Freeden and Helga Nutz
Geomathematics Group
University of Kaiserslautern
MPI-Gebäude, Paul-Ehrlich-Str. 26
D-67663 Kaiserslautern, Germany

Michael Schreiner
Institute for Computational Engineering
University of Buchs
Buchs, Switzerland

Handbook of Mathematical Geodesy
Geosystems Mathematics, 605–639

Parameter Choices for Fast Harmonic Spline Approximation

Martin Gutting

Abstract. The approximation by harmonic trial functions allows the construction of the solution of boundary value problems in geoscience where the boundary is often the known surface of the Earth itself. Using harmonic splines such a solution can be approximated from discrete data on the surface. Due to their localizing properties regional modeling or the improvement of a global model in a part of the Earth's surface is possible with splines.

Fast multipole methods have been developed for some cases of the occurring kernels to obtain a fast matrix-vector multiplication. The main idea of the fast multipole algorithm consists of a hierarchical decomposition of the computational domain into cubes and a kernel approximation for the more distant points. This reduces the numerical effort of the matrix-vector multiplication from quadratic to linear in reference to the number of points for a prescribed accuracy of the kernel approximation. In combination with an iterative solver this provides a fast computation of the spline coefficients.

The application of the fast multipole method to spline approximation which also allows the treatment of noisy data requires the choice of a smoothing parameter. We summarize several methods to (ideally automatically) choose this parameter with and without prior knowledge of the noise level.

Keywords. Spline approximation, fast multipole methods.

1. Introduction

Spherical splines have been developed by Freeden [38] and independently by Wahba [134] for interpolation and smoothing/approximation problems on the sphere and have been generalized to harmonic splines by Freeden [37, 39, 40]. These harmonic splines can be used for interpolation/approximation on regular surfaces, but in particular for the solution of boundary value problems where the boundary is a regular surface. Convergence theorems exist for both spherical splines and harmonic splines (cf. [41, 42, 43, 44]), and the different types of spline spaces and their

reproducing kernels have been investigated (cf. [49] and the references therein). In geosciences they have found many applications (see, e.g., [45, 46, 48, 107] and the references therein).

Splines lead to a system of linear equations which in case of harmonic splines has to be densely populated. This makes the solution expensive in terms of the numerical effort. Iterative solvers require fast summation methods corresponding to the reproducing kernels of these splines to be truly efficient. On the sphere there are several possible ways to achieve a fast summation (the problem that is considered determines which method should be preferred): spherical panel clustering (cf. [45, 52] and the references therein), spherical FFT for gridded data points or spherical NFFT for non-equispaced data (cf. [116, 82]).

[61, 62, 63, 124] have introduced the fast multipole method (FMM) in two and three dimensions for fast evaluation of potentials corresponding to the Laplace operator (generalizations to further operators have been introduced later). Since such potentials are closely related to certain reproducing kernels of our splines, the FMM allows fast summation of harmonic splines as well as spherical splines. Such a combination is used in [57] to solve problems of satellite geodesy with harmonic splines corresponding to the singularity kernel. We have extended this to the Abel–Poisson kernel and use the accelerated version of the FMM that was first introduced in [24, 64]. This approach has also been applied to the oblique boundary value problem of potential theory in [67, 68].

We consider the following (generalized) interpolation problems:

Problem 1.1 (Interpolation on a regular surface). Let Σ be a $C^{(0)}$-regular surface (see Definition 2.1 below for details). Let a finite set of points $\{x_1, \dots, x_N\} \subset \Sigma$ on the surface and data F_i, $i = 1, \dots, N$ corresponding to these points be given. The aim is to find a function F in a function space of choice such that $F(x_i) = F_i$, $i = 1, \dots, N$.

If the data F_i are error-affected and strict interpolation is no longer desirable and the interpolation conditions are reduced to $F(x_i) \approx F_i$, $i = 1, \dots, N$, and F has to minimize a functional that balances closeness to the data and smoothness of F, usually with one (or several) parameters.

Closely related to this interpolation/approximation problem is the discrete version of the Dirichlet boundary value problem which requires only the values of the boundary function in a finite set of points on the surface. This is also called a generalized interpolation problem.

Problem 1.2. Let Σ be a $C^{(k)}$-regular surface with $k \geq 2$ (see again Definition 2.1 below for details). Let $\{x_1, \dots, x_N\} \subset \Sigma$ be a discrete set of N points on the surface. For each point x_i let $F_i = U(x_i)$ be given, where $i = 1, \dots, N$.
The task is to determine the potential $U \in C^{(0)}\left(\overline{\Sigma_{\text{ext}}}\right) \cap C^{(2)}\left(\Sigma_{\text{ext}}\right)$ which is harmonic in Σ_{ext}, i.e., the exterior of the surface Σ, and regular at infinity, i.e., for

$|x| \to \infty$,

$$|U(x)| = \mathcal{O}\left(|x|^{-1}\right), \tag{1}$$

$$|\nabla U(x)| = \mathcal{O}\left(|x|^{-2}\right), \tag{2}$$

or an approximation U_N to it which fits the data, i.e., for $i = 1, \ldots, N$,

$$U_N(x_i) = F_i = U(x_i). \tag{3}$$

As before the interpolation conditions (3) are relaxed in case of error-affected data/measurements.

The outline of this paper is as follows: Section 2 summarizes the theory of harmonic splines and spline approximation. In Section 3 we establish the connection between harmonic splines and the sums that can be computed by the fast multipole method, we introduce the adaptive construction of the decomposition of the computational domain and provide our version of the fast multipole algorithm for harmonic splines. Section 4 gives an overview of suitable methods to choose the smoothing parameter of the approximating splines if the data are afflicted with (stochastic) noise.

2. Preliminaries

Spherical harmonics, which we denote by $Y_{n,m}$ (with degree $n \in \mathbb{N}_0$, order $m = -n, \ldots, n$), are known to form a complete orthonormal basis of the space $L^2(\mathbb{S}^2)$ of square integrable functions on the unit sphere $\mathbb{S}^2$ (see, e.g., [30, 47, 130]). The spherical harmonics $\{Y_{n,m}\}_{n\in\mathbb{N}_0, m=-n,\ldots,n}$ also form a closed system in $C(\mathbb{S}^2)$ and are closed and complete in $L^2(\mathbb{S}^2)$. This allows the representation of square-integrable functions on any sphere $\mathbb{S}^2_R$ of radius $R > 0$ by their Fourier series, where the Fourier coefficients of $F \in L^2(\mathbb{S}^2_R)$ are denoted by

$$F^\wedge(n,m) = \int_{\mathbb{S}^2_R} F(x)\, \frac{1}{R} Y_{n,m}\left(\tfrac{x}{R}\right)\, dS_R(x). \tag{4}$$

2.1. Regular surfaces and Runge spheres

Due to the Runge–Walsh approximation theorem, we can use functions which possess a larger domain of harmonicity to approximate the solution of a problem which requires harmonicity only outside the Earth's surface (see [45, 48] for an extensive introduction of this technique). Harmonic splines as introduced in [37, 39, 40, 125] are constructed in such a way that they are subspaces of the space of harmonic functions on a sphere situated inside the Earth, the so-called Runge (or Krarup) sphere (see [109]).

The Earth's surface is considered to be regular as by the following definition.

Definition 2.1. A $C^{(k)}$-regular surface $\Sigma \subset \mathbb{R}^3$ is a surface in $\mathbb{R}^3$ which has to fulfill the following properties:

(i) Σ divides $\mathbb{R}^3$ into the interior Σ_{int} and the exterior Σ_{ext}, where Σ_{int} is a bounded region and Σ_{ext} is an unbounded region.

(ii) The origin is contained in Σ_{int}.
(iii) Σ is closed (and therefore compact) and free of double points.
(iv) Σ is a $C^{(k)}$-surface, i.e., for each $x \in \Sigma$ there exists a neighborhood $U \subset \mathbb{R}^3$ of x such that $\Sigma \cap U$ possesses a $C^{(k)}$-parametrization.

We can also define $C^{(k,\lambda)}$-regular surfaces $\Sigma \subset \mathbb{R}^3$ with $\lambda \in (0, 1)$ as a $C^{(k)}$-regular surface where every point $x \in \Sigma$ possesses a neighborhood U such that $\Sigma \cap U$ can locally be parameterized by a k-times λ-Hölder continuously differentiable parametrization. Such surfaces are required for oblique derivative boundary value problems as in [68], but not here. Note that any sphere $\mathbb{S}_R^2$ of radius $R > 0$ is obviously a $C^{(\infty)}$-regular surface.

For regular surfaces Σ we can define Runge spheres (cf. [45, 46, 48], see also [109] where it is called Krarup sphere).

Definition 2.2. The *Runge sphere* $\mathbb{S}_R^2$ is a sphere of radius R around the origin such that the exterior of the Runge sphere, i.e., $\mathbb{S}_{R,\text{ext}}^2$, contains the exterior of the regular surface Σ, i.e., $\overline{\Sigma_{\text{ext}}} \subset \mathbb{S}_{R,\text{ext}}^2$.

2.2. Sobolev spaces

Now we briefly introduce the Sobolev spaces of the form $\mathcal{H} = \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)$ using the Runge sphere $\mathbb{S}_R^2$. For more details, the reader is referred to [45, 48, 107] and the references in these books.

Definition 2.3. Let $\{A_n\}_{n\in\mathbb{N}_0} \subset \mathbb{R}$ be a sequence which satisfies the summability condition

$$\sum_{n=0}^{\infty} \frac{2n+1}{4\pi A_n^2} < \infty. \tag{5}$$

The Sobolev space $\mathcal{H} = \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)$ is defined by

$$\mathcal{H} = \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right) = \overline{\mathcal{E}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)}^{\|\cdot\|_{\mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)}}, \tag{6}$$

where $\mathcal{E}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right) \subset C^{(\infty)}\left(\overline{\mathbb{S}_{R,\text{ext}}^2}\right)$ is the set of all functions that are harmonic in $\mathbb{S}_{R,\text{ext}}^2$, infinitely often differentiable on the Runge sphere $\mathbb{S}_R^2$ and regular at infinity (i.e., (1) and (2) hold) and whose Fourier coefficients $F^\wedge(n,m)$ with respect to $L^2(\mathbb{S}_R^2)$ (as defined in (4)) fulfill

$$\|F\|_{\mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^2 \left(F^\wedge(n,m)\right)^2 < \infty\,. \tag{7}$$

$\mathcal{H}$ is a Hilbert space with the inner product defined by

$$\langle F, G\rangle_{\mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^2\, F^\wedge(n,m) G^\wedge(n,m) \tag{8}$$

for $F, G \in \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}_{R,\text{ext}}^2}\right)$.

It is well known (cf. [45, 107] and the references therein) that such a space possesses a so-called reproducing kernel (see [3] for an overview on reproducing kernels in general).

Definition 2.4. Let U be a non-empty set and $(X, \langle\cdot,\cdot\rangle_X)$ be a separable Hilbert space of real-valued functions on U. Let $\{B_n\}_{n\in\mathbb{N}_0}$ be a complete orthonormal system in $(X, \langle\cdot,\cdot\rangle_X)$. Any function $K : U \times U \longrightarrow \mathbb{R}$ of the form

$$K(x,y) = \sum_{n=0}^{\infty} K^{\wedge}(n) B_n(x)\, B_n(y) \tag{9}$$

with $x, y \in U$ and $K^{\wedge}(n) \in \mathbb{R}$ for $n \in \mathbb{N}_0$ is called an *X-product kernel* (briefly an X-kernel).

An X-kernel $K(\cdot,\cdot) : U \times U \longrightarrow \mathbb{R}$ is called a *reproducing kernel* (or shortly *repro-kernel*) for $(X, \langle\cdot,\cdot\rangle_X)$ if:

(i) $K(x,\cdot) \in X$ for all $x \in U$.
(ii) $\langle K(x,\cdot), F\rangle_X = F(x)$ for all $x \in U$ and all $F \in X$.

If there exists such a repro-kernel in X, then X is called a reproducing kernel Hilbert space and the repro-kernel is unique (cf. [3]). In the space $\mathcal{H} = \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}^2_{R,\mathrm{ext}}}\right)$ with a summable sequence $\{A_n\}$ the repro-kernel (9) can be represented by its expansion in Legendre polynomials due to the well-known addition theorem for spherical harmonics:

$$K_{\mathcal{H}}(x,y) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi A_n^2} \frac{1}{|x||y|} \left(\frac{R^2}{|x||y|}\right)^n P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right). \tag{10}$$

2.3. Harmonic splines

We use the reproducing kernels of Section 2.2 to define harmonic splines.

Definition 2.5. Let $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\} \subset \mathcal{H}^*$ be a set of N linearly independent bounded linear functionals on the reproducing kernel Hilbert space $\mathcal{H}$. Then any function S of the form

$$S = \sum_{i=1}^{N} a_i \mathcal{L}_i K_{\mathcal{H}}(\cdot,\cdot) \tag{11}$$

with a set of so-called *spline coefficients* $\{a_1, \ldots, a_N\} \subset \mathbb{R}$ is called an *$\mathcal{H}$-spline* relative to $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\}$. The function space of all $\mathcal{H}$-splines relative to $\{\mathcal{L}_1, \ldots, \mathcal{L}_N\}$ is denoted by $\mathcal{S}_{\mathcal{H}}(\mathcal{L}_1, \ldots, \mathcal{L}_N)$.

By construction any $\mathcal{H}$-spline is a harmonic function. The interpolating spline S^F for the function $F \in \mathcal{H}$ has to fulfill the interpolation conditions

$$\mathcal{L}_i S^F = \mathcal{L}_i F \quad \text{for } i = 1, \ldots, N. \tag{12}$$

The interpolation conditions (12) can be rewritten as a system of linear equations for the spline coefficients a_i:

$$\sum_{i=1}^{N} a_i \mathcal{L}_i \mathcal{L}_j K_{\mathcal{H}}(\cdot,\cdot) = \mathcal{L}_j F, \quad j = 1, \dots, N, \tag{13}$$

whose corresponding matrix possesses the entries $\mathcal{L}_i \mathcal{L}_j K_{\mathcal{H}}(\cdot,\cdot)$ and is symmetric and positive definite (for linear functionals $\mathcal{L}_1, \dots, \mathcal{L}_N \in \mathcal{H}^*$ which are linearly independent).

In this paper, we consider only evaluation functionals $\mathcal{L}_x$, i.e., $\mathcal{L}_x F = F(x)$ where $x \in \overline{\Sigma_{\text{ext}}}$. Furthermore, $\mathcal{L}_1, \dots, \mathcal{L}_N$ are given by $\mathcal{L}_i F = F(x_i)$ where $x_i \in \Sigma$. For other types of functionals see [43, 44] or [67, 68] (for the case of oblique derivatives). In the following theorem we summarize the properties of $\mathcal{H}$-splines.

Theorem 2.6. *Let $F \in \mathcal{H}$ and let $\{\mathcal{L}_1, \dots, \mathcal{L}_N\} \subset \mathcal{H}^*$. Then the $\mathcal{H}$-spline interpolation problem with the interpolation conditions* (12) *is uniquely solvable and its solution $S^F \in \mathcal{S}_{\mathcal{H}}(\mathcal{L}_1, \dots, \mathcal{L}_N)$ possesses the following properties:*

(i) *S^F is the $\mathcal{H}$-orthogonal projection of F onto $\mathcal{S}_{\mathcal{H}}(\mathcal{L}_1, \dots, \mathcal{L}_N)$.*

(ii) $\left\|S^F\right\|_{\mathcal{H}} \leq \|F\|_{\mathcal{H}}$.

(iii) *If $G \in \mathcal{H}$ also satisfies the interpolation conditions* (12)*, then the first minimum property holds:*

$$\|G\|_{\mathcal{H}}^2 = \left\|S^F\right\|_{\mathcal{H}}^2 + \left\|G - S^F\right\|_{\mathcal{H}}^2, \tag{14}$$

i.e., S^F is the interpolating function of F in $\mathcal{H}$ with minimal norm.

(iv) *If $S \in \mathcal{S}_{\mathcal{H}}(\mathcal{L}_1, \dots, \mathcal{L}_N)$ and $G \in \mathcal{H}$ also satisfies the interpolation conditions* (12)*, then the second minimum property holds:*

$$\|S - G\|_{\mathcal{H}}^2 = \left\|S^F - G\right\|_{\mathcal{H}}^2 + \left\|S - S^F\right\|_{\mathcal{H}}^2. \tag{15}$$

For the proof and for further details on splines, the reader is referred to [37, 43, 45, 51, 107] and the references therein.

Example 2.7. The choice $A_n = h^{-\frac{n}{2}}$, $h \in (0,1)$, fulfills (5) and provides us with the reproducing kernel called *Abel–Poisson kernel* which is given by

$$K_{\mathcal{H}}(x,y) = \frac{1}{4\pi} \frac{|x|^2|y|^2 - h^2R^4}{(|x|^2|y|^2 + h^2R^4 - 2hR^2 x \cdot y)^{\frac{3}{2}}}. \tag{16}$$

Example 2.8. The sequence $A_n = (n + \frac{1}{2})^{\frac{1}{2}} h^{-\frac{n}{2}}$, $h \in (0,1)$, also satisfies (5) and leads to the *singularity kernel* given by

$$K_{\mathcal{H}}(x,y) = \frac{1}{2\pi} \frac{1}{(|x|^2|y|^2 + h^2R^4 - 2hR^2 x \cdot y)^{\frac{1}{2}}}. \tag{17}$$

In [48] the existence of approximations fulfilling interpolation conditions is shown by the Runge–Walsh approximation theorem and an extension of Helly's theorem (cf. [138]). Convergence results for harmonic splines (cf. [43, 44]) can be

derived that show the convergence to the solution of the Dirichlet boundary value problem for an increasing density of data points, i.e., if the largest data gap goes to zero (cf. [43]).

We consider two specific types of splines (using Abel–Poisson and singularity kernels) and propose the fast multipole method to quickly compute the sums $\sum a_i K_{\mathcal{H}}(x_i, y_j)$ for many points in Section 3. This can be used to solve the systems of linear equations (13) that occur in the solution of the interpolation problems using harmonic splines.

2.4. Spline approximation

For noisy data, i.e., $F_i = U(x_i) + \delta_i$, $i = 1, \ldots, N$, where the noise δ_i is modeled by some stochastic process, e.g., white noise (see Section 4.1 for details), in Problem 1.2, it makes no sense to compute an interpolation problem. We look for an approximation to U which can be interpreted as a smoothing of the data (see [38, 51, 50, 135] for the spherical spline approximation, [37, 45] for the case of harmonic spline approximation). Minimizing the following functional

$$\mu(S) = \sum_{i=1}^{N} \sum_{j=1}^{N} (\mathcal{L}_i S - F_i) C_{ij} (\mathcal{L}_j S - F_j) + \beta \|S\|_{\mathcal{H}} \tag{18}$$

in the reproducing kernel Hilbert space $\mathcal{H} = \mathcal{H}\left(\{A_n\}; \overline{\mathbb{S}^2_{R,\text{ext}}}\right)$ yields the desired smoothed approximation of the data. $C = (C_{ik}) \in \mathbb{R}^{N \times N}$ denotes a positive definite matrix which allows us to include covariance information on the data if available. $\beta > 0$ is a constant smoothing parameter which balances closeness to the data and smoothing. The following theorem of [45, 107] (see also the references therein) summarizes the existence and uniqueness of a spline approximation.

Theorem 2.9. *Let F_i, $i = 1, \ldots, N$, correspond to a set of linearly independent bounded linear functionals $\mathcal{L}_1, \ldots, \mathcal{L}_N \in \mathcal{H}^*$.*
Then there exists a unique element $S \in \mathcal{S}_{\mathcal{H}}(\mathcal{L}_1, \ldots, \mathcal{L}_N)$ such that

$$\mu(S) \leq \mu(F) \quad \text{for all } F \in \mathcal{H} \tag{19}$$

and $\mu(S) = \mu(F)$ if and only if $S = F$. This element is called the smoothing spline *or* approximating spline. *Its spline coefficients a_i, $i = 1, \ldots, N$, are uniquely determined by the system of linear equations*

$$\sum_{i=1}^{N} a_i \left(\mathcal{L}_i \mathcal{L}_j K_{\mathcal{H}}(\cdot, \cdot) + \beta (C^{-1})_{ij} \right) = \mathcal{L}_j F, \quad j = 1, \ldots, N. \tag{20}$$

The matrix in (20) corresponds to the sum of the matrix in (13) and βC^{-1}. It is still positive definite. If C is the identity matrix, there is only the one smoothing parameter β. Using a diagonal matrix as C it is possible to introduce weights for the data F_i and include additional information on the noise of the data. The choice of the smoothing parameter(s) can be interpreted as the application of a parameter choice method in the regularization theory of ill-posed problems (see Section 4).

3. The fast multipole method for splines

The interpolation conditions (12) as well as the minimization of the smoothing functional (18) lead to a system of linear equations (13), or (20) respectively, with a dense matrix whose size is the number of data points. This matrix can be large and the solution of the corresponding system of linear equations becomes difficult.

Reproducing kernels of a reproducing kernel Hilbert space $\mathcal{H}$ defined by the summable sequence $\{A_n\}$ can be expanded in terms of Legendre polynomials as in (10). The singularity kernel (17) and the Abel–Poisson kernel (16) possess a representation as an elementary function and both kernels are closely related to the single pole $\frac{1}{|x-y|}$ by the Kelvin transform. Therefore, we can use the fast multipole method (FMM), which has been introduced by Greengard [61], Greengard and Rokhlin [62, 63], and Rokhlin [124], for the fast summation of harmonic splines, i.e., of the sum $\sum a_i K_{\mathcal{H}}(x_i, \cdot)$. This corresponds to the matrix-vector products occurring in an iterative solver for (13) or (20).

The FMM creates a hierarchical subdivision of the computational domain into nested cubes that are organized in an octtree data structure. Instead of single point interaction the cubes summarize the part of the kernel sum corresponding to the points they contain and interact with other cubes via the coefficients of truncated inner/outer harmonics expansions. This kernel approximation is applied as often as possible and on the coarsest possible level of the tree data structure. Direct evaluation is used only for the closest cubes where the approximation is not accurate enough. The algorithm has been improved several times to increase its efficiency (cf., e.g., [24, 64, 137]). We summarize our implementation and show the application of the FMM to harmonic splines (see [67, 69] for a more detailed analysis).

3.1. Kelvin transform of reproducing kernels

The Kelvin transform yields the connection between the kernels (17) and (16) and the fundamental solution of the Laplace equation, i.e., the single pole. The transform can be seen as a reflection on a sphere around the origin and we choose the Runge sphere $\mathbb{S}^2_R$ of Definition 2.2 for this (see also text books on potential theory, e.g., [46, 83]).

Definition 3.1. Let $\Gamma \subseteq \mathbb{R}^3$ be a domain, $W : \Gamma \longrightarrow \mathbb{R}$ a function. Let the reflection of Γ on the sphere $\mathbb{S}^2_R$ be given by

$$\Gamma^{\mathrm{KT}} = \left\{ x^{\mathrm{KT}} \in \mathbb{R}^3 : \frac{R^2}{|x^{\mathrm{KT}}|^2} x^{\mathrm{KT}} = x \in \Gamma \right\}. \tag{21}$$

The function

$$W^{\mathrm{KT}} : \Gamma^{\mathrm{KT}} \longrightarrow \mathbb{R},$$

$$x^{\mathrm{KT}} \mapsto W^{\mathrm{KT}}(x^{\mathrm{KT}}) = \frac{R}{|x^{\mathrm{KT}}|} W\left(\frac{R^2}{|x^{\mathrm{KT}}|^2} x^{\mathrm{KT}} \right) = \frac{R}{|x^{\mathrm{KT}}|} W(x), \tag{22}$$

is called the *Kelvin transform* of W with respect to the sphere of radius R.

The Kelvin transform is applied to the reproducing kernels with respect to one argument (the other is kept fixed). The Kelvin transform $K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}})$ of the singularity kernel (17) can be computed, e.g., by its expansion

$$\begin{aligned} K_{\mathcal{H}}(x,y) &= \sum_{n=0}^{\infty} \frac{h^n}{n+\frac{1}{2}} \frac{2n+1}{4\pi|x||y|} \left(\frac{R^2}{|x||y|}\right)^n P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right) \\ &= \frac{1}{2\pi|y|} \sum_{n=0}^{\infty} \frac{(h|y^{\mathrm{KT}}|)^n}{|x|^{n+1}} P_n\left(\frac{x}{|x|}\cdot\frac{y^{\mathrm{KT}}}{|y^{\mathrm{KT}}|}\right) \\ &= \frac{1}{2\pi|y|}\frac{1}{|x-hy^{\mathrm{KT}}|} = \frac{|y^{\mathrm{KT}}|}{R} K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}}), \end{aligned} \tag{23}$$

where $y^{\mathrm{KT}} = \frac{R^2}{|y|^2} y$ and

$$K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}}) = \frac{1}{2\pi R}\frac{1}{|x-hy^{\mathrm{KT}}|}. \tag{24}$$

The Kelvin transform $K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}})$ of the Abel–Poisson kernel (16) is given by

$$\begin{aligned} K_{\mathcal{H}}(x,y) &= \frac{1}{4\pi} \frac{|x|^2|y|^2 - h^2R^4}{(|x|^2|y|^2 + h^2R^4 - 2hR^2 x\cdot y)^{\frac{3}{2}}} \\ &= \frac{|y^{\mathrm{KT}}|}{R}\frac{1}{4\pi R}\frac{|x|^2 - h^2|y^{\mathrm{KT}}|^2}{|x-hy^{\mathrm{KT}}|^3} = \frac{|y^{\mathrm{KT}}|}{R} K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}}), \end{aligned} \tag{25}$$

which is related to (24) by

$$K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}}) = \frac{1}{2\pi R}\left(-x\cdot\nabla_x - \tfrac{1}{2}\mathrm{Id}\right)\frac{1}{|x-hy^{\mathrm{KT}}|}. \tag{26}$$

We summarize both (24) and (26) by use of the operator D_x such that

$$K_{\mathcal{H}}^{\mathrm{KT}}(x, y^{\mathrm{KT}}) = \frac{1}{2\pi R} D_x \frac{1}{|x-hy^{\mathrm{KT}}|}, \tag{27}$$

where $D_x = \mathrm{Id}$ (singularity kernel) or $D_x = -x\cdot\nabla_x - \frac{1}{2}\mathrm{Id}$ (Abel–Poisson kernel).

3.2. Adaptive decomposition of the domain

Now we consider the evaluation of the sum

$$\begin{aligned} \sum_{i=1}^{N} a_i K_{\mathcal{H}}(x_i, y) &= \sum_{i=1}^{N} a_i \frac{|y^{\mathrm{KT}}|}{R} K_{\mathcal{H}}^{\mathrm{KT}}(x_i, y^{\mathrm{KT}}) \\ &= \frac{|y^{\mathrm{KT}}|}{R} \sum_{i=1}^{N} \frac{a_i}{2\pi R} D_x \left.\frac{1}{|x-hy^{\mathrm{KT}}|}\right|_{x=x_i} \end{aligned} \tag{28}$$

at the points y_j, $j = 1, \ldots, M = \mathcal{O}(N)$, which can be the points x_i as in the systems of linear equations (13) or (20). To better distinguish the points x_i and y_j we call the Kelvin transformed points $hy_j^{\mathrm{KT}} = h\frac{R^2}{|y_j|^2} y_j$ which are used in (28) *targets*.

In the beginning, a bounding cube is determined which is large enough such that it contains all points and all targets. This single cube forms level 0 of the octtree structure and is subdivided into eight equally sized cubes of half its edge length which then form level 1. Each cube is adaptively divided into nested cubes where a cube of level l has half the edge length of a cube of level $l-1$ as proposed, e.g., by Cheng et al. [24]. Points and targets are sorted into the currently available cubes. If a cube contains more than the prescribed maximal number of points or targets m, it is split into eight smaller cubes of the next level. All its points/targets are redistributed into these eight cubes. We summarize the necessary vocabulary in some definitions (see also [24, 64]).

Definition 3.2.

(a) A cube C is called *child* of the cube B if C results from a single subdivision of B which in return is named the *parent* of C.
(b) A cube that is not further subdivided is called *childless* or a *leaf*.
(c) Cubes are said to be *neighbors* if they are of the same size (same level) and share at least one boundary point. Each cube is a neighbor of itself.
(d) If two cubes are at the same level, but are no neighbors, they are called *well-separated*, i.e., between these cubes exists at least one cube of their size.

Each cube carries the relevant information about other cubes, in particular its neighbors, in four lists as suggested by Cheng et al. [24]. Figures 1 and 2 give a two-dimensional illustration.

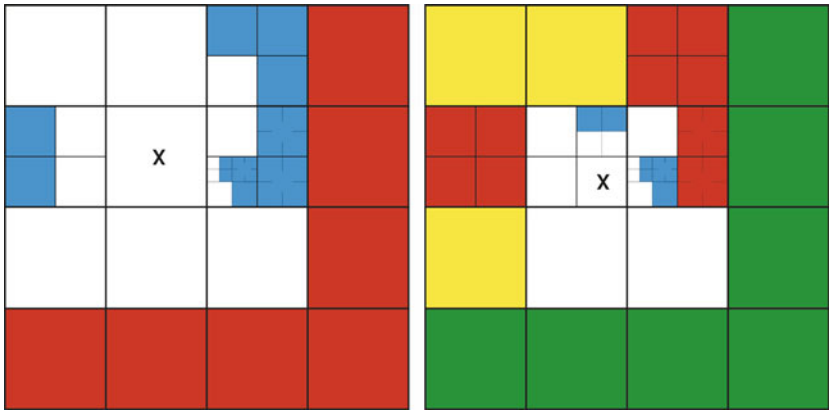


FIGURE 1. Two-dimensional illustration of an adaptive decomposition for levels 2 (left) and 3 (right). If the marked square is a leaf, the white cubes form its list 1 and the blue cubes correspond to list 3. If the marked square is not a leaf, all white and all blue cubes are treated at the next level. List 2 cubes are red and list 4 cubes are yellow. Cubes in green have been handled at coarser levels.

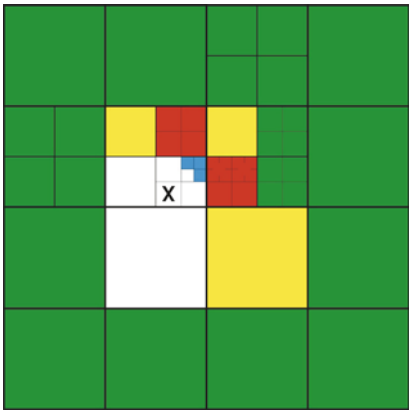


FIGURE 2. Two-dimensional illustration of an adaptive decomposition for level 4. If the marked square is a leaf, the white cubes form its list 1 and the blue cubes correspond to list 3. If the marked square is not a leaf, all white and all blue cubes are treated at the next level. List 2 cubes are red and list 4 cubes are yellow. Cubes in green have been handled at coarser levels.

Definition 3.3.

(a) In *list* 1 of the childless cube X are all childless cubes directly adjacent to X. List 1 only contains any cubes if X is a leaf. In this case it always contains at least X itself.
(b) *List* 2 of a cube X consists of all children of neighbors of the parent cube of X which are well separated from X. The cube X does not need to be childless.
(c) Children of neighbors of the leaf X (or smaller cubes descending from neighbors of X) which do not have any point in common with X form *list* 3. Their parents have to be adjacent to X. If X is not childless, then *list* 3 is empty.
(d) *List* 4 consists of childless cubes which are neighbors of the parent cube of X, but these childless cubes are not adjacent to X.

Notice the following observations:

(i) List 1 is the list of all neighbors.
(ii) All cubes in list 2 of a cube X are of the same size as X and well separated from X.
(iii) The elements of list 3 are all smaller than X and the distance between them and X is at least their side length and at most the side length of X.
(iv) List 4 of a cube X only contains cubes that are larger than X. They are separated from X by a distance that is at least the side length of X and at most their own edge length.
(v) A cube X is in list 3 of a cube Y if and only if Y is in list 4 of X.
(vi) All members of list 1 and list 4 are leaves and list 1 as well as list 3 of a cube X remain empty if X is not childless.

After finishing the adaptive construction of the octtree and sorting all points and targets into cubes, the algorithm removes childless cubes that contain neither points nor targets and are no longer required.

3.3. Single pole expansion

In addition to the decomposition of the domain, the other part of the FMM is the kernel approximation of the single pole by a truncated expansion and the use of translation theorems to shift the expansion center and to change the type of expansion. Similar to (23) we use the generating function of the Legendre polynomials P_n and the addition theorem of spherical harmonics (see, e.g., [1, 47, 102]) to expand the single pole.

$$\begin{aligned}\frac{1}{|x-y|} &= \frac{1}{|y-x_0-(x-x_0)|} \\ &= \sum_{n=0}^{\infty} \frac{|x-x_0|^n}{|y-x_0|^{n+1}} P_n\left(\frac{y-x_0}{|y-x_0|}\cdot\frac{x-x_0}{|x-x_0|}\right) \\ &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} I^*_{n,m}(x-x_0) O_{n,m}(y-x_0), \end{aligned} \tag{29}$$

where $|y-x_0| > |x-x_0|$ for the expansion center $x_0 \in \mathbb{R}^3$. The upper star $*$ in (29) denotes the complex conjugate. Thereby, we use the (complex-valued) outer and inner harmonics for $n \in \mathbb{N}_0$, $m = -n, \ldots, n$:

$$O_{n,m}(x) = \sqrt{\frac{4\pi}{2n+1}} \frac{\sqrt{(n+m)!(n-m)!}}{|x|^{n+1}} Y_{n,m}\left(\tfrac{x}{|x|}\right), \tag{30}$$

$$I_{n,m}(x) = \sqrt{\frac{4\pi}{2n+1}} \frac{|x|^n}{\sqrt{(n+m)!(n-m)!}} Y_{n,m}\left(\tfrac{x}{|x|}\right), \tag{31}$$

where $\vartheta \in [0,\pi]$, $\varphi \in [0, 2\pi)$ are the usual spherical coordinates of $\frac{x}{|x|}$ and $Y_{n,m} : \mathbb{S}^2 \to \mathbb{C}$ with

$$Y_{n,m}(\xi) = (-1)^m \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_{n,m}(\cos(\vartheta)) e^{im\varphi}, \qquad \xi \in \mathbb{S}^2, \tag{32}$$

are complex-valued fully normalized spherical harmonics of degree n and order m. $P_{n,m} : [-1,1] \to \mathbb{R}$ are the *associated Legendre functions* with

$$P_{n,m}(t) = \frac{1}{2^n n!}(1-t^2)^{\frac{m}{2}} \frac{d^{n+m}}{dt^{n+m}}\left((t^2-1)^n\right), \quad m = 0, \ldots, n. \tag{33}$$

The symmetry relation $P_{n,-m}(t) = (-1)^m \frac{(n-m)!}{(n+m)!} P_{n,m}(t)$ extends them for negative orders (cf., e.g., [16, 30]).

Well-known translation theorems for these outer and inner harmonics allow to shift the expansion center (see, e.g., [34] for a detailed derivation).

Theorem 3.4 (Translation Theorem for Outer Harmonics). *Let $x, y \in \mathbb{R}^3$ such that $|x| > |y|$. Then the outer harmonic of degree $n \in \mathbb{N}_0$ and order $m \in \mathbb{Z}$, $-n \leq m \leq n$, at $x - y$ can be expanded in terms of inner and outer harmonics as follows*

$$O_{n,m}(x-y) = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} I^*_{n',m'}(y) O_{n+n',m+m'}(x) \tag{34}$$

$$= \sum_{n'=n}^{\infty} \sum_{m'=-n'}^{n'} I^*_{n'-n,m'-m}(y) O_{n',m'}(x). \tag{35}$$

Note that in (35) we make use of the convention that $I_{n,m} = 0$ if $|m| > n$. Obviously, this infinite series as well as the infinite sum in (29) have to be truncated for the algorithm which are sources of (truncation) errors. Error estimates for the truncation errors can be found in [61, 63, 64] and the references therein. Another approach which investigates the combined error of both truncations in (29) and (35) is considered in [67].

Theorem 3.5 (Translation Theorem for Inner Harmonics). *Let $x, y \in \mathbb{R}^3$. Then the inner harmonic of degree $n \in \mathbb{N}_0$ and order $m \in \mathbb{Z}$, $-n \leq m \leq n$, at $x - y$ can be expanded in a finite sum of inner harmonics*

$$I_{n,m}(x-y) = \sum_{n'=0}^{n} \sum_{m'=-n'}^{n'} (-1)^{n'} I_{n',m'}(y) I_{n-n',m-m'}(x). \tag{36}$$

For orders with $|m| > n$ we have again by convention $I_{n,m} = 0$. Note that no truncation is necessary for this translation theorem, i.e., no truncation errors occur.

By applying (35) of Theorem 3.4 we can translate an outer harmonics expansion with expansion center x_0 such as

$$F(x) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} F^{\wedge,\mathrm{O}}_{x_0}(n,m) O_{n,m}(x - x_0) \tag{37}$$

which converges uniformly for $x \in \mathbb{S}^2_{r_0,\mathrm{ext}}(x_0)$ with some $r_0 > 0$. $\mathbb{S}^2_{r_0,\mathrm{ext}}(x_0)$ denotes the exterior of the sphere of radius r_0 around x_0. The outer harmonics series resulting from the translation possesses the expansion center x_1 and the coefficients

$$F^{\wedge,\mathrm{O}}_{x_1}(n',m') = \sum_{n=0}^{n'} \sum_{m=-n}^{n} F^{\wedge,\mathrm{O}}_{x_0}(n,m) I^*_{n'-n,m'-m}(x_0 - x_1). \tag{38}$$

This expansion converges uniformly for $x \in \mathbb{S}^2_{r_1,\mathrm{ext}}(x_1)$ where

$$\mathbb{S}^2_{r_1,\mathrm{ext}}(x_1) \subset \mathbb{S}^2_{r_0,\mathrm{ext}}(x_0).$$

This translation of coefficients is called *multipole to multipole translation* (M2M). Note that one can show that no further errors arise if the series in (37) has already been truncated before the translation.

By using formulation (34) of Theorem 3.4 we also find that the outer harmonics expansion with expansion center x_1 can be translated into an inner harmonics series centered around x_2 which converges uniformly for $x \in \mathbb{S}^2_{r_2,\mathrm{int}}(x_2)$ if the new ball of convergence is situated completely in $\mathbb{S}^2_{r_1,\mathrm{ext}}(x_1)$, i.e., $\overline{\mathbb{S}^2_{r_1,\mathrm{int}}(x_1)} \cap \overline{\mathbb{S}^2_{r_2,\mathrm{int}}(x_2)} = \emptyset$. The resulting coefficients of the inner harmonic expansion are

$$F^{\wedge,\mathrm{I}}_{x_2}(n',m') = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} F^{\wedge,\mathrm{O}}_{x_1}(n,m)(-1)^{n'+m} O^*_{n+n',m'-m}(x_2 - x_1) \tag{39}$$

and this translation is named *multipole to local translation* (M2L).

Finally, Theorem 3.5 lets us shift the expansion center of such inner harmonics expansions to the new center x_3 which possesses the coefficients

$$F^{\wedge,\mathrm{I}}_{x_3}(n',m') = \sum_{n=n'}^{\infty} \sum_{m=-n}^{n} F^{\wedge,\mathrm{I}}_{x_2}(n,m) I_{n-n',m-m'}(x_3 - x_2). \tag{40}$$

and converges uniformly for $x \in \mathbb{S}^2_{r_3,\mathrm{int}}(x_3) \subset \mathbb{S}^2_{r_2,\mathrm{int}}(x_2)$. This translation step is called *local to local translation* (L2L). For further details we refer to [67] and the references therein, in particular [34].

3.4. The fast multipole algorithm

To start the algorithm a first set of multipole expansion coefficients for each cube containing any points has to be computed. We consider only the part of the spline related to a single cube X, i.e., the kernel functions $K_{\mathcal{H}}(x_i, \cdot)$, where $x_i \in X$ and $y \in \overline{\Sigma_{\mathrm{ext}}}$:

$$F(y) = \sum_{\substack{i=1 \\ x_i \in X}}^{N} a_i K_{\mathcal{H}}(x_i, y) = \sum_{\substack{i=1 \\ x_i \in X}}^{N} a_i \left(\frac{|y^{\mathrm{KT}}|}{R} \frac{1}{2\pi R} D_x \frac{1}{|x - hy^{\mathrm{KT}}|} \right) \Bigg|_{x=x_i}. \tag{41}$$

We find the following expansion for $|hy^{\mathrm{KT}} - x_0| > |x_i - x_0|$, $x_i \in X$, i.e., if x_0 is the center of the cube X, the targets hy^{KT} and the cube X need to fulfill a distance requirement, i.e., targets must be contained in a well-separated cube.

$$\begin{aligned} F(y) &= \frac{|y^{\mathrm{KT}}|}{R} \sum_{\substack{i=1 \\ x_i \in X}}^{N} \frac{a_i}{2\pi R} \left(D_x \sum_{n=0}^{\infty} \sum_{m=-n}^{n} I^*_{n,m}(x - x_0) O_{n,m}(hy^{\mathrm{KT}} - x_0) \right) \Bigg|_{x=x_i} \\ &= \frac{|y^{\mathrm{KT}}|}{R} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} F^{\wedge,O}_{x_0}(n,m) O_{n,m}(hy^{\mathrm{KT}} - x_0) \end{aligned} \tag{42}$$

where the multipole coefficients $F^{\wedge,O}_{x_0}(n,m)$ of the cube X are given by

$$F^{\wedge,O}_{x_0}(n,m) = \sum_{\substack{i=1 \\ x_i \in X}}^{N} \frac{a_i}{2\pi R} \left(D_x I^*_{n,m}(x - x_0) \right) \Bigg|_{x=x_i}. \tag{43}$$

This first step is called *point to multipole* (P2M) step where the infinite sum in (42) has to be truncated at degree p. The degree p essentially determines the accuracy of the algorithm. The coefficients $F_{x_0}^{\wedge,O}(n,m)$ can be translated to other cubes via relations (38), (39) as well as (40) as long as the distance requirements are fulfilled by the construction of the decomposition of the domain into nested cubes.

At the end of the fast multipole cycle, i.e., after several M2M-, M2L-, L2L-translations, each cube Y possesses an inner harmonics expansion centered around the center of the cube. This expansion has to be evaluated at the targets contained by Y. This evaluation is called the *local to targets* (L2T) *step*:

$$\mathcal{L}_j F = F(y_j) = \left(\frac{|y^{\mathrm{KT}}|}{R} \sum_{n=0}^{p} \sum_{m=-n}^{n} F_{x_0}^{\wedge,I}(n,m) I_{n,m}(hy^{\mathrm{KT}} - x_0) \right) \Bigg|_{y=y_j}, \qquad (44)$$

where the variable y is hidden by $y^{\mathrm{KT}} = \frac{R^2}{|y|^2} y$.

Now we briefly summarize the fast multipole algorithm (see, e.g., [20, 24] or [67, 69] for our specific implementation).

Algorithm 3.6 (Fast Multipole Algorithm).
Input:

- A set of points $x_i \in \overline{\Sigma_{\mathrm{ext}}}$ (often $x_i \in \Sigma$), $i = 1, \ldots, N$,
- a set of coefficients a_i, $i = 1, \ldots, N$,
- the choice of the type of the reproducing kernel $K_{\mathcal{H}}$ (singularity or Abel–Poisson with the parameter h and the radius of the Runge sphere R),
- a set of evaluation points $y_j \in \overline{\Sigma_{\mathrm{ext}}}$, $j = 1, \ldots, M$, where $M = \mathcal{O}(N)$,
- the degree of the multipole expansion p,
- the maximal number of points per cube m.

Aim: compute the sum

$$F(y_j) = \sum_{i=1}^{N} a_i K_{\mathcal{H}}(x_i, y_j) \quad \text{for each } j = 1, \ldots, M. \qquad (45)$$

Initialization:

- Compute the targets $hy_j^{\mathrm{KT}} = h\frac{R^2}{|y_j|^2} y_j$, $j = 1, \ldots, M$.
- Create a bounding box that contains all points and all targets, build the adaptive octtree and sort in all points and targets. Set L as the maximum level, eliminate all empty cubes.
- Determine list 1 to list 4 of Definition 3.2. Create a list of all cubes of level l for each level $l = 0, \ldots, L$. Collect all leaves in a list.
- Allocate memory for the different expansion coefficients of each cube X: multipole expansion (coefficient vector M_X), local expansion (coefficient vector L_X).

Fast multipole cycle:

1. Generation of the multipole coefficients:
For all leaves X: P2M, i.e., compute the multipole coefficients M_X of the multipole

expansion up to degree p around the center of X from the points in X as in (43). For level $l = L-1, \ldots, 2$: M2M, i.e., translate the multipole coefficients of the children of X to X itself for all cubes X of level l via (38).

2. Interaction phase for list 4:
For level $l = 2, \ldots, L$: for all cubes X of level l: compute the expansion coefficients of an inner harmonics expansion around the center of X from the points in Y for all cubes Y of list 4 of X and add them to L_X – or use direct evaluation of the kernel sum corresponding to the points in Y to obtain the result at the targets in X if the number of targets in $X \leq p^2$ and X is a leaf.

3. Multipole to local translation:
For level $l = 2, \ldots, L$: for all cubes X of level l: use (39) to translate M_X to L_Y for all cubes Y in list 2 of X.

4. Translation of the inner harmonics expansions:
For level $l = 2, \ldots, L-1$: L2L, i.e., translate the local coefficients L_X to the children of X (if there are any) via (40) and add the resulting coefficients to L_Z where Z denotes the corresponding child of X for all cubes X of level l.

5. Evaluation of the expansions and direct interaction:
For all leaves X: L2T, i.e., evaluate the inner harmonics expansion of X at all targets in X as in (44). Store the result in F.
For all cubes Y in list 1 of X: P2T, i.e., add the direct evaluation of the kernel sum corresponding to the points in Y at the targets in X to F.
For all cubes Y in list 3 of X: evaluate the multipole expansion around the center of Y (coefficients M_Y) at the targets in X and add the results to F – or use direct evaluation of the kernel sum corresponding to the points in Y to add the result at the targets in X to F if the number of points in $Y \leq p^2$ and Y is a leaf.

6. Reverse the effects of the Kelvin transformation:
$\widetilde{F}_j = \frac{|y_j^{\mathrm{KT}}|}{R} F_j$ for $j = 1, \ldots, M$.
Return the result $\widetilde{F}$.

For the computation of the spline coefficients of the spline approximation of Section 2.4 we consider the system of linear equations (20) instead of (13). This means that we have to add $\beta \sum_{i=1}^{N} a_i (C^{-1})_{ij}$ to the matrix-vector product that is computed by the FMM. In order to keep a fast algorithm the matrix C^{-1} has to allow a fast summation method or C has to be a sparse matrix. The trivial cases where C is a diagonal matrix can also be included in the direct evaluation step of the fast multipole algorithm.

3.5. Acceleration of the translations

Newer iterations of the FMM include several ideas to reduce the numerical effort of the translations from the original $\mathcal{O}(p^4)$ to $\mathcal{O}(p^3)$ or even $\mathcal{O}(p^2)$ per translation operation. This includes the ideas of [137] (see also [24, 64]) for the multipole to multipole (M2M) and the local to local (L2L) steps using Wigner rotation

matrices (cf., e.g., [16, 25, 30, 130]). The main point is to rotate the coordinate system such that the shift direction becomes the ε^3-axis, shift there and rotate back. This reduces the numerical costs from $\mathcal{O}(p^4)$ in the M2M- and L2L-steps to $\mathcal{O}(p^3)$, since each rotation as well as the shift along the ε^3-axis requires an effort of $\mathcal{O}(p^3)$. For a detailed description we refer to [137] or [67] with all technical details of our implementation.

For the M2L translation, [64, 24] have replaced this step with *exponential translations* which are based on the numerical integration of the integral representation

$$\begin{aligned}\frac{1}{|x-y|} &= \frac{1}{2\pi}\int_0^\infty e^{-\lambda(x_3-y_3)}\int_0^{2\pi} e^{i\lambda((x_1-y_1)\cos\alpha)+(x_2-y_2)\sin\alpha)}\,d\alpha\,d\lambda \\ &= \sum_{k=1}^{s(\varepsilon)}\frac{w_k}{M_k}\sum_{j=1}^{M_k} e^{-\lambda_k(x_3-y_3)}e^{i\lambda_k((x_1-y_1)\cos\alpha_{j,k})+(x_2-y_2)\sin\alpha_{j,k})} + \mathcal{O}(\varepsilon) \quad (46)\end{aligned}$$

for points x, y whose Cartesian coordinates satisfy $1 \le x_3 - y_3 \le 4$ as well as $0 \le \sqrt{(x_1-y_1)^2+(x_2-y_2)^2} \le 4\sqrt{2}$. Details as well as integration points λ_k, weights w_k and numbers of points M_k for the trapezoidal rule applied to the inner integral can be found in [24, 64, 139]. The accuracy ε of the numerical integration is determined by the discretization parameter $s(\varepsilon) = \mathcal{O}(p)$ of the outer integral in (46). The total number of numerical integration points, i.e., the number of exponential functions and coefficients, is supposed to be $\mathcal{O}(p^2)$.

By Hobson's formula (cf. [76]) a multipole expansion of F is transformed by (46) into a series of exponentials (multipole to exponential step, briefly M2X), these exponentials can be translated efficiently by the exponential to exponential shift (X2X). Afterwards the local coefficients are computed from the exponential coefficients (X2L). The restrictions on the positions of x and y mean that the exponential translations are applicable for cubes in list 2 (see Definition 3.3) that are situated above the current cube with another cube in between.

However, by combining rotations of the multipole expansion using again the Wigner rotation matrices, the exponential translation can substitute the M2L translation for all cubes in list 2. Therefore, the list of all well-separated cubes (list 2) is split into 6 directional lists (up, down, North, South, East and West) and instead of M2L the following sequence of transformations is used: (rotation), M2X, X2X, X2L, (inverse rotation).

Each exponential shift requires numerical costs of $\mathcal{O}(p^2)$ and the rotations can be applied using $\mathcal{O}(p^3)$ operations (as do the M2X and X2L steps). Thus, this improves the performance compared to the M2L step's $\mathcal{O}(p^4)$ effort. Moreover, we can save translations by recombination (see [24, 64, 67, 69] for more on the technical details). It should also be noted that there are several symmetries in the coefficients of the exponential expansion since we are dealing with a real-valued function F. These symmetries can be used to further reduce the constant of the numerical costs (cf. [24, 64]).

3.6. Parameters and results of our FMM implementation

Here we present just a few of the parameters that we use in our implementation of the FMM with exponential translations. For more detailed investigations of our version (in particular of the recombination of exponential translations mentioned at the end of Section 3.5) we refer to [67, 68, 69].

The truncation degree p is investigated for different accuracies of the exponential translation $s(\varepsilon)$. We increase p while $s(\varepsilon)$ is kept fixed and determine when the integration error of the numerical integration in the exponential translation (46) dominates the truncation error. This leads to the choices of p for different levels of $s(\varepsilon)$ given by Table 1. Note that the kernels of Abel–Poisson type require a slightly higher degree and therefore a bit more numerical effort.

$s(\varepsilon)$	Singularity kernel	Abel–Poisson kernel
8	4	5
17	12	13
26	23	25

TABLE 1. Resulting truncation degrees p for different $s(\varepsilon)$ for the two types of kernels.

The maximal number of points or targets per cube m has a strong influence on the adaptive octtree construction and the performance of the FMM. If m is too small, there are many cubes each containing only very few points. Thus, the kernel expansion coefficients no longer combine the information of enough points to be efficient. If m is too large, there are only few cubes each with a large number of points. This means that far too often instead of kernel expansion direct interaction is used. Therefore, m can be used to balance the effort of the direct interaction and the kernel approximation. It has been suggested to choose $m = \mathcal{O}(p^{3/2})$ (cf. [24, 64]) which may serve as a guideline. Many empirical tests (cf. [67]) led us to the conclusion that the choices for m given by Table 2 provide a good performance

$s(\varepsilon)$	Singularity kernel	Abel–Poisson kernel
8	85	75
17	130	140
26	380	240

TABLE 2. Chosen maximal numbers of points m per cube for the singularity kernel and the Abel–Poisson kernel and the different error levels.

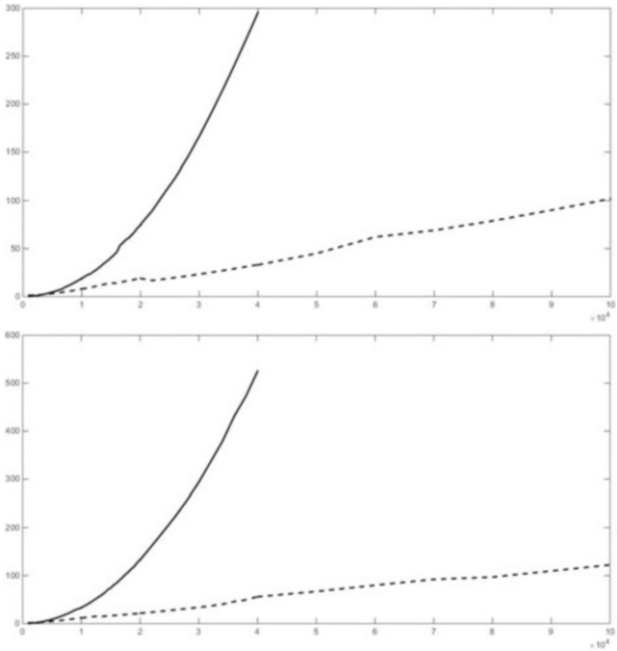


FIGURE 3. Break-even points by comparison of computation times for direct (solid line) and FMM accelerated (dashed line) computation (top: Singularity kernel, bottom: Abel–Poisson kernel), the number of points forms the abscissae.

in our implementation. Note again that there are remarkable differences between the two different types of kernels under investigation. Using these optimizations of the parameters of the FMM we compare its performance with direct computation and find the break-even points of our implementation, i.e., the minimal number of points that is necessary for our algorithm to be faster than the direct approach (see Table 3).

$s(\varepsilon)$	Singularity kernel	Abel–Poisson kernel
8	530	360
17	1,160	960
26	2,670	2,250

TABLE 3. Break-even points for the singularity kernel and the Abel–Poisson kernel.

Note that such results are always very dependent on the implementation. Our implementation turns out to be efficient even for rather small problem sizes. In general, the Abel–Poisson kernel requires some more computational time since it leads to a more difficult P2M-step. Finally, we show the linear asymptotic behavior which we expect from the FMM in Figure 3 compared to the quadratic behavior of the direct approach.

4. Parameter choice methods for spline approximation

In this section we summarize several parameter choice methods for the determination of the smoothing parameter in (18) and (20) that are known from the theory of regularization of inverse problems. For this context the reader is referred to [9, 14] where also many numerical tests for a wide range of inverse problems are presented.

We consider the problem of solving $Ax = y$ where $A : \mathcal{X} \to \mathcal{Y}$ is a linear operator. In our case $\mathcal{X} = \mathcal{Y} = \mathbb{R}^N$ and A is the matrix of the linear system (13) (in [9, 14] this is called case C2). This is in fact not an ill-posed problem, but can be severely ill-conditioned (depending on the distribution of the data points) and as seen in Section 2.4 noisy data lead to the use of a form of Tikhonov regularization, i.e., the linear system (20) with the smoothing parameter $\beta > 0$. It should be noted that we want to use the FMM for a fast matrix-vector-multiplication, i.e., A is never given as an actual matrix. This excludes some of the parameter choice methods in order to keep reasonably low computational costs. Moreover, it should be pointed out that the use of an iterative solver (e.g., cg-method or GMRES) requires a stopping criterion and the number of iterations can be seen as another regularization parameter which needs to be chosen in some way.

The smoothing parameter $\beta > 0$ is discretized exponentially, i.e., $\beta_n = \beta_0 q_\beta^n$ with $q_\beta \in (0, 1)$ and $n \in \mathbb{N}$ (actually only $n = 1, \ldots, n_{\max}$). The use of a discrete set of regularization parameters with a fine enough resolution does not alter the behavior of most parameter choice methods. For the efficient implementation of these methods, it is useful to have a bound on the value of the optimal parameter (i.e., a maximal regularization parameter), especially if the method minimizes some function (see also [9, 14] and Section 4.3). We denote the vector of noisy data by y^δ (see Section 4.1) and $x_n^\delta \in \mathcal{X}$ is the vector of spline coefficients resulting from the minimization of the smoothing functional (18) using β_n and y^δ, i.e., x_n^δ solves the system of linear equations (20):

$$x_n^\delta = (A + \beta_n I)^{-1} y^\delta = A_n^{-1} y^\delta, \tag{47}$$

where $A_n = A + \beta_n I$. Furthermore, $x_n^0 = A_n^{-1} y$ with noise-free data $y \in \mathcal{Y}$.

4.1. Noise models

We investigate additive noise models, i.e., $y^\delta = y + \delta\xi$, where ξ is a normalized noise element and $\delta > 0$ is the noise level. The most common noise model in the classical inverse problems literature is *deterministic noise* (cf. [33]), where $\xi \in \mathcal{Y}$

with $\|\xi\| \leq 1$, so $\|y^\delta - y\| \leq \delta$. This models discretization errors, but only poorly represents random measurement errors arising in practice.

A *stochastic noise* model for a discrete data vector $y^\delta \in \mathbb{R}^N$ (see [135]) uses ξ, where the components ξ_i are i.i.d. random variables with mean $\mathbb{E}\xi_i = 0$ and variance $\mathbb{E}\xi_i^2 = 1$. δ is the standard deviation of each error component $\delta\xi_i$ and $\mathbb{E}\|y^\delta - y\|^2 = \delta^2\mathbb{E}\|\xi\|^2 = N\delta^2$. Note that for the number of data $N \to \infty$ this is unbounded. For correlated errors, $\delta\xi_i$ possesses the covariance matrix C. If known, this matrix can be used in (20). We restrict ourselves to diagonal C here. If $y^\delta = y + \delta\xi$ with $\xi \sim N(0, I)$, the noise model corresponds to Gaussian white noise. For colored noise, if the entries C_{kk} are increasing, it is called blue noise, and, if they are decreasing, it is called red noise (see [9] and the references therein for more details).

4.2. Parameter choice methods

A parameter choice method is a rule that assigns a value for the regularization/smoothing parameter. In case of a discrete set of parameters, the method selects a value for the index, which will be denoted by n_*. Parameter choice methods can be classified as three basic types by their input (see, e.g., [8, 33]):

- a priori method, i.e., n_* is a function of δ and information about x which is not known in practice. Thus, we do not discuss such methods here.
- a posteriori method, i.e., $n_* = n_*(\delta, y^\delta)$ requiring the noise level δ. If δ is not known, an estimate is used.
- data-driven method, i.e., $n_* = n_*(y^\delta)$ requiring only the data y^δ as input (sometimes called "heuristic method").

If y^δ contains stochastic noise, then n_* is a random variable. n_* is defined as either the point at which a function $F(n)$ falls below a threshold (Type 1) or the minimizer of F (Type 2). Some methods need (sensitive) tuning parameters (in particular methods of type 1).

It should be pointed out that [4] states that, for an ill-posed problem, a parameter choice rule that does not explicitly use the noise level (e.g., data-driven methods) cannot yield a regularization method such that the worst case error converges to 0 as $\delta \to 0$. This Bakushinskii veto is important for deterministic noise, but it is not really appropriate for stochastic noise (cf. [8, 15]). There are data-driven rules yielding regularization methods that converge and perform very well in practice (see also [9]).

Since $x = A^{-1}y$ is unknown (noise-free data y is not available), a practical parameter choice method must use some other known or easily computed/estimated quantities such as, e.g., the norm of the residual defined as $\|y^\delta - Ax_n^\delta\|$. In our case this is a Euclidean norm which is easily computed. Splitting the error $\|x - x_n^\delta\|$ such that

$$\|x - x_n^\delta\| \leq \|x - x_n^0\| + \|x_n^0 - x_n^\delta\|, \tag{48}$$

the first term (regularization error) is usually bounded by a decreasing function $\varphi(n)$ reflecting smoothness properties and the so-called qualification of the regularization method (see [9, 33] and the references therein). The second term (propagated noise error) on the right-hand side of (48) can often be bounded for regularization methods as

$$\|x_n^0 - x_n^\delta\| \leq \delta\varrho(n), \tag{49}$$

where ϱ is a known increasing function of n, indicating that, with less smoothing, there is more influence of the data noise (cf. [33]).

In the case of stochastic noise, the risk, i.e., the expected squared error $\mathbb{E}\|x - x_n^\delta\|^2$, is considered. For noise with zero mean, instead of (48), the risk can be decomposed exactly into a sum of squared bias $\|x - x_n^0\|^2$ and variance terms $\mathbb{E}\|x_n^0 - x_n^\delta\|^2$, i.e.

$$\mathbb{E}\|x - x_n^\delta\|^2 = \|x - x_n^0\|^2 + \mathbb{E}\|x_n^0 - x_n^\delta\|^2. \tag{50}$$

The squared bias can be bounded as before and, under suitable assumptions, the variance can be expressed as $\delta^2\varrho^2(n)$ for some increasing function $\varrho(n)$. For white noise, our Tikhonov regularized solution (47) has variance

$$\delta^2\varrho^2(n) = \delta^2\mathbb{E}\|A_n^{-1}\xi\|^2 = \delta^2\sum_k \frac{1}{(\sigma_k + \beta_n)^2}, \tag{51}$$

where σ_k are the singular values of A (in our case they are eigenvalues of A). A much more detailed discussion of the above errors (including, e.g., minimax results) in various situations can be found in [5, 9, 23, 26, 33, 78, 95, 103, 105]. For some methods, there are stronger results involving oracle inequalities (see [12, 19, 21, 22]), which provide, for any noise level, a bound on the risk $\mathbb{E}\|x - x_n^\delta\|^2$ relative to the smallest possible value of the risk, and allow the classification of methods as asymptotically optimal.

4.3. Maximal regularization parameter

Some parameter choice methods, e.g., the balancing principle (Section 4.4.2), require a maximal index $n_{\max}$ as essential input in the algorithm itself. [9] suggest to define the *maximal index* as

$$n_{\max} = \max\{n|\varrho(n) < 0.5\varrho(\infty)\}, \tag{52}$$

where $\mathbb{E}\|x_n^0 - x_n^\delta\|^2 = \delta^2\varrho^2(n)$ and $\delta^2\varrho^2(\infty)$ is the supremum of the variance. This allows us to expect that the optimal parameter index is smaller than $n_{\max}$. Also note that $n_{\max}$ should not be too large for an efficient computation.

We obtain $n_{\max}$ either by an analytic expression for $\delta^2\varrho^2(n)$, as in (51) for white noise, or by a good estimate of it. If there are several independent data sets available, then a good estimate (for any noise color) is

$$\delta^2\varrho^2(n) \approx 2^{-1}\text{mean}\{\|x_{n,i}^\delta - x_{n,j}^\delta\|^2, i \neq j\}. \tag{53}$$

Often two sets of data are sufficient (see [5] for further details). If only a single data set is available, then it may not be possible to estimate $\delta^2\varrho^2(n)$ if the noise is correlated with unknown covariance. Then one can define a maximal index $\tilde{n}_{\max}$ by

$\beta_{\tilde{n}_{\max}} \approx \sigma_{\tilde{n}_{\max}}$ for our Tikhonov regularization if there are at least good estimates of the eigenvalues σ_k of A available. However, methods that perform much worse without the use of the maximal index $n_{\max}$, may yield different results for $\tilde{n}_{\max}$ instead of $n_{\max}$ (see [9, 14]).

4.4. Description and evaluation of methods

In this section, we describe the origin and idea of the method, list the input of the method and the algorithm that we use. We also give a brief discussion of known theoretical and practical issues about the method, including the most relevant references. Several of the methods use a tuning parameter or some other parameter that must be chosen (see [9, 14] and the references therein for further details).

4.4.1. Discrepancy principle. The discrepancy principle of [115, 110, 111] is one of the oldest and most widely used parameter choice procedures (cf. [33] and references therein). Its idea is that for a good regularized solution, the norm of the residual should match the noise level δ of the data. The method needs the following input:

- Norms of residuals $\{Ax_n^\delta - y^\delta\}_{n \le n_{\max}}$ until a certain bound is satisfied.
- Noise level δ.
- Tuning parameter $\tau \ge 1$.

In a deterministic setting with $\|y^\delta - y\| \le \delta$, the parameter choice n_* is the first n such that $\|Ax_n^\delta - y^\delta\| \le \tau\delta$. In a stochastic setting, with the error in each element of $y^\delta \in \mathbb{R}^N$ having standard deviation δ, the choice n_* is the first n such that

$$\|Ax_n^\delta - y^\delta\| \le \tau\delta\sqrt{N}. \tag{54}$$

Originating from a deterministic setting, the discrepancy principle has also been studied in stochastic settings (see, e.g., [17, 29, 97, 132]) and for many regularization methods and many inverse problems. There are many results on convergence properties of this method for both settings (see, e.g., [33, 65, 79, 104, 111, 112, 114] for the deterministic case and [29, 97, 98, 132] for the stochastic case).

The discrepancy principle is one of the fastest methods available, since one only needs to compute the residuals until the bound (54) is satisfied which allows the use of the FMM of Section 3. However, its drawback is the requirement of an accurate estimate of the noise level. Estimations that are just slightly off can lead to very poor solutions (see [73, Chap. 7]).

There are also many variants of the method such as the transformed discrepancy principle (cf. [119, 120, 71]), the modified discrepancy principle (MD rule) (cf. [32, 53, 117, 118]), or the varying discrepancy principle (cf. [17, 94]). Their main drawback is that they are no longer easily compatible with the FMM. For comparative studies in the context of inverse problems with stochastic noise we refer, e.g., to [9, 14].

4.4.2. Balancing principle. The balancing principle of [90] was originally derived for statistical estimation from direct observations in a white noise model. Since then it has been developed further for regularization of linear and nonlinear inverse problems (see, e.g., [6, 13, 11, 58, 129, 105, 106]) in deterministic and stochastic settings. The idea is to balance the known propagated noise error bound $\delta\varrho(n)$ in (49) with the unknown regularization error (48) by an adaptive procedure that employs a collection of differences of regularized solutions. As input the balancing principle needs:

- Maximal index $n_{\max}$, e.g., as defined in (52).
- All regularized solutions $\{x_n^\delta\}_{n\le n_{\max}}$ up to the index $n_{\max}$.
- An upper bound $\delta\varrho(n)$ for the propagated noise error $\|x_n^0 - x_n^\delta\|$ or a bound or estimate $\delta^2\varrho^2(n)$ of the variance $\mathbb{E}\|x_n^0 - x_n^\delta\|^2$.
- Noise level δ (and the covariance in the stochastic setting if known). Then one can use known expressions for $\delta\varrho(n)$. Alternatively, if one has two or more independent sets of data y_i^δ, then $\mathbb{E}\|x_n^0 - x_n^\delta\|^2$ can be estimated by (53).
- Tuning constant κ, typically $\kappa \in [0.5, 1.5]$ (cf. [9] and the references therein).

The balancing functional is defined by

$$b(n) = \max_{n<k\le n_{\max}} \left\{4^{-1}\|x_n - x_k\|/(\delta\varrho(k))\right\}. \tag{55}$$

The smoothed balancing functional (which is monotonously decreasing) is defined as $B(n) = \max_{n\le k\le n_{\max}}\{b(k)\}$. Then the parameter choice n_* is the first n such that $B(n) \le \kappa$.

The balancing principle is one of the few parameter choice methods for which oracle inequalities for the error are known (cf. [121, 13]), i.e., there are stronger results than rates of convergence alone. For variants we refer to [10, 105] and for comparative studies we mention [9, 14].

One variant should be noted: The hardened balancing principlewhich is a modified version of the balancing principle in the stochastic setting first proposed in [5]. The input is the same as before, but without the tuning parameter and the noise level. Furthermore, an expression or approximation of the scaled variance $\varrho^2(n) = \delta^{-2}\mathbb{E}\|x_n^0 - x_n^\delta\|^2$, or any scalar multiple of this (so δ can be unknown), is required. The balancing functional $b(n)$ is defined as in (55) and the smoothed balancing functional $B(n) = \max_{n\le k\le n_{\max}}\{b(k)\}$. The parameter choice is

$$n_* = \operatorname{argmin}_{n\le n_{\max}} \left\{B(n)\sqrt{\varrho(n)}\right\}, \tag{56}$$

where any scalar multiple of $\varrho(n)$ gives the same choice. The method has the advantage that it does not require a tuning parameter. Numerical experiments in [5, 12, 9, 14] indicate that the method is very stable even for colored noise.

4.4.3. Quasi-optimality criterion. The quasi-optimality criterion by Tikhonov and Arsenin [127], Tikhonov and Glasko [128] (see also [77]) is one of the oldest and simplest available parameter choice methods. An overview of the method and its

history can be found, e.g., in [8]. As input for the minimization the following is required:

- Maximal index $n_{\max}$, e.g., as defined in (52).
- All regularized solutions $\{x_n^\delta\}_{n\le n_{\max}}$ up to $n_{\max}$.

The noise level does not need to be known, and there is no tuning parameter. The parameter choice n_* is defined simply as

$$n_* = \operatorname{argmin}_{n\le n_{\max}} \left\{ \|x_n^\delta - x_{n+1}^\delta\| \right\}. \tag{57}$$

The well-known continuous version for Tikhonov regularization defines the parameter choice by $\beta_* = \operatorname{argmin} \left\| \beta \frac{d}{d\beta} x_\beta^\delta \right\|$. Using a difference quotient in place of the derivative for the discrete parameters $\beta_n = \beta_0 q_\beta^n$ we obtain (57).

For a discrete set of regularization parameters, the use of a suitable maximal index $n_{\max}$ is essential, because the method is based on a discrete evaluation of a differential. Hence is very sensitive to a situation where the regularization operators A_n^{-1} are formally different, but are practically the same (cf. [9] and the references therein). Convergence results for the Tikhonov regularization with the quasi-optimality criterion for different settings can be found in [89, 56], for further convergence properties see [7, 84, 85, 113].

4.4.4. L-curve method. The L-curve method, proposed by Hansen [72, 73] and Hansen and O'Leary [74], is based on the long-known fact that a log-log parametric plot of $(\|Ax_n^\delta - y^\delta\|, \|x_n^\delta\|)$ often has a distinct L-shape (cf. [88]). Points on the vertical part correspond to large n (under-smoothed solutions) and those on the horizontal part correspond to small n (over-smoothed solutions), which suggests that the "corner point" of the L-curve should define a good value of the parameter n. It is usually applied manually and can provide good results then whereas finding the L-curve corner is hard to automate. As input to minimize a certain function the following is used:

- Norms of all residuals $\{Ax_n^\delta - y^\delta\}_{n\le n_{\max}}$.
- Norms of the regularized solutions $\{x_n^\delta\}_{n\le n_{\max}}$.

The noise level does not need to be known. The parameter choice can be defined by the product of the norms of the residual and regularized solution, i.e.,

$$n_* = \operatorname{argmin}_{n\le n_{\max}} \left\{ \|Ax_n^\delta - y^\delta\| \cdot \|x_n^\delta\| \right\}. \tag{58}$$

Here the "corner point" is defined by the slope of its "tangent" being -1 as in [122] (see also [33]). The generalizations minimize $\|Ax_n^\delta - y^\delta\| \|x_n^\delta\|^\tau$ (see [122, 93]), where τ is a tuning constant.

Since "corner point" is not a well-defined notion, several algorithms have evolved with different definitions (see [74, 18, 75, 122]). [93] derived first rigorous optimality results for the L-curve criterion. In many (but not all) problems, variants of the L-curve method has been observed to give a reasonably good parameter choice which can deal with correlated errors. See [9] for an overview of references where the method works or runs into severe limitations.

4.4.5. Generalized cross-validation. Generalized cross-validation (GCV), due to [133], is a popular method for problems with discrete data and stochastic noise as (20). It goes back to ordinary cross-validation, where the idea is to consider all the "leave-one-out" regularized solutions and choose the parameter that minimizes the average of the squared prediction errors using each solution to predict the missing data value. These calculations do not require the computation of all the regularized solutions. Weighting the prediction errors, [27, 60, 133, 135] derived the GCV method, which has is invariant under orthogonal transformations of the data. Some other parameter choice methods proposed in the literature have been shown to be closely related to GCV, in particular the Akaike information criterion (AIC) of [2, 35]. As input to minimize a certain function we need:

- Sums of squares of all the residuals $\{Ax_n^\delta - y^\delta\}_{n\le n_{\max}}$ where $y^\delta \in \mathbb{R}^N$.
- The trace of the influence matrix AA_n^{-1} mapping y^δ to Ax_n^δ.

The noise level does not need to be known. The GCV parameter estimate is defined by

$$n_* = \operatorname{argmin}_{n\le n_{\max}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(N^{-1}\operatorname{tr}(I - AA_n^{-1}))^2} \right\}. \tag{59}$$

GCV is closely related to and behaves like the unbiased prediction risk method (also known as Mallows C_p or C_L; see [31, 92, 135]). It is asymptotically optimal with respect to the prediction risk as $N \to \infty$ for stochastic white noise and the Tikhonov regularization (cf. [66, 91, 96, 132, 136]). The GCV method has been used widely and performs very well for reasonably large data sets with uncorrelated errors (white noise). However, it is known (see, e.g., [31, 86, 87, 99, 101, 126, 135]) that for smaller data sets or correlated errors of red noise type, the method is rather unstable, often resulting in under-smoothing.

The term $\operatorname{tr}(AA_n^{-1})$ in the GCV function is a measure of the degrees of freedom in the regularized solution. For its fast computation making use of the FMM trace estimation methods are needed that use stochastic (Monte-Carlo) algorithms (cf. [54, 55, 59, 80, 81]).

In order to overcome the instability of GCV, several variants have evolved. The robust GCV (RGCV) method has been developed and investigated in [99, 100, 123]. It needs the same input as for GCV and additionally:

- The trace of the square of the influence matrix $(AA_n^{-1})^2$.
- A robustness parameter $\gamma \in (0, 1)$. Note that with $\gamma = 1$ the RGCV method is just GCV.

The RGCV parameter estimate is defined by minimizing a certain function:

$$n_* = \operatorname*{argmin}_{n\le n_{\max}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(N^{-1}\operatorname{tr}(I - AA_n^{-1}))^2} \left(\gamma + (1-\gamma)N^{-1}\operatorname{tr}((AA_n^{-1})^2)\right) \right\}. \tag{60}$$

The family of robust GCV methods developed in [100] also includes the strong robust GCV method, denoted $\mathrm{R_1GCV}$. As input one needs the same as for GCV as well as

- The trace of $A_n^{-1\,*}A_n^{-1}$.
- A robustness parameter $\gamma \in (0,1)$. Note that for $\gamma = 1$ the R_1GCV method is just GCV.

The R_1GCV parameter estimate is defined by minimizing a certain function:

$$n_* = \underset{n \le n_{\max}}{\operatorname{argmin}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(N^{-1}\operatorname{tr}(I - AA_n^{-1}))^2} \left(\gamma + (1-\gamma)N^{-1}\operatorname{tr}((A_n^{-1\,*}A_n^{-1})\right) \right\}. \quad (61)$$

The modified GCV method involves a simple modification of the GCV function that is designed to stabilize the method (cf. [28, 131]). The inputs are the same as for GCV plus:

- A stabilization parameter $c > 1$. For $c = 1$ the method reduces to GCV.

The noise level does not need to be known. The modified GCV estimate is defined by

$$n_* = \operatorname{argmin}_{n \le n_{\max}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(N^{-1}\operatorname{tr}(I - cAA_n^{-1}))^2} \right\}. \quad (62)$$

For comparative studies of these variants and further details we refer to [9, 14] and the references therein.

5. Conclusion

Using the FMM in an iterative algorithm like, e.g., conjugate gradients or GMRES is an efficient solution strategy that can treat interpolation problems and Dirichlet boundary value problems with many data points on regular surfaces (e.g., the actual topography of the Earth) (see [67, 68, 69]). It should be pointed out that this spline approach is not restricted to a global treatment, but also applies to regional domains (cf. [67, 68]). This can lead to a local improvement of the gravitational field in areas of particular interest. The approach can be extended to spline approximation (in particular for diagonal covariance matrices) as seen in Section 2.4 and the end of Section 3.4).

The smoothing parameter(s) plays a crucial role in this approach and must be chosen very carefully or a lot of information is lost to oversmoothing, in particular the high-frequent details of the signal. We have presented several parameter choice methods the can be used without losing the advantages of the FMM. Their performance for the regularization of inverse problems has been investigated in several studies with different solution techniques (see, e.g., [9, 14, 70] and the references therein). Tests of the combination of the FMM with these parameter choice methods (cf. [9, 14, 70] and the references therein) are an interesting challenge for the future. In particular the interaction with stopping criteria for iterative solvers needs further investigation. Note that often the solution (even using the FMM) requires much more computational effort than the evaluation of the parameter choice. It can be advisable to apply several parameter choice methods to find the best choice of the parameter.

For highly irregular distributions of data points, the spline approach reaches its limits due to large data gaps which result in severe ill-conditioning. Even smoothing splines cannot completely bridge this gap so far though further investigation is required. However, functional matching pursuit methods (RFMP or ROFMP) can result in better approximations (see [36, 70, 108] and the references therein), but so far these algorithms require high numerical costs. These methods are also iterative regularizations and the combination of stopping criteria and regularization parameters has been investigated for a class of ill-posed problems in [70].

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, Inc., New York, 1972.

[2] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor,* 1971), pages 267–281. Akadémiai Kiadó, Budapest, 1973.

[3] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[4] A.B. Bakushinskii. Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. *U.S.S.R. Comput. Math. Math. Phys.*, 24:181–182, 1984.

[5] F. Bauer. Some considerations concerning regularization and parameter choice algorithms. *Inverse Prob.*, 23(2):837–858, 2007.

[6] F. Bauer and T. Hohage. A Lepskij-type stopping rule for regularized Newton methods. *Inverse Prob.*, 21:1975–1991, 2005.

[7] F. Bauer and S. Kindermann. The quasi-optimality criterion for classical inverse problems. *Inverse Prob.*, 24:035002, 20 pp., 2008.

[8] F. Bauer and S. Kindermann. Recent results on the quasi-optimality principle. *J. Inverse Ill-Posed Prob.*, 17(1):5–18, 2009.

[9] F. Bauer and M.A. Lukas. Comparing parameter choice methods for regularization of ill-posed problems. *Math. Comput. Simul.*, 81(9):1795–1841, 2011.

[10] F. Bauer and A. Munk. Optimal regularization for ill-posed problems in metric spaces. *J. Inverse Ill-Posed Probl.*, 15(2):137–148, 2007.

[11] F. Bauer and S. Pereverzev. Regularization without preliminary knowledge of smoothness and error behavior. *European J. Appl. Math.*, 16(3):303–317, 2005.

[12] F. Bauer and M. Reiß. Regularization independent of the noise level: an analysis of quasi-optimality. *Inverse Prob.*, 24:055009, 16 pp., 2008.

[13] F. Bauer, T. Hohage, and A. Munk. Iteratively regularized Gauss–Newton method for nonlinear inverse problems with random noise. *SIAM J. Numer. Anal.*, 47(3): 1827–1846, 2009.

[14] F. Bauer, M. Gutting, and M.A. Lukas. Evaluation of parameter choice methods for regularization of ill-posed problems in geomathematics. In W. Freeden, M.Z.

Nashed, and T. Sonar, editors, *Handbook of Geomathematics. 2nd ed.*, pages 1713–1774. Springer, Heidelberg, 2015.

[15] S.M.A. Becker. Regularization of statistical inverse problems and the Bakushinkii veto. *Inverse Prob.*, 27:115010, 22 pp., 2011.

[16] L.C. Biedenharn and J.D. Louck. *Angular Momentum in Quantum Physics* (*Theory and Application*). Encyclopedia of Mathematics and its Applications. Addison-Wesley, Reading, Massachusetts, 1981.

[17] G. Blanchard and P. Mathé. Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. *Inverse Prob.*, 28:115011, 23 pp., 2012.

[18] D. Calvetti, P.C. Hansen, and L. Reichel. L-curve curvature bounds via Lanczos bidiagonalization. *Electron. Trans. Numer. Anal.*, 14:20–35, 2002.

[19] E.J. Candès. Modern statistical estimation via oracle inequalities. *Acta Numerica*, 15:257–325, 2006.

[20] J. Carrier, L. Greengard, and V. Rokhlin. A fast adaptive multipole algorithm for particle simulations. *SIAM J. Sci. Stat. Comput.*, 9(4):669–686, 1988.

[21] L. Cavalier. Nonparametric statistical inverse problems. *Inverse Prob.*, 24(3): 034004, 19 pp., 2008.

[22] L. Cavalier, G.K. Golubev, D. Picard, and A.B. Tsybakov. Oracle inequalities for inverse problems. *Ann. Statist.*, 30(3):843–874, 2002.

[23] L. Cavalier, Y. Golubev, O. Lepski, and A. Tsybakov. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Theory Probab. Appl.*, 48(3):426–446, 2004.

[24] H. Cheng, L. Greengard, and V. Rokhlin. A fast adaptive multipole algorithm in three dimensions. *J. Comput. Phys.*, 155:468–498, 1999.

[25] C.H. Choi, J. Ivanic, M.S. Gordon, and K. Ruedenberg. Rapid and staple determination of rotation matrices between spherical harmonics by direct recursion. *J. Chem. Phys.*, 111(19):8825–8831, 1999.

[26] D.D. Cox. Approximation of method of regularization estimators. *Ann. Statist.*, 16 (2):694–712, 1988.

[27] P. Craven and G. Wahba. Smoothing noisy data with spline functions. *Numer. Math.*, 31:377–403, 1979.

[28] D.J. Cummins, T.G. Filloon, and D. Nychka. Confidence intervals for nonparametric curve estimates: Toward more uniform pointwise coverage. *J. Amer. Statist. Assoc.*, 96(453):233–246, 2001.

[29] A.R. Davies and R.S. Anderssen. Improved estimates of statistical regularization parameters in Fourier differentiation and smoothing. *Numer. Math.*, 48:671–697, 1986.

[30] A.R. Edmonds. *Drehimpulse in der Quantenmechanik*. Bibliographisches Institut, Mannheim, 1964.

[31] B. Efron. Selection criteria for scatterplot smoothers. *Ann. Statist.*, 29(2):470–504, 2001.

[32] H.W. Engl and H. Gfrerer. A posteriori parameter choice for general regularization methods for solving linear ill-posed problems. *Appl. Numer. Math.*, 4(5):395–417, 1988.

[33] H.W. Engl, H. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.

[34] M.A. Epton and B. Dembart. Multipole translation theory for the three-dimensional Laplace and Helmholtz equations. *SIAM J. Sci. Comput.*, 16 (4): 865–897, 1995.

[35] R.L. Eubank. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York, 1988.

[36] D. Fischer and V. Michel. Sparse regularization of inverse gravimetry – case study: Spatial and temporal mass variation in South America. *Inverse Prob.*, 28:065012, 34 pp., 2012.

[37] W. Freeden. On approximation by harmonic splines. *Manuscripta Geod.*, 6:193–244, 1981a.

[38] W. Freeden. On spherical spline interpolation and approximation. *Math. Method. Appl. Sci.*, 3:551–575, 1981b.

[39] W. Freeden. Interpolation and best approximation by harmonic spline functions. *Boll. Geod. Sci. Aff.*, 1:105–120, 1982a.

[40] W. Freeden. On spline methods in geodetic approximation problems. *Math. Method. Appl. Sci.*, 4:382–396, 1982b.

[41] W. Freeden. Spherical spline interpolation: Basic theory and computational aspects. *J. Comput. Appl. Math.*, 11:367–375, 1984a.

[42] W. Freeden. Ein Konvergenzsatz in sphärischer Spline-Interpolation. *Z. f. Vermessungswes.* (*ZfV*), 109:569–576, 1984b.

[43] W. Freeden. A spline interpolation method for solving boundary value problems of potential theory from discretely given data. *Numer. Methods Partial Differential Equations*, 3:375–398, 1987a.

[44] W. Freeden. Harmonic splines for solving boundary value problems of potential theory. In J.C. Mason and M.G. Cox, editors, *Algorithms for Approximation*, pages 507–529. The Institute of Mathematics and its Applications, Conference Series, Vol. 10, Clarendon Press, Oxford, 1987b.

[45] W. Freeden. *Multiscale Modelling of Spaceborne Geodata*. B.G. Teubner, Stuttgart, Leipzig, 1999.

[46] W. Freeden and C. Gerhards. *Geomathematically Oriented Potential Theory*. Chapman & Hall/CRC, Boca Raton, 2013.

[47] W. Freeden and M. Gutting. *Special Functions of Mathematical (Geo-)Physics*. Birkhäuser, Basel, 2013.

[48] W. Freeden and V. Michel. *Multiscale Potential Theory* (*with Applications to Geoscience*). Birkhäuser, Boston, Basel, Berlin, 2004.

[49] W. Freeden and M. Schreiner. Special functions in mathematical geosciences: An attempt at a categorization. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics.* 1*st ed.*, pages 925–948. Springer, Heidelberg, 2010.

[50] W. Freeden, M. Schreiner, and R. Franke. A survey on spherical spline approximation. *Surv. Math. Ind.*, 7:29–85, 1997.

[51] W. Freeden, T. Gervens, and M. Schreiner. *Constructive Approximation on the Sphere* (*with Applications to Geomathematics*). Oxford Science Publications, Clarendon, 1998a.

[52] W. Freeden, O. Glockner, and M. Schreiner. Spherical panel clustering and its numerical aspects. *J. Geodesy*, 72:586–599, 1998b.

[53] H. Gfrerer. An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Math. Comput.*, 49(180):507–522, 1987.

[54] D. Girard. A fast "Monte-Carlo cross-validation" procedure for large least squares problems with noisy data. *Numer. Math.*, 56(1):1–23, 1989.

[55] D. Girard. Asymptotic comparison of (partial) cross-validation, gcv and randomized gcv in nonparametric regression. *Ann. Statist.*, 26(1):315–334, 1998.

[56] V. Glasko and Y. Kriksin. On the quasioptimality principle for linear ill-posed problems in Hilbert space. *Vychisl. Mat. Mat. Fiz.*, 24:1603–1613, 1984.

[57] O. Glockner. *On Numerical Aspects of Gravitational Field Modelling From SST and SGG by Harmonic Splines and Wavelets* (*with Application to CHAMP Data*). Ph.D.-thesis, Geomathematics Group, Department of Mathematics, University of Kaiserslautern. Shaker, Aachen, 2002.

[58] A. Goldenshluger and S. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Prob. Th. Rel. Fields*, 118(2): 169–186, 2000.

[59] G.H. Golub and U. von Matt. Generalized cross-validation for large scale problems. *J. Comput. Graph. Statist.*, 6(1):1–34, 1997.

[60] G.H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.

[61] L. Greengard. *The Rapid Evaluation of Potential Fields in Particle Systems*. MIT Press, Cambridge, MA, 1988.

[62] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *J. Comput. Phys.*, 73(1):325–348, 1987.

[63] L. Greengard and V. Rokhlin. Rapid evaluation of potential fields in three dimensions. In C. Anderson and L. Greengard, editors, *Vortex Methods*, pages 121–141. Springer, 1988.

[64] L. Greengard and V. Rokhlin. A new version of the fast multipole method for the Laplace equation in three dimensions. *Acta Numerica*, 6:229–269, 1997.

[65] C.W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman, Boston, 1984.

[66] C. Gu. *Smoothing Spline ANOVA Models*. Springer, New York, 2002.

[67] M. Gutting. *Fast Multipole Methods for Oblique Derivative Problems*. Ph.D.-thesis, Geomathematics Group, Department of Mathematics, University of Kaiserslautern. Shaker, Aachen, 2007.

[68] M. Gutting. Fast multipole accelerated solution of the oblique derivative boundary value problem. *GEM – Int. J. Geom.*, 3(2):223–252, 2012.

[69] M. Gutting. Fast spherical/harmonic spline modeling. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics. 2nd ed.*, pages 2711–2746. Springer, Heidelberg, 2015.

[70] M. Gutting, B. Kretz, V. Michel, and R. Telschow. Study on parameter choice methods for the RFMP with respect to downward continuation. *Front. Appl. Math. Stat.*, 3, doi:10.3389/fams.2017.00010, 17 pp., 2017.

[71] U. Hämarik and T. Raus. On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data. *J. Inverse Ill-Posed Probl.*, 14(3):251–266, 2006.

[72] P.C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, 34(4):561–580, 1992.

[73] P.C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems.* SIAM, Philadelphia, 1998.

[74] P.C. Hansen and D.P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, 14(6):1487–1503, 1993.

[75] P.C. Hansen, T.K. Jensen, and G. Rodriguez. An adaptive pruning algorithm for the discrete L-curve criterion. *J. Comput. Appl. Math.*, 198(2):483–492, 2007.

[76] E.W. Hobson. *The Theory of Spherical and Ellipsoidal Harmonics* (*Second Reprint*). Chelsea Publishing Company, New York, 1965.

[77] B. Hofmann. *Regularization of Applied Inverse and Ill-Posed Problems.* Teubner, Leipzig, 1986.

[78] B. Hofmann and P. Mathé. Analysis of profile functions for general linear regularization methods. *SIAM J. of Numer. Anal.*, 45(3):1122–1141, 2007.

[79] T. Hohage. Regularization of exponentially ill-posed problems. *Numer. Funct. Anal. Optim.*, 21:439–464, 2000.

[80] M. Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun. Stat. Simulation Comput.*, 18(3):1059–1076, 1989.

[81] M.F. Hutchinson and F.R. de Hoog. Smoothing noisy data with spline functions. *Numer. Math.*, 47:99–106, 1985.

[82] J. Keiner, S. Kunis, and D. Potts. Fast summation of radial functions on the sphere. *Computing*, 78:1–15, 2006.

[83] O.D. Kellogg. *Foundation of Potential Theory.* Springer, Berlin, Heidelberg, New York, 1967.

[84] S. Kindermann. Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems. *Electron. Trans. Numer. Anal.*, 38: 233–257, 2011.

[85] S. Kindermann and A. Neubauer. On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization. *Inverse Probl. Imaging*, 2(2):291–299, 2008.

[86] R. Kohn, C.F. Ansley, and D. Tharm. The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *J. Amer. Statist. Assoc.*, 86:1042–1050, 1991.

[87] S.C. Kou and B. Efron. Smoothers and the C_p, generalized maximum likelihood, and extended exponential criteria: a geometric approach. *J. Amer. Statist. Assoc.*, 97(459):766–782, 2002.

[88] C.L. Lawson and R.J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, Englewood Cliffs, N.J., 1974.

[89] A. Leonov. Justification of the choice of regularization parameter according to quasi-optimality and quotient criteria. *U.S.S.R. Comput. Math. Math. Phys.*, 18(6):1–15, 1979.

[90] O. Lepskij. On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, 35(3):454–466, 1990.

[91] K.-C. Li. Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing. *Ann. Statist.*, 14:1101–1112, 1986.

[92] K.-C. Li. Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Ann. Statist.*, 15:958–975, 1987.

[93] S. Lu and P. Mathé. Heuristic parameter selection based on functional minimization: optimality and model function approach. *Math. Comput.*, 82(283):1609–1630, 2013.

[94] S. Lu and P. Mathé. Discrepancy based model selection in statistical inverse problems. *J. Complexity*, 30(3):290–308, 2014.

[95] M.A. Lukas. Convergence rates for regularized solutions. *Math. Comput.*, 51(183): 107–131, 1988.

[96] M.A. Lukas. Asymptotic optimality of generalized cross-validation for choosing the regularization parameter. *Numer. Math.*, 66(1):41–66, 1993.

[97] M.A. Lukas. On the discrepancy principle and generalized maximum likelihood for regularization. *Bull. Aust. Math. Soc.*, 52(3):399–424, 1995.

[98] M.A. Lukas. Comparisons of parameter choice methods for regularization with discrete noisy data. *Inverse Prob.*, 14(1):161–184, 1998.

[99] M.A. Lukas. Robust generalized cross-validation for choosing the regularization parameter. *Inverse Prob.*, 22(5):1883–1902, 2006.

[100] M.A. Lukas. Strong robust generalized cross-validation for choosing the regularization parameter. *Inverse Prob.*, 24:034006, 16 pp., 2008.

[101] M.A. Lukas. Robust GCV choice of the regularization parameter for correlated data. *J. Integral Equations Appl.*, 22(3):519–547, 2010.

[102] W. Magnus, F. Oberhettinger, and R.P. Soni. *Formulas and Theorems for the Special Functions of Mathematical Physics*, volume 52 of *Die Grundlehren der mathematischen Wissenschaften*. Springer, New York, 3rd edition, 1966.

[103] B.A. Mair and F.H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56(5):1424–1444, 1996.

[104] P. Mathé. What do we learn from the discrepancy principle? *Z. Anal. Anwend.*, 25 (4):411–420, 2006.

[105] P. Mathé and S.V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert spaces. *Inverse Prob.*, 19(3):789–803, 2003.

[106] P. Mathé and S.V. Pereverzev. Regularization of some linear ill-posed problems with discretized random noisy data. *Math. Comput.*, 75(256):1913–1929, 2006.

[107] V. Michel. *Lectures on Constructive Approximation – Fourier, Spline, and Wavelet Methods on the Real Line, the Sphere, and the Ball*. Birkhäuser, Boston, 2013.

[108] V. Michel. RFMP – an iterative best basis algorithm for inverse problems in the geosciences. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*, 2nd ed., pages 2121–2147. Springer, Heidelberg, 2015.

[109] H. Moritz. Classical physical geodesy. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics. 1st ed.*, pages 127–158. Springer, Heidelberg, 2010.

[110] V.A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414–417, 1966.

[111] V.A. Morozov. *Methods for Solving Incorrectly Posed Problems*. Springer-Verlag, New York, 1984.

[112] M.T. Nair, E. Schock, and U. Tautenhahn. Morozov's discrepancy principle under general source conditions. *Z. Anal. Anwend.*, 22:199–214, 2003.

[113] A. Neubauer. The convergence of a new heuristic parameter selection criterion for general regularization methods. *Inverse Prob.*, 24(5):055005, 10 pp., 2008.

[114] S. Pereverzev and E. Schock. Morozov's discrepancy principle for tikhonov regularization of severely ill-posed problems in finite dimensional subspaces. *Numer. Funct. Anal. Optim.*, 21:901–916, 2000.

[115] D. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9:84–97, 1962.

[116] D. Potts and G. Steidl. Fast summation at nonequispaced knots by NFFTs. *SIAM J. Sci. Comput.*, 24(6):2013–2037, 2003.

[117] T. Raus. On the discrepancy principle for the solution of ill-posed problems. *Uch. Zap. Tartu. Gos. Univ.*, 672:16–26, 1984.

[118] T. Raus. The principle of the residual in the solution of ill-posed problems with nonselfadjoint operator. *Tartu Riikl. Ül. Toimetised*, 715:12–20, 1985.

[119] T. Raus. An a posteriori choice of the regularization parameter in case of approximately given error bound of data. In Pedas, A., editor, *Collocation and Projection Methods for Integral Equations and Boundary Value Problems*, pages 73–87. Tartu University, Tartu, 1990.

[120] T. Raus. About regularization parameter choice in case of approximately given error bounds of data. In G. Vainikko, editor, *Methods for solution of integral equations and ill-posed problems*, pages 77–89. Tartu University, Tartu, 1992.

[121] T. Raus and U. Hämarik. On the quasioptimal regularization parameter choices for solving ill-posed problems. *J. Inverse Ill-Posed Probl.*, 15(4):419–439, 2007.

[122] T. Reginska. A regularization parameter in discrete ill-posed problems. *SIAM J. Sci. Comput.*, 17(3):740–749, 1996.

[123] T. Robinson and R. Moyeed. Making robust the cross-validatory choice of smoothing parameter in spline smoothing regression. *Commun. Stat. Theory Methods*, 18 (2):523–539, 1989.

[124] V. Rokhlin. Rapid solution of integral equations of classical potential theory. *J. Comput. Phys.*, 60:187–207, 1985.

[125] L. Shure, R.L. Parker, and G.E. Backus. Harmonic splines for geomagnetic modelling. *Phys. Earth Planet. Inter.*, 28:215–229, 1982.

[126] A.M. Thompson, J.W. Kay, and D.M. Titterington. A cautionary note about cross-validatory choice. *J. Statist. Comput. Simul.*, 33:199–216, 1989.

[127] A. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems*. Wiley, New York, 1977.

[128] A. Tikhonov and V. Glasko. Use of the regularization method in non-linear problems. *U.S.S.R. Comput. Math. Math. Phys.*, 5(3):93–107, 1965.

[129] A. Tsybakov. On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci., Paris, Sér. I, Math.*, 330(9):835–840, 2000.

[130] D.A. Varšalovič, A.N. Moskalev, and V.K. Chersonskij. *Quantum Theory of Angular Momentum*. World Scientific, Singapore, 1988.

[131] R. Vio, P. Ma, W. Zhong, J. Nagy, L. Tenorio, and W. Wamsteker. Estimation of regularization parameters in multiple-image deblurring. *Astron. Astrophys.*, 423: 1179–1186, 2004.

[132] C.R. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, 2002.

[133] G. Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14(4):651–667, 1977.

[134] G. Wahba. Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Stat. Comput.*, 2:5–16. Also errata: SIAM J. Sci. Stat. Comput., 3, 385–386, 1981.

[135] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.

[136] G. Wahba and Y.H. Wang. When is the optimal regularization parameter insensitive to the choice of the loss function? *Comm. Statist. Theory Methods*, 19(5):1685–1700, 1990.

[137] C.A. White and M. Head-Gordon. Rotating around the quartic angular momentum barrier in fast multipole method calculations. *J. Chem. Phys.*, 105(12):5061–5067, 1996.

[138] H. Yamabe. On an extension of the Helly's theorem. *Osaka Math. J.*, 2(1):15–17, 1950.

[139] N. Yarvin and V. Rokhlin. Generalized Gaussian quadratures and singular value decomposition of integral equations. *SIAM J. Sci. Comput.*, 20(2):699–718, 1998.

Martin Gutting
Geomathematics Group
University of Siegen
Walter-Flex-Straße 3
D-57068 Siegen, Germany

Handbook of Mathematical Geodesy
Geosystems Mathematics, 641–685

Inverse Gravimetry as an Ill-Posed Problem in Mathematical Geodesy

Willi Freeden and M. Zuhair Nashed

Abstract. The gravitational potential of (a part of) the Earth is assumed to be available, the inverse gravimetry problem is to determine the density contrast function inside (the specified part of) the Earth from known potential values. This paper deals with the characteristic ill-posed features of transferring input gravitational information in the form of Newtonian volume integral values to geological output characteristics of the density contrast function. Some properties of the Newton volume integral are recapitulated. Different methodologies of the resolution of the inverse gravimetry problem and their numerical implementations are examined dependent on the data source. Three cases of input information may be distinguished, namely internal (borehole), terrestrial (surface), and/or external (spaceborne) gravitational data sets. Singular integral theory based inversion of the Newtonian integral equation such as Haar-type solutions are proposed in a multiscale framework to decorrelate specific geological signal signatures with respect to inherently available features. Reproducing kernel Hilbert space regularization techniques are studied (together with their transition to mollified variants) to provide geological contrast density distributions by "downward continuation" from terrestrial and/or spaceborne data. Finally, reproducing kernel Hilbert space solutions are formulated for use of gravimeter data, independent of a specifically chosen input area, i.e., in whole Euclidean space $\mathbb{R}^3$.

Keywords. Newton potential, Poisson differential equation, density modeling, inverse gravimetry, multiscale decorrelation, multiscale spline modeling.

1. Introduction

Gravimetry is a central research area of geodesy, geophysics, and geoexploration. It is a potential field technique which reflects variations in the Earth's gravitational field. These variations are caused by density contrasts inside the Earth. Gravimetric surveys are carried out by use of extremely sensitive instruments capable of measuring tiny variations in the gravitational field. A gravimeter is a type of an

accelerometer. There are essentially two types of gravimeters, namely relative and absolute gravimeters. Absolute gravimeters measure the local gravity in absolute units. They work by directly measuring the acceleration of a mass during free fall in a vacuum. A new measurement technique is provided by atomic gravimeters. Relative gravimeters compare the value of gravity at one point with another. Most common relative gravimeters are spring-based. A spring-based relative gravimeter is basically a weight on a spring, and by measuring the amount by which the weight stretches the spring, gravitation becomes available via Hooke's Law in linearized form. On global scale gravimetric datasets are used in gravity surveys for establishing the figure of the geoid. Locally micro-gravimeters are in use, e.g., for geodetic and geophysical research, geothermal exploration, petroleum and mineral recovery.

In applied mathematics, inverse gravimetry (in its conventional form) may be reduced to the following situation: The *Newtonian potential* of a density contrast function F is defined as an improper integral over a volume $\mathcal{G} \subset \mathbb{R}^3$, namely

$$A[F](x) = \int_{\mathcal{G}} G(\Delta; |x-y|)\ F(y)\ dy = V(x),\ x \in \mathbb{R}^3, \tag{1.1}$$

formed by convolution of F with a kernel function $G(\Delta; \cdot)$ having a singularity, namely the *Newtonian kernel*

$$G(\Delta; |x-y|) = \frac{1}{4\pi} \frac{1}{|x-y|}, \quad x \in \mathbb{R}^3 \backslash \{y\}. \tag{1.2}$$

which apart of a minus sign represents the *fundamental solution of the Laplace equation*, i.e.,

$$\Delta_x G(\Delta; |x-y|) = 0, x \neq y$$

(or in distributional jargon, $-\Delta_x G(\Delta; |x-y|) = \delta(|x-y|)$, where $\delta(\cdot)$ is the Dirac distribution). In the region $\mathcal{G}$ the volume integral constitutes an operator A that acts as the inverse to the negative Laplace operator,

$$-\Delta_x V(x)\ = -\Delta_x\ A[F](x)\ =\ -\ \Delta_x \int_{\mathcal{G}} G(\Delta; |x-y|)\ F(y)\ dy = F(x),\ x \in \mathcal{G}, \tag{1.3}$$

which is to say (at least if the function F is Hölder continuous) that the operation of taking the Newtonian potential of a function is an inverse operation to the application of the negative Laplace operator.

Note that the integral (1.1) is named for I. Newton (1642–1720), who first discovered it and later pioneered the work of P.-S. Laplace (1749–1829) about harmonic functions. Indeed, the setting (1.1) serves as the fundamental gravitational potential in Newton's Law of Gravitation (1687).

Seen from potential theoretic perspective (see, e.g., [29]), Equation (1.1) is in close relation to the Third Green Theorem

$$\alpha(x)\ P(x) = -\int_{\mathcal{G}} G(\Delta; |x-y|)\ \Delta P(y)\ dy \tag{1.4}$$
$$+\int_{\partial\mathcal{G}} \left(G(\Delta; |x-y|)\frac{\partial}{\partial\nu(y)}P(y) - P(y)\frac{\partial}{\partial\nu(y)}G(\Delta; |x-y|)\right) d\omega(y),$$

that holds true for all twice continuously differentiable functions P on $\overline{\mathcal{G}}$, where $\alpha(x)$ is the solid angle subtended by the surface $\partial\mathcal{G}$ at the point $x \in \mathbb{R}^3$. It should be mentioned that

$$\alpha(x) = \begin{cases} 1, & x \in \mathcal{G}, \\ \dfrac{1}{2}, & x \in \partial\mathcal{G}, \\ 0, & x \in \mathcal{G}^c, \end{cases}$$

if the boundary surface $\partial\mathcal{G}$ is (locally) continuously differentiable. As an immediate consequence we may expect that the discussion of $A[F](x)$, $x \in \mathbb{R}^3$, actually has to be split into three cases, dependent on the location of $x \in \mathbb{R}^3$ as a point of the inner space $\mathcal{G}$, outer space $\mathcal{G}^c$, or the boundary $\partial\mathcal{G}$, i.e., the internal, surface (terrestrial), and external (spaceborne) input data $V(x)$. Moreover, a solution of (1.1) in $\mathcal{G}$ is not unique, since the addition of any harmonic function to V will not affect the equation (1.3). In potential theory this observation for the inner space $\mathcal{G}$ can be used to prove existence and uniqueness of solutions to the Dirichlet problem for the Poisson equation in suitably regular domains and for suitably well-behaved functions: One first applies a Newtonian potential to obtain a solution, and then adjusts by adding a harmonic function to get the correct boundary data.

Our intent in this paper is different from solving a boundary-value problem: We are interested in the inverse gravimetry problem, i.e., the extraction of information from the gravitational potential V known for certain locations to determine the density contrast F inside $\mathcal{G}$. In the language of functional analysis, we have to solve a Fredholm integral equation of the first kind (1.1) that canonically leads to the framework of the theory of ill-posed problems (as presented by the authors in another chapter of this handbook). The main difficulty, however, is that the input data of the inverse problem are not canonically given in the inner space $\mathcal{G}$, but usually in $\overline{\mathcal{G}^c}$. As a matter of fact, until now in physical geodesy, only measurements are taken on the surface $\partial\mathcal{G}$ (terrestrial measurements) and/or in the outer space $\mathcal{G}$ (spaceborne measurements), i.e., in the set $\overline{\mathcal{G}^c}$. Only in exceptional cases, e.g., in the neighborhood of "boreholes" of geothermal projects, the gravitational potential V and the target function F are given inside $\mathcal{G}$, so that the use of the Poisson differential equation (1.3) becomes applicable in the inversion process.

Typically, for inverse problems, there will also be certain *physical constraints* which will be necessary to impose on the potential pattern so that the wanted geological pattern of the density distribution can be approximated in some acceptable manner. Such constraints are usually referred to as conditions reflecting

realizability conditions. They will be represented in our mathematical framework by requiring the density functions to lie in some appropriate subset of the output space. Under certain conditions these realizability constraints will serve to regularize the originally ill-posed problem, while in others, they will dictate compromises that will have to be made between requirements for accuracy of the approximating functions and the demands of meeting such *a priori* constraints. In this contribution we are essentially interested in regularization procedures based on mollifier techniques. Different types of mollifiers will be studied in more detail, namely Haar-type and singular integral-type mollifiers.

Not only in physical geodesy, but also in inverse problem theory, there is a huge literature about the character and role of gravimetry in the framework of ill-posed and inverse problems, from which we only mention a limited selection: [3, 9–11, 29, 34, 59, 71–75, 90, 93, 99, 100, 107, 114] (for further details the reader is referred to the references therein). Our paper, however, follows a different approach, whose point of departure is the introductory chapter [27] of the "Handbook of Geomathematics".

2. Newton volume integral

In order to handle the inverse gravimetry problem some potential-theoretic preliminaries are needed: Let $\mathcal{G}$ be a *regular region* in $\mathbb{R}^3$, i.e., a bounded region $\mathcal{G}$ dividing $\mathbb{R}^3$ uniquely into the inner space $\mathcal{G}$ and the outer space $\mathcal{G}^c = \mathbb{R}^3 \backslash \overline{\mathcal{G}}$, $\overline{\mathcal{G}} = \mathcal{G} \cup \partial\mathcal{G}$, such that the boundary $\partial\mathcal{G}$ is an orientable Lipschitzian manifold of dimension 2 (for example, ball, ellipsoid, cube and other polyhedra, spheroid, telluroid, geoid, (actual) Earth or appropriate parts of it).

A real-valued function P is called *harmonic in* $\mathcal{G} \subset \mathbb{R}^3$ if P is of class $C^{(2)}(\mathcal{G})$ of functions with continuous second-order partial derivatives and satisfies the Laplace equation

$$\Delta P(x) = \left(\left(\frac{\partial}{\partial x_1} \right)^2 + \left(\frac{\partial}{\partial x_2} \right)^2 + \left(\frac{\partial}{\partial x_3} \right)^2 \right) P(x_1, x_2, x_3) = 0, \quad x = (x_1, x_2, x_3)^T, \tag{2.1}$$

for all $x \in \mathcal{G}$.

2.1. Basics of potential theory

Some important examples of harmonic functions are given below in the classical nomenclature of potential theory (see [29]).

(a) *Potential of a mass point:* According to *Newton's Law of Gravitation* two points x, y with masses M_x, M_y attract each other with a force given by

$$-\frac{\gamma}{4\pi} \frac{M_x M_y}{|x-y|^3}(x-y), \quad x, y \in \mathbb{R}^3,\ x \neq y. \tag{2.2}$$

The force is directed along the line connecting the two points x, y . The constant γ denotes Newton's gravitational constant (note that γ can be assumed to be equal to one in the theoretical part, but not in numerical applications).

Although the masses M_x, M_y attract each other in symmetric way, it is convenient to call one of them the *attracting mass* and the other one the *attracted mass*. Conventionally the attracted mass is set equal to unity and the attracting mass is denoted by M:

$$v(x) = -\frac{\gamma}{4\pi}\frac{M}{|x-y|^3}(x-y), \quad x \in \mathbb{R}^3\backslash\{y\}. \tag{2.3}$$

The formula (2.3) expresses the force exerted by the mass M on a unit mass located at the distance $|x-y|$ from M. Obviously, the intensity $|v(x)|$ of the force $v(x)$ is given by

$$|v(x)| = \frac{\gamma}{4\pi}\frac{M}{|x-y|^2}, \quad x \in \mathbb{R}^3\backslash\{y\}. \tag{2.4}$$

The scalar function V defined by

$$V(x) = \gamma\ M\ G(\Delta; |x-y|) = \gamma\ M\ \frac{1}{4\pi}\frac{1}{|x-y|}, \quad x \in \mathbb{R}^3\backslash\{y\} \tag{2.5}$$

is called the *potential of gravitation* at y. The force vector $v(x)$ is the gradient vector of the scalar $V(x)$:

$$v(x) = \nabla V(x), \quad x \in \mathbb{R}^3\backslash\{y\}. \tag{2.6}$$

Calculating the divergence $\nabla\cdot$ of the gradient field v, it readily follows that

$$\nabla \cdot v(x) = \nabla \cdot \nabla\ V(x) = \Delta V(x) = 0, \quad x \in \mathbb{R}^3\backslash\{y\}. \tag{2.7}$$

(b) *Potential of a finite mass point system:* The potential for N points x_i with masses M_i, $i = 1, \ldots, N$, is the sum of the individual contributions

$$V(x) = \gamma\ \sum_{i=1}^{N} M_i\ G(\Delta; |x-y_i|), \quad x \in \mathbb{R}^3\backslash\{y_1, \ldots, y_n\}. \tag{2.8}$$

Clearly we have

$$\Delta V(x) = 0, \quad x \in \mathbb{R}^3\backslash\{y_1, \ldots, y_N\}. \tag{2.9}$$

(c) *Potential of a volume:* Let $\mathcal{G} \subset \mathbb{R}^3$ be a *regular region*. The point masses are distributed continuously over $\overline{\mathcal{G}} \subset \mathbb{R}^3$ with density F. Then the discrete sum (2.8) becomes a continuous sum, i.e., an integral over the body $\mathcal{G}$:

$$V(x) = \gamma\ \int_{\mathcal{G}} G(\Delta; |x-y|)F(y)\ dy. \tag{2.10}$$

Obviously,

$$\Delta V(x) = 0, \quad x \in \mathbb{R}^3\backslash\overline{\mathcal{G}}. \tag{2.11}$$

Note that V is defined on the whole space $\mathbb{R}^3$, however, $\Delta V(x)$ may not be obtained easily by interchanging the Laplace operator and the integral over $\mathcal{G}$ for all points x inside $\mathcal{G}$. At infinity the potential behaves like

$$|V(x)| = O\left(\frac{1}{|x|}\right), \qquad |x| \to \infty, \tag{2.12}$$

uniformly with respect to all directions (note that $|y| \leq \frac{|x|}{2}$ implies $|x-y| \geq ||x|-|y|| \geq \frac{1}{2}|x|$), i.e., V is regular at infinity.

2.2. Properties of the Newton integral

The Newton (volume) integral (2.10) over a regular region $\mathcal{G}$ corresponding to a mass density distribution F satisfies the Laplace equation in the outer space $\mathcal{G}^c = \mathbb{R}^3 \backslash \overline{\mathcal{G}}$. Clearly, this property is an immediate consequence of the harmonicity of the fundamental solution for the Laplace operator (in what follows we restrict ourselves to a Newton integral (2.10) with γ chosen equal to 1).

Harmonicity in $\mathcal{G}^c$. Let $F : \overline{\mathcal{G}} \to \mathbb{R}$ be an integrable bounded function. Then

$$V(x) = \int_{\mathcal{G}} G(\Delta; |x-y|) F(y)\, dy, \quad x \in \mathcal{G}^c \tag{2.13}$$

satisfies

$$\Delta_x \int_{\mathcal{G}} G(\Delta; |x-y|) F(y)\, dy = 0 \tag{2.14}$$

for all $x \in \mathcal{G}^c$, i.e., V *is harmonic in* $\mathcal{G}^c$.

Properties in $\mathcal{G}$. By one-dimensional Taylor linearization (cf. [27, 38]) we obtain

$$\frac{1}{\sqrt{u}} = \frac{1}{\sqrt{u_0}} - \frac{1}{2}\frac{1}{u_0^{\frac{3}{2}}}(u-u_0) + \frac{3}{8}\frac{1}{(u_0+\theta(u-u_0))^{\frac{5}{2}}}(u-u_0)^2 \tag{2.15}$$

for some $\theta \in (0,1)$. Setting $u = r^2$ and $u_0 = \rho^2$ we therefore find

$$\frac{1}{r} = \frac{1}{2\rho}\left(3 - \frac{r^2}{\rho^2}\right) + \frac{3}{8}\frac{1}{(\rho^2+\theta(r^2-\rho^2))^{\frac{5}{2}}}(r^2-\rho^2)^2. \tag{2.16}$$

In other words, by letting $r = |x-y|$ we are able to give a simple example for the "mollification" of the fundamental solution of the Laplace equation

$$G(\Delta; r) = \frac{1}{4\pi r}, \qquad r > 0, \tag{2.17}$$

by

$$G_\rho^H(\Delta; r) = \begin{cases} \dfrac{1}{8\pi\rho}\left(3 - \dfrac{1}{\rho^2}r^2\right), & r \leq \rho \\ \dfrac{1}{4\pi r}, & r > \rho. \end{cases} \tag{2.18}$$

such that $G_\rho^H(\Delta;\cdot)$ is continuously differentiable for all $r \geq 0$. Obviously, $G(\Delta;r) = G_\rho^H(\Delta;r)$ for all $r > \rho$. As a consequence,

$$G(\Delta;|x-y|) = \frac{1}{4\pi}\frac{1}{|x-y|}, \qquad |x-y| \neq 0, \tag{2.19}$$

admits a "mollification" of the form

$$G_\rho^H(\Delta;|x-y|) = \begin{cases} \dfrac{1}{8\pi\rho}\left(3 - \dfrac{1}{\rho^2}|x-y|^2\right), & |x-y| \leq \rho \\ \dfrac{1}{4\pi|x-y|}, & \rho < |x-y|. \end{cases} \tag{2.20}$$

Let $F : \overline{\mathcal{G}} \to \mathbb{R}$ be of class $\mathrm{C}^{(0)}(\overline{\mathcal{G}})$. We set

$$V(x) = \int_{\mathcal{G}} G(\Delta;|x-y|)F(y)\, dy \tag{2.21}$$

and

$$V_\rho^H(x) = \int_{\mathcal{G}} G_\rho^H(\Delta;|x-y|)F(y)\, dy. \tag{2.22}$$

The integrands of V and V^ρ differ only in the ball $\mathcal{B}_\rho(x) = \{y \in \mathbb{R}^3 : |x-y| < \rho\}$ around the point x with radius ρ. Because of its continuity the function $F : \overline{\mathcal{G}} \to \mathbb{R}$ is uniformly bounded on $\overline{\mathcal{G}}$. This fact shows that

$$\begin{aligned} \left|V(x) - V_\rho^H(x)\right| &= O\left(\int_{\mathcal{B}_\rho(x)} |G(\Delta;|x-y|) - G_\rho^H(\Delta;|x-y|)|\, dy\right) \\ &= O(\rho^2), \quad \rho \to 0. \end{aligned} \tag{2.23}$$

Therefore, V is of class $\mathrm{C}^{(0)}(\overline{\mathcal{G}})$ as the limit of a uniformly convergent sequence of continuous functions on $\overline{\mathcal{G}}$. Furthermore, we let

$$v(x) = \int_{\mathcal{G}} \nabla_x G(\Delta;|x-y|)F(y)\, dy \tag{2.24}$$

and

$$v_\rho^H(x) = \int_{\mathcal{G}} \nabla_x G_\rho^H(\Delta;|x-y|)F(y)\, dy. \tag{2.25}$$

Because of the fact

$$|\nabla_x G(\Delta;|x-y|)| = O(|x-y|^{-2}), \tag{2.26}$$

the integrals v and v_ρ^H exist for all $x \in \overline{\mathcal{G}}$. It is not hard to see that

$$\sup_{x\in\overline{\mathcal{G}}} \left|v(x) - v_\rho^H(x)\right| = \sup_{x\in\overline{\mathcal{G}}} |\nabla_x V(x) - \nabla_x V_\rho^H(x)| = O(\rho), \quad \rho \to 0. \tag{2.27}$$

Consequently, v is a continuous vector field on $\overline{\mathcal{G}}$. Moreover, as the relation (2.27) holds uniformly on $\overline{\mathcal{G}}$, we obtain

$$v(x) = \nabla V(x) = \int_{\mathcal{G}} \nabla_x G(\Delta;|x-y|)F(y)\, dy. \tag{2.28}$$

Altogether, we are allowed to formulate the following properties:

Let $\mathcal{G}$ be a regular region. Let $F : \overline{\mathcal{G}} \to \mathbb{R}$ be of class $C^{(0)}(\overline{\mathcal{G}})$. Then V, V_ρ^H as defined by (2.21), (2.22), *respectively, are of class $C^{(1)}(\overline{\mathcal{G}})$, such that*

$$\lim_{\rho \to 0} \sup_{x \in \overline{\mathcal{G}}} |V(x) - V_\rho^H(x)| = 0. \tag{2.29}$$

Furthermore, ∇V is of class $C^{(0)}(\overline{\mathcal{G}})$, such that

$$\nabla_x V(x) = \int_{\mathcal{G}} F(y)\ \nabla_x G(\Delta; |x-y|)\ dy, \qquad x \in \overline{\mathcal{G}}. \tag{2.30}$$

and

$$\lim_{\rho \to 0} \sup_{x \in \overline{\mathcal{G}}} |\nabla_x V(x) - \nabla_x V_\rho^H(x)| = 0. \tag{2.31}$$

2.3. Poisson's differential equation

We come to the *Poisson differential equation* under the assumption of μ-Hölder continuity, $\mu \in (0, 1]$, imposed on the function F on $\overline{\mathcal{G}}$. For that purpose we note that the Taylor linearization of $s^{-\frac{3}{2}}$ is given by $s_0^{\frac{3}{2}} - \frac{3}{2} s_0^{-\frac{5}{2}}(s - s_0)$. Hence, by letting $s = r^2$ and $s_0 = \rho^2$, we are able to replace r^{-3} by $\frac{1}{2\rho^3}(5 - \frac{3}{\rho^2} r^2)$. As a consequence, as "mollification" of

$$Z\,(\Delta; |x-y|) = \frac{1}{4\pi |x-y|^3}, \qquad |x-y| \neq 0, \tag{2.32}$$

we are able to introduce

$$Z_\rho(\Delta; |x-y|) = \begin{cases} \dfrac{1}{8\pi\rho^3} \left(5 - \dfrac{3}{\rho^2} |x-y|^2 \right), & |x-y| \leq \rho \\ \dfrac{1}{4\pi |x-y|^3}, & \rho < |x-y|. \end{cases} \tag{2.33}$$

The function $r \mapsto Z_\rho(\Delta; r), r \geq 0$, is continuously differentiable. Moreover, by the same arguments as above, it can be shown that the vector field

$$z_\rho(x) = -\int_{\mathcal{G}} Z_\rho(\Delta; |x-y|)(x-y) F(y)\ dy \tag{2.34}$$

converges uniformly on $\overline{\mathcal{G}}$ with $\rho \to 0$ to the limit field

$$v(x) = \nabla V_x(x) = \int_{\mathcal{G}} \nabla_x G(\Delta; |x-y|) F(y)\ dy. \tag{2.35}$$

For all $x \in \mathcal{B}_\rho(x)$ we obtain by a simple calculation

$$\nabla_x \cdot (\ Z_\rho(\Delta; |x-y|)(x-y)) = \frac{15}{8\pi} \left(\frac{1}{\rho^3} - \frac{|x-y|^2}{\rho^5} \right). \tag{2.36}$$

Furthermore, an easy calculation shows that

$$\int_{\mathcal{B}_\rho(x)} \nabla_x \cdot (Z_\rho(\Delta; |x-y|)\ (x-y))\ dy = 1. \tag{2.37}$$

Hence, under the additional assumption of μ-Hölder continuity, $\mu \in (0,1]$, for the function F on $\overline{\mathcal{G}}$, i.e., $|F(x) - F(y)| \leq C\ |x-y|^{\mu}$ for all $x, y \in \overline{\mathcal{G}}$, we obtain

$$\begin{aligned}\nabla_x \cdot z_\rho(x) &= -\ \nabla_x \cdot \int_{\mathcal{G}} Z_\rho(\Delta; |x-y|)(x-y)F(y)\ dy \qquad (2.38)\\ &= -\int_{\mathcal{B}_\rho(x)} \nabla_x \cdot (Z_\rho(\Delta; |x-y|)(x-y))\, F(y)\ dy\\ &= -\ \alpha(x)\ F(x) + \int_{\mathcal{B}_\rho(x)} (F(x) - F(y))\ \nabla_x \cdot (Z_\rho(\Delta|x-y|)(x-y))\ dy.\end{aligned}$$

Thus, the μ-Hölder continuity of F guarantees the estimate

$$\sup_{x \in \overline{\mathcal{G}}} |\nabla_x \cdot z_\rho(x)\ +\ \alpha(x)\ F(x)| = O(\rho^{\mu}), \quad \rho \to 0, \qquad (2.39)$$

uniformly with respect to $x \in \overline{\mathcal{G}}$, where $\alpha(x)$ is the solid angle at x subtended by the surface $\partial\mathcal{G}$. In an analogous way, we are able to show that the first partial derivatives of (2.34) converge uniformly to continuous limit fields. Again, the uniform convergence shows that ∇V is differentiable in $\overline{\mathcal{G}}$, and we have

$$\nabla_x \cdot v(x)\ = \Delta V(x)\ =\ \Delta_x \int_{\mathcal{G}} G(\Delta; |x-y|)F(y)\ dy = -\alpha(x)\ F(x), \quad x \in \overline{\mathcal{G}}. \qquad (2.40)$$

It should be noted that the assumption of μ-Hölder continuity of F, $\mu \in (0,1]$, is needed for the proof of (2.40). Indeed, Petrini [91] showed that the μ-Hölder continuity of F, $\mu \in (0,1]$, is necessary to imply the second continuous differentiability of the Newton volume potential.

Let $\mathcal{G}$ be a regular region. If F is of class $C^{(0,\mu)}(\overline{\mathcal{G}})$, $\mu \in (0,1]$, then the Poisson differential equation

$$-\Delta_x \int_{\mathcal{G}} F(y)\ G(\Delta; |x-y|)\ dV(y) = \alpha(x)\ F(x) \qquad (2.41)$$

holds true for all $x \in \overline{\mathcal{G}}$, where $\alpha(x)$ is the solid angle subtended by the surface $\partial\mathcal{G}$ at x.

3. Ill-posedness of the gravimetry problem

Contrary to the case of $L^2(\partial\mathcal{G})$ (see [23] for its verification), the class $L^2(\mathcal{G})$ of square-integrable functions on a regular region $\mathcal{G}$ is not obtainable only by the L^2-completion of a countable harmonic function system. In addition, we have to take into account a so-called "anharmonic function system" (see, e.g., [9, 34, 71, 109]). This observation should be studied here in a more detailed way, since it explains the ill-posedness of the gravimetry problem.

Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region. By $E(\mathcal{G})$ we denote the space of all infinitely differentiable functions F in $\mathbb{R}^3$ possessing a compact support $\mathrm{supp}(F)$ in $\mathcal{G}$. We equip $E(\mathcal{G})$ with the following topology: a sequence $\{\phi_n\} \subset E(\mathcal{G})$ is called convergent to zero if and only if (1) there exists a bounded $\mathcal{B} \subset \mathbb{R}^3$ such that ϕ_n

vanishes outside $\mathcal{B}$, (2) for every differential operator ∇^α the sequence $\{\nabla^\alpha \phi_n\}$ is convergent to zero with respect to the norm $\|\cdot\|_{C^{(0)}(\mathcal{G})}$. Members of $E(\mathcal{G})$ are called *test functions*. Elements of the dual space $E^*(\mathcal{G})$, i.e., continuous linear functionals $\mathcal{F} : E(\mathcal{G}) \to \mathbb{R}$, are called *distributions* (or *generalized functions*). Clearly, multiplication (by a scalar) and addition are defined canonically for members of the class $E^*(\mathcal{G})$, hence, they are in use for distributions in the same way, too. More details can be found in any textbook on distributions, e.g., [58].

Let $\mathcal{F} \in E^*(\mathcal{G})$ be a given distribution. Assume that there exists a function $F : \mathcal{G} \to \mathbb{R}$ that is locally integrable, i.e., F is integrable on every compact subset of $\mathcal{G}$, such that $\mathcal{F}(\phi) = \int_{\mathcal{G}} F(x)\phi(x)\ dV(x)$ holds for all test functions $\phi \in E(\mathcal{G})$. Then $\mathcal{F}$ is called a *regular distribution*. If $\mathcal{F} \in E^*(\mathcal{G})$ is a regular distribution, then the associated function F is uniquely determined (except on a set of Lebesgue measure zero). Note that a well-known distribution that is not regular is the *delta distribution* δ given by $\delta(\phi) = \phi(0)$.

A sequence $\{\mathcal{F}_n\} \subset E^*(\mathcal{G})$ is called convergent to $\mathcal{F} \in E^*(\mathcal{G})$ if and only if $\lim_{n\to\infty} \mathcal{F}_n(\phi) = \mathcal{F}(\phi)$ for all $\phi \in E(\mathcal{G})$. This definition helps to introduce derivatives of distributions: If, for a given distribution $\mathcal{F} \in E^*(\mathcal{G})$, there exists a distribution $\tilde{\mathcal{F}} \in E^*(\mathcal{G})$ such that $\tilde{\mathcal{F}}(\phi) = (-1)^{[\alpha]}\mathcal{F}(\nabla^\alpha \phi)$, $\alpha \in \mathbb{N}_0^3$, $[\alpha] = \alpha_1 + \alpha_2 + \alpha_3$, for every $\phi \in E(\mathcal{G})$, then we set $\tilde{\mathcal{F}} = \nabla^\alpha \mathcal{F}$. In our potential theoretic approach we are particularly interested in Laplace derivatives: A functional $\mathcal{F} \in E^*(\mathcal{G})$ is called *distributionally harmonic* if and only if $\Delta\mathcal{F} = 0$. The set of all regular harmonic $L^2(\mathcal{G})$-distributions in $E^*(\mathcal{G})$ is denoted by $DistHarm(\mathcal{G})$. The space $DistHarm(\mathcal{G})$ apparently represents a generalization of the set $Harm(\mathcal{G})$ of harmonic functions in $\mathcal{G}$. Indeed, the following characterization is valid (see, e.g., [16]): The set $DistHarm(\mathcal{G})$ of all regular harmonic $L^2(\mathcal{G})$-distributions is a closed subspace of $L^2(\mathcal{G})$.

It is known from the theory of distributions that the set

$$Harm(\mathcal{G}) = \{H \in C^{(2)}(\mathcal{G}) : \int_{\mathcal{G}} (H(x))^2\ dx < \infty,\ \Delta H(x) = 0,\ x \in\ \mathcal{G}\} \tag{3.1}$$

is a subset of $DistHarm(\mathcal{G})$. Moreover, the so-called Weyl Lemma (see, e.g., [58]) tells us that $\int_{\mathcal{G}} F(x)\Delta\Phi(x)\ dx = 0$ for all $\Phi \in E(\mathcal{G})$ implies $F \in Harm(\mathcal{G})$, i.e., $DistHarm(\mathcal{G}) \subset Harm(\mathcal{G})$. As a consequence, we are led to the following remarkable result:

Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region. Then,

$$DistHarm(\mathcal{G}) = Harm(\mathcal{G}). \tag{3.2}$$

$Harm(\mathcal{G})$ is a closed linear subspace of $L^2(\mathcal{G})$. Thus, a well-known result of functional analysis on orthogonal decompositions (see, e.g., [16, 112]) enables us to formulate the decomposition of $L^2(\mathcal{G})$ in the form:

$$L^2(\mathcal{G}) = \overline{Harm(\mathcal{G})}^{\|\cdot\|_{L^2(\mathcal{G})}} \oplus \left(\overline{Harm(\mathcal{G})}^{\|\cdot\|_{L^2(\mathcal{G})}}\right)^{\perp}. \tag{3.3}$$

The following convention is in use (see [109], and also [9] and [71]):

$$AnHarm(\mathcal{G}) = \left(\overline{Harm(\mathcal{G})}^{\|\cdot\|_{L^2(\mathcal{G})}}\right)^{\perp}. \tag{3.4}$$

The members of $AnHarm(\mathcal{G})$ are called *anharmonic functions* in $\mathcal{G}$.

Summarizing our results, we finally arrive at the following *decomposition theorem*:

If $\mathcal{G} \subset \mathbb{R}^3$ is a regular region, then

$$L^2(\mathcal{G}) = Harm(\mathcal{G}) \oplus AnHarm(\mathcal{G}). \tag{3.5}$$

3.1. Hadamard's classification of the gravimetry problem

In classical nomenclature of physical geodesy, the inversion of "Newton's Law of Gravitation" (1.1) from terrestrial and spaceborne gravitational data, i.e., the determination of the internal density contrast function from potential data on and outside the boundary $\partial\mathcal{G}$ is known as the *gravimetry problem*. In other words, for a regular region $\mathcal{G} \subset \mathbb{R}^3$, we are interested in the problem of determining the density function $F \in L^2(\mathcal{G})$ from (information of) the gravitational potential V on $\overline{\mathcal{G}^c}$ in accordance with the integral equation

$$V(x) = A[F](x) = \int_{\mathcal{G}} F(y)\ G(\Delta; |x-y|)\ dy, \quad x \in \mathbb{R}^3. \tag{3.6}$$

In the sequel we denote the image of $X = L^2(\mathcal{G})$ under the operator A by Y, i.e.,

$$Y := A[L^2(\mathcal{G})] = \{V : V = A[F] = \int_{\mathcal{G}} G(\Delta; |\cdot - y|)F(y)\ dy,\ F \in L^2(\mathcal{G})\}. \tag{3.7}$$

Furthermore, for any subset $\mathcal{H} \subset \mathbb{R}^3$, we introduce the operator

$$A_{\mathcal{H}} : X = L^2(\mathcal{G}) \to Y|\mathcal{H}. \tag{3.8}$$

(more accurately, $A^{\mathcal{G}}_{\mathcal{H}}$) with $Y|\mathcal{H}$ consisting of all $A_{\mathcal{H}}[F]$ given by

$$\mathcal{H} \ni x \mapsto A_{\mathcal{H}}[F](x) = \int_{\mathcal{G}} G(\Delta; |x-y|)\ F(y)\ dy,\ F \in L^2(\mathcal{G}) \tag{3.9}$$

(note that A may be formally understood as $A_{\mathbb{R}^3}$). Clearly, $Y|\mathcal{H}$ forms a set of harmonic functions in $\mathcal{H}$, provided that $\mathcal{H}$ is a subset of $\mathcal{G}^c$.

In shorthand notation, the (*terrestrial/spaceborne gravimetry problem*) (TSGP) of classical physical geodesy can be formulated as follows:

(TSGP)*: Given $V \in L^2(\mathcal{G}^c)$, find $F \in L^2(\mathcal{G})$ with $A_{\overline{\mathcal{G}^c}}[F] = V$.*

In accordance with *Hadamard's classification*, TSGP violates all criteria, viz. uniqueness, existence, and stability:

(i) A solution of the gravimetry problem exists only if V belongs to the space $Y|\overline{\mathcal{G}^c}$. However, it should be pointed out that this restriction does not cause any numerical difficulty since, in practice, the information of V is only finite-dimensional.

(ii) The most serious problem is the non-uniqueness of the solution: The associated Fredholm integral operator $A_{\overline{\mathcal{G}^c}}$ has a kernel (null space) which is already known to coincide with the $L^2(\mathcal{G})$-orthogonal space of the closed linear subspace of all harmonic functions on $\mathcal{G}$. Unfortunately, the orthogonal complement, i.e., the class of anharmonic functions, is infinite-dimensional.

More precisely, if F is a member of class $L^2(\mathcal{G})$, then $A_{\overline{\mathcal{G}^c}} : L^2(\mathcal{G}) \to Y|\mathcal{G}^c$ given by

$$V = A_{\overline{\mathcal{G}^c}}[F] = \int_{\mathcal{G}} G(\Delta; |\cdot - y|)\ F(y)\ \ dy \Big|_{\mathcal{G}^c}, \quad F \in L^2(\mathcal{G}), \tag{3.10}$$

defines a linear operator such that $A_{\overline{\mathcal{G}^c}}[F]$ is harmonic in $\mathcal{G}^c$ and regular at infinity. The operator $A_{\overline{\mathcal{G}^c}}$ as defined by (3.10) is surjective, but it is not injective. Indeed, the null space (kernel) of $A_{\overline{\mathcal{G}^c}}$

$$\mathcal{N}(A_{\overline{\mathcal{G}^c}}) = AnHarm(\mathcal{G}) \tag{3.11}$$

consists of all functions in $L^2(\mathcal{G})$ that are orthogonal to harmonic functions in $\mathcal{G}$. $\mathcal{N}(A_{\overline{\mathcal{G}^c}})$ is the space of anharmonic functions in $\mathcal{G}$: Let $\mathcal{G} \subset \mathbb{R}^3$ be a regular region. Then we have

$$L^2(\mathcal{G}) = \overline{Harm(\mathcal{G})}^{\|\cdot\|_{L^2(\mathcal{G})}} \oplus \left(\overline{Harm(\mathcal{G})}^{\|\cdot\|_{L^2(\mathcal{G})}}\right)^{\perp}, \tag{3.12}$$

hence,

$$L^2(\mathcal{G}) = Harm(\mathcal{G}) \oplus AnHarm(\mathcal{G}) = Harm(\mathcal{G}) \oplus \mathcal{N}(A_{\overline{\mathcal{G}^c}}). \tag{3.13}$$

(iii) Restricting the operator $A_{\overline{\mathcal{G}^c}}$ to $Harm(\mathcal{G})$ leads to an injective mapping which has a discontinuous inverse.

Concerning the historical background, the problem of non-uniqueness has been discussed extensively in literature. This problem can be resolved by imposing some reasonable additional condition on the density. As we already saw, a suitable condition, suggested by the mathematical structure of the Newton potential operator A is to require that the density be harmonic. In fact, the approximate calculation of the harmonic density has already been implemented in several papers (see, e.g., [79] and the references therein), whereas the problem of determining the anharmonic part seems to be still a great challenge.

3.2. Spectral inversion procedure for balls

The set $Harm(\mathcal{B}_\beta(0))$ of harmonic functions in the ball $\mathcal{B}_\beta(0)$ with radius β around the origin 0 is a closed subspace of $L^2(\mathcal{B}_\beta(0))$ (note that β can be chosen, for example, to be the radius of a Runge (Bjerhammar) sphere (see, e.g., [29]) or the (mean) Earth's radius). Moreover, the *inner harmonics* $\{H_{n,j}(\beta;\cdot)\}_{n=0,1,\ldots;j=1,\ldots,2n+1}$ given by

$$H_{n,j}(\beta; x) = \sqrt{\frac{2n+3}{\beta^3}} \left(\frac{|x|}{\beta}\right)^n Y_{n,j}\left(\frac{x}{|x|}\right),\ x \in \overline{\mathcal{B}_\beta(0)}, \tag{3.14}$$

constitute a complete orthonormal system in the Hilbert space

$$\left(Harm(B_{\mathcal{B}_\beta(0)}), \langle\cdot,\cdot\rangle_{L^2(\mathcal{B}_\beta(0))}\right)$$

provided that $\{Y_{n,j}\}_{n=0,1,\ldots;j=1,\ldots,2n+1}$ is a complete system of spherical harmonics on the unit sphere (see, e.g., [38]).

The set of square-integrable harmonic functions on the outer space

$$\mathcal{B}^c_\beta(0) = \mathbb{R}^3 \backslash \overline{\mathcal{B}_\beta(0)}$$

of a sphere $\partial\mathcal{B}_\beta(0)$ given by $Harm(\mathcal{B}^c_\beta(0))$ is a closed subspace of $L^2(\mathcal{B}^c_\beta(0))$. Moreover, the *outer harmonics* $\{H^{\text{ext}}_{-n-1,j}(\beta;\cdot)\}_{n=1,2,\ldots;j=1,\ldots,2n+1}$ given by

$$H_{-n-1,j}(\beta;x) = \sqrt{\frac{2n-1}{\beta^3}} \left(\frac{\beta}{|x|}\right)^{n+1} Y_{n,j}\left(\frac{x}{|x|}\right), \quad x \in \overline{\mathcal{B}^c_\beta(0)} \tag{3.15}$$

form a complete orthonormal system in the Hilbert space

$$\left(Harm(\mathcal{B}^c_\beta(0)), \langle\cdot,\cdot\rangle_{L^2(\mathcal{B}^c_\beta(0))}\right).$$

It should be remarked that an outer harmonic of degree $n = 0$ is proportional to

$$\frac{1}{|x|} Y_{0,1}\left(\frac{x}{|x|}\right) = \frac{1}{\sqrt{4\pi}|x|}, \; x \in \overline{\mathcal{B}^c_\beta(0)}. \tag{3.16}$$

This function, however, is not an element of $L^2(\mathcal{B}^c_\beta(0))$.

Harmonic Case. The operator $A_{\overline{\mathcal{B}^c_\beta(0)}}$, given by

$$A_{\overline{\mathcal{B}^c_\beta(0)}}[F](x) = \int_{\mathcal{B}_\beta(0)} G(\Delta; |x-y|)\; F(y)\; dy, \quad x \in \overline{\mathcal{B}^c_\beta(0)}, \tag{3.17}$$

has the null space

$$\mathcal{N}\left(A_{\overline{\mathcal{B}^c_\beta(0)}}\right) = AnHarm(\mathcal{B}_\beta(0)). \tag{3.18}$$

For any $F \in L^2(\mathcal{B}_\beta(0))$, there exists a unique orthogonal decomposition

$$F = P_{Harm(\mathcal{B}_\beta(0))}[F] + P_{AnHarm(\mathcal{B}_\beta(0))}[F], \tag{3.19}$$

where $P_{Harm(\mathcal{B}_\beta(0))}[F] \in Harm(\mathcal{B}_\beta(0))$ and $P_{\text{AnHarm}(\mathcal{B}_\beta(0))}[F] \in AnHarm(\mathcal{B}_\beta(0))$. We are allowed to represent $P_{Harm(\mathcal{B}_\beta(0))}[F]$ as a Fourier series in terms of inner harmonics

$$P_{Harm(\mathcal{B}_\beta(0))}[F] = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \langle F, H_{n,j}(\beta;\cdot)\rangle_{L^2(\mathcal{B}_\beta(0))}\, H_{n,j}(\beta;\cdot) \tag{3.20}$$

with respect to the topology of $L^2(\mathcal{B}_\beta(0))$. Suppose that $y \in \overline{\mathcal{B}^c_\beta(0)}$ is arbitrary but fixed. Then the potential at y corresponding to the mass density distribution

F can be represented in the form

$$A_{\overline{\mathcal{B}_\beta^c(0)}}[F](y) = \beta^2 \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \frac{4\pi}{2n+1} \frac{1}{\sqrt{(2n-1)(2n+3)}} \left\langle H_{n,j}\left(\beta;\cdot\right), F\right\rangle_{L^2(\mathcal{B}_\beta(0))} \; H_{-n-1,j}\left(\beta;y\right). \tag{3.21}$$

A harmonic solution $F \in Harm(B_{\mathcal{B}_\beta(0)})$ of the problem

$$A_{\overline{\mathcal{B}_\beta^c(0)}}[F] = P, \quad P \in Harm(\mathcal{B}_\beta^c(0)) \tag{3.22}$$

is unique and is given via its Fourier coefficients

$$\left\langle F, H_{n,j}(\beta;\cdot)\right\rangle_{L^2(\mathcal{B}_\beta(0))} = \frac{2n+1}{4\pi\beta^2}\sqrt{(2n-1)(2n+3)}\left\langle P, H_{-n-1,j}(\beta;\cdot)\right\rangle_{L^2(\mathcal{B}_\beta^c(0))}, \tag{3.23}$$

$n \in \mathbb{N}$, $j \in \{1,\dots,2n+1\}$, and

$$\left\langle F, H_{0,1}(\beta;\cdot)\right\rangle_{L^2(\mathcal{B}_\beta(0))} = 0, \tag{3.24}$$

In other words, we have

$$F = \sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} \frac{2n+1}{4\pi\beta^2}\sqrt{(2n-1)(2n+3)}\left\langle P, H_{-n-1,j}(\beta;\cdot)\right\rangle_{L^2(\mathcal{B}_\beta^c(0))} \; H_{n,j}(\beta;\cdot) \tag{3.25}$$

in the sense of $L^2(\mathcal{B}_\beta(0))$.

In accordance with the Picard condition (that appeared in book form in the chapter [84]) the equation $A_{\overline{\mathcal{B}_\beta^c(0)}}[F] = P$ is solvable if and only if P is harmonic and the following series is finite, i.e.,

$$\sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} n^4 \left\langle P, H_{-n-1,j}(\beta;\cdot)\right\rangle^2_{L^2(\mathcal{B}_\beta^c(0))} < \infty. \tag{3.26}$$

Note that

$$\begin{aligned} &\sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} (2n+1)^2(2n-1)(2n+3)\left\langle P, H_{-n-1,j}(\beta;\cdot)\right\rangle^2_{L^2(\mathcal{B}_\beta^c(0))} < \infty \\ &\qquad \Leftrightarrow \quad \sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} n^4 \left\langle P, H_{-n-1,j}(\beta;\cdot)\right\rangle^2_{L^2(\mathcal{B}_\beta^c(0))} < \infty. \end{aligned} \tag{3.27}$$

This condition can be also motivated within the framework of harmonics by observing

$$F \in Harm(B_{\mathcal{B}_\beta(0)}) = \overline{\left\{H_{n,j}(\beta;\cdot) : n \in \mathbb{N}_0; j \in \{1,\dots,2n+1\}\right\}}^{\|\cdot\|_{L^2(\mathcal{B}_\beta(0))}}, \tag{3.28}$$

that implies

$$\sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \left\langle F, H_{n,j}(\beta;\cdot)\right\rangle^2_{L^2(\mathcal{B}_\beta(0))} < \infty \ . \tag{3.29}$$

Operators of the type $A_{\overline{\mathcal{B}_\beta^c(0)}}$ are compact, hence, we are confronted with the fact that the restricted operator

$$A_{\overline{\mathcal{B}_\beta^c(0)}}\big|Harm(B_{\mathcal{B}_\beta(0)}) : Harm(B_{\mathcal{B}_\beta(0)}) \to A_{\overline{\mathcal{B}_\beta^c(0)}}\big(Harm(B_{\mathcal{B}_\beta(0)})\big)$$

is invertible, but its inverse operator $\left(A_{\overline{\mathcal{B}_\beta^c(0)}}\big|_{Harm(B_{\mathcal{B}_\beta(0)})}\right)^{-1}$ is discontinuous.

Anharmonic Case. An orthogonal basis for $AnHarm(\mathcal{B}_\beta(0))$ (with respect to the space $L^2(\mathcal{B}_\beta(0))$) can be found in [9]. A different non-orthogonal anharmonic basis has been developed in [71, 72]:

(a) A complete $L^2(\mathcal{B}_\beta(0))$-orthogonal system in $AnHarm(\mathcal{B}_\beta(0))$ is given by

$$\left\{x \mapsto |x|^n P_{k,n}(|x|^2)\; Y_{n,j}\left(\frac{x}{|x|}\right)\right\}_{\substack{k\in\mathbb{N};\; n\in\mathbb{N}_0\\ j\in\{1,\ldots,2n+1\}}}, \tag{3.30}$$

where $\{P_{k,n}\}_{k\in\mathbb{N};n\in\mathbb{N}_0}$ is a system of polynomials defined by

$$P_{k,n}(t) = \sqrt{\frac{2}{\beta^{2n+3}}}\; G_k\left(n+\frac{3}{2}, n+\frac{3}{2}; \frac{t}{\beta^2}\right). \tag{3.31}$$

Here, the functions $G_k, k \in \mathbb{N}_0$, are the Jacobi polynomials, which are the only polynomials on $[0,1]$ to satisfy the following conditions for all $n, m \in \mathbb{N}_0$:

(i) $G_n(a,b;\cdot)$ is a polynomial of degree n on $[0,1]$.

(ii) $G_n(a,b;0) = 1$.

(iii) $\int_0^1 x^{a-1}(1-x)^{b-a} G_n(a,b;x)\; G_m(a,b;x)\; dx = 0$ for $n \neq m$,

provided that $a > 0$ and $b > a - 1$.

(b) A closed system in $\text{Anharm}(\overline{\Sigma_{\text{int}}})$ is given by

$$\left\{x \mapsto \left(|x|^{n+2k} - \frac{(2n+3)\beta^{2k}}{2n+2k+3}\; |x|^n\right) Y_{n,j}\left(\frac{x}{|x|}\right)\right\}_{\substack{k\in\mathbb{N};\; n\in\mathbb{N}_0\\ j\in\{1,\ldots,2n+1\}}}. \tag{3.32}$$

Moreover, the basis functions are polynomials of degree $\leq N \in \mathbb{N}\backslash\{1\}$ if and only if the index triple (k,n,j) is within the range $n \in \{0,\ldots,N-2\}$, $j \in \{1,\ldots,2n+1\}$, $k \in \left\{1,\ldots,\left[\frac{N-n}{2}\right]\right\}$, where $[\,\cdot\,]$ is the Gauss bracket, defined by $[x] = \max\{\nu \in \mathbb{Z} : \nu \leq x\}$, $x \in \mathbb{R}$. The set of anharmonic polynomials with degrees $\leq N$ possesses the dimension $\frac{1}{6}N^3 - \frac{1}{6}N$.

The obvious advantage of the system in (a) is its orthogonality. On the other hand, the system described in (b) has a radial part (see also [34]), which is explicitly given, whereas the radial part of the orthogonal system has to be calculated iteratively by means of recurrence formulas.

The important role of the anharmonic functions in the theory of the gravimetry problem is also stressed if we investigate a radially symmetric density distribution which is given for the mantle and the outer and inner core of the Earth. Such a structure of spherical layers does not give any information in the gravitational potential and, therefore, cannot be recovered by means of harmonic functions. The Ph.D.-thesis of V. Michel [71], indeed, shows that a reconstruction of the (deep)

Earth's interior with a harmonic function system makes no sense. Therefore, a reliable method for the (global) approximation of the density distribution of the Earth requires a treatment of both orthogonal projections: the harmonic part and the anharmonic part.

Moreover, we recall that the contribution of $H_{-1,1}$ to an (outer) gravitational (disturbance) potential can be neglected when applying an appropriate coordinate transformation (see, e.g., [81] for more details). This operation can, therefore, physically be interpreted as filtering out the contribution of the radially symmetric density structures in the Earth's interior (note that the total mass of an anharmonic density function is zero).

3.3. Spectral inversion procedure for regular regions

The above results will now be extended to the investigation of the inverse problem $A_{\overline{\mathcal{G}^c}}[F] = V$, where $A_{\overline{\mathcal{G}^c}}[F]$ is the gravitational potential of a regular region $\mathcal{G} \subset \mathbb{R}^3$ and $F \in L^2(\mathcal{G})$ is the desired mass density distribution F. As already known from (3.11), the null space of the operator $A_{\overline{\mathcal{G}^c}}$ is given by $\mathcal{N}\left(A_{\overline{\mathcal{G}^c}}\right) = AnHarm(\mathcal{G})$.

A general complete orthonormal basis system for the harmonic functions inside or outside an arbitrary regular region is not available. This is the reason why the following setting is useful: Let the families of functions

$$\{H_{n,j}(\mathcal{G};\cdot)\}_{\substack{n\in\mathbb{N}_0;\\ j=1,\ldots,2n+1}} \quad \text{and} \quad \{H_{-n-1,j}(\mathcal{G}^c;\cdot)\}_{\substack{n\in\mathbb{N};\\ j=1,\ldots,2n+1}} \tag{3.33}$$

be complete orthonormal systems of the Hilbert spaces

$$\left(Harm(\mathcal{G}), \langle\cdot,\cdot\rangle_{L^2(\mathcal{G})}\right) \quad \text{and} \quad \left(Harm(\mathcal{G}^c), \langle\cdot,\cdot\rangle_{L^2(\mathcal{G}^c)}\right), \tag{3.34}$$

respectively, and $\{k_{\mathcal{G}}^{\wedge}(n)\}_{n\in\mathbb{N}_0}$ be the symbol of

$$A_{\overline{\mathcal{G}^c}} : L^2(\mathcal{G}) \to Y|\overline{\mathcal{G}^c} = \mathcal{R}(A_{\overline{\mathcal{G}^c}}) = A_{\overline{\mathcal{G}^c}}(L^2(\mathcal{G}), \tag{3.35}$$

given by

$$\begin{aligned} &A_{\overline{\mathcal{G}^c}}[F](x)\\ &= \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} k_{\mathcal{G}}^{\wedge}(n)\,\langle F, H_{n,j}(\mathcal{G};\cdot)\rangle_{L^2(\mathcal{G})}\, H_{-n-1,j}(\overline{\mathcal{G}^c};x),\ x\in\overline{\mathcal{G}^c},\ F\in L^2(\mathcal{G}), \end{aligned} \tag{3.36}$$

where $H_{-1,1}(\mathcal{G};\cdot)$ is not an element of $L^2\left(\mathcal{G}^c\right)$). We assume that $k_{\mathcal{G}}^{\wedge}(n) \neq 0$ for all $n \in \mathbb{N}_0$.

If $\partial\mathcal{G}$ is a sphere with radius β around the origin, we let

$$H_{n,j}(\mathcal{G};\cdot) := H_{n,j}(\beta;\cdot); \qquad n\in\mathbb{N}_0,\, j\in\{1,\ldots,2n+1\}; \tag{3.37}$$

$$H_{-n-1,j}(\overline{\mathcal{G}^c};\cdot) := H_{-n-1,j}(\beta;\cdot); \quad n\in\mathbb{N},\ \ j\in\{1,\ldots,2n+1\}\ . \tag{3.38}$$

Moreover, we set

$$k_{\mathcal{G}}^{\wedge}(n) = k_{\beta}^{\wedge}(n) = \frac{4\pi}{2n+1}\,\frac{\beta^2}{\sqrt{(2n-1)(2n+3)}}\ . \tag{3.39}$$

The *inverse problem* $A_{\overline{\mathcal{G}^c}}[F] = V$ *with* $F \in Harm(\mathcal{G})$ *unknown,* is solvable if and only if $V \in Harm(\mathcal{G})$ with

$$\sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} \left(\frac{\left\langle V, H_{-n-1,j}(\overline{\mathcal{G}^c}; \cdot) \right\rangle_{L^2(\mathcal{G}^c)}}{k_{\mathcal{G}}^{\wedge}(n)} \right)^2 < \infty \ . \tag{3.40}$$

In this case, the harmonic solution $F \in Harm(\mathcal{G})$ is uniquely determined and spectrally given by

$$\langle F, H_{0,1}(\mathcal{G}; \cdot) \rangle_{L^2(\mathcal{G})} = 0, \tag{3.41}$$

$$\langle F, H_{n,j}(\mathcal{G}; \cdot) \rangle_{L^2(\mathcal{G})} = \frac{\langle V, H_{-n-1,j}(\overline{\mathcal{G}^c}; \cdot) \rangle_{L^2(\mathcal{G}^c)}}{k_{\mathcal{G}}^{\wedge}(n)}, \tag{3.42}$$

for $n \in \mathbb{N}, j \in \{1, \ldots, 2n+1\}$.

As already known, the inverse operator $(A_{\overline{\mathcal{G}^c}} | Harm(\mathcal{G}))^{-1}$, defined on the image $Y|\overline{\mathcal{G}^c}$, is unbounded. Due to unavoidable errors in the measurements of the gravitational field the application of this inverse operator to the observed potential for a direct reconstruction of the mass density distribution is not senseful. Therefore, we have to take into account suitable regularizations. Indeed, the results as presented here enable us to apply projection-, multiscale-, and iteration regularization techniques in the way indicated, e.g., in our contribution [35].

4. Mollifier methods

Next we deal with space regularization methods for the Newton volume integral involving singular integral mollification.

4.1. Haar-type mollifier method

We start from the differential equation

$$\Delta_y \ G_\rho^H(\Delta; |y-z|) = \ - \ H_\rho(|y-z|) \tag{4.1}$$

with

$$G_\rho^H(\Delta; |y-z|) = \begin{cases} \frac{1}{8\pi\rho}(3 - \frac{|y-z|^2}{\rho^2}) & , \quad |y-z| \le \rho, \\ \frac{1}{4\pi|y-z|} & , \quad |y-z| > \rho, \end{cases} \tag{4.2}$$

where

$$H_\rho(|y-z|) = \begin{cases} \frac{3}{4\pi\rho^3} & , \quad |y-z| \le \rho, \\ 0 & , \quad |y-z| > \rho \end{cases} \tag{4.3}$$

is the so-called *Haar kernel* (note that $\|\mathcal{B}_\rho(0)\| = \frac{4}{3}\pi\rho^3$).

It is well known (see, e.g., [29]) that the *Haar singular integral* $\{I_\rho\}_{\rho>0}$ defined by

$$I_\rho[F] = F_\rho^H = \int_{\mathcal{G}} H_\rho(|\cdot - z|) \ F(z) \ dz, \tag{4.4}$$

with the *Haar kernel as mollifier* satisfies the limit relation $\lim_{\rho\to 0+} I_\rho[F] = F, F \in L^2(\mathcal{G})$, in the topology of $L^2(\mathcal{G})$. Moreover, we have

$$\lim_{\rho\to 0+} I_\rho[F](x) = \alpha(x)\ F(x),\ x \in \overline{\mathcal{G}},\ F \in C^{(0)}(\overline{\mathcal{G}}), \tag{4.5}$$

where $\alpha(x)$ is the solid angle at x subtended by the surface $\partial\mathcal{G}$.

In constructive approximation, locally supported functions

$$y \mapsto H_\rho(|x-y|), \rho > 0, \quad x \in \mathbb{R}^3,$$

are nothing new, with one-dimensional counterparts having been discussed already by Haar (1910). The primary importance of locally supported Haar kernels in the classical one-dimensional Euclidean space is that they led to the "birth" to an entire "basis family" by means of two operations, viz. dilations and translations. In other words, an entire set of approximants is available from the single locally supported "Haar mother kernel", and this basis family provides useful "building block functions" that enable the multiscale modeling and the decorrelation of data.

Internal/Terrestrial Gravimetry Problem (ITGP). Correspondingly to $\{I_\rho\}_{\rho>0}$ we introduce the family $\{A_\rho^H\}_{\rho>0}$ given by

$$A_\rho^H[F] = V_\rho^H = \int_{\mathcal{G}} G_\rho^H(\Delta; |\cdot - z|)\ F(z)\ dz, \tag{4.6}$$

such that

$$\begin{aligned}\Delta A_\rho^H[F] &= \Delta \int_{\mathcal{G}} G_\rho^H(\Delta; |\cdot - z|)\ F(z)\ dz \\ &= -I_\rho[F] = -F_\rho^H = -\int_{\mathcal{G}} H_\rho(|\cdot - z|)\ F(z)\ dz.\end{aligned} \tag{4.7}$$

Multiscale mollifier approximation. Next we are interested in applying the multiscale "Haar philosophy" to an approximate determination of the mass density distribution inside $\mathcal{G}$ (cf. [29]): Suppose that $\{\rho_j\}_{j\in\mathbb{N}_0}$ is a positive, monotonously decreasing sequence with $\lim_{j\to\infty} \rho_j = 0$, for example, the dyadic sequence given by $\rho_j = 2^{-j}$. For $j \in \mathbb{N}_0$, we consider the differences

$$\Psi_{G_{\rho_j}^H}(\Delta; |\cdot - y|) = G_{\rho_{j+1}}^H(\Delta; |\cdot - y|) - G_{\rho_j}^H(\Delta; |\cdot - y|) \tag{4.8}$$

and

$$\Psi_{H_{\rho_j}}(|\cdot - y|) = H_{\rho_{j+1}}(|\cdot - y|) - H_{\rho_j}(|\cdot - y|). \tag{4.9}$$

$\Psi_{G_{\rho_j}^H}(\Delta; \cdot)$ and $\Psi_{H^{\rho_j}}$ are called *"ρ_j-fundamental wavelet function"* and *"ρ_j-Haar wavelet function"*, respectively. The associated *"ρ_j-potential wavelet functions"* and the *"ρ_j-density wavelet functions"* are given by

$$(WV)_{\rho_j}^H = \int_{\mathcal{G}} \Psi_{G_{\rho_j}^H}(\Delta; |\cdot - y|)\ F(y)\ dy \tag{4.10}$$

and

$$(WF)^H_{\rho_j} = \int_{\mathcal{G}} \Psi_{H_{\rho_j}}(|\cdot - y|)\ F(y)\ dy, \tag{4.11}$$

respectively. The ρ_j-potential wavelet functions and the ρ_j-density wavelet functions, respectively, characterize the successive *detail information* contained in $V^H_{\rho_{j+1}} - V^H_{\rho_j}$ and $F^H_{\rho_{j+1}} - F^H_{\rho_j}$, $j \in \mathbb{N}_0$. In other words, we are able to decorrelate the potential V and the "density signature" F, respectively, in form of "band structures"

$$(WV)^H_{\rho_j} = V^H_{\rho_{j+1}} - V^H_{\rho_j}, \tag{4.12}$$

and

$$(WF)^H_{\rho_j} = F^H_{\rho_{j+1}} - F^H_{\rho_j}. \tag{4.13}$$

The essential problem to be solved in multiscale extraction of geological features is to identify those detail information, i.e., band structures in (4.12), which contain specifically desired geological (density) characteristics in (4.13). Seen from a numerical point of view, it is remarkable that both wavelet functions $y \mapsto \Psi_{G^H_{\rho_j}}(\Delta; |\cdot - y|)$ and $y \mapsto \Psi_{H_{\rho_j}}(|\cdot - y|)$ vanish outside a ball around the center x due to their construction, i.e., these functions are spacelimited showing a ball as local support. Furthermore, the ball becomes smaller with increasing scale parameter j, so that more and more high frequency phenomena can be highlighted without changing the features outside the balls. Forming the telescoping sums

$$\sum_{j=0}^{J-1} (WV)^H_{\rho_j} = \sum_{j=0}^{J-1} \left(V^H_{\rho_{j+1}} - V^H_{\rho_j} \right), \tag{4.14}$$

and

$$\sum_{j=0}^{J-1} (WF)^H_{\rho_j} = \sum_{j=0}^{J-1} \left(F_{\rho_{j+1}} - F_{\rho_j} \right), \tag{4.15}$$

we are easily led to the identities

$$V^H_{\rho_J} = V^H_{\rho_0} + \sum_{j=0}^{J-1} (WV)^H_{\rho_j} \tag{4.16}$$

and

$$F^H_{\rho_J}(x) = F^H_{\rho_0} + \sum_{j=0}^{J-1} (WF)^H_{\rho_j}. \tag{4.17}$$

Thus we finally end up with the following multiscale relations

$$\lim_{J\to\infty} V^H_{\rho_J} = V^H_{\rho_0} + \sum_{j=0}^{\infty} (WV)^H_{\rho_j} \tag{4.18}$$

and

$$\lim_{J\to\infty} F^H_{\rho_J} = F^H_{\rho_0} + \sum_{j=0}^{\infty} (WF)^H_{\rho_j} = \lim_{J\to\infty} \Delta V^H_{\rho_J} = \Delta V^H_{\rho_0} + \sum_{j=0}^{\infty} \Delta (WV)^H_{\rho_j}. \tag{4.19}$$

Altogether, the potential V as well as the "density signature" F can be expressed in additive way as a low-pass filtered signals $V^H_{\rho_0}$ and $F^H_{\rho_0}$ and successive band-pass filtered signals $(WV)^H_{\rho_j}$ and $(WF)^H_{\rho_j}$, $j = 0, 1, \ldots$, respectively.

Multiscale mollifier numerics. For a sufficiently large integer J, it follows from (4.5) that

$$\begin{aligned} \alpha(x)\ F(x) &\simeq I_{\rho_J}[F](x) = F^H_{\rho_J}(x) \\ &= \int_{\mathcal{G}} H_{\rho_J}(|x-y|) F(y)\ dy, \quad x \in \overline{\mathcal{G}}, F \in C^{(0)}(\overline{\mathcal{G}}) \end{aligned} \tag{4.20}$$

("$\simeq$" means that the error is negligible). From (4.1) we obtain

$$\begin{aligned} \Delta_x A_{\rho_J}[F](x) &= \Delta_x \int_{\mathcal{G}} G^H_{\rho_J}(\Delta; |x-y|) F(y)\ dV(y) \\ &= -\int_{\mathcal{G}} H_{\rho_J}(|x-y|) F(y)\ dy \simeq\ -\ \alpha(x)\ F(x), \end{aligned} \tag{4.21}$$

where we are aware of the fact that

$$V(x) \simeq \int_{\mathcal{G}} G^H_{\rho_J}(\Delta; |x-y|) F(y)\ dy, \quad x \in \overline{\mathcal{G}}, \tag{4.22}$$

with negligible error. In order to realize a fully discrete approximation of F we have to apply approximate integration formulas leading to

$$V(x) \simeq \sum_{i=1}^{N_J} G^H_{\rho_J}(\Delta; |x - y_i^{N_J}|)\ w_i^{N_J} F(y_i^{N_J}), \quad x \in \overline{\mathcal{G}}, \tag{4.23}$$

where $w_i^{N_J} \in \mathbb{R}$, $y_i^{N_J} \in \overline{\mathcal{G}}$, $i = 1, \ldots, N_J$, are the known weights and knots, respectively.

For numerical realization of mass density modeling by means of Haar kernels we notice that all coefficients

$$a_i^{N_J} = w_i^{N_J} F(y_i^{N_J}),\ i = 1, \ldots, N_J, \tag{4.24}$$

are unknown. Then we have to solve a linear system, namely

$$V(x_k^{T_J}) = \sum_{i=1}^{N_J} G^H_{\rho_J}(\Delta; |x_k^{T_J} - y_i^{N_J}|)\ a_i^{N_J}, \quad x_k^{T_J} \in \overline{\mathcal{G}},\ k = 1, \ldots, T_J, \tag{4.25}$$

to determine $a_i^{N_J}, i = 1, \ldots, N_J$, from known gravitational values $V(x_k^{T_J})$ at knots $x_k^{T_J} \in \overline{\mathcal{G}}$, $k = 1, \ldots, T_J$.

Once all density values $F(y_i^{N_J})$, $i = 1, \ldots, N_J$, are available (note that the integration weights $w_i^{N_J}, i = 1, \ldots, N_J$, are known), the density distribution F can be obtained from the formula

$$F(x) \simeq F_{\rho_J}^H(x) = \sum_{i=1}^{N_J} H_{\rho_J}(|x - y_i^{N_J}|) \underbrace{w_i^{N_J} F(y_i^{N_J})}_{=a_i^{N_J}}, \quad x \in \mathcal{G}. \tag{4.26}$$

Even better, fully discrete Haar filtered versions of F at lower scales can be derived in accordance with the approximate integration rules

$$\int_{\mathcal{G}} H_{\rho_j}(|x - z|)\ F(z)\ dV(z) \simeq \sum_{i=1}^{N_j} H_{\rho_j}(|x - y_i^{N_j}|)\ w_i^{N_j}\ F(y_i^{N_j}) \tag{4.27}$$

for $j = J_0, \ldots, J$, where $w_i^{N_j}$, $y_i^{N_j}$, $i = 1, \ldots, N_j$, are known weights and knots, respectively, such that $\{y_1^{N_j}, \ldots, y_{N_j}^{N_j}\} \subset \{y_1^{N_J}, \ldots, y_{N_J}^{N_J}\} \subset \overline{\mathcal{G}}$, i.e., the sequence of knots $\{y_1^{N_J}, \ldots, y_{N_J}^{N_J}\} \subset \overline{\mathcal{G}}$ shows a hierarchical positioning. Altogether, our approach yields Haar filtered versions (4.27) establishing a (space-based) multiscale decomposition $F_{\rho_J}^H, \ldots, F_{\rho_{J_0}}^H$ of the density distribution F, such that an entire set of approximations is available from a single locally supported "mother function", i.e., the Haar kernel function (4.3), and this set provides useful "building block functions", which enable decorrelation of the density signatures and suitable storage and fast decorrelation of density data.

Moreover, fully discrete Haar filtered versions of F at lower scales can be derived in accordance with the approximate integration rules

$$F_{\rho_j}^H(x) = \int_{\mathcal{G}} H_{\rho_j}(|x-y|)\ F(y)\ dy \simeq \sum_{i=1}^{N_j} H_{\rho_j}(|x-y_i^{N_j}|)\ w_i^{N_j} F(y_i^{N_j}), \quad x \in \mathcal{G}, \tag{4.28}$$

for $j = J_0, \ldots, J$, where $w_i^{N_j}$, $y_i^{N_j}$, $i = 1, \ldots, N_j$, are known weights and knots, respectively, such that we can take advantage of the fact that $\{y_1^{N_j}, \ldots, y_{N_j}^{N_j}\} \subset \{y_1^{N_J}, \ldots, y_{N_J}^{N_J}\} \subset \overline{\mathcal{G}}$.

The serious problem of our multiscale approach, however, is that measurements of gravitation are only available in the interior $\mathcal{G}$ in exceptional cases, for example, locally in geothermal boreholes. However, we are able to take into account surface measurements on $\partial\mathcal{G}$, but it may be questioned that deep geological formations can be detected by an exclusive use of terrestrial gravitational data. Nevertheless, the multiscale method as explained above is an important postprocessing method to improve the interpretability of already available geological models as well as (wavelet) decorrelation mechanisms to extract certain local features of practical relevance in density band signatures (see [15, 78]).

4.2. De la Vallée Poussin-type mollifier method

The critical point in the Haar-type approach is the discontinuity of the Laplace derivative of $G_\rho^H(\Delta;\cdot)$, i.e., the ordinary Haar function H_ρ. In what follows we are therefore interested in a smoothed Haar kernel variant, called de la Vallée Poussin kernel.

For $x, y \in \mathbb{R}^3$ we define the de la Vallée Poussin kernel $VP_\rho = [0,\infty) \to \mathbb{R}$, $\rho > 0$, by

$$VP_\rho(r) = \frac{1}{C_\rho^{VP}} \begin{cases} (1 - \frac{r^2}{\rho^2}) & , \quad r \le \rho \\ 0 & , \quad r > \rho, \end{cases} \tag{4.29}$$

where the normalization constant $C_\rho^{VP} = \frac{8\pi}{15}\rho^3$ is chosen in such a way that

$$\int_{\mathbb{R}^3} VP_\rho(|x-y|)\, dy = 4\pi \int_0^\rho VP_\rho(r)\, r^2\, dr = 1. \tag{4.30}$$

It is easy to see that $r \mapsto -\frac{1}{6}r^2 + \frac{1}{20\rho^2}r^4, r \ge 0$, satisfies

$$-\tfrac{1}{r^2}\tfrac{d}{dr} r^2 \tfrac{d}{dr} \left(-\tfrac{1}{6}r^2 + \tfrac{1}{20\rho^2}r^4\right) = 1 - \tfrac{r^2}{\rho^2}, \quad r \ge 0,\ \rho > 0. \tag{4.31}$$

As a consequence, it follows that

$$G_\rho^{VP}(\Delta; |x-y|) = \begin{cases} \frac{1}{C_\rho^{VP}} \left(-\frac{1}{6}|x-y|^2 + \frac{1}{20\rho^2}|x-y|^4\right) & , \quad |x-y| \le \rho \\ 0 & , \quad |x-y| > \rho \end{cases} \tag{4.32}$$

satisfies

$$-\Delta_x G_\rho^{VP}(\Delta; |x-y|) = VP_\rho(|x-y|), \quad x, y \in \mathbb{R}^3. \tag{4.33}$$

An elementary calculation yields $-\frac{1}{r^2}\frac{d}{dr} r^2 \frac{d}{dr}(1 - \frac{r^2}{\rho^2}) = \frac{6}{\rho^2}$, so that

$$-\Delta_x VP_\rho(\Delta; |x-y|) = D_\rho^{VP}(|x-y|), \quad x, y \in \mathbb{R}^3, \tag{4.34}$$

where

$$D_\rho^{VP}(|x-y|) = \begin{cases} \frac{1}{C_\rho^{VP}} \frac{6}{\rho^2} = \frac{8\pi}{3}\rho & , \quad |x-y| \le \rho \\ 0 & , \quad |x-y| > \rho. \end{cases} \tag{4.35}$$

Clearly, all methodological concepts developed for the Haar case together with its multiscale settings remain valid. Their formulations are straightforward. The following result, however, serves as strategic basis for our approach to density feature extraction in specific representation within the de la Vallée Poussin framework.

Theorem 4.1. *The "ρ-de la Vallée Poussin potential functions"*

$$V_\rho^{VP}(x) = \int_{\mathcal{G}} G_\rho^{VP}(\Delta; |x-y|) F(y)\, dy \tag{4.36}$$

and the "ρ-de la Vallée Poussin density function"

$$F_\rho^{VP}(x) = \int_{\mathcal{G}} VP_\rho(|x-y|) F(y)\, dy \tag{4.37}$$

satisfy the relations

$$\sup_{x\in\overline{\mathcal{G}}}|V(x)-V_\rho^{VP}(x)|=O(\rho^2),\quad \rho\to 0 \tag{4.38}$$

and

$$\lim_{\rho\to 0}\sup_{x\in\overline{\mathcal{G}}}|\alpha(x)F(x)-F_\rho^{VP}(x)|=0, \tag{4.39}$$

where $\alpha(x)$ *is the solid angle subtended by the boundary* $\partial\mathcal{G}$ *at* $x\in\overline{\mathcal{G}}$.

Unfortunately, de la Vallée Poussin potentials V_ρ^{VP} do not generally show a faster convergence to V than V_ρ^H.

Approximate mollifier solution. In similarity to our previous Haar considerations we use the operators

$$A_\rho^{VP}[F]=V_\rho^{VP}=\int_{\mathcal{G}}G_\rho^{VP}(\Delta;|\cdot-z|)\ F(z)\ dz,\ F\in L^2(\mathcal{G}), \tag{4.40}$$

and

$$I_\rho^{VP}[F]=F_\rho^{VP}=\int_{\mathcal{G}}VP_\rho(|\cdot-z|)\ F(z)\ dz,\ F\in L^2(\mathcal{G}). \tag{4.41}$$

We denote the image of $X=L^2(\mathcal{G})$ under the operator A_ρ^{VP} by Y_ρ^{VP}. So, instead of discussing the integral $A[F](x)=\int_{\mathcal{G}}G(\Delta;|x-y|)\ F(y)\ dy$ we choose $A_\rho^{VP}[F]$, $F\in L^2(\mathcal{G})$, for some sufficiently small $\rho>0$. We take advantage of the fact that

$$\int_{\mathcal{G}}G(\Delta;|x-z|)\ D_\rho^{VP}(|y-z|)\ dz=VP_\rho(|x-y|),\quad x,y\in\overline{\mathcal{G}}. \tag{4.42}$$

Note that

$$\Delta_x\int_{\mathcal{G}}G(\Delta;|x-z|)\ D_\rho^{VP}(|y-z|)\ dz=\Delta_xVP_\rho(|x-y|)=D_\rho^{VP}(|x-y|),\quad x,y\in\overline{\mathcal{G}}. \tag{4.43}$$

After these preliminaries we are able to conclude that

$$\begin{aligned}
I_\rho^{VP}[F](x)=F_\rho^{VP}(x)&=\int_{\mathcal{G}}VP_\rho(|x-w|)F(w)\ dw\\
&=\int_{\mathcal{G}}\left(\int_{\mathcal{G}}G(\Delta;|w-z|)\ D_\rho^{VP}(|x-z|)\right)\ dz\ F(w)\ dw\\
&=\int_{\mathcal{G}}D_\rho^{VP}(|x-z)\left(\int_{\mathcal{G}}G(\Delta;|w-z|)F(w)\ dw\right)\ dz\\
&=\int_{\mathcal{G}}D_\rho^{VP}(|x-z|)\ A[F](z)\ dz=\int_{\mathcal{G}}D_\rho^{VP}(|x-z|)\ V(z)\ dz.
\end{aligned}$$

holds true for $x\in\overline{\mathcal{G}}$, so that

$$F_\rho^{VP}(x)=\int_{\mathcal{G}}D_\rho^{VP}(|x-z|)\ V(z)\ dz,\ \ x\in\mathcal{G}. \tag{4.44}$$

The right-hand side of (4.44) is given analytically when the parameter ρ is chosen appropriately. So, if we define the operator $S_\rho : Y_\rho^{VP} \to X$ in the form

$$F_\rho^{VP} = S_\rho[V] = \int_{\mathcal{G}} D_\rho^{VP}(|\cdot - z|)\, V(z)\, dz, \quad x \in \mathcal{G}, \tag{4.45}$$

then, by (4.44), this operator maps the gravitational potential to mollified solutions of (ITGP). This property motivates the term *mollified inverse* of A used for S_ρ. The discretization of the identity (4.45) given by

$$F_\rho^{VP}(x) \simeq \sum_{i=1}^{N} w_i D_\rho^{VP}(|x - z_i^N|)\, V(z_i^N), \quad z_i^N \in \overline{\mathcal{G}}, \quad x \in \mathcal{G} \tag{4.46}$$

may serve as an alternative to improve local density knowledge from given internal (e.g., borehole) data $V(z_i^N), i = 1, \ldots, N$, where $w_i, i = 1, \ldots, N$, are the known integration weights.

Finally, it should be noted that, more generally, any singular integral (cf. [76, 77]) can be chosen in analogy to the de la Vallée Poussin kernel, i.e., smoothed Haar kernel, as far as its Laplace derivative takes a reasonable role in the mollification context.

4.3. Singular integral-type mollifier method

First we recapitulate the concept of a *singular integral*: Let $\{K_\rho\}_{\rho>0}$ be a family of functions $r \mapsto K_\rho(r), r \geq 0$, satisfying the following conditions:

(i) $K_\rho(r) = 0, \quad r > \rho,$
(ii) $K_\rho(r) \geq 0, \quad r \geq 0,$
(iii) $K_\rho|[0,\rho]$ is of class $C^{(\infty)}$,
(iii) $-\frac{1}{r^2}\frac{d}{dr} r^2 \frac{d}{dr} K_\rho(r)|_{r\in[0,\rho]} \neq 0,$
(iv) $4\pi \int_0^\rho K_\rho(r)\, r^2\, dr = 1.$

Then the family $\{I_\rho\}_{\rho>0}$ of operators $I_\rho : F \mapsto I_\rho[F],\ F \in X, (X = C^{(0)}(\mathbb{R}^3)$ or $X = L^2(\mathbb{R}^3))$, given by

$$I_\rho[F](x) = F_\rho(x) = \int_{\mathbb{R}^3} K_\rho(|x-y|)\, F(y)\, dy = \int_{\mathcal{B}_\rho(x)} K_\rho(|x-y|)\, F(y)\, dy \tag{4.47}$$

is called a *singular integral in* X, if the following approximate identity relation holds true

$$\lim_{\rho\to 0} \|I_\rho[F] - F\|_X = 0 \tag{4.48}$$

for all $F \in X$.

Obviously, an example of a singular integral of the aforementioned type is given by the de la Vallée Poussin kernel.

Let $\mathcal{G}$ be a regular region. Suppose that $\{K_\rho\}_{\rho>0}$ is a kernel constituting a singular integral in L^2-metric, then it is not difficult to show (see, e.g., [76, 77])

that the limit relation

$$\lim_{\rho\to 0}\left(\int_{\mathcal{G}}|I_\rho[F](x)-F(x)|^2\ dx\right)^{\frac{1}{2}}=0 \tag{4.49}$$

holds true for all $F\in L^2(\mathcal{G})$, while, for all $F\in C^{(0)}(\mathcal{G})$, we have

$$\lim_{\rho\to 0}\sup_{x\in\overline{\mathcal{G}}}|I_\rho[F](x)-\alpha(x)\ F(x)|=0, \tag{4.50}$$

where $\alpha(x)$ is the solid angle subtended by the boundary $\partial\mathcal{G}$ at the point $x\in\overline{\mathcal{G}}$.

Correspondingly to the family $\{K_\rho\}_{\rho>0}$ we are led to families $\{G_\rho\}_{\rho>0}$ and $\{D_\rho\}_{\rho>0}$ such that

$$-\Delta_x G_\rho(\Delta;|x-y|)=K_\rho(|x-y|),\quad x,y\in\mathbb{R}^3 \tag{4.51}$$

and

$$-\Delta_x K_\rho(|x-y|)=D_\rho(|x-y|),\quad x,y\in\mathbb{R}^3. \tag{4.52}$$

Our interest now is in the *terrestrial gravimetry problem* (TGP), that may be regarded as particularly relevant problem in geoscientific practice (our considerations, however, remain valid for (ITGP)). We start from known values $V(x_i), x_i\in\partial\mathcal{G}, i=1,\dots,N$, given by

$$A[F](x_i)=\int_{\mathcal{G}}G(\Delta;|x_i-z|)\ F(z)\ dz=V(x_i),\ x_i\in\partial\mathcal{G},\ i=1,\dots,N, \tag{4.53}$$

which can be thought of as resulting from moment discretization of the gravimetry integral equation (cf. (1.1))

$$A[F](x)=\int_{\mathcal{G}}G(\Delta;|x-z|)\ F(z)\ dz=V(x),\ x\in\partial\mathcal{G},\ F\in L^2(\mathcal{G}). \tag{4.54}$$

(TGP) aims at determining an approximation of the function $F\in L^2(\mathcal{G})$ from the N equations (4.53). Introducing the following settings

$${}^{(N)}A[F]:=\left(\int_{\mathcal{G}}G(\Delta;|x_1-y|)\ F(y)\ dy,\ \dots,\int_{\mathcal{G}}G(\Delta;|x_N-y|)\ F(y)\ dy\right), \tag{4.55}$$

$${}^{(N)}v:=(V(x_1),\dots,V(x_N))^T, \tag{4.56}$$

we are able to rewrite the equations (4.53) in operator form as follows:

$${}^{(N)}A:L^2(\mathcal{G})\to\mathbb{R}^N,\ F\mapsto{}^{(N)}v={}^{(N)}A\ [F]. \tag{4.57}$$

We look for an approximate inverse ${}^{(N)}S:\mathbb{R}^N\to L^2(\mathcal{G})$ for ${}^{(N)}A$ in the form

$${}^{(N)}S\ t:=\sum_{i=1}^{N}\underbrace{V(x_i)}_{=t_i}\ D(|x_i-\cdot|) \tag{4.58}$$

in terms of functions $D(|x_i - \cdot|) \in L^2(\mathcal{G}),\ i = \ldots, N$, satisfying

$$\begin{aligned} {}^{(N)}S\ {}^{(N)}A\ [F] &= \sum_{i=1}^{N} \int_{\mathcal{G}} G(\Delta; |x_i - z|) F(z)\ dz\ D(|x_i - \cdot|) \\ &= \int_{\mathcal{G}} F(z) \sum_{k=1}^{N} G(\Delta; |x_i - z|)\ D(|x_i - \cdot|)\ dz. \end{aligned} \tag{4.59}$$

Now the stage is set for explaining the mollifier philosophy, i.e., the sum

$$\sum_{i=1}^{N} G(\Delta; |x - x_i|)\ D(|x_i - y|) \tag{4.60}$$

is understood as discrete version of the "continuous expression"

$$\int_{\mathcal{G}} G(\Delta; |x - z|)\ D(|z - y|)\ dz\ \simeq\ \delta(|x - y|) \tag{4.61}$$

whose "mollifier version" for some family $\{K_\rho\}_{\rho>0}$ constituting a singular integral is given by

$$\int_{\mathcal{G}} G(\Delta; |x - z|)\ D_\rho(|z - y|)\ dz = K_\rho(|x - y|), \tag{4.62}$$

with sufficiently small $\rho > 0$. This observation leads to the sum

$${}^{(N)}S_\rho\ t = \sum_{i=1}^{N} V(x_i)\ D_\rho(|x_i - \cdot|) \tag{4.63}$$

and

$$\begin{aligned} {}^{(N)}S_\rho\ {}^{(N)}A\ [F] &= \sum_{i=1}^{N} \int_{\mathcal{G}} G(\Delta; |x_i - z|) F(z)\ dz\ D_\rho(|x_i - \cdot|) \\ &= \int_{\mathcal{G}} F(z) \sum_{k=1}^{N} G(\Delta; |x_i - z|)\ D_\rho(|x_i - \cdot|)\ dz. \end{aligned} \tag{4.64}$$

as approximations to ${}^{(N)}S\ t$ and ${}^{(N)}S\ {}^{(N)}A\ [F]$, respectively.

Moment method. Next we mention the *finite moment problem for* (TGP). For that purpose we assume that the N potential (volume integral) values

$$\int_{\mathcal{G}} G(\Delta; |x_i - y|)\ F(y)\ dy = V(x_i), \quad x_i \in \partial\mathcal{G},\ i = 1, \ldots, N. \tag{4.65}$$

are known.

The standard solution process (see, e.g., [19, 60]) consists of finding a linear combination in terms of the functions $x \mapsto G(\Delta; |x_i - x|), x \in \mathcal{G}, x_i \in \partial\mathcal{G},\ i = 1, \ldots, N$. In other words, the moment method looks for a function $F \in X_N$ satisfying the conditions (4.65), where X_N is given by

$$X_N := \text{span}_{i=1,\ldots,N}\ G(\Delta; |x_i - \cdot|). \tag{4.66}$$

As a consequence, the moment solution is a harmonic function inside $\mathcal{G}$.

More formally, consider again a semi-discrete observation operator ${}^{(N)}A : L^2(\mathcal{G}) \to \mathbb{R}^N,\ F \mapsto {}^{(N)}v = {}^{(N)}A\,[F]$, of type (4.55), (4.56). Remembering $F \in X_N$ and choosing F as the linear combination

$$F = \sum_{k=1}^{N} \beta_i\ G(\Delta; |x_i - \cdot|) \tag{4.67}$$

we are led to a (uniquely solvable) linear system in the unknowns $\beta_1, \ldots, \beta_N$, viz.

$$\sum_{k=1}^{N} \beta_i \int_{\mathcal{G}} G(\Delta; |x_i - y|)\ G(\Delta; |x_j - y|)\ dy\ =\ V(x_j),\quad j = 1, \ldots, N, \tag{4.68}$$

that turns out to play a central role in the context of minimum norm (spline) interpolation in reproducing kernel Hilbert spaces as discussed in Section 5.

4.4. Backus–Gilbert method

The concept originally proposed by Backus and Gilbert (cf. [5–7]) is that one does not primarily wish to solve the finite moment problem as explained above, but rather one is interested in how well all possible candidates for solution can be recovered pointwise. More specifically, the Backus–Gilbert method is based on a *pointwise minimization criterion*: Keep $y \in \mathcal{G}$ fixed and determine the numbers $\mu_i (= \mu_i(y))$, $i = 1, \ldots, N$, as the solution of the following minimization problem:

$$\int_{\mathcal{G}} |z - y|^2 \left| \sum_{i=1}^{N} \mu_i\ G(\Delta; |x_i - z|) \right|^2 dz \quad \to min. \tag{4.69}$$

subject to $\mu \in \mathbb{R}^N, \mu = (\mu_1, \ldots, \mu_N)^T$ with

$$\int_{\mathcal{G}} \sum_{i=1}^{N} \mu_i\ G(\Delta; |x_i - z|)\ dz = 1. \tag{4.70}$$

It should be remarked that the factor $z \mapsto |z - y|^2, z \in \overline{\mathcal{G}}$, in the integrand of (4.69) is a measure for the concentration of the sum $\sum_{i=1}^{N} \mu_i\, G(\Delta; |x_i - y|)$ around the point $y \in \mathcal{G}$ under consideration. In the literature (see, e.g., [66, 92]), more generally, the term $z \mapsto |z - y|^{2\nu}, z \in \overline{\mathcal{G}}, \nu \geq 1$, is sometimes chosen. In this case, the larger ν, the more concentrated is the sum $\sum_{i=1}^{N} \mu_i\ G(\Delta; |x_i - y|)$ around $y \in \mathcal{G}$.

In matrix-vector nomenclature (thereby omitting the dependence on the fixed, but arbitrary point $y \in \mathcal{G}$) we are able to rewrite the quadratic optimization problem (4.69), (4.70), in the form

$$\mu \cdot\ Q\ \mu \quad \to \min, \tag{4.71}$$

subject to

$$\kappa \cdot \mu\ =\ 1, \tag{4.72}$$

where

$$(Q)_{i,j} := \int_{\mathcal{G}} |z-y|^2 \; G(\Delta; |x_i - z|) \; G(\Delta; |x_j - z|) \; dz, \; i,j = 1, \ldots, N, \tag{4.73}$$

and

$$\kappa_j := \int_{\mathcal{G}} G(\Delta; |x_j - z|) \; dz, \quad j = 1, \ldots, N. \tag{4.74}$$

In fact, (4.71) and (4.72) is a quadratic minimization problem with only one linear equation constraint. We may assume that $\kappa = (\kappa_1, \ldots, \kappa_N)^T$ is different from 0, since otherwise the constraint (4.72) cannot be valid. The introduction of a Lagrange multiplier well known from optimization theory (see, e.g., [110]) can be used to characterize the solvability of the resulting linear $Q\mu - \lambda\kappa = 0$ under the constraint $\kappa \cdot \mu = 1$, i.e., existence and uniqueness. In more detail, from the integral in (4.69), we see that $\mu \cdot Q \, \mu \geq 0$ and $\mu \cdot Q \, \mu = 0$ implies $\sum_{i=1}^{N} \mu_i \, G(\Delta; |x_i - \cdot|) = 0$, so that the linear independence of the system $\{G(\Delta; |x_i - \cdot|)\}_{i=1,\ldots,N}$ shows that Q is positive definite.

Summarizing our results we therefore obtain the following statement:

The symmetric matrix $Q \in \mathbb{R}^{N \times N}$ as defined by (4.73)*is positive definite for every $y \in \mathcal{G}$. The quadratic minimization problem* (4.71) *and* (4.72) *is uniquely solvable. The vector μ is the unique solution of* (4.71) *and* (4.72) *if and only if there exist a real number λ (the Lagrange multiplier) so that $(\mu, \lambda) \in \mathbb{R}^{N+1}$ solves the linear system $Q\mu - \lambda\kappa = 0$ under the constraint $\kappa \cdot \mu = 1$.*

The Lagrange multiplier $\lambda = \mu \cdot Q \, \mu$ represents the minimal value of the quadratic minimization problem.

Consider the unique solution $\mu \in \mathbb{R}^N, \mu = (\mu_1, \ldots, \mu_N)^T, \mu_i = M_i(y), \; i = 1, \ldots, N$, of the quadratic minimization problem (4.71) and (4.72). The *Backus–Gilbert solution* F_N of the discrete version of (TGP)

$$\int_{\mathcal{G}} G(\Delta; |x_i - y|) \; F_N(y) \; dy = V(x_i), \quad x_i \in \partial\mathcal{G}, \; i = 1, \ldots, N \tag{4.75}$$

is defined by

$$F_N(y) = \sum_{i=1}^{N} V(x_i) \; \mu_i, \quad y \in \mathcal{G}. \tag{4.76}$$

The minimal value λ (more accurately, $\lambda(y)$) is called the *spread.*

As already mentioned, the Backus–Gilbert solution (4.76) generally is not a solution of the finite moment problem (4.65). This observation is certainly a disadvantage. Therefore, the question arises if the error may be estimated in an appropriate way (see [60] for related work in one-dimensional context): Let $F \in L^2(\mathcal{G})$ be any solution of the finite moment problem (4.65). Suppose that F_N given by (4.76) is the Backus–Gilbert solution. Then, in connection with (4.70), it

follows that

$$F_N(y) - F(y) = \sum_{i=1}^{N} V(x_i)\ \mu_i \ - \ F(y) \int_{\mathcal{G}} \sum_{i=1}^{N} \ \mu_i\ G(\Delta; |x_i - z|)\ dz$$
$$= \sum_{i=1}^{N} \int_{\mathcal{G}} G(\Delta; |x_i - z|)\ (F(z) - F(y))\ \mu_i\ dz \tag{4.77}$$

holds true. Consequently, we obtain

$$|F_N(y) - F(y)| \leq \int_{\mathcal{G}} \left| \sum_{i=1}^{N} G(\Delta; |x_i - z|)\ \mu_i) \right| \ |F(z) - F(y)|\ dz. \tag{4.78}$$

Under the assumption of Lipschitz-continuity of F in $\overline{\mathcal{G}}$, i.e., the existence of a constant C_F so that

$$|F(z) - F(y)| \ \leq \ C_F\ |z - y|, \quad y, z \in \overline{\mathcal{G}}, \tag{4.79}$$

we are able to deduce that

$$|F_N(y) - F(y)| \leq C_F \int_{\mathcal{G}} \left| \sum_{i=1}^{N} G(\Delta; |x_i - z|)\ \mu_i \right| \ |z - y|\ dz. \tag{4.80}$$

By virtue of the Cauchy–Schwarz inequality we therefore obtain from (4.80)

$$|F_N(y) - F(y)| \leq C_F \int_{\mathcal{G}} 1 \cdot \left| \sum_{i=1}^{N} G(\Delta; |x_i - z|)\ \mu_i \right| \ |z - y|\ dz$$
$$\leq C_F \sqrt{\|\mathcal{G}\|} \ \left(\int_{\mathcal{G}} \left| \sum_{i=1}^{N} G(\Delta; |x_i - z|)\ \mu_i \right|^2 \ |z - y|^2\ dz \right)^{1/2}. \tag{4.81}$$

For $N \in \mathbb{N},\ y \in \mathcal{G}$, we set

$$e_N^2(y) \ := \min \left\{ \int_{\mathcal{G}} \ |Z_N(z)|^2 \ |z - y|^2\ dz : Z_N \in X_N,\ \int_{\mathcal{G}} \ Z_N(z)\ dz = 1 \right\}. \tag{4.82}$$

Thus, we finally arrive at

$$|F_N(y) - F(y)| \leq \ C_F\ \sqrt{\|\mathcal{G}\|}\ e_N(y) \tag{4.83}$$

as pointwise *error estimate of the difference of the solution of the finite moment problem* (4.65) *and the Backus–Gilbert solution* (4.76).

We conclude our considerations with the question if the Backus–Gilbert method admits a *relation to the mollifier method*: Once again, consider the semi-discrete observation operator

$${}^{(N)}A : L^2(\mathcal{G}) \to \mathbb{R}^N,\ F \mapsto {}^{(N)}v = {}^{(N)}A\ [F], \tag{4.84}$$

where

$$ {}^{(N)}A[F] := \left(\int_{\mathcal{G}} G(\Delta; |x_1 - y|)\ F(y)\ dy,\ \ldots, \int_{\mathcal{G}} G(\Delta; |x_N - y|)\ F(y)\ dy \right), \tag{4.85} $$

$$ {}^{(N)}v := (V(x_1), \ldots, V(x_N))^T. \tag{4.86} $$

By virtue of the operator ${}^{(N)}S$ given by

$$ \left({}^{(N)}S\ v \right)(y) = \sum_{k=1}^{N} V(x_i)\ \mu_i(y), \qquad y \in \mathcal{G}, \tag{4.87} $$

we have constructed a left inverse ${}^{(N)}S : \mathbb{R}^N \to L^2(\mathcal{G})$ such that

$$ \begin{aligned} {}^{(N)}S\ {}^{(N)}A\ [F](y) &= \sum_{i=1}^{N} \int_{\mathcal{G}} G(\Delta; |x_i - z|) F(z)\ dz\ \mu_i(y) \\ &= \int_{\mathcal{G}} F(z) \underbrace{\left(\sum_{i=1}^{N} G(\Delta; |x_i - z|)\ \mu_i(y) \right)}_{\simeq\ \delta(|z-y|)} dz, \\ &\simeq F(y). \end{aligned} \tag{4.88} $$

Note that we are formally allowed (in distributional context) to formulate

$$ \begin{aligned} F(y) &= \int_{\mathcal{G}} F(z)\ \delta(|z - y|)\ dz \\ &\simeq \int_{\mathcal{G}} F(z) \int_{\mathcal{G}} G(\Delta; |x - z|)\ M(|x - y|)\ dx\ dz, \end{aligned} \tag{4.89} $$

where, in analogy to (4.52), we have

$$ -\Delta_z\ \delta(z - y|) = M(|z - y|) \simeq -\Delta_z \sum_{i=1}^{N} G(\Delta; |x_i - z|)\ \mu_i(y). \tag{4.90} $$

5. Reproducing Kernel Hilbert Space (RKHS) Methods

Next we consider reproducing kernel Hilbert space solutions. First we discuss the classical geodetic External/Terrestrial Gravimetry Problem (ETGP). Then we go over to the Internal/TerrestrialExternal Gravimetry Problem (ITEGP), i.e., the gravimetry problem in whole Euclidean space $\mathbb{R}^3$.

5.1. External/terrestrial RKHS for regular regions

Let $P_{Harm(\mathcal{G})}$ and $P_{AnHarm(\mathcal{G})}$ be the orthogonal projector of the space $L^2(\mathcal{G})$ to $Harm(\mathcal{G})$ and $\mathcal{N}(A_{\overline{\mathcal{G}^c}}) = AnHarm(\mathcal{G})$, respectively. Then, every function F of the Hilbert space $L^2(\mathcal{G})$ can be uniquely decomposed in the form

$$ F = P_{Harm(\mathcal{G})}[F] + P_{AnHarm(\mathcal{G})}[F] \tag{5.1} $$

such that

$$A_{\overline{\mathcal{G}^c}}[F] = A_{\overline{\mathcal{G}^c}}\left[P_{Harm(\mathcal{G})}[F]\right] + \underbrace{A_{\overline{\mathcal{G}^c}}\left[P_{AnHarm(\mathcal{G})}[F]\right]}_{=0} = A_{\overline{\mathcal{G}^c}}\left[P_{Harm(\mathcal{G})}[F]\right]. \quad (5.2)$$

Furthermore, it is clear that

$$\|F\|^2_{L^2(\mathcal{G})} = \left\|P_{Harm(\mathcal{G})}[F]\right\|^2_{L^2(\mathcal{G})} + \quad \left\|P_{AnHarm(\mathcal{G})}[F]\right\|^2_{L^2(\mathcal{G})}. \quad (5.3)$$

In conclusion, $A_{\overline{\mathcal{G}^c}}[P_{Harm(\mathcal{G})}[F]]$ is that function of class $L^2(\mathcal{G})$ that has the smallest $L^2(\mathcal{G})$-norm among all (density) functions F in $L^2(\mathcal{G})$ generating the same potential in the space $Y|\overline{\mathcal{G}^c} = A_{\overline{\mathcal{G}^c}}(L^2(\mathcal{G}))$. Consequently, to every $P \in Y|\overline{\mathcal{G}^c}$, there corresponds a unique $F \in Harm(\mathcal{G})$ such that

$$A_{\overline{\mathcal{G}^c}}[F] = A_{\overline{\mathcal{G}^c}}[P_{Harm(\mathcal{G})}[F]] = P. \quad (5.4)$$

The restriction $A_{\overline{\mathcal{G}^c}}|Harm(\mathcal{G})$ is a linear bijective operator, i.e., to every $P \in Y|\overline{\mathcal{G}^c}$ there exists a unique $F \in Harm(\mathcal{G})$ such that $A_{\overline{\mathcal{G}^c}}|Harm(\mathcal{G})[F] = P$.

On the space $Y|\overline{\mathcal{G}^c}$ we are able to impose an inner product $\langle\cdot,\cdot\rangle_{Y|\overline{\mathcal{G}^c}}$ by defining

$$\left\langle A_{\overline{\mathcal{G}^c}}|Harm(\mathcal{G})[F], A_{\overline{\mathcal{G}^c}}|Harm(\mathcal{G})[G]\right\rangle_{Y|\overline{\mathcal{G}^c}} = \langle F, G\rangle_{L^2(\mathcal{G})}, \quad (5.5)$$

where $F, G \in L^2(\mathcal{G})$. $Y|\overline{\mathcal{G}^c}$ equipped with the inner product $\langle\cdot,\cdot\rangle_{Y|\overline{\mathcal{G}^c}}$ is a Hilbert space. $A_{\overline{\mathcal{G}^c}}|Harm(\mathcal{G})$ is an isometric operator relating $L^2(\mathcal{G})$ to $Y|\overline{\mathcal{G}^c}$. Our goal is to show that $(Y|\overline{\mathcal{G}^c}, \langle\cdot,\cdot\rangle_{Y|\overline{\mathcal{G}^c}})$ is a reproducing kernel Hilbert space, i.e., a Hilbert space equipped with the reproducing kernel $K_{Y|\overline{\mathcal{G}^c}}(\cdot,\cdot)$. It is clear that, for every $x \in \overline{\mathcal{G}^c}$, $G(\Delta; |x-\cdot|)$ is an element of $Harm(\mathcal{G})$. From well-known reproducing Hilbert space theory (see, e.g., [4]), it follows that any given potential $P \in Y|\overline{\mathcal{G}^c}$ can be represented in the form

$$P(x) = A_{\overline{\mathcal{G}^c}}|Harm(\mathcal{G})[F](x) = \langle\, G(\Delta; |x-\cdot|), F\rangle_{L^2(\mathcal{G})}, \quad x \in \overline{\mathcal{G}^c},\ F \in Harm(\mathcal{G}). \quad (5.6)$$

For $x \in \overline{\mathcal{G}^c}$, the evaluation functional $\mathcal{E}_x[P] = P(x)$ is a bounded functional on $\overline{\mathcal{G}^c}$. Indeed, from the Cauchy–Schwarz inequality applied to (5.6) we get

$$|\mathcal{E}_x[P]| = |P(x)| \le ||F||_{L^2(\mathcal{G})}||G(\Delta; |x-\cdot|)||_{L^2(\mathcal{G})}. \quad (5.7)$$

Consequently, we have

$$|\mathcal{E}_x[P]| = P(x)| \le C_x\ \|P\|_{Y|\overline{\mathcal{G}^c}}, \quad P \in Y|\overline{\mathcal{G}^c},\ x \in \overline{\mathcal{G}^c}. \quad (5.8)$$

Thus, a necessary and sufficient condition for the Hilbert space $Y|\overline{\mathcal{G}^c}$ to possess a reproducing kernel (see, e.g., [4]) is fulfilled. Even more, we are able to find the explicit expression of the reproducing kernel $K_{Y|\overline{\mathcal{G}^c}}(\cdot,\cdot) : \overline{\mathcal{G}^c} \times \overline{\mathcal{G}^c} \to \mathbb{R}$ for the Hilbert space $Y|\overline{\mathcal{G}^c}$ such that, for every $P \in Y|\overline{\mathcal{G}^c}$, the reproducing property

$$P(x) = \left\langle P, K_{Y|\overline{\mathcal{G}^c}}(x,\cdot)\right\rangle_{Y|\overline{\mathcal{G}^c}}, \quad x \in \overline{\mathcal{G}^c}, \quad (5.9)$$

is valid. For $x \in \overline{\mathcal{G}^c}$ and $F \in Harm(\mathcal{G})$ such that $A_{\overline{\mathcal{G}^c}}[F] = P$, we obtain

$$\begin{aligned} P(x) = \langle F, G(\Delta; |x - \cdot|)\rangle_{L^2(\mathcal{G})} &= \langle A_{\overline{\mathcal{G}^c}}[F], A_{\overline{\mathcal{G}^c}}[G(\Delta; |x - \cdot|)]\rangle_{Y|\overline{\mathcal{G}^c}} \\ &= \langle P, A_{\overline{\mathcal{G}^c}}[G(\Delta; |x - \cdot|)]\rangle_{Y|\overline{\mathcal{G}^c}}. \end{aligned} \tag{5.10}$$

Hence, $K_{Y|\overline{\mathcal{G}^c}}(x, \cdot) = A_{\overline{\mathcal{G}^c}}[G(\Delta; |x - \cdot|)]$, i.e., we have for $x, y \in \overline{\mathcal{G}^c}$:

The integral

$$K_{Y|\overline{\mathcal{G}^c}}(x, y) = \langle G(\Delta; |x - \cdot|), G(\Delta; |y - \cdot|)\rangle_{L^2(\mathcal{G})} = \frac{1}{(4\pi)^2} \int_{\mathcal{G}} \frac{1}{|x - z||y - z|}\, dz \tag{5.11}$$

represents the (unique) reproducing kernel of $Y|\overline{\mathcal{G}^c}$.

Clearly, for "geoscientifically relevant geometries" $\mathcal{G}$ such as geoid, real Earth, etc. the integral (5.11) has to be determined by approximate integration rules.

Summarizing our considerations we end up with the following result:

$\left(Y|\overline{\mathcal{G}^c}, \langle\cdot,\cdot\rangle_{Y|\overline{\mathcal{G}^c}}\right)$ *is a Hilbert space possessing the reproducing kernel* (5.11)

$$\begin{aligned} K_{Y|\overline{\mathcal{G}^c}}(x, y) &= \int_{\mathcal{G}} G(\Delta; |x - z|)\; G(\Delta; |y - z|)\; dz \\ &= \frac{1}{(4\pi)^2} \int_{\mathcal{G}} \frac{1}{|x - z||y - z|}\, dz. \end{aligned} \tag{5.12}$$

Equation (5.12) formally states that, for every fixed $x \in \overline{\mathcal{G}^c}$, the function $K_{Y|\overline{\mathcal{G}^c}}(x, \cdot) = A_{\overline{\mathcal{G}^c}}[G(\Delta; |x - \cdot|)]$ is the Newtonian potential corresponding to the "density function" $G(\Delta; |x - \cdot|)$.

5.2. External/terrestrial RKHS for balls

For the special case of a ball $\mathcal{B}_\beta(0)$ of radius β around the origin the kernel $K_{Y|\mathcal{B}^c_\beta(0)}(\cdot,\cdot)$ given by

$$K_{Y|\mathcal{B}^c_\beta(0)}(x, y) = \frac{1}{(4\pi)^2} \int_{\mathcal{B}_\beta(0)} \frac{1}{|x - z||y - z|}\, dz, \tag{5.13}$$

can be expressed as series representation by use of the expansion (see, e.g., [38])

$$G(\Delta; |x - y|) = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{|y|^n}{|x|^{n+1}} P_n\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right),\ |y| < |x|,$$

where P_n is the Legendre polynomial of degree n.

$$K_{Y|\mathcal{B}^c_\beta(0)}(x, y) = \frac{\beta}{4\pi} \sum_{n=0}^{\infty} \frac{1}{(2n + 1)(2n + 3)} \left(\frac{\beta^2}{|x||y|}\right)^{n+1} P_n\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right). \tag{5.14}$$

We are interested in an explicit expression of the infinite Legendre sum (5.14). To this end, we have a closer look at the term

$$\frac{1}{(2n+1)(2n+3)}$$

that can be decomposed via partial fraction decomposition in the form

$$\frac{1}{(2n+1)(2n+3)} = \frac{1}{2(2n+1)} - \frac{1}{2(2n+3)}.$$

As a consequence, the reproducing kernel can be rewritten in the form

$$\begin{aligned} K_{Y|\mathcal{B}^e_\beta(0)}(x,y) = {} & \frac{\beta^3}{8\pi|x||y|}\sum_{n=0}^{\infty}\frac{1}{2n+1}\left(\frac{\beta^2}{|x||y|}\right)^n P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right) \\ & - \frac{\beta^3}{8\pi|x||y|}\sum_{n=0}^{\infty}\frac{1}{2n+3}\left(\frac{\beta^2}{|x||y|}\right)^n P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right). \end{aligned} \tag{5.15}$$

We only consider the Legendre expansions

$$\Phi_1(h,t) = \frac{\beta^3}{8\pi|x||y|}\sum_{n=0}^{\infty}\frac{1}{2n+1}h^{2n}P_n(t),$$

$$\Phi_2(h,t) = \frac{\beta^3}{8\pi|x||y|}\sum_{n=0}^{\infty}\frac{1}{2n+3}h^{2n}P_n(t)$$

for $h := \frac{\beta}{\sqrt{|x||y|}} < 1$ and $t := \frac{x}{|x|}\cdot\frac{y}{|y|} \in (-1,1)$. The remaining cases follow accordingly. Recalling the generating series in terms of the Legendre polynomials (see, e.g., [1, 68])

$$\sum_{n=0}^{\infty} h^{2n}P_n(t) = \frac{1}{\sqrt{1+h^4-2h^2t}}, \quad h\in[0,1),\, t\in[-1,1]. \tag{5.16}$$

we obtain by integration of both sides of (5.16) with respect to h

$$\sum_{n=0}^{\infty}\frac{1}{2n+1}h^{2n+1}P_n(t) = \int\frac{1}{\sqrt{1+h^4-2h^2t}}\,dh = \mathcal{F}_1(h,t), \tag{5.17}$$

where $\mathcal{F}_1(h,t)$ is the abbreviation given by

$$\begin{aligned} \mathcal{F}_1(h,t) = {} & -\frac{i\sqrt{\frac{h^2}{-t+\sqrt{t^2-1}}+1}\sqrt{-\frac{h^2}{t+\sqrt{t^2-1}}+1}}{\sqrt{\frac{1}{-t+\sqrt{t^2-1}}}\sqrt{1+h^4-2h^2t}} \\ & \times\; E_1\left(i\sinh^{-1}\left(h\sqrt{\frac{1}{-t+\sqrt{t^2-1}}},\frac{t-\sqrt{t^2-1}}{t+\sqrt{t^2-1}}\right)\right) \end{aligned} \tag{5.18}$$

and E is an elliptic integral of first kind. Then the sum Φ_1 is given by

$$\Phi_1(h,t) = \frac{\beta^3}{8\pi|x||y|}\frac{1}{h}\mathcal{F}_1(h,t).$$

For the determination of the sum Φ_2 we multiply Equation (5.16) by h^2:

$$\sum_{n=0}^{\infty} h^{2n+2} P_n(t) = h^2 \sum_{n=0}^{\infty} h^{2n} P_n(t) = \frac{h^2}{\sqrt{1+h^4-2h^2t}}.$$

Integrating the last equation with respect to h we obtain

$$\sum_{n=0}^{\infty} \frac{1}{2n+3} h^{2n+3} P_n(t) = \int \frac{h^2}{\sqrt{1+h^4-2h^2t}}\, dh = \mathcal{F}_2(h,t),$$

where $\mathcal{F}_2(h,t)$ is given by

$$\begin{aligned}\mathcal{F}_2(h,t) \;=\; & \frac{(\sqrt{t^2-1}+t)\sqrt{\frac{h^2}{-t+\sqrt{t^2-1}}+1}\sqrt{-\frac{h^2}{t+\sqrt{t^2-1}}+1}}{\sqrt{\frac{1}{-t+\sqrt{t^2-1}}}\sqrt{1+h^4-2h^2t}} \\ & \times \left\{ E_2\left(i\sinh^{-1}\left(h\sqrt{\frac{1}{-t+\sqrt{t^2-1}}}\right), \frac{t-\sqrt{t^2-1}}{t+\sqrt{t^2-1}}\right)\right. \\ & \left. -\; E_3\left(i\sinh^{-1}\left(h\sqrt{\frac{1}{-t+\sqrt{t^2-1}}}\right), \frac{t-\sqrt{t^2-1}}{t+\sqrt{t^2-1}}\right)\right\},\end{aligned}$$

and E_2 and E_3 are elliptic integrals of first and second kind, respectively. Hence, it follows that

$$\Phi_2(h,t) = \frac{\beta^3}{8\pi|x||y|}\frac{1}{h^3}\mathcal{F}_2(h,t).$$

Altogether, for

$$\frac{\beta}{\sqrt{|x||y|}} < 1 \quad \text{and} \quad \frac{x}{|x|}\cdot\frac{y}{|y|} \in (-1,1),$$

the reproducing kernel $K_{Y|\mathcal{B}_\beta^c(0)}(x,y),\ x,y \in \mathcal{B}_\beta^c(0)$, can be represented in the form

$$\begin{aligned}K_{Y|\mathcal{B}_\beta^c(0)}(x,y) \;=\; & \frac{\beta^3}{8\pi|x||y|}\left(\frac{\sqrt{|x||y|}}{\beta}\mathcal{F}_1\left(\frac{\beta}{\sqrt{|x||y|}}, \frac{x}{|x|}\cdot\frac{y}{|y|}\right)\right. \\ & \left. -\; \frac{\sqrt{|x||y|}^3}{\beta^3}\mathcal{F}_2\left(\frac{\beta}{\sqrt{|x||y|}}, \frac{x}{|x|}\cdot\frac{y}{|y|}\right)\right). \qquad (5.19)\end{aligned}$$

5.3. External/terrestrial/internal RKHS for regular regions

Looking at the kernel given by

$$(x,y) \mapsto \int_{\mathcal{G}} G(\Delta;|x-z|)\; G(\Delta;|z-y|)\; dz \qquad (5.20)$$

we see that it is defined for all $x,y \in \mathbb{R}^3$ (with $x,y \in \mathcal{G}^c$ constituting even a regular integral expression). Furthermore, to every $F \in L^2(\mathcal{G})$ there exists a unique $V \in Y$ of the form

$$V(x) = A[F](x) = \int_{\mathcal{G}} G(\Delta;|x-y|)F(y)\; dy, \quad x \in \mathbb{R}^3. \qquad (5.21)$$

On the space Y we are able to impose an inner product $\langle\cdot,\cdot\rangle_Y$ by setting

$$\langle A[F], A[G]\rangle_Y = \langle F, G\rangle_{L^2(\mathcal{G})}, \tag{5.22}$$

where $F, G \in L^2(\mathcal{G})$. Y equipped with the inner product $\langle\cdot,\cdot\rangle_Y$ is a Hilbert space.

For all $x \in \mathbb{R}^3$, the Cauchy–Schwarz inequality yields the estimate

$$|V(x)| \leq \sqrt{\int_{\mathcal{G}} |\ G(\Delta; |x-y|)|^2\ dy}\ \sqrt{\int_{\mathcal{G}} |F(y)|^2\ dy}, \tag{5.23}$$

where we already know that there exists a constant C_x such that

$$|V(x)| \leq C_x\ \sqrt{\int_{\mathcal{G}} |F(y)|^2\ dy}, \tag{5.24}$$

holds true for all $x \in \overline{\mathcal{G}^c}$. Moreover, for all $x \in \mathcal{G}$ and some $R \geq d$ with $d = \mathrm{diam}(\mathcal{G}) = \max_{x,y\in\overline{\mathcal{G}}} |x-y|$ we are able to see that

$$\begin{aligned}
\int_{\mathcal{G}} |G(\Delta; |x-y|)|^2 dy &= \frac{1}{(4\pi)^2} \int_{\mathcal{G}} \frac{1}{|x-y|^2}\ dy \\
&\leq \frac{1}{(4\pi)^2} \int_{\mathcal{B}_R(x)} \frac{1}{|x-y|^2}\ dy \\
&= \frac{1}{(4\pi)^2} \int_0^R \int_{|x-y|=r} \frac{1}{|x-y|^2} dS(y)\ dr \\
&= \frac{R}{4\pi}.
\end{aligned} \tag{5.25}$$

All in all, for each fixed $x \in \mathbb{R}^3$, the evaluation functional $\mathcal{E}_x$ is bounded. Hence, a necessary and sufficient condition that $(Y, \langle\cdot,\cdot\rangle_Y)$ be a reproducing kernel Hilbert space (see, e.g., [4, 16]) is satisfied. In fact, for $x \in \mathcal{G}$ and $F \in L^2(\mathcal{G})$, we obtain

$$\begin{aligned}
V(x) &= \langle G(\Delta; |x-\cdot|), F\rangle_{L^2(\mathcal{G})} \\
&= \langle A[G(\Delta; |x-\cdot|)], A[F]\rangle_Y \\
&= \langle A[G(\Delta; |x-\cdot|)], V\rangle_Y,
\end{aligned} \tag{5.26}$$

so that

$$\begin{aligned}
K_Y(x,y) &= \int_{\mathcal{G}} G(\Delta; |x-z|)\ G(\Delta; |z-y|)\ dz \\
&= \frac{1}{(4\pi)^2} \int_{\mathcal{G}} \frac{1}{|x-z|} \frac{1}{|z-y|}\ dz, \quad x, y \in \mathbb{R}^3
\end{aligned} \tag{5.27}$$

is the unique reproducing kernel of Y. Summarizing our considerations we are finally allowed to formulate the following statement [98]:

The image space $Y = A[L^2(\mathcal{G})]$ is a reproducing kernel Hilbert space processing the reproducing kernel

$$K_Y(x,y) = \int_{\mathcal{G}} G(\Delta; |x-z|)\ G(\Delta; |z-y|)\ dz, \quad x, y \in \mathbb{R}^3. \tag{5.28}$$

Mollifier realization. Denoting by $Y_\rho, \rho > 0$, a space of all *mollified singular integral-type Newton integrals* $A_\rho[F]$ given by

$$A_\rho[F] = \int_{\mathcal{G}} G_\rho(\Delta; |x-y|)\ F(y)\ dy, \quad F \in L^2(\mathcal{G}), \tag{5.29}$$

with $G_\rho(\Delta; |\cdot - \cdot|)$ given by either (2.20) or (4.51), so that $Y_\rho = A_\rho[L^2(\mathcal{G})]$, we are led to an analogous result in the framework of singular integral-type mollification:

The image space $Y_\rho = A_\rho[L^2(\mathcal{G})]$ is a reproducing kernel Hilbert space possessing the reproducing kernel

$$K_{Y_\rho}(x,y) = \int_{\mathcal{G}} G_\rho(\Delta; |x-z|)\ G_\rho(\Delta; |z-y|)\ dz, \quad x, y \in \mathbb{R}^3. \tag{5.30}$$

Finally, it should be mentioned that

$$-\Delta_x\ K_{Y_\rho}(x,y) = \int_{\mathcal{G}} K_\rho(|x-z|)\ G_\rho(\Delta; |z-y|)\ \ dz, \quad x, y \in \mathbb{R}^3. \tag{5.31}$$

Remark. The mathematical structures and results developed for the gravimetry problem enable us to apply a large variety of ideas and concepts known from the theory of ill-posed problems (see, e.g., [35] for a geodetically relevant approach). In our work, we do not consider the details.

However, it should be remarked that reproducing kernel Hilbert space structure is of particular importance in the inversion of Newton's Law of Gravitation, since the reproducing property makes a numerical computation efficient and economical (as we shall see from the following gravimetric spline context).

5.4. External/terrestrial/internal spline theory for regular regions

Let $\mathcal{G}$ be a regular region. Suppose that $\{x_1, \ldots, x_N\}, x_i \neq x_j, i \neq j$, is a discrete set of N given points in $\mathbb{R}^3$. Assume that the values $\gamma_i = V(x_i)$, $x_i \in \mathbb{R}^3$, $i = 1, \ldots, N$, constitute a given data set from the Newton potential (1.1). We want to find an approximation S_N^V to the potential V such that

$$S_N^V(x_i) = V(x_i) = \gamma_i, \quad i = 1, \ldots, N. \tag{5.32}$$

(If the data are noisy, interpolation should be replaced by smoothing (see, e.g., [41] and the references therein).) A functional value $V(x)$ at a point $x \in \mathbb{R}^3$ can be identified with an evaluation functional

$$\mathcal{E}_x : V \mapsto \mathcal{E}_x[V] = V(x), \quad V \in Y(\overline{\mathcal{G}^c}). \tag{5.33}$$

For each $x \in \mathbb{R}^3$, the linear functional $\mathcal{E}_x$ defined by $\mathcal{E}_x : V \mapsto \mathcal{E}_x[V] = V(x), V \in Y$, is bounded on Y, i.e, $|\mathcal{E}_x[V]| = |V(x)| \leq C_x\ \|V\|_Y$. Moreover, for $x \in \partial\mathcal{G}$ and for all $V \in Y$ we have $\mathcal{E}_x[V] = V(x) = (V, K_Y(x, \cdot))_{Y)}$.

Spline method. The Newton potential V, from which the discrete data are known, is considered as an element of the Hilbert space Y possessing the reproducing kernel $K_Y(\cdot, \cdot)$, while the observed values at the points $x_1, \ldots, x_N \in \mathbb{R}^3$ are assumed to be

associated with linearly independent bounded functionals $\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}$. In doing so, we are able to find a minimum norm solution $S_N^V \in Y$ as a linear combination of the representers $\mathcal{E}_{x_i}[K_Y(\cdot,\cdot)]$ to the functionals $\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}$, i.e., S_N^V is meant as the projection of V to the N-dimensional linear subspace spanned by the linearly independent representers $\mathcal{E}_{x_i}[K_Y(\cdot,\cdot)]$, $i = 1, \ldots, N$ (see, e.g., [16]).

Let $\{x_1, \ldots, x_N\} \subset \mathbb{R}^3$ be a point system, such that the evaluation functionals $\mathcal{E}_{x_1}, \ldots, \mathcal{E}_{x_N}$ are linearly independent. Then, within the set

$$\mathcal{I}^V_{\mathcal{E}_{x_1},\ldots,\mathcal{E}_{x_N}} = \{U \in Y : \mathcal{E}_{x_i}[U] = \mathcal{E}_{x_i}[V] = \gamma_i,\ i = 1, \ldots, N\}, \tag{5.34}$$

the minimum norm interpolation problem of finding S_N^V that satisfies

$$\|S_N^V\|_Y = \inf_{U \in \mathcal{I}^V_{\mathcal{E}_{x_1},\ldots,\mathcal{E}_{x_N}}} \|U\|_Y \tag{5.35}$$

is well posed, i.e., its solution exists, is unique and depends continuously on the data $\gamma_1, \ldots, \gamma_N$. The uniquely determined solution S_N^V is given in the explicit form

$$S_N^V(x) = \sum_{i=1}^{N} a_i^N \mathcal{E}_{x_i}[K_Y(x,\cdot)], \quad x \in \mathbb{R}^3, \tag{5.36}$$

where the coefficients $a_1^N, \ldots, a_N^N$ are determined by solving the linear system of equations

$$\sum_{i=1}^{N} a_i^N \mathcal{E}_{x_i}\mathcal{E}_{x_j}[K_Y(\cdot,\cdot)] = \gamma_j, \quad j = 1, \ldots, N. \tag{5.37}$$

As a consequence of the interpolation procedure, the density inside $\mathcal{G}$ is obtained as linear combination in terms of fundamental solutions with singularities in the points $\{x_1, \ldots, x_N\}$:

$$\begin{aligned} S_N^F(x) = -\Delta_x S_N^V(x) &= -\sum_{i=1}^{N} a_i^N \mathcal{E}_{x_i}[G(\Delta; |x - \cdot|)] \\ &= -\sum_{i=1}^{N} a_i^N G(\Delta; |x - x_i|), \quad x \in \mathcal{G}\backslash\{x_1, \ldots, x_N\}. \end{aligned} \tag{5.38}$$

As a consequence, S_N^F is a harmonic function provided that $\{x_1, \ldots, x_N\} \subset \overline{\mathcal{G}^c}$.

Spline mollifier method. For sufficiently small ρ, an approximate version of the kernel

$$K_Y(x,y) = \int_{\mathcal{G}} G(\Delta; |x - z|)\ G(\Delta; |y - z|)\ dz \tag{5.39}$$

is given by

$$K_{Y_\rho}(x,y) = \int_{\mathcal{G}} G_\rho(\Delta; |x - z|)\ G_\rho(\Delta; |y - z|)\ dz. \tag{5.40}$$

Note that, from the integral in (5.40), we see that

$$\sum_{k=1}^{N}\sum_{i=1}^{N} a_k\ a_i\ \int_{\mathcal{G}} G_\rho(\Delta;|x_k-z|)\ G_\rho(\Delta;|x_i-z|)\ dz$$
$$= \int_{\mathcal{G}} \left| \sum_{k=1}^{N} a_k G_\rho(\Delta;|x_k-z|) \right|^2 dz \geq 0. \tag{5.41}$$

Moreover, the linear independence of the system $\{G_\rho(\Delta;|x_i-\cdot|)\}_{i=1,\ldots,N}$ implies that the Gram matrix

$$\left(\int_{\mathcal{G}} G_\rho(\Delta;|x_i-z|)\ G_\rho(\Delta;|x_k-z|)\ dz \right)_{k,i=1,\ldots,N} \tag{5.42}$$

is positive definite, so that $K_{Y_\rho}(\cdot,\cdot)$ is a positive definite kernel.

In other words, the integral (5.40) defines a Hilbert space $\left(Y_\rho, \langle\cdot,\cdot\rangle_{Y_\rho}\right)$ possessing (5.40) as the reproducing kernel. In the space Y_ρ minimum norm (spline) interpolation as described above can be performed in analogous way:

Let $\{x_1,\ldots,x_N\} \subset \mathbb{R}^3$ be a point system, such that the evaluation functionals $\mathcal{E}_{x_1},\ldots,\mathcal{E}_{x_N}$ are linearly independent. Then, within the set

$$\mathcal{I}^V_{\mathcal{E}_{x_1},\ldots,\mathcal{E}_{x_N}} = \{U \in Y^H_{\rho_J}\left(\overline{\mathcal{G}^c}\right) : \mathcal{E}_{x_i}[U] = \mathcal{E}_{x_i}[V] = \gamma_i,\ i=1,\ldots,N\}, \tag{5.43}$$

the minimum norm interpolation problem of finding S^V_N that satisfies

$$\|S^V_N\|_{Y_\rho} = \inf_{U \in \mathcal{I}^V_{\mathcal{E}_{x_1},\ldots,\mathcal{E}_{x_N}}} \|U\|_{Y_\rho} \tag{5.44}$$

is well posed, i.e., its solution exists, is unique and depends continuously on the data $\gamma_1,\ldots,\gamma_N$. The uniquely determined solution S^V_N is given in the explicit form

$$S^V_N(x) = \sum_{i=1}^{N} a^N_i \mathcal{E}_{x_i}[K_{Y_\rho}(x,\cdot)], \quad x \in \mathbb{R}^3, \tag{5.45}$$

where the coefficients $a^N_1,\ldots,a^N_N$ are determined by solving the linear system of equations

$$\sum_{i=1}^{N} a^N_i \mathcal{E}_{x_i}\mathcal{E}_{x_j}[K_{Y_\rho}(\cdot,\cdot)] = \gamma_j, \quad j=1,\ldots,N. \tag{5.46}$$

In this case, we obtain an approximation of the density distribution as a linear combination of singular integral-type kernels (5.31) which are not harmonic.

6. Concluding remarks

Beside gravimetry the (Newton) volume potential (2.10) turns out to appear from different points of view in the context of mathematical geodesy (see, e.g., [26, 29, 31, 32, 34, 38, 52, 55–57, 61, 62, 80, 81, 94]).

Terrestrial measurements of the gravitational force intensity $|v(x)| = |\nabla V(x)|$ typically lead to an *oblique derivative problem for the Earth's gravitational potential* V. The obliqueness is a result of the fact that the real Earth's surface does not coincide with the geoidal surface (except over certain parts on oceans).

Satellite measurements on orbits lead to vectorial and/or tensorial derivatives of first and second order. Depending on the type of measurements, modern satellite problems for the determination of the Earth's external gravitational potential are categorized as *Satellite-to-Satellite Tracking* (SST) problems (i.e., ∇V is derivable from orbit deviations) or *Satellite Gravity Gradiometry* (SGG) problems (i.e., $\nabla^2 V = (\nabla \otimes \nabla) V$ is available by gradiometer measurements on the orbit).

References

[1] Abramowitz, M., Stegun, I.A.: *Handbook of Mathematical Functions*, Dover Publications, Inc., New York, 1964

[2] Anger, G.: A Characterization of Inverse Gravimetric Source Problem through Extremal Measures. *Rev. Geophys. Space Phys.*, **19**:299–306, 1981.

[3] Anger, G.: *Inverse Problems in Differential Equations.* Akademie-Verlag, Berlin, 1990.

[4] Aronszajn, N.: Theory of Reproducing Kernels. *Trans. Am. Math. Soc.*, **68**:337–404, 1950.

[5] Backus, G.E., Gilbert, F.: Numerical Applications of a Formalism for Geophysical Inverse Problems. *Geophys. J.R. Astron. Soc.*, **13**:247–276, 1967.

[6] Backus, G.E., Gilbert, F.: The Resolving Power of Gross Earth Data. *Geophys. J.R. Astron. Soc.*, **16**:169–205, 1968.

[7] Backus, G.E., Gilbert, F.: Uniqueness of the Inversion of Inaccurate Gross Earth Data. *Philos. Trans. R. Soc. London* ., **226**:123–197, 1970.

[8] Ballani, L.: Solving the Inverse Gravimetric Problem: On the Benefit of Wavelets. In: *Geodetic Theory Today, Proceedings of the* 3*rd Hotine–Marussi Symposium on Mathematical Geodesy 1994* (Sansò, F. Ed.), 151–161, Springer, Berlin, 1995.

[9] Ballani, L., Engels, J., Grafarend, E.W.: Global Base Functions for the Mass Density in the Interior of a Massive Body (Earth). *Manuscr. Geod.*, **18**:99–114, 1993.

[10] Ballani, L., Stromeyer, D.: The Inverse Gravimetric Problem: A Hilbert Space Approach. In: *Proceedings of the International Symposium Figure of the Earth, the Moon, and Other Planets* 1982 (Holota, P. Ed.), 359–373, Prague, 1983.

[11] Ballani, L., Stromeyer, D., Barthelmes, F.: Decomposition Principles for Linear Source Problems. In: *Inverse Problems: Principles and Applications in Geophysics, Technology, and Medicine, Math. Res.* 47 (Anger, G., Gorenflo, R., Jochmann, H., Moritz, H., Webers, W., Eds.), Akademie-Verlag, Berlin, 1993.

[12] Barzaghi, R., Sansò, F.: Remarks on the Inverse Gravimetric Problem. *Boll. Geod. Scienze Affini*, **45**:203–216, 1986.

[13] Berkel, P., Fischer, D., Michel, V. : Spline Multiresolution and Numerical Results for Joint Gravitation and Normal Mode Inversion with an Outlook on Sparse Regularization. *Int. J. Geomath.*, **1**:167–204, 2010.

[14] Blakely, R.J.: *Potential Theory in Gravity and Magnetic Applications*. Cambridge University Press, Cambridge, 1996.

[15] Blick, C.: Multiscale Potential Methods in Geothermal Research: Decorrelation Reflected Post-Processing and Locally Based Inversion. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern, 2015.

[16] Davis, P.J.: *Interpolation and Approximation*. Blaisdell, New York, 1963.

[17] Dunford N., Schwartz, J.: *Linear Operators*, Vol. II, Wiley-Interscience, New York, 1963.

[18] Engl, H.: *Integralgleichungen*. Springer Lehrbuch Mathematik, Wien, 1997.

[19] Engl, H.W., Hanke, M., Neubauer, A.: *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.

[20] Engl, H., Louis, A.K., Rundell, W. (Eds.): *Inverse Problems in Geophysical Applications*. SIAM, Philadelphia, 1997, **231**:244–261, 1992.

[21] Folland, G.G.: *Introduction to Partial Differential Equations*, Princeton University Press, Princeton, 1995.

[22] Fredholm, I.: Sur une nouvelle méthode pour la résolution du problème de Dirichlet, *Översigt Kongl. Vetenskaps-Akademiens Förhandlingar* 57, 39–46, 1900.

[23] Freeden, W.: On the Approximation of External Gravitational Potential With Closed Systems of (Trial) Functions. *Bull. Géod.*, **54**: 1–20, 1980.

[24] Freeden, W.: On Approximation by Harmonic Splines. *Manuscr. Geod.*, **6**:193–244, 1981.

[25] Freeden, W.: Interpolation by Multidimensional Periodic Splines, *J. Approx. Theory*, **55**:104–117, 1988.

[26] Freeden, W.: *Multiscale Modelling of Spaceborne Geodata*, B.G. Teubner, Stuttgart, Leipzig, 1999.

[27] Freeden, W.: Geomathematics: Its Role, its Aim, and its Potential. In: *W. Freeden, Z. Nashed, and T. Sonar (Eds.), Handbook of Geomathematics*. 2nd edition, Vol. 1, Chapter 1, 3–78, Springer, New York, Heidelberg, 2015.

[28] Freeden, W., Blick, C.: Signal Decorrelation by Means of Multiscale Methods. *World of Mining*, **65**: 1–15, 2013.

[29] Freeden, W., Gerhards, C.: *Geomathematically Oriented Potential Theory*. Chapman and Hall, CRC Press, Boca Raton, London, New York, 2013.

[30] Freeden, W., Gutting, M.: *Special Functions of Mathematical (Geo-)Physics*. Birkhäuser, Basel, 2013.

[31] Freeden, W., Kersten, H.: *The Geodetic Boundary-Value Problem Using the Known Surface of the Earth*. Veröff. Geod. Inst. RWTH Aachen, **29**, 1980.

[32] Freeden, W., Kersten, H.: A Constructive Approximation Theorem for the Oblique Derivative Problem in Potential Theory. Math. Meth. Appl. Sci., **4**:104–114, 1981.

[33] Freeden, W., Maier, T.: Spectral and Multiscale Signal-to-Noise Thresholding of Spherical Vector Fields. *Computational Geosciences*, **7**(3):215–250,2003.

[34] Freeden, W., Michel, V.: *Multiscale Potential Theory* (*with Applications to Geoscience*). Birkhäuser, Boston, 2004.

[35] Freeden, W., Nashed, M.Z.: Operator-Theoretic and Regularization Approaches to Ill-Posed Problems. *GEM Int. J. Geomath.*, https://doi.org/10.1007/s13137-017-0100-0, 2017.

[36] Freeden, W., Nutz, H.: Satellite Gravity Gradiometry as Tensorial Inverse Problem., *Int. J. Geomath.*, **2**:177–218, 2012.

[37] Freeden, W., Schneider, F.: Regularization Wavelets and Multiresolution, *Inverse Problems*, **14**:493–515, 1998.

[38] Freeden, W., Schreiner, M.: *Spherical Functions of Mathematical Geosciences – A Scalar, Vectorial, and Tensorial Setup*. Springer, Heidelberg, 2009.

[39] Freeden, W., Schreiner, M.: Satellite Gravity Gradiometry (SGG): From Scalar to Tensorial Solution. In: *Handbook of Geomathematics, Vol.* 1 (Freeden, W., Nashed, M.Z., Sonar,T., (Eds.)), 269–302, Springer, Heidelberg, 2010.

[40] Freeden, W., Schreiner, M.: Satellite Gravity Gradiometry (SGG): From Scalar to Tensorial Solution. In: Handbook of Geomathematics, Freeden, W., Nashed, M.Z., and Sonar, T., Eds., Springer, 2nd edition, New York, Heidelberg, 2nd edition, 2015.

[41] Freeden, W. and Witte, B.: A Combined (Spline-)Interpolation and Smoothing Method for the Determination of the Gravitational Potential from Heterogeneous Data. *Bull. Geod.*, **56**:53–62, 1982.

[42] Freeden, W., Michel, V., Nutz, H.: Satellite-to-Satellite Tracking and Satellite Gravity Gradiometry (Advanced Techniques for High-Resolution Geopotential Field Determination). *J. Eng. Math.*, **43**: 19–56, 2002.

[43] Freeden, W., Schneider, F., Schreiner, M.: Gradiometry – An Inverse Problem in Modern Satellite Geodesy, in: Engl, H.W., Louis, A., Rundell, W. (Eds.), *GAMM-SIAM Symposium on Inverse Problems: Geophysical Applications*, 179–239, 1997.

[44] Glockner, O.: *On Numerical Aspects of Gravitational Field Modelling from SST and SGG by Harmonic Splines and Wavelets* (*with Application to CHAMP Data*), Ph.D.-thesis, Geomathematics Group, University of Kaiseslautern, 2001.

[45] Grafarend, E.W.: Six Lectures on Geodesy and Global Geodynamics. In: *Proceedings of the Third International Summer School in the Mountains* (Moritz, H., Sünkel, H., Eds.), 531–685, 1982.

[46] Groetsch, C.W.: *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman, London-Boston, 1984.

[47] Groetsch, C.W.: *Inverse Problems in the Mathematical Science*. Vieweg, Braunschweig, 1993.

[48] Haar, A.: Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.*, **69**:331–371, 1910.

[49] Hadamard, J.: Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton Univ. Bull.*, **13**:49–52, 1902.

[50] Hadamard, J.: *Lectures on the Cauchy Problem in Linear Partial Differential Equations*, Yale University Press, New Haven, 1923.

[51] Hanson, R.J.: A Numerical Methods for Solving Fredholm Integral Equations of the First Kind. *SIAM J. Numer. Anal.* **8**:616–662, 1971.

[52] Heiskanen, W.A., Moritz, H.: *Physical Geodesy*. Freeman, San Francisco, 1967.

[53] Helmert, F.: *Die Mathematischen und Physikalischen Theorien der Höheren Geodäsie* 2. B.G. Teubner, Leipzig, 1884.

[54] Hille, E.: Introduction to the General Theory of Reproducing Kernels, *Rocky Mountain, J. Math.*, **2**:321–368, 1972.

[55] Hörmander, L.: Pseudodifferential Operators. *Comm. Pure Appl. Math.*, **18**:501–517, 1965.

[56] Hörmander, L.: The Boundary Problems of Physical Geodesy. *Archive for Rational Mechanics and Analysis*, **62**:1–52, 1976.

[57] Hofmann-Wellenhof, B., Moritz, H.: *Physical Geodesy*, Springer, Wien, New York, 2005.

[58] Jantscher, L.: *Distributionen*. Walter de Gruyter, Berlin, New York, 1971.

[59] Kellogg, O.D.: *Foundations of Potential Theory*. Frederick Ungar Publishing Company, New York, 1929.

[60] Kirsch, A.: *An Introduction to the Mathematical Theory of Inverse Problems*. 2nd ed., Springer, Heidelberg, 1996.

[61] Koch, K.R., Pope, A.J.: Uniqueness and Existence for the Geodetic Boundary Value Problem Using the Known Surface of the Earth. *Bulletin Géodésique*, **106**:467–476, 1972.

[62] Krarup, T.: *A Contribution to the Mathematical Foundation of Physical Geodesy*. Danish Geodetic Institute, Report No. 44, Copenhagen, 1969.

[63] Lavrentiev, M.M.: *Some Improperly Posed Problems of Mathematical Physics*, Izdat. Sibirsk. Otdel, Akad. Nauk. SSSR, Novosibirsk, 1962, *English Transl., Springer Tracts in Natural Philosophy*, Vol. 11, Springer-Verlag, Berlin, 1967.

[64] Lieusternik, L.A., Sobolev, V.J.: *Elements of Functional Analysis*. New York: Ungar, 1961.

[65] Locker J., Prenter, P.M.: Regularization with Differential Operators. *J. Math. Anal. Appl.*, **74**:504–529, 1980.

[66] Louis, A.K.: *Inverse und schlecht gestellte Probleme*. Teubner, Stuttgart, 1989.

[67] Louis, A.K., Maass, P.: A Mollifier Method for Linear Equations of the First Kind. *Inverse Problems*, **6**:427–440, 1989.

[68] Magnus, W., Oberhettinger, F., Soni, R.P.: *Formulas and Theorems for the Special Functions of Mathematical Physics*. In: Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Band 52, Springer, Berlin, 3. Auflage, 1966.

[69] Meissl, P.A.: *A Study of Covariance Functions Related to the Earth's Disturbing Potential*. Department of Geodetic Science, No. 151, The Ohio State University, Columbus, OH, (1971).

[70] Meissl, P.A.: Hilbert Spaces and Their Applications to Geodetic Least Squares Problems. *Boll. Geod. Sci. Aff.*, N. 1, 181–210, 1976.

[71] Michel, V.: *A Multiscale Method for the Gravimetry Problem: Theoretical and Numerical Aspects of Harmonic and Anharmonic Modelling*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen, 1999.

[72] Michel, V.: Scale Continuous, Scale Discretized and Scale Discrete Harmonic Wavelets for the Outer and the Inner Space of a Sphere and Their Application to an Inverse Problem in Geomathematics. *Appl. Comp. Harm. Anal.* (*ACHA*), **12**:77–99, 2002.

[73] Michel, V.: *A Multiscale Approximation for Operator Equations in Separable Hilbert Spaces – Case Study: Reconstruction and Description of the Earth's Interior*. Habilitation Thesis, University of Kaiserslautern, Geomathematics Group, Shaker, Aachen, 2002a.

[74] Michel, V: Regularized Wavelet-Based Multiresolution Recovery of the Harmonic Mass Density Distribution from Data of the Earth's Gravitational Field at Satellite Height. *Inverse Problems*, **21**:997–1025, 2005.

[75] Michel, V., Fokas, A.S.: A Unified Approach to Various Techniques for the Non-Uniqueness of the Inverse Gravimetric Problem and Wavelet-Based Methods. *Inverse Problems*, **24**: doi:10.1088/0266–5611/24/4/045019, 2008.

[76] Michlin, S.G.: *Multidimensional Singular Integrals and Integral Equations*. Pergamon Press, New York, 1965.

[77] Michlin, S.G.: *Lehrgang der Mathematischen Physik*. 2nd edition, Akademie-Verlag, Berlin, 1975.

[78] Möhringer, S.: Decorrelation of Gravimetric Data. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern, 2014.

[79] Moritz, H.: *Advanced Physical Geodesy*. Herbert Wichmann Verlag, Karlsruhe, Abacus Press, Tunbridge, 1980.

[80] Moritz, H.: *The Figure of the Earth. Theoretical Geodesy of the Earth's Interior*. Wichmann Verlag, Karlsruhe, 1990.

[81] Moritz, H.: Classical Physical Geodesy. In: *W. Freeden, Z. Nashed, and T. Sonar, Eds., Handbook of Geomathematics*. Springer, Heidelberg, 253–290, 2015.

[82] Morozov, V.A.: *Methods for Solving Incorrectly Posed Problems*. Springer, New York, 1984.

[83] Nashed, M.Z.: Generalized Inverses, Normal Solvability and Iteration for Singular Operator Equations. In: *Nonlinear Functional Analysis and Applications*, L.B. Rall, ed., Academic, New York, 311–359, 1971.

[84] Nashed, M.Z.: Aspects of Generalized Inverses in Analysis and Regularization. In: *Generalized Inverses and Applications*. Academic Press, New York, 193–244, 1976.

[85] Nashed, M.Z.: New Applications of Generalized Inverses in System and Control Theory. In: *Proc.* 1980 *Conf. on Information Sciences and Systems, J.B. Thomas. ed. Princeton.* NJ, Princeton, 353–358, 1980

[86] Nashed, M.Z.: Operator-Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory. *IEEE Trans. Antennas Propagation*, **29**:220–231, 1981.

[87] Nashed, M.Z.: A New Approach to Classification and Regularization of Ill-Posed Operator Equations. In: H. Engl and C.W. Groetsch, Eds., Inverse and Ill-Posed Problems, Band 4, *Notes and Reports in Mathematics and Science and Engineering*, Academic Press, Boston, 1987.

[88] Nashed, M.Z.: *Inverse Problems, Moment Problems and Signal Processing: Un Menage a Trois, Mathematics in Science and Technology*. Siddiqi, A.H., Singh, R.C. Manchanda, P. (Eds.), World Scientific, New Jersey, 1–19, 2010.

[89] Nashed, M.Z., Votruba, F.G.: A Unified Operator Theory of Generalized Inverses. In: *Generalized Inverses and Applications*. M.Z. Nashed ed., Academic Press, New York, 1–109, 1976.

[90] Parker, R.L.: The Theory of Ideal Bodies for Gravity Interpretation. *Geophys. J. R. Astr. Soc.*, **42**:315–334, 1975.

[91] Petrini, H.: Sur l'existence des dérivées secondes du potentiel. *C.R. Acad. Sci. Paris*, **130**:233–235, 1900.

[92] Rieder, A.: *Keine Probleme mit Inversen Problemen*. Vieweg 2003.

[93] Rummel, R.: *Geodesy*, In: Encyclopedia of Earth System Science, Volume 2, 253–262, Academic Press, 1992.

[94] Rummel, R.: GOCE: Gravitational Gradiometry in a Satellite. In: *Handbook of Geomathematics, Vol.* 1. (Freeden, W., Nashed, M.Z., Sonar,T., (Eds.)), 98–105, Springer, Heidelberg, 2010.

[95] Rummel R., Balmino, G., Johannessen, J., Visser, P., Woodworth P.: Dedicated Gravity Field Missions – Principles and Aims. *J. Geodyn.*, **33**:3–20, 2002.

[96] Rummel, R., van Gelderen, M.: Meissl Scheme – Spectral Characteristics of Physical Geodesy. Manuscr. Geod. **20**:379–385, 1995.

[97] Saitoh, S.: *Theory of Reproducing Kernels and its Applications*. Longman, New York, 1988.

[98] Sansò, F., *Internal Collocation*. Atti Della Accademia Nazionale Dei Lincei, Serie VIII, Volume XVI, 2–52, 1980.

[99] Sansò, F., Rummel, R. (Eds.): *Geodetic Boundary Value Problems in View of the One Centimeter Geoid*. Lecture Notes in Earth Sciences, Vol. 65, Springer, Berlin, Heidelberg, 1997.

[100] Sansò, F., Tscherning, C.C.: The Inverse Gravimetric Problem in Gravity Modelling. In: *Festschrift to Torben Krarup* (Kejlsø, E., Poder, K., Tscherning, C.C., Eds.), 299–334, Geodaetisk Institute, Copenhagen, 1989.

[101] Sansò, F., Barzaghi, R., Tscherning, C.C.: Choice of Norm for the Density Distribution of the Earth. *Geophys. J. R. Astr. Soc.* **87**:123–141, 1986.

[102] Shure, L., Parker, R.L., Backus, G.E.: Harmonic Splines for Geomagnetic Modelling. *Phys. Earth Planet. Inter.* **28**: 215–229, 1982.

[103] Skorvanek, M.: The Inverse Gravimetric Problem for the Earth. In: *Proceedings of the* 4*th International Symposium on Geodesy and Physics of the Earth* 1980, 464–475, *Veröff. Zentralinst. Physik der Erde*, Vol. **63**, 1981.

[104] Tikhonov, A.N.: On the Stability of Inverse Problems. *Dokl. Akad. Nauk SSSR*, **39**:195–198, 1943.

[105] Tikhonov, A.N.: On the Solution of Incorrectly Formulated Problems and the Regularization Method, *Dokl. Akad Nauk SSSR*, **151**:501–504, 1963.

[106] Tscherning, C.C.: Analytical and Discrete Inversion Applied to Gravity Data. In: *Proceedings of the Interdisciplinary Inversion Workshop* 1*, Methodology and Application Perspectives in Geophysics, Astronomy and Geodesy.* (Holm Jacobsen, B., (Ed.)), 5–8, Aarhus, 1992.

[107] Tscherning, C.C., Strykowski, G.: Quasi-Harmonic Inversion of Gravity Field Data, Model Optimization in Exploration Geophysics 2. In: *Proceedings of the* 5*th International Mathematical Geophysics Seminar* (Vogel, A., Ed.), 137–154, Vieweg, Braunschweig, Wiesbaden, 1987.

[108] Vogel, C.R.: *Computational Methods for Inverse Problems.* SIAM, Philadelphia, 2002.

[109] Weck, N.: Zwei inverse Probleme in der Potentialtheorie. In: *Mitt. Inst. Theor. Geodäsie, Universität Bonn*, **4**:27–36, 1972.

[110] Werner, J.: *Optimization Theory and Applications.* Vieweg-Verlag, Braunschweig, Wiesbaden 1984.

[111] Xia X.G., Nashed M.Z.: The Backus–Gilbert Method for Signals in Reproducing Hilbert Spaces and Wavelet Subspaces. *Inverse Problems*, **10**:785–804, 1994.

[112] Yosida, K.: *Functional Analysis.* 5th ed., Springer, Berlin, 1965.

[113] Zidarov, D.P.: Some Uniqueness Conditions for the Solution of the Inverse Gravimetric Problem. *Comptes rendus de l'Académie bulgare des Sciences*, **33**:909–912, 1980.

[114] Zidarov, D.P.: Conditions for Uniqueness of Self-Limiting Solutions of the Inverse Problems. *Comptes rendus de l'Académie bulgare des Sciences*, **39**:57–60, 1986.

[115] Zidarov, D.P.: *Inverse Gravimetric Problem in Geoprospecting and Geodesy.* Developments in Solid Earth Geophysics, **19**, Elsevier, Amsterdam, 1990.

Willi Freeden
Geomathematics Group
University of Kaiserslautern
MPI-Gebäude, Paul-Ehrlich-Str. 26
D-67663 Kaiserslautern, Germany

M. Zuhair Nashed
Mathematics Department
University of Central Florida
Orlando, USA

Handbook of Mathematical Geodesy
Geosystems Mathematics, 687–751

Gravimetry and Exploration

C. Blick, W. Freeden, and H. Nutz

Abstract. In this work we are especially concerned with the "mathematization" of gravimetric exploration and prospecting. We investigate the extractable information of the Earth's gravitational potential and its observables obtained by gravimetry for gravitational modeling as well as geological interpretation. More explicitly, local gravimetric data sets are exploited to visualize multiscale reconstruction and decorrelation features to be found in geophysically and geologically relevant signature bands.

Keywords. Absolute and relative gravimetry, Newtonian gravitational approach, multiscale density field modeling.

Introduction

Newton's famous law about the mutual attraction of two masses formulated in "De mundi systemate" 1715 tells us that the attractive force, called *gravitation*, is directed along the line connecting the two centers of mass of the objects and is proportional to two masses as well as to the squared inverse of the distance between the objects. If the Earth had a perfectly spherical shape and if the mass inside the Earth were distributed homogeneously or rotationally symmetric, the line along which an apple fell would indeed be a straight line, directed radially and going exactly through the Earth's center of mass. The gravitational field obtained in this way would be perfectly spherically symmetric. In reality, however, the situation is more complex. The topographic features, mountains and valleys, are very irregular. The actual gravitational field is influenced by strong irregularities in density within the Earth. As a result, the gravitational force deviates from one place to the other from that of a homogeneous sphere. More explicitly, internal density signatures are reflected in gravitational field signatures, and orthogonal coefficients in terms of spherical harmonics of gravitational field signatures smooth out exponentially with increasing distance from the Earth's body. As a consequence, positioning systems are ideally located as far as possible from the Earth, whereas gravity field sensors are ideally located as close as possible to the Earth. Following these basic

principles, various positioning and gravity field determination techniques have been designed. Sensors may be sensitive to local or global features of the gravity field. Considering the spatial location of the data points, we may differentiate between terrestrial (surface), airborne, and spaceborne methods.

Concerning gravity on a global scale, e.g., for global geoid determination (that will not be investigated here), it should be pointed out (see, e.g., [6–8]) that the terrestrial distribution of Earth's gravity data is far from being homogeneous with large gaps, in particular over oceans but also over land. In addition, the quality of the data is very distinct. Thus, global terrestrial gravity data coverage now and in the foreseeable future is far from being satisfactory. This is the reason why spaceborne measurements come into play for global gravity determination. Until now, the relatively poor precision of satellite-only spaceborne gravity measurements has hindered a wider use of this type of measurements for local purposes. Seen from future exploration aspects, however, it must be remarked that only coordinated research will provide a breakthrough in modeling and understanding significant structures and processes in the Earth's interior. In fact, the authors are convinced that the way forward, even in global modeling, has to be based on two requirements:

i) combining data from different sensors and sources,
ii) multiscale modeling, i.e., "zooming-in downward continuation" of the different data sources starting from globally available spaceborne data as means for an appropriate trend solution via more accurate (regional) airborne data down to (local) high-precision gravimetric data sets.

In this contribution, we are especially concerned with the mathematical study of gravimetry in exploration, in particular postprocessing of all already available models. To this end we briefly explain the status quo of gravimetric observation and standard modeling. On the basis of these results we present new multiscale methods by means of geoscientifically relevant wavelets for the decorrelation of signatures inherent in geological information.

1. Gravity, gravitation, and gravimetry

The force of gravity provides a directional structure to the space above the Earth's surface. It is tangential to the vertical plumb lines and perpendicular to all (level) equipotential surfaces. Any water surface at rest is part of a level surface. (Level) equipotential surfaces are ideal reference surfaces, for example, for heights. The geoid is defined as that level surface of the gravity field which best fits the mean sea level.

The direction of the gravity vector can be obtained by astronomical positioning. Measurements are possible on the Earth's surface. Observations of the gravity vector are converted into so-called vertical deflections by subtracting a corresponding reference direction derived from a simple gravity field model associated to, e.g., an ellipsoidal surface. Vertical deflections are tangential fields

of the anomalous potential. Due to the high measurement effort required to acquire these types of data compared to a gravity measurement, the data density of vertical deflections is much less than that of gravity anomalies. Gravitational field determination based on the observation of deflections of the vertical and combined with gravity is feasible in smaller areas with good data coverage.

1.1. Gravitational, centrifugal, and gravity acceleration

The *gravity acceleration* (*gravity*) w is the resultant of the *gravitation* v and the *centrifugal acceleration* c such that

$$w = v + c. \tag{1.1}$$

The centrifugal force c arises as a result of the rotation of the Earth about its axis. In this work concerned with local gravity exploration we are allowed to assume a rotation of constant angular velocity ω. The centrifugal acceleration acting on a unit mass is directed outward perpendicularly to the spin axis (see Figure 1.1). Introducing the so-called *centrifugal potential* C, such that $c = \nabla C$, the function C turns out to be non-harmonic. The direction of the gravity w is known as the direction of the *plumb line*, the quantity $|w|$ is called the *gravity intensity* (often also just called *gravity* and denoted in the geodetic jargon by g). Altogether, the *gravity potential of the Earth* can be expressed in the form

$$W = V + C, \tag{1.2}$$

and the gravity acceleration w is given by

$$w = \nabla W = \nabla V + \nabla C. \tag{1.3}$$

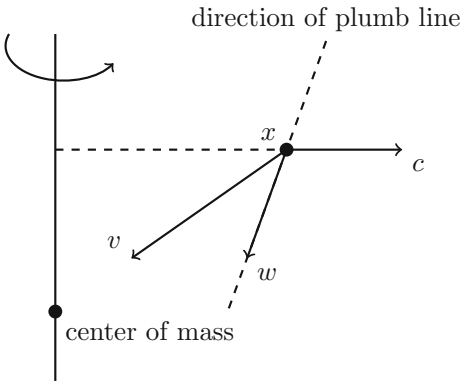


FIGURE 1.1. Gravitation v, centrifugal acceleration c, gravity acceleration w.

As already pointed out, the surfaces of constant gravity potentials, i.e., $W =$ const., are designated as *equipotential* (*level,* or *geopotential*) *surfaces of gravity* (for more details, the reader is referred to monographs in physical geodesy, e.g., [27, 32, 51, 69]).

The actual Earth's surface (globally available from modern spaceborne techniques such as DOPPLER, GNSS, LASER, VLBI, etc.) does not coincide with an equipotential surface (i.e., a level surface). The force of gravity is generally not perpendicular to the actual Earth's surface (see Figure 1.2). However, we are confronted with the gravity intensity as an oblique derivative on the Earth's surface. The gravity vector is an oblique vector at any point on the Earth's surface and generally not the normal vector.

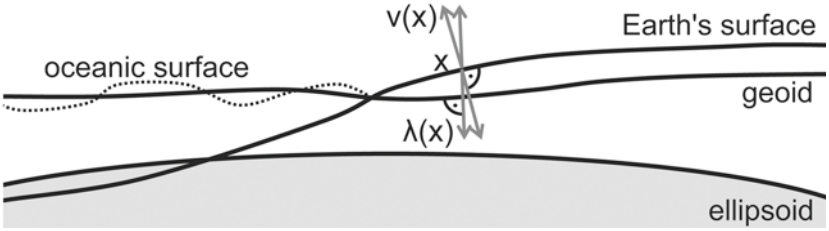


FIGURE 1.2. Earth's surface, geoid, ellipsoid ($\lambda = \frac{w}{|w|}$ oblique unit gravity vector, normal vector to the geoid, but usually not normal to the Earth's surface).

The determination of equipotential surfaces of the potential W is strongly related to the knowledge of the potential V. The gravity vector w given by $w = \nabla W$ is normal to the equipotential surface passing through the same point. Thus, equipotential surfaces such as the geoid intuitively express the notion of tangential surfaces, as they are normal to the plumb lines given by the direction of the gravity vector.

1.2. Gravimeter and gravimetry

Essentially, the Earth is a "spheroid", with a slight flattening (0.35%) at the poles, a mean radius of 6368km, and a mean mass of $5.98 \cdot 10^{24}$kg. At the surface of the Earth, its mean value of gravity intensity is given by $9.80\frac{m}{s^2}$. At the equator, it reduces to $9.78\frac{m}{s^2}$; at the poles, it increases to about $9.83\frac{m}{s^2}$, reflecting the flattening.

Gravimeters are typically designed to measure very tiny fractional changes of the Earth's gravity, caused by nearby geologic structures or the shape of the Earth. There are two types of gravimeters, viz. *relative* and *absolute gravimeters*. Absolute gravimeters measure the local gravity and are directly based on measuring the acceleration of free fall (for example, of a test mass in a vacuum tube). Relative gravimeters compare the value of gravity at one point with another. They must be calibrated at a location, where the gravity is known accurately and measure the ratio of the gravity at the two points. Most common relative gravimeters are spring-based. By determining the amount by which the weight stretches the spring, gravity becomes available via Hooke's law (see Figure 1.3). The highest possible accuracy of relative gravity measurements are conducted at the Earth's surface. Measurements on ships and in aircrafts deliver reasonably good data only after the removal of inertial noise. In addition, when interested in gravimetric exploration,

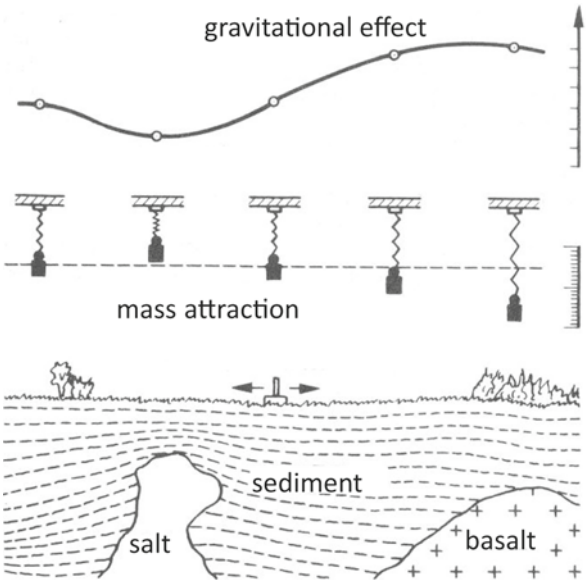


FIGURE 1.3. The principle of gravimetry (with kind permission of Teubner-publishing taken from [38] in modified form).

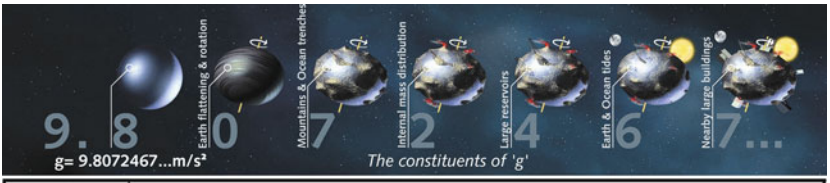
it should be noted that a high measurement accuracy of at least 0.1mGals, but more adequately 0.01mGals (1 mGal $= 10^{-5}\frac{m}{s^2}$, cf. Table 1) has to be achieved (cf. Figure 1.4).

SI Units	Traditional Units
$10^{-2}ms^{-2}$	1 Gal
$10^{-5}ms^{-2}$	1 mGal
$10^{-8}ms^{-2}$	1 μGal

TABLE 1. Traditional units for gravimetric measurements and their SI unit complement.

By *gravimetry*, we denote the determination of the Earth's mass density distribution from data of the gravitational potential or related quantities. Clearly, for purposes of exploration, it is obvious that the determination of gravity intensities as well as gravity anomalies of dimension very much larger than the gravity anomalies caused by, e.g., aquifers, oil and gas structures are of less significance. The fundamental interest in gravimetric methods in exploration is based on the measurements of small variations.

Gravity prospecting has been first used in the case of strong density contrasts in a geological structure, and the usual approach is to measure differences in gravity from place to place. Today, the interpretation of gravimetric data is done by comparing the shape and size of gravity disturbances and anomalies to those



10^{0}	Spherical Earth
10^{-3}	Flattening & Centrifugal Acceleration
10^{-4}	Mountains, Valleys, Oceanic Ridges, Subduction
10^{-5}	Density Variations in Crust and Mantle
10^{-6}	Salt Domes, Sedimentary Basins, Deposits
10^{-7}	Tide, Atmospherical Pressure
10^{-8}	Temporal Variations: Oceans, Hydrology
10^{-9}	Ocean Topography, Movement of Poles

FIGURE 1.4. Illustration of the components of the gravity acceleration (ESA medialab, ESA communication production SP–1314).

caused by bodies of various geometrical shapes at different depths and differing densities.

The observed gravity depends on the following effects to be removed (for more detailed studies, see, e.g., [53, 59, 64]): attraction of the reference ellipsoid, elevation above sea level, topography, time dependent variations (tidal), (Eőtvős) effect of a moving platform, isostatic balance on the lower lithosphere, density variations inside the upper crust. To isolate the effects of local density variations from all other contributions, it is necessary to apply a series of reductions: The attraction of, e.g., the reference ellipsoid or another reference surface has to be subtracted from the measured values. An elevation correction must be done, i.e., the vertical gradient of gravity is multiplied by the elevation of the station and the result is added. With increasing elevation of the Earth, there is usually an additional mass between the reference level and the actual level. This additional mass itself exerts a positive gravitational attraction. Bouguer correction and terrain correction are applied to correct for the attraction of the slab of material between the observation point and the geoid. A terrain correction accounts for the effect of nearby masses above or mass deficiencies below the station. Isostatic correction accounts for the isostatic roots (Moho). Other corrections have to be applied to the data in order to account for effects not related to the subsurface: Drift corrections are necessary, since each gravimeter suffers mechanical changes over time, and so does its output measurement. This change is generally assumed

to be linear. Tidal corrections have to be imposed, i.e., the attraction of the Sun and Moon has to be calculated and subtracted from the measurements. In case of acquisition on a moving platform, the motion relative to the surface of the Earth implies a change in centrifugal acceleration. The Eőtvős correction depends on the latitude and velocity vector of the moving platform. It should be observed that free air anomaly does not correct for the first two effects which could mask the gravity anomalies related to the Bouguer density contrasts in the crust. Complete Bouguer correction effectively remove the gravity anomalies due to bathymetry, but still contain the gravity effect of the Moho. Isostatics contain the gravity effect of the Moho. For more details the reader is referred to geodetic textbooks such as [32, 36, 68] and to the literature concerned with prospecting and exploration (see, e.g., [53, 54, 64] and the references therein).

Gravity prospecting can be done over land or sea areas using different techniques and equipment. Terrestrial gravimetry was first applied *to prospect for salt domes* (*e.g., in the Gulf of Mexico*) (an example of the Eastern part of Germany is shown in Figure 1.5), and later for looking for *anticlines in continental areas* (see, e.g., [53, 54], and the references therein). Nowadays, gravimetry is in use all over the world in diverse applications:

(1) Gravimetric surveys serve regional geological mapping.
(2) Gravimetry is helpful in different phases of the oil exploration and production processes.
(3) Gravimetric surveys are employed in mineral exploration, for example, to detect mineral deposits (see Figure 1.5) of economic interest (such as metals, salt, coal).
(4) Archaeological and geotechnical studies aim at the mapping of subsurface voids and overburden variations.
(5) Gravimetric campaigns may be applied for groundwater and environmental studies. They help to map aquifers to provide formations and/or structural control.
(6) Gravimetric studies give information about tectonically derived changes and volcanological phenomena.
(7) Gravimetric studies provide useful information on changes in the level of water in geothermal reservoirs and therefore on the longevity of a geothermal resource.

It is surprising that the use of gravimetry is in infancy in the German geothermal scene, although it has much to offer. Due to (regional) airborne and (global) spaceborne gravity information such as satellite-to-satellite tracking (SST) and/or satellite gravity gradiometry (SGG), new promising components in gravimetrically oriented modeling can be expected in the future, for example, based on multiscale modeling providing reconstruction and decomposition of geological signatures, where seismic modeling is difficult or impossible because of anthropogenic activities, e.g., in mining areas.

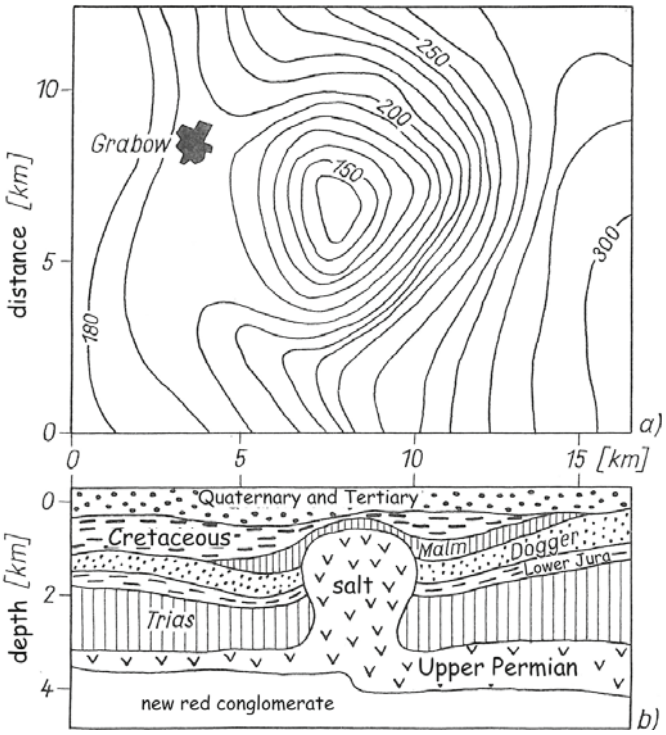


FIGURE 1.5. Top: Gravity effect in $[\mu m \cdot s^{-1}]$ of the salt dome Werle (Mecklenburg, Germany); bottom: Geological vertical profile (with kind permission of Teubner-publishing taken from [38] in modified form).

All in all, nowadays the main applications of gravimetry can be listed as follows:

(i) definition of geological structural settings,
(ii) faults delineation,
(iii) recovery of salt bodies, metal deposits,
(iii) detection of heap of coal, ore, etc.,
(iv) 2D/3D forward modeling, inversion, and postprocessing to assist seismic modeling,
(v) combination with geomagnetic interpretation,
(vi) 4D monitoring, etc.

Figure 1.6 shows the gravity as well as the vertical/horizontal gradient curves induced by a simple geological structure in sectional illustration (for similar illustrations, the reader is referred to, e.g., [5, 26, 39, 48, 53, 54, 64]). It is remarkable that the vertical/horizontal gradient curves show significant interactions on density variations.

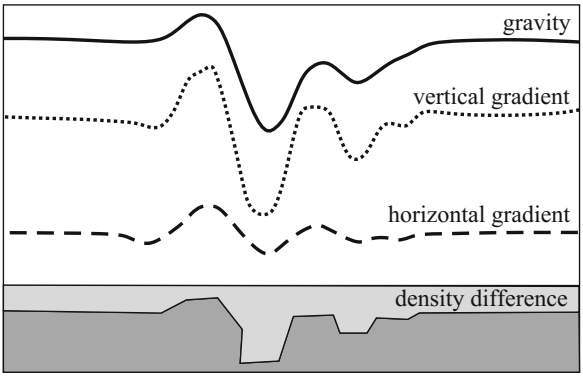


FIGURE 1.6. Schematic diagram of the horizontal/vertical gradients vs. the gravity potential (cf. [54]).

The knowledge of horizontal/vertical derivatives of the gravity potential is therefore a useful addendum to prospecting and exploration. This is the reason why we are interested in discussing these derivatives in more detail (based on ideas and concepts developed in [15]).

2. Surface horizontal/vertical derivatives of the gravity potential

The lines that intersect all equipotential surfaces orthogonally are not exactly straight but slightly curved (cf. Figure 2.1). They are called lines of gravity force or *plumb lines*. The gravity vector at any point is tangential to the plumb line. Hence, "direction of the gravity vector", "vertical" and "direction of the plumb line" are synonymous. As the equipotential surfaces are, so to speak, "horizontal", i.e., orthogonal to the plumb lines, they play an important part in our daily life (e.g., in civil engineering for the purpose of height determination). Equipotential surfaces of the Earth's gravity potential W allow, in general, no simple mathematical representation. This is the reason why physical geodesy and geophysics choose a suitable reference surface for modeling the geoid, i.e., the equipotential surface at sea level. The reference surface is constructed as an equipotential surface of an artificial *normal gravity potential* U . Its gradient field, i.e., $u = \nabla U$, is called *normal gravity*. For reasons of simplicity, physical geodesy usually uses an ellipsoid of revolution in such a way that a good adaption to the Earth's surface is guaranteed. Closed representations of normal gravity potentials, in consideration of the centrifugal force, can be found extensively in the geodetic literature (cf. [25, 27, 32, 36, 45, 51, 65]), and the references therein). The deviations of the gravity field of the Earth from the normal field of such an ellipsoid are small. The remaining parts of the gravity field are gathered in a so-called *disturbing gravity field* ∇T corresponding to the *disturbing potential* $T = W - U$.

2.1. Gravity anomalies, gravity disturbances, and vertical deflections

Knowing the gravity potential, all equipotential surfaces (including the geoid) are given by an equation of the form $W(x) = \text{const}$. By introducing U as the normal gravity potential corresponding to the ellipsoidal field, the disturbing potential T is the difference of the gravity potential W and the normal gravity potential U, i.e., we are led to a decomposition of the gravity potential in the form $W = U + T$. According to the concept developed by Stokes [65], Helmert [33], and Pizzetti [56, 57] we may assume that

(a) the center of the ellipsoid coincides with the center of gravity of the Earth,
(b) the difference of the mass of the Earth and the mass of the reference body (ellipsoid) is zero.

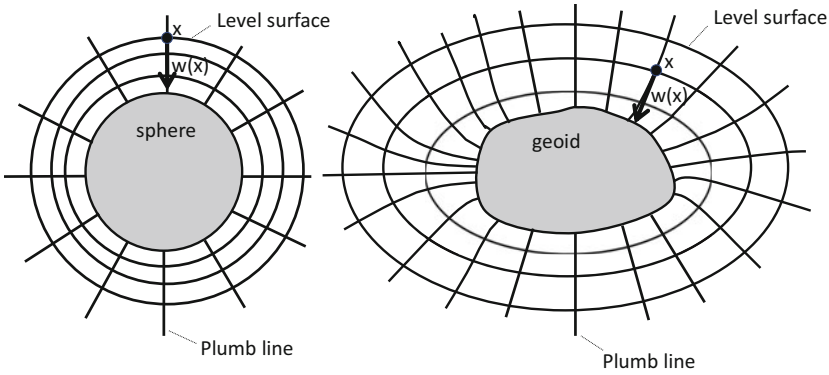


FIGURE 2.1. Level surfaces and plumb lines for a homogeneous ball (left) and an Earth-like body (right) (from [15]).

A point x of the geoid can be projected onto its associated point y of the ellipsoid by means of the ellipsoidal normal. The distance $N(x)$ between x and y is called the *geoidal height* or *geoidal undulation* in x (cf. Figure 2.2). The *gravity anomaly vector* $a(x)$ at the point x of the geoid is defined as the difference between the gravity vector $w(x)$ and the normal gravity vector $u(y)$, i.e.,

$$a(x) = w(x) - u(y). \tag{2.1}$$

Another possibility is to form the difference between the vectors w and u at the same point x such that we get the *gravity disturbance vector* $d(x)$ defined by

$$d(x) = w(x) - u(x). \tag{2.2}$$

In geodesy, several basic mathematical relations between the scalar fields $|w|$ and $|u|$ as well as between the vector fields a and d are known. In the following, we only describe the fundamental relations heuristically (see also [27, 32]).

The point of departure for our excursion into geodesy is the observation that the gravity disturbance vector $d(x)$ at the point x on the geoid can be written as

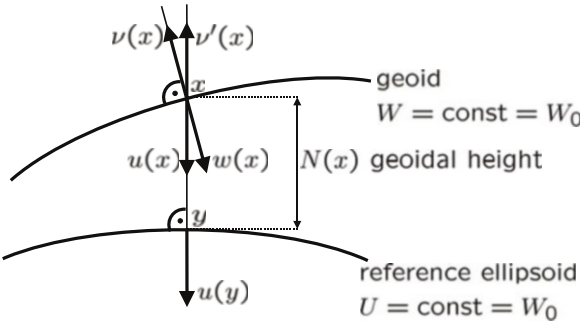


FIGURE 2.2. Illustration of the gravity vector $w(x)$, the normal gravity vector $u(x)$, and the geoidal height $N(x)$. Here, ν and ν' denote the normal to the geoid and the reference ellipsoid, respectively (following [32]).

follows:

$$d(x) = w(x) - u(x) = \nabla\left(W(x) - U(x)\right) = \nabla T(x). \tag{2.3}$$

According to Taylor's formula, $U(y)+\frac{\partial U}{\partial \nu'}(y)N(x)$ is the linearization of $U(x)$, i.e., by expanding the potential U at the point x and truncating the Taylor series at the linear term, we get

$$U(x) \simeq U(y) + \frac{\partial U}{\partial \nu'}(y)N(x), \tag{2.4}$$

where

$$\nu'(y) = -\frac{u(y)}{|u(y)|} \tag{2.5}$$

is the ellipsoidal normal at y and the geoidal undulation $N(x)$ is the aforementioned distance between x and y (note that the symbol '$\simeq$' means that the error between the left- and the right-hand side may be assumed to be insignificantly small). Using the fact that $T(x) = W(x) - U(x)$ and observing the relations

$$|u(y)| = -\nu'(y)\cdot u(y) = -\nu'(y)\cdot \nabla U(y) = -\frac{\partial U}{\partial \nu'}(y), \tag{2.6}$$

we obtain under the assumption of (2.4) that

$$N(x) = \frac{U(y) - U(x)}{|u(y)|} = \frac{T(x) - (W(x) - U(y))}{|u(y)|}. \tag{2.7}$$

Finally, considering $U(y) = W(x) = \text{const.} = W_0$, we end up with the so-called *Bruns formula* (cf. [4])

$$N(x) = \frac{T(x)}{|u(y)|}. \tag{2.8}$$

This formula relates the physical quantity $T(x)$ to the geometric quantity $N(x)$ for points x on the geoid.

It is helpful to study the vector field $\nu(x)$ in more detail:

$$\nu(x) = -\frac{w(x)}{|w(x)|}. \tag{2.9}$$

Due to the definition of the normal vector field (2.9), we obtain the following identity

$$w(x) = \nabla W(x) = \; - \, |w(x)| \; \nu(x). \tag{2.10}$$

In an analogous way we obtain

$$u(x) = \nabla U(x) = \; - \, |u(x)| \; \nu'(x). \tag{2.11}$$

The *vertical deflection* $\Theta(x)$ at the point x on the geoid is understood to be the angular (i.e., tangential) difference between the directions $\nu(x)$ and $\nu'(x)$. More concretely, the vertical deflection is determined by the angle between the plumb line and the ellipsoidal normal through the same point:

$$\Theta(x) = \nu(x) - \nu'(x) - ((\nu(x) - \nu'(x)) \cdot \nu(x)) \, \nu(x). \tag{2.12}$$

According to its construction, the vertical deflection $\Theta(x)$ at x is orthogonal to the normal vector field $\nu(x)$, i.e., $\Theta(x) \cdot \nu(x) = 0$. Since the plumb lines are orthogonal to the equipotential surfaces of the geoid and the ellipsoid, respectively, the vertical deflection gives briefly spoken a measure of the gradient of the equipotential surfaces (cf. [32]). From (2.10), in connection with (2.12), it follows that

$$w(x) = -|w(x)| \, (\Theta(x) + \nu'(x) + ((\nu(x) - \nu'(x)) \cdot \nu(x)) \, \nu(x)) \, . \tag{2.13}$$

Using Eqs. (2.11) and (2.13) we finally obtain for the gravity disturbing vector $d(x)$ at the point x

$$\begin{aligned} d(x) = \nabla T(x) &= w(x) - u(x) \qquad (2.14)\\ &= -|w(x)| \, (\Theta(x) + \nu'(x) + ((\nu(x) - \nu'(x)) \cdot \nu(x)) \, \nu(x)) - (-|u(x)| \nu'(x))\\ &= -|w(x)| \, (\Theta(x) + ((\nu(x) - \nu'(x)) \cdot \nu(x)) \, \nu(x)) - (|w(x)| - |u(x)|) \, \nu'(x). \end{aligned}$$

The quantity

$$D(x) = |w(x)| - |u(x)| \tag{2.15}$$

is called the *gravity disturbance*, whereas

$$A(x) = |w(x)| - |u(y)| \tag{2.16}$$

is called the *gravity anomaly*.

Splitting the gradient $\nabla T(x)$ of the disturbing potential T at x into a normal part (pointing into the direction of $\nu(x)$) and an angular (tangential) part (using the representation of the surface gradient ∇^*), we have

$$\nabla T(x) = \nu(x) \frac{\partial T}{\partial \nu}(x) + \frac{1}{|x|} \nabla^* T(x), \tag{2.17}$$

where ∇^* is the surface gradient.

Since the gravity disturbances represent at most a factor 10^{-4} of the Earth's gravitational force (for more details see [32]), the error between $\nu(x)\frac{\partial T}{\partial \nu}(x)$ and $\nu'(x)\frac{\partial T}{\partial \nu'}(x)$ has no (computational) significance. Consequently, we may assume

$$d(x) \simeq \nu'(x)\frac{\partial T}{\partial \nu'}(x) + \frac{1}{|x|}\nabla^* T(x). \tag{2.18}$$

Moreover, the scalar product $(\nu(x) - \nu'(x)) \cdot \nu(x)$ can also be neglected. Thus, in connection with (2.14), we obtain

$$d(x) \simeq -|w(x)|\ \Theta(x) - D(x)\nu'(x). \tag{2.19}$$

By comparison of (2.18) and (2.19), we therefore get

$$D(x) = -\frac{\partial T}{\partial \nu'}(x) = -\nu'(x) \cdot d(x), \tag{2.20}$$

$$|w(x)|\ \Theta(x) = -\frac{1}{|x|}\nabla^* T(x). \tag{2.21}$$

In other words, the gravity disturbance $D(x)$, beside being the difference in magnitude of the actual and the normal gravity vector, is also the normal component of the gravity disturbance vector $d(x)$. In addition, we are led to the angular differential equation (2.21).

Applying Bruns' formula (2.8) to Eqs. (2.20) and (2.21) we obtain

$$D(x) = |w(x)| - |u(x)| = -|u(y)|\ \frac{\partial N}{\partial \nu'}(x) \tag{2.22}$$

for the gravity disturbance and

$$|w(x)|\ \Theta(x) = -\frac{1}{|x|}\ \nabla^* T(x) = -\frac{1}{|x|}|u(y)|\ \nabla^* N(x) \tag{2.23}$$

for the vertical deflections. Note that $\Theta(x)$ may be multiplied (without loss of (computational) precision) either by $|w(x)|$ or by $|u(x)|$ since it is a small quantity.

Turning over to the gravity anomalies $A(x)$, it follows from the identity (2.20) by linearization that

$$-\frac{\partial T}{\partial \nu'}(x) = D(x) \simeq A(x) - \frac{\partial |u(y)|}{\partial \nu'} N(x). \tag{2.24}$$

Using Bruns' formula (2.8), we obtain for the gravity anomalies that

$$A(x) = -\frac{\partial T}{\partial \nu'}(x) + \frac{1}{|u(y)|}\frac{\partial |u(y)|}{\partial \nu'} T(x). \tag{2.25}$$

Summing up our results (2.20) for the gravity disturbance $D(x)$ and (2.25) for the gravity anomaly $A(x)$, we are led to the so-called *fundamental equations of*

physical geodesy:

$$D(x) = |w(x)| - |u(x)| = -\frac{\partial T}{\partial \nu'}(x), \tag{2.26}$$

$$A(x) = |w(x)| - |u(y)| = -\frac{\partial T}{\partial \nu'}(x) + \frac{1}{|u(y)|}\ \frac{\partial |u(y)|}{\partial \nu'}\ T(x). \tag{2.27}$$

Eqs. (2.26) and (2.27) show the relation between the disturbing potential T and the gravity disturbance D and the gravity anomaly A, respectively, on the geoid (see, for example, [27, 32, 45]). They are used as boundary conditions in boundary-value problems.

Remark 2.1. Following [32], the geoidal heights N, i.e., the deviations of the equipotential surface on the mean ocean level from the reference ellipsoid, are extremely small. Their order is of only a factor 10^{-5} of the Earth's radius (see [32] for more details). Even more, the reference ellipsoid only differs from a sphere Ω_R with (mean Earth's) radius R in the order of the flattening of about $3 \cdot 10^{-3}$. Therefore, since the time of [65], it is common use that, in theory, an ellipsoidal reference surface should be taken into account. However, in numerical practice, the reference ellipsoid is treated as a sphere and the Equations (2.22) and (2.23) are solved in spherical approximation. In doing so, a relative error of the order of the flattening of the Earth's body at the poles, i.e., a relative error of 10^{-3}, is accepted in all equations containing the disturbing potential. Considering appropriately performed reductions in numerical calculations, this error seems to be quite permissible (cf. [32] and the remarks in [24, 25] for comparison with ellipsoidal approaches), and this is certainly the case if local exploration is under consideration. For local purposes as discussed in this contribution, the problem of non-ellipticity seems to be obsolete.

Remark 2.2. According to the Pizzetti assumptions (see [56, 57]), it follows that the first moment integrals of the disturbing potential vanish, i.e.,

$$\int_{\Omega_R} T(y) H^R_{-n-1,k}(y)\, d\omega(y) = 0, \tag{2.28}$$

for $n = 0, 1,\ k = 1, \ldots, 2n+1$, where $H^R_{-n-1,k}$ denotes the system of outer harmonics and $d\omega$ is the surface element in $\mathbb{R}^3$. More concretely, if the Earth's center of gravity is the origin, there are no first-degree terms in the spherical harmonic expansion of T. If the mass of the spherical Earth and the mass of the normal ellipsoid is equal, there is no zero term. In this way, together with the indicated processes in gravitational modeling, formulas and structures are obtained that are rigorously valid for the sphere.

In the well-known spherical nomenclature, involving a sphere Ω_R as reference surface (R being the mean Earth's radius) with a mass M distributed homoge-

neously in its interior, we are simply led to (cf. [32])

$$U(y) = \frac{\gamma M}{|y|}, \qquad u(y) = \nabla U(y) = -\frac{\gamma M}{|y|^2}\frac{y}{|y|}, \tag{2.29}$$

where γ is the gravitational constant ($\gamma = 6.6742 \cdot 10^{-11}\mathrm{m}^3\ \mathrm{kg}^{-1}\ \mathrm{s}^{-2}$). Hence, we obtain

$$|u(y)| = \frac{\gamma M}{|y|^2}, \tag{2.30}$$

$$\frac{\partial |u(y)|}{\partial \nu'} = -\frac{u(y)}{|u(y)|} \cdot \nabla |u(y)| = -2\frac{\gamma M}{|y|^3}, \tag{2.31}$$

$$\frac{1}{|u(y)|}\frac{\partial |u(y)|}{\partial \nu'} = -\frac{2}{|y|}, \tag{2.32}$$

where x is on the geoid and y is on the reference surface (cf. Fig. 2.2). Furthermore, in spherical nomenclature, i.e., $x \in \Omega_R$, we obviously have

$$-\frac{\partial T}{\partial \nu'}(x) = -\frac{x}{|x|} \cdot \nabla T(x). \tag{2.33}$$

Therefore, we end up with the formulation of the *fundamental equations of physical geodesy for the sphere*:

$$D(x) = -\frac{x}{|x|} \cdot \nabla T(x), \tag{2.34}$$

$$A(x) = -\frac{x}{|x|} \cdot \nabla T(x) - \frac{2}{|x|} T(x). \tag{2.35}$$

In addition, in a vector spherical context (see also [22]), we obtain for the differential equation (2.21)

$$-\nabla^* T(x) = \frac{\gamma M}{R} \Theta(x), \tag{2.36}$$

and, by virtue of Bruns' formula (2.8), we finally find that

$$-\nabla^* N(x) = R\Theta(x). \tag{2.37}$$

Remark 2.3. In physical geodesy (see, e.g., [32, 36]), a componentwise scalar determination of the vertical deflection is usually used. Our work prefers the vectorial framework, i.e., the vector equation (2.36). In doing so, we are concerned with an isotropic vector approach by means of the fundamental solution with respect to the Beltrami operator (see also [22]) instead of the conventional anisotropic scalar decomposition into vector components due to [44].

The disturbing potential enables us to make the following *geophysical interpretations* (for more details the reader is referred, e.g., [25, 41, 60, 63], and the references therein): Gravity disturbances D and gravity anomalies A (Figure 2.3) represent a relation between the real Earth and an ellipsoidal Earth model. In accordance with Newton's Law of Gravitation they therefore show the imbalance of forces in the interior of the Earth due to the irregular density distribution

inside the Earth. Clearly, gravity anomalies and/or gravity disturbances do not determine uniquely the interior density distribution of the Earth. They may be interpreted as certain filtered signatures, which give major weight to the density contrasts close to the surface and simultaneously suppress the influence of deeper structures inside the Earth.

Geoid undulations provide a measure for the perturbations of the Earth from a hydrostatic equilibrium. They form the deviations of the equipotential surfaces at mean sea level from the reference ellipsoid. Geoid undulations show no essential correlation to the distributions of the continents. They seem to be generated by density contrasts much deeper inside the Earth.

As already explained, the task of determining the disturbing potential T from gravity disturbances or gravity anomalies, respectively, leads to boundary-value problems usually corresponding to a spherical boundary. Numerical realizations of such boundary-value problems have a long tradition, starting from [65] and [55]. Nonetheless, our work presents some new aspects in their potential theoretic treatment by proposing appropriate space-regularization techniques applied to the resulting integral representations of their solutions. For both boundary-value problems, viz. the Neumann and the Stokes problem, we are able to present two solution methods: The disturbing potential may be either solved by a Fourier (orthogonal) expansion method in terms of spherical harmonics or it can be described by a singular integral representation over the boundary Ω_R.

Remark 2.4. So far, much more data on gravity anomalies $A(x) = |w(x)| - |u(y)|$ are available than on gravity disturbances $D(x) = |w(x)| - |u(x)|$. However, by modern GNSS-technology (see, e.g., [52]), the point x on the geoid is rather determined than y on the reference ellipsoid. Therefore, in future, it can be expected that D will become more important than A (as [36] point out in their monograph on physical geodesy). This is the reason why we continue to work with D. Nevertheless, the results of our (multiscale) approach applied to A are of significance. Therefore, the key ideas and concepts concerning A can be treated in parallel (see [9, 15, 73] for explicit details).

In order to formulate some results in the language of potential theory, we first introduce the potential space $\mathrm{Pot}^{(1)}\left(\mathbb{R}^3\backslash\mathcal{B}_R(0)\right)$, where $\mathcal{B}_R(0)$ is the (open) ball of radius R around the origin. More concretely, we let $\mathrm{Pot}(\mathbb{R}^3\backslash\overline{\mathcal{B}_R(0)})$ be the space of all functions $F : \mathbb{R}^3\backslash\overline{\mathcal{B}_R(0)} \to \mathbb{R}$ satisfying

(i) $F|_{\mathbb{R}^3\backslash\overline{\mathcal{B}_R(0)}}$ is a member of $C^{(2)}(\mathbb{R}^3\backslash\overline{\mathcal{B}_R(0)})$,
(ii) F satisfies Laplace's equation $\Delta F(x) = 0$, $x \in \mathbb{R}^3\backslash\overline{\mathcal{B}_R(0)}$.
(iii) F is regular at infinity, i.e., $F(x) = O(|x|^{-1})$, $|x| \to \infty$.

$\mathrm{Pot}^{(1)}(\mathbb{R}^3\backslash\mathcal{B}_R(0))$ is formally understood to be the space

$$\mathrm{Pot}^{(1)}\left(\mathbb{R}^3\backslash\mathcal{B}_R(0)\right) = C^{(1)}\left(\mathbb{R}^3\backslash\mathcal{B}_R(0)\right) \cap \mathrm{Pot}(\mathbb{R}^3\backslash\overline{\mathcal{B}_R(0)}). \tag{2.38}$$

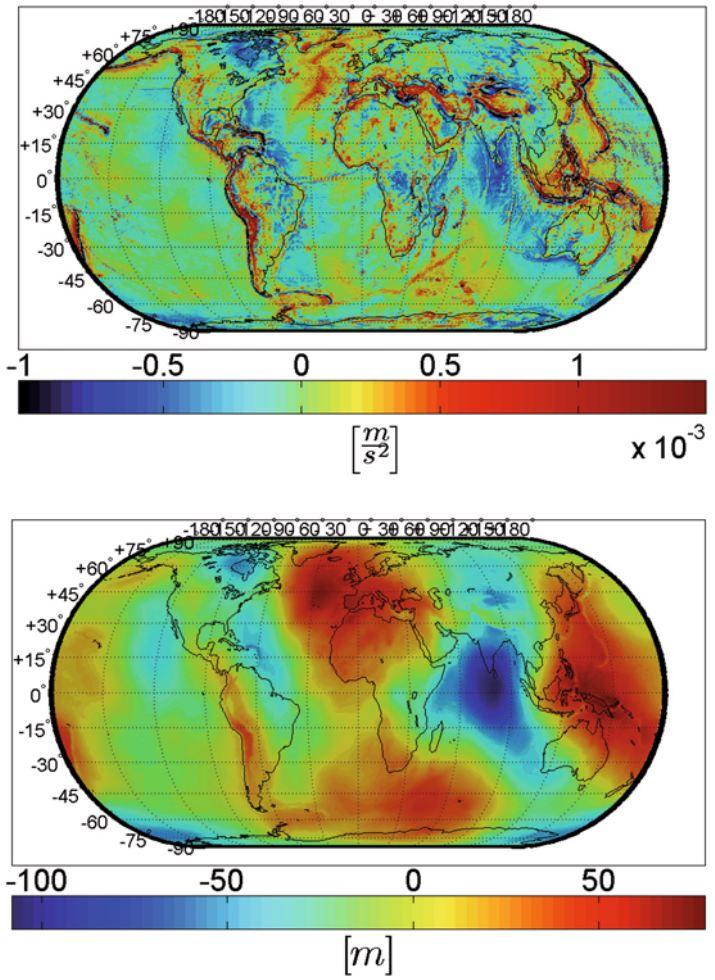


FIGURE 2.3. EIGEN-GL04C derived gravity anomalies and geoidal undulations (reconstructed by use of (spherical) smoothed Haar scaling functions from [22] based on computations by Mathar [43]).

In the language of potential theory, the *exterior Neumann boundary-value problem* corresponding to known gravity disturbances D (compare (2.34)) reads as follows:

(ENPPG) Let D be a continuous function on $\Omega_R = \partial\mathcal{B}_R(0)$, i.e., $D \in C^{(0)}(\Omega_R)$ with

$$\int_{\Omega_R} D(y) H^R_{-n-1,k}(y)\, d\omega(y) = 0, \tag{2.39}$$

for $n = 0, 1,\ k = 1, \dots, 2n+1$. Find $T \in \mathrm{Pot}^{(1)}\left(\mathbb{R}^3 \backslash \mathcal{B}_R(0)\right)$, such that the boundary condition $D = \frac{\partial T}{\partial \nu}\big|_{\Omega_R}$ holds true and the potential T fulfills the conditions

$$\int_{\Omega_R} T(y) H^R_{-n-1,k}(y)\, d\omega(y) = 0 \tag{2.40}$$

for $n = 0, 1,\ k = 1, \dots, 2n+1$.

It is known (see, e.g., [15]) that the solution of the boundary-value problem (ENPPG) can be represented in the form

$$T(x) = \frac{1}{4\pi R} \int_{\Omega_R} D(y)\ N(x,y)\ d\omega(y), \qquad x \in \mathbb{R}^3 \backslash \mathcal{B}_R(0), \tag{2.41}$$

where the *Neumann kernel* $N(\cdot,\cdot)$ in (2.41) possesses the spherical harmonic expansion

$$N(x,y) = \sum_{n=2}^{\infty} \left(\frac{R^2}{|x||y|}\right)^{n+1} \frac{2n+1}{n+1} P_n\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right). \tag{2.42}$$

By well-known manipulations, the series in terms of Legendre polynomials can be expressed as an elementary function leading to the integral representation

$$T(x) = \frac{1}{4\pi R} \int_{\Omega_R} D(y) \left(\frac{2R}{|x-y|} + \ln\left(\frac{|y| + \left|y - \frac{R^2}{|x|^2}x\right| - \frac{R^2}{|x|}}{|y| + \left|y - \frac{R^2}{|x|^2}x\right| + \frac{R^2}{|x|}}\right)\right) d\omega(y). \tag{2.43}$$

It is not difficult to see that for $x \in \Omega_R$, the integral (2.43) is equivalent to

$$T(x) = \frac{1}{4\pi R} \int_{\Omega_R} D(y) \left(\frac{2R}{|x-y|} + \ln\left(\frac{|y| + |x-y| - R}{|y| + |x-y| + R}\right)\right) d\omega(y). \tag{2.44}$$

Written out in spherical nomenclature $x = R\frac{x}{|x|}$, $y = R\frac{y}{|y|}$, $x \neq y$ on Ω_R, we find

$$N\left(R\frac{x}{|x|}, R\frac{y}{|y|}\right) = \frac{2}{\left|\frac{x}{|x|} - \frac{y}{|y|}\right|} + \ln\left(\frac{R\left|\frac{x}{|x|} - \frac{y}{|y|}\right|}{2R + R\left|\frac{x}{|x|} - \frac{y}{|y|}\right|}\right). \tag{2.45}$$

If we use

$$\left|\frac{x}{|x|} - \frac{y}{|y|}\right| = \left(2 - 2\frac{x \cdot y}{|x|\ |y|}\right)^{\frac{1}{2}}, \tag{2.46}$$

then, for $x \neq y$, we are led to the identity

$$\begin{aligned} N\left(R\frac{x}{|x|}, R\frac{y}{|y|}\right) &= N\left(\frac{x}{|x|}, \frac{y}{|y|}\right) \\ &= \frac{\sqrt{2}}{\sqrt{1 - \frac{x}{|x|} \cdot \frac{y}{|y|}}} - \ln\left(1 + \frac{\sqrt{2}}{\sqrt{1 - \frac{x}{|x|} \cdot \frac{y}{|y|}}}\right). \end{aligned} \tag{2.47}$$

Consequently, for points $x \in \Omega_R$, we (formally) get the so-called *Neumann formula*, which constitutes an improper integral over Ω_R:

$$T\left(R\frac{x}{|x|}\right) = \frac{1}{4\pi R}\int_{\Omega_R} D\left(R\frac{y}{|y|}\right) N\left(\frac{x}{|x|}, \frac{y}{|y|}\right)\, d\omega(y), \tag{2.48}$$

where the Neumann kernel constitutes a radial basis function due to (2.47).

Once more, in accordance with the conventional approach of physical geodesy, the Neumann formula (2.48) is valid under the following constraints (see also [23, 32, 51]):

(i) the mass within the reference ellipsoid is equal to the mass of the Earth,
(ii) the center of the reference ellipsoid coincides with the center of the Earth,
(iii) the formulation is given in the spherical context to guarantee economical and efficient numerics.

Note that we are able to set $N(R\xi, R\eta) = N(\xi, \eta) = N(\xi \cdot \eta)$ in terms of the unit vectors $\xi = \frac{x}{|x|}$ and $\eta = \frac{y}{|y|}$ which simplifies our notation: If we define the single-layer kernel $S : [-1, 1) \to \mathbb{R}$ by

$$S(t) = \frac{\sqrt{2}}{\sqrt{1-t}}, \qquad t \in [-1, 1), \tag{2.49}$$

the Neumann kernel is the zonal function of the form

$$N(\xi \cdot \eta) = S(\xi \cdot \eta) - \ln\left(1 + S(\xi \cdot \eta)\right), \qquad 1 - \xi \cdot \eta > 0. \tag{2.50}$$

An equivalent formulation of the improper integral (2.48) over the unit sphere $\Omega = \partial\mathcal{B}_1(0)$ is then given by

$$T(R\xi) = \frac{R}{4\pi}\int_{\Omega} D(R\eta) N(\xi \cdot \eta)\, d\omega(\eta). \tag{2.51}$$

It should be remarked that the exterior Stokes boundary value problem of determining the disturbing potential from known gravity anomalies can be handled in a quite analogous way (see [9, 15, 73]), providing the so-called Stokes integral associated to the radially symmetric Stokes kernel as an improper integral on Ω_R.

Next we deal with the vertical deflections Θ (cf. [15, 21]). Suppose that T fulfills the conditions (2.40). We consider the differential equation (compare Eq. (2.36))

$$\nabla^*_\xi T(R\xi) = -\frac{\gamma M}{R}\Theta(R\xi), \tag{2.52}$$

where $T(R\cdot)$ represents the disturbing potential and $\Theta(R\cdot)$ denotes the vertical deflection. The differential equation (2.52) can be solved in a unique way by means of the fundamental solution with respect to the Beltrami operator

$$T(R\xi) = \frac{\gamma M}{R}\int_{\Omega} \Theta(R\eta) \cdot \nabla^*_\eta G\left(\Delta^*; \xi \cdot \eta\right)\, d\omega(\eta), \tag{2.53}$$

where $(\xi,\eta) \mapsto G(\Delta^*;\xi\cdot\eta)$, $1-\xi\cdot\eta \neq 0$, is the fundamental solution of the Beltrami equation on the unit sphere Ω (see, e.g., [11]) given by

$$G(\Delta^*;\xi\cdot\eta) = \frac{1}{4\pi}\ln(1-\xi\cdot\eta) + \frac{1}{4\pi}(1-\ln(2)). \tag{2.54}$$

The identity (2.53) immediately follows from the Third Green Theorem (cf. [15, 22]) for ∇^* on Ω in connection with (2.40). By virtue of the identity

$$\nabla^*_\eta G(\Delta^*;\xi\cdot\eta) = -\frac{\xi-(\xi\cdot\eta)\eta}{4\pi(1-\xi\cdot\eta)}, \quad \xi\neq\eta, \tag{2.55}$$

the integral (2.53) can be written in the form

$$T(R\xi) = \frac{R}{4\pi}\int_\Omega \Theta(R\eta)\cdot g(\Delta^*;\xi,\eta)\; d\omega(\eta), \tag{2.56}$$

where the vector kernel $g(\Delta^*;\xi,\eta),\ \xi\neq\eta$, is given by

$$g(\Delta^*;\xi,\eta) = -\frac{\gamma M}{R^2}\frac{\xi-(\xi\cdot\eta)\eta}{1-\xi\cdot\eta}. \tag{2.57}$$

Again we are confronted with a representation of the disturbing potential T as an improper integral over the sphere Ω_R.

All our settings leading to the disturbing potential on the sphere Ω_R turn out to be improper integrals. As we have shown they have either the singularity behavior of the single-layer kernel S (cf. Eq. (2.49)) or the characteristic logarithmic singularity of the fundamental solution with respect to the Beltrami operator $G(\Delta^*;\cdot,\cdot)$ (cf. (2.54)). Indeed, the fundamental solution and the single-layer kernel are interrelated (see [15]) by the identities

$$S(\xi\cdot\eta) = \sqrt{2}\; e^{-2\pi G(\Delta^*;\xi\cdot\eta)+\frac{1}{2}} \tag{2.58}$$

and

$$G(\Delta^*;\xi\cdot\eta) = -\frac{1}{2\pi}\ln(S(\xi\cdot\eta)) - \frac{1}{4\pi}(1-2\ln(2)). \tag{2.59}$$

Therefore, we are confronted with the remarkable situation that a (Taylor) regularization of the single-layer kernel implies a regularization of the fundamental solution, and vice versa.

2.2. Zooming-in localization of signature bands

Next, we present multiscale representations for the Neumann kernel N (cf. Eq. (2.50)). Note that all modern multiscale approaches have a conception of wavelets as constituting multiscale building blocks in common, which provide a fast and efficient way to decorrelate a given signal data set.

The characterizing properties of the multiscale approach (basis property, decorrelation, and efficient algorithms) are common features of all wavelets and these attributes form the key for a variety of applications (see, e.g., [12, 15, 20]), particularly for signal reconstruction and decomposition, thresholding, data compression, denoising by, e.g., multiscale signal-to-noise ratio, etc. and, in particular, decorrelation.

Next, we follow the (taylorized) regularization methods presented in [23] for linear regularization of the single-layer kernel S and [21] for linear regularization of the fundamental solution $G(\Delta^*;\cdot,\cdot)$ of the Beltrami equation. For higher-order approximations, the reader is referred to the Ph.D.-theses [9] and [73].

The essential idea is to regularize the single-layer kernel function

$$S(t) = \frac{\sqrt{2}}{\sqrt{1-t}} \tag{2.60}$$

by replacing it by a Taylor linearization. To this end, we notice that the first derivative of the kernel S is given by

$$S'(t) = \frac{1}{\sqrt{2}(1-t)^{\frac{3}{2}}}, \qquad t \in [-1,1). \tag{2.61}$$

Consequently, we obtain as (Taylor) linearized approximation corresponding to the expansion point $1-\frac{\tau^2}{2R^2},\ \tau \in (0,2R]$,

$$S(t) = S\left(1-\frac{\tau^2}{2R^2}\right) + S'\left(1-\frac{\tau^2}{2R^2}\right)\left(t-\left(1-\frac{\tau^2}{2R^2}\right)\right) + \cdots. \tag{2.62}$$

In more detail, the kernel S is replaced by its (Taylor) linearized approximation S^τ at the point $1-\frac{\tau^2}{2R^2},\ \tau \in (0,2R]$, given by

$$S^\tau(t) = \begin{cases} \frac{R}{\tau}\left(3-\frac{2R^2}{\tau^2}(1-t)\right), & 0 \le 1-t \le \frac{\tau^2}{2R^2}, \\ \frac{\sqrt{2}}{\sqrt{1-t}}, & \frac{\tau^2}{2R^2} < 1-t \le 2. \end{cases} \tag{2.63}$$

Note that the expansion point $1-\tau^2/(2R^2),\ \tau \in (0,2R]$, is chosen in consistency with the notation in the initial paper [21] and the subsequent papers [14] and [23]. A graphical illustration of the original kernel $S(t)$ and a τ-scale dependent version of its linear space-regularized kernel $S^\tau(t)$ is shown in Figure 2.4.

Clearly, the function S^τ is continuously differentiable on the interval $[-1,1]$, and we have

$$\left(S^\tau\right)'(t) = \begin{cases} \frac{2R^3}{\tau^3}, & 0 \le 1-t \le \frac{\tau^2}{2R^2}, \\ \frac{1}{\sqrt{2}(1-t)^{\frac{3}{2}}}, & \frac{\tau^2}{2R^2} < 1-t \le 2. \end{cases} \tag{2.64}$$

Furthermore, the functions S and S^τ are monotonously increasing on the interval $[-1,1)$, such that $S(t) \ge S^\tau(t) \ge S(-1) = S^\tau(-1) = 1$ holds true on the interval $[-1,1)$. Considering the difference between the kernel S and its linearly regularized version S^τ, we find

$$S(t) - S^\tau(t) = \begin{cases} \frac{\sqrt{2}}{\sqrt{1-t}} - \frac{R}{\tau}\left(3-\frac{2R^2}{\tau^2}(1-t)\right), & 0 < 1-t \le \frac{\tau^2}{2R^2}, \\ 0, & \frac{\tau^2}{2R^2} < 1-t \le 2. \end{cases} \tag{2.65}$$

By elementary manipulations of one-dimensional analysis we readily obtain

$$\int_{-1}^{1} |S(t) - S^\tau(t)|\, dt \;=\; O(\tau). \tag{2.66}$$

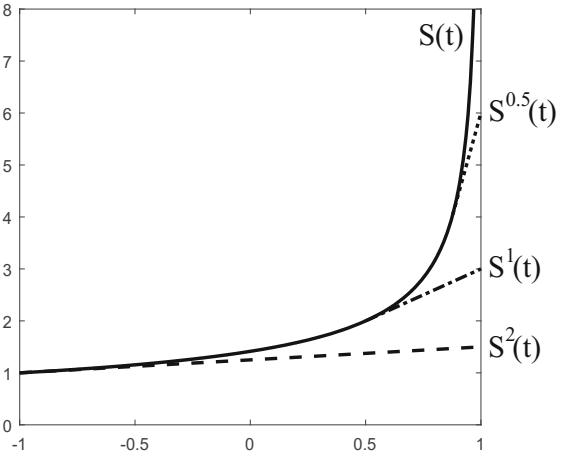


FIGURE 2.4. Single-layer kernel $S(t)$ (continuous black line) and its Taylor linearized regularization $S^\tau(t)$, for $R = 1$ and $\tau = \frac{1}{2}, 1, 2$ (dotted lines).

As a consequence, we have

Lemma 2.5. *For $F \in C^{(0)}(\Omega)$ and S^τ defined by* (2.63) *the limit relation*

$$\lim_{\tau\to 0+} \sup_{\xi\in\Omega} \left| \int_\Omega S(\xi\cdot\eta)F(\eta)\, d\omega(\eta) - \int_\Omega S^\tau(\xi\cdot\eta)(\xi\cdot\eta)F(\eta)\, d\omega(\eta) \right| = 0 \tag{2.67}$$

holds true.

In a similar way, by some elementary calculations, one can find the following relations that are also of importance for the Stokes boundary value problem (see also the Ph.D.-theses [9, 73]).

Lemma 2.6. *Let S be the single-layer kernel given by* (2.60) *and let $S^\tau, \tau \in (0, 2R]$, be the corresponding (Taylor) linearized regularized kernel defined by* (2.63)*. Then*

$$\lim_{\tau\to 0+} \int_{-1}^{1} |\ln(1+S(t)) - \ln(1+S^\tau(t))|\, dt = 0, \tag{2.68}$$

$$\lim_{\tau\to 0+} \int_{-1}^{1} \left| \ln\left(\frac{1}{S(t)} + \frac{1}{(S(t))^2}\right) - \ln\left(\frac{1}{S^\tau(t)} + \frac{1}{(S^\tau(t))^2}\right) \right| dt = 0, \tag{2.69}$$

$$\lim_{\tau\to 0+} \int_{-1}^{1} \left((S(t))^2 - (S^\tau(t))^2 \right) \sqrt{1-t^2}\, dt = 0. \tag{2.70}$$

To study the surface gradient and the surface curl gradient, we let F be of class $C^{(1)}(\Omega)$. Letting $\mathbf{t}_\xi \in \mathbb{R}^{3\times 3}$ be the orthogonal matrix (with $\det(\mathbf{t}_\xi) = 1$) leaving ε^3 fixed such that $\mathbf{t}_\xi \xi = \varepsilon^3$, we get

$$\nabla_\xi^* \int_\Omega S(\xi\cdot\eta)F(\eta)\, d\omega(\eta) = \int_\Omega S(\eta_3)\nabla_\xi^* F(\mathbf{t}_\xi^T\eta)\, d\omega(\eta) \tag{2.71}$$

for $\xi \in \Omega$ and $\eta = (\eta_1, \eta_2, \eta_3)^T$. By regularizing the single-layer kernel, we obtain

$$\int_\Omega \nabla_\xi^* S^\tau(\xi \cdot \eta) F(\eta)\ d\omega(\eta) = \int_\Omega S^\tau(\eta_3) \nabla_\xi^* F(\mathbf{t}_\xi^T \eta)\ d\omega(\eta) \tag{2.72}$$

for $\xi \in \Omega$. The same argumentation holds true for the operator L^*. Therefore, Lemma 2.5 leads us to the following limit relations (see [22]).

Lemma 2.7. *Let F be of class $C^{(1)}(\Omega)$. Let S^τ be given by* (2.63)*. Then*

$$\lim_{\tau \to 0+} \sup_{\xi \in \Omega} \left| \int_\Omega \nabla_\xi^* S^\tau(\xi \cdot \eta) F(\eta) d\omega(\eta) - \nabla_\xi^* \int_\Omega S(\xi \cdot \eta) F(\eta) d\omega(\eta) \right| = 0, \tag{2.73}$$

$$\lim_{\tau \to 0+} \sup_{\xi \in \Omega} \left| \int_\Omega \mathrm{L}_\xi^* S^\tau(\xi \cdot \eta) F(\eta) d\omega(\eta) - \mathrm{L}_\xi^* \int_\Omega S(\xi \cdot \eta) F(\eta) d\omega(\eta) \right| = 0. \tag{2.74}$$

Using the kernel $G^\tau(\Delta^*; \cdot)$, given by (see Eq. (2.59))

$$G^\tau(\Delta^*; t) = -\frac{1}{2\pi} \ln(S^\tau(t)) - \frac{1}{4\pi}(1 - 2\ln(2)), \quad -1 \le t \le 1, \tag{2.75}$$

as *"single-layer kernel regularization" of the fundamental solution* $G(\Delta^*; \cdot)$, we are led to the following integral relations.

Lemma 2.8. *For $F \in C^{(0)}(\Omega)$ and $G^\tau(\Delta^*; \cdot)$ defined by* (2.75)*, we have*

$$\lim_{\tau \to 0+} \sup_{\xi \in \Omega} \left| \int_\Omega G(\Delta^*; \xi \cdot \eta) F(\eta)\ d\omega(\eta) - \int_\Omega G^\tau(\Delta^*; \xi \cdot \eta) F(\eta)\ d\omega(\eta) \right| = 0, \tag{2.76}$$

and

$$\lim_{\tau \to 0+} \sup_{\xi \in \Omega} \left| \int_\Omega \nabla_\xi^* G^\tau(\Delta^*; \xi \cdot \eta) F(\eta)\ d\omega(\eta) - \nabla_\xi^* \int_\Omega G(\Delta^*; \xi \cdot \eta) F(\eta)\ d\omega(\eta) \right| = 0, \tag{2.77}$$

$$\lim_{\tau \to 0+} \sup_{\xi \in \Omega} \left| \int_\Omega \mathrm{L}_\xi^* G^\tau(\Delta^*; \xi \cdot \eta) F(\eta)\ d\omega(\eta) - \mathrm{L}_\xi^* \int_\Omega G(\Delta^*; \xi \cdot \eta) F(\eta)\ d\omega(\eta) \right| = 0. \tag{2.78}$$

Remark 2.9. Numerical implementations and computational aspects of the Taylor regularization techniques as presented here have been applied (even for subsets of Ω_R) to different fields of physical geodesy (see, e.g., [13–15, 21–23] and the references therein).

The regularization techniques enable us to formulate multiscale solutions for the disturbing potential from gravity disturbances or vertical deflections (note that we need higher-order regularizations whenever gravitational observables containing second or higher-order derivatives come into play; an example is gravity gradiometry, which will not be discussed here).

As point of departure for our considerations serves the special case study of the linear regularization of the single-layer kernel in the integral representation of the solution of the Neumann boundary-value problem (ENPPG).

Disturbing Potential from Gravity Disturbances. As we already know, the solution of the (Earth's) disturbing potential $T \in Pot^{(1)}\left(\mathbb{R}^3 \backslash \mathcal{B}_R(0)\right)$ from known vertical

derivatives, i.e., gravity disturbances $D = \frac{\partial T}{\partial \nu}\big|_{\Omega_R}$, satisfying the conditions (2.40) on the sphere Ω_R, can be formulated as an improper integral (see Eq. (2.51))

$$T(R\xi) = \frac{R}{4\pi} \int\limits_{\Omega} D(R\eta)\ N(\xi \cdot \eta)\, d\omega(\eta),\ \xi \in \Omega, \tag{2.79}$$

with the Neumann kernel N (cf. (2.50)). Our interest is to formulate regularizations of the disturbing potential T by use of the (Taylor) linearized approximation of the singe-layer kernel $S^\tau : [-1,1] \to \mathbb{R},\ \tau \in (0, 2R]$, introduced in (2.63). As a result, we obtain the regularized Neumann kernels

$$\begin{aligned} N^\tau(\xi \cdot \eta) &= \begin{cases} S^\tau(\xi \cdot \eta) - \ln\left(1 + S^\tau(\xi \cdot \eta)\right), & 0 \le 1 - \xi \cdot \eta \le \frac{\tau^2}{2R^2}, \\ S(\xi \cdot \eta) - \ln(1 + S(\xi \cdot \eta)), & \frac{\tau^2}{2R^2} < 1 - \xi \cdot \eta \le 2, \end{cases} \\ &= \begin{cases} \frac{R}{\tau}\left(3 - \frac{2R^2}{\tau^2}(1 - \xi \cdot \eta)\right) - \ln\left(1 + \frac{R}{\tau}\left(3 - \frac{2R^2}{\tau^2}(1 - \xi \cdot \eta)\right)\right), & \\ \qquad\qquad 0 \le 1 - \xi \cdot \eta \le \frac{\tau^2}{2R^2}, & \\ \frac{\sqrt{2}}{\sqrt{1 - \xi \cdot \eta}} - \ln\left(1 + \frac{\sqrt{2}}{\sqrt{1 - \xi \cdot \eta}}\right), & \frac{\tau^2}{2R^2} < 1 - \xi \cdot \eta \le 2. \end{cases} \end{aligned} \tag{2.80}$$

In doing so, we are immediately led to the regularized representation of the disturbing potential T corresponding to the known gravity disturbances:

$$\begin{aligned} T^\tau(R\xi) &= \frac{R}{4\pi} \int\limits_{\eta \in \Omega} D(R\eta) N^\tau(\xi \cdot \eta)\, d\omega(\eta) \qquad (2.81) \\ &= \frac{R}{4\pi} \int\limits_{\substack{\eta \in \Omega; \\ 1 - \xi \cdot \eta > \frac{\tau^2}{2R^2}}} D(R\eta) N(\xi \cdot \eta)\, d\omega(\eta) + \frac{R}{4\pi} \int\limits_{\substack{\eta \in \Omega; \\ 1 - \xi \cdot \eta \le \frac{\tau^2}{2R^2}}} D(R\eta) N^\tau(\xi \cdot \eta)\, d\omega(\eta). \end{aligned}$$

The representation (2.81) is remarkable, since the integrands of T and T^τ only differ on the spherical cap

$$\Gamma_{\tau^2/(2R^2)}(\xi) = \left\{ \eta \in \Omega\ :\ 1 - \xi \cdot \eta \le \frac{\tau^2}{2R^2} \right\}. \tag{2.82}$$

By aid of Lemma 2.5 and Lemma 2.6, we obtain

Theorem 2.10. *Suppose that T is the solution of the Neumann boundary-value problem* (ENPPG) *of the form* (2.79). *Let T^τ, $\tau \in (0, 2R]$, represent its regularization* (2.81). *Then*

$$\lim_{\tau \to 0+} \sup_{\xi \in \Omega} |T(R\xi) - T^\tau(R\xi)| = 0. \tag{2.83}$$

For numerical applications, we have to go over to scale-discretized approximations of the solution to the boundary-value problem (ENPPG). For that purpose, we choose a monotonously decreasing sequence $\{\tau_j\}_{j \in \mathbb{N}_0}$, such that

$$\lim_{j \to \infty} \tau_j = 0,\ \tau_0 = 2R. \tag{2.84}$$

A particularly important example, that we use in our numerical implementations below, is the dyadic sequence with

$$\tau_j = 2^{1-j}R, \quad j \in \mathbb{N}_0. \tag{2.85}$$

It is easy to see that $2\tau_{j+1} = \tau_j$, $j \in \mathbb{N}_0$, is the relation between two consecutive elements of the sequence. In correspondence to the sequence $\{\tau_j\}_{j\in\mathbb{N}_0}$, a sequence $\{N^{\tau_j}\}_{j\in\mathbb{N}_0}$ of discrete versions of the regularized Neumann kernels (2.80), so-called *Neumann scaling functions*, is available. Figure 2.5 (left) shows a graphical illustration of the regularized Neumann kernels for different scales j.

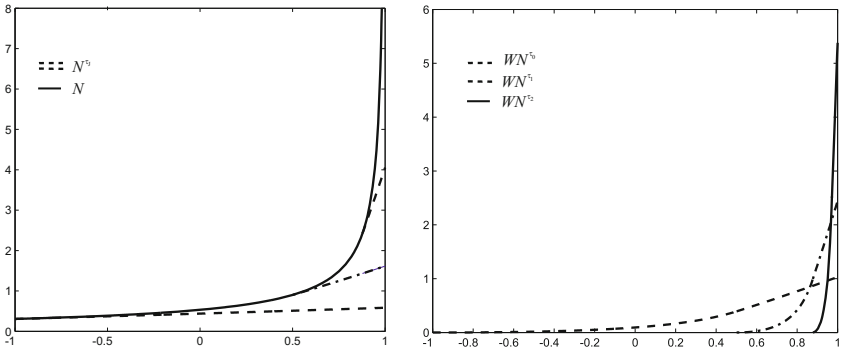


FIGURE 2.5. Illustration of the Neumann kernel $N(t)$ (left, continuous black line) and its Taylor linearized regularization $N^{\tau_j}(t)$, $j = 0, 1, 2$, $\tau_j = 2^{1-j}R$ and $R = 1$ (left, dotted lines). The corresponding Taylor linearized Neumann wavelets $WN^{\tau_j}(t)$ for scales $j = 0, 1, 2$, are shown on the right.

The *regularized Neumann wavelets*, forming the sequence $\{WN^{\tau_j}\}_{j\in\mathbb{N}_0}$, are understood to be the difference of two consecutive regularized Neumann scaling functions, respectively,

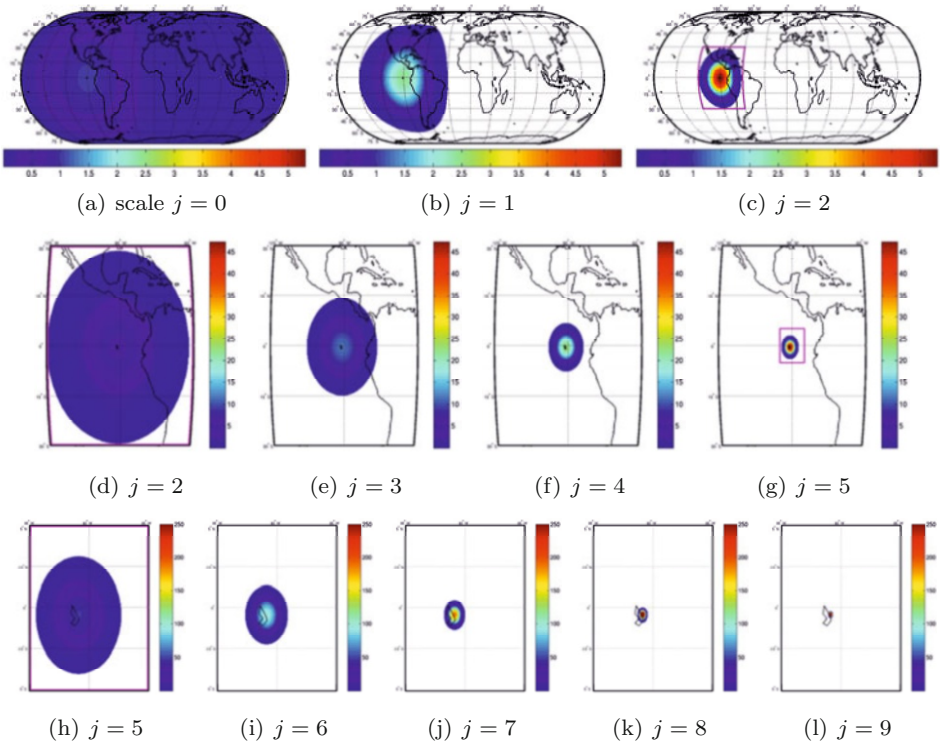
$$WN^{\tau_j} = N^{\tau_{j+1}} - N^{\tau_j}, \quad j \in \mathbb{N}_0. \tag{2.86}$$

The Neumann wavelets are illustrated in Figure 2.5 (right). These wavelets possess the numerically important property of a local support. More concretely, $\eta \mapsto WN^{\tau_j}(\xi\cdot\eta)$, $\eta \in \Omega$, vanishes everywhere outside the spherical cap $\Gamma_{\tau_j^2/(2R^2)}(\xi)$.

Let $J \in \mathbb{N}_0$ be an arbitrary scale. Suppose that N^{τ_J} is the regularized Neumann scaling function at scale J. Furthermore, let WN^{τ_j}, $j = 0, \ldots, J$, be the regularized Neumann wavelets as given by (2.86). Then, we obviously have

$$N^{\tau_J} = N^{\tau_0} + \sum_{j=0}^{J-1} WN^{\tau_j}. \tag{2.87}$$

The local support of the Neumann wavelets within the framework of (2.87) should be studied in more detail: We start with the globally supported scaling kernel $N^{\tau_0} = N^{2R}$. Then we add more and more wavelet kernels WN^{τ_j}, $j = 0, \ldots, J-1$,



(a) scale $j = 0$ (b) $j = 1$ (c) $j = 2$

(d) $j = 2$ (e) $j = 3$ (f) $j = 4$ (g) $j = 5$

(h) $j = 5$ (i) $j = 6$ (j) $j = 7$ (k) $j = 8$ (l) $j = 9$

FIGURE 2.6. Illustration of the regularized Neumann wavelets $\eta \mapsto WN^{\tau_j}(\xi \cdot \eta)$ for scales $j = 0, \ldots, 9$ to visualize the local supports $\Gamma_{\tau_j^2/(2R^2)}(\xi)$ for a fixed ξ (cf. [73]) "zooming in" to the hotspot of the Galapagos islands.

to achieve the scaling kernel N^{τ_J}. It is of particular importance that the kernel functions $\eta \mapsto WN^{\tau_j}(\xi \cdot \eta)$, $\xi \in \Omega$ fixed, are ξ-zonal functions with local support (spherical caps). Figure 2.6 illustrates the computationally relevant regions for the different wavelet scales j (more detailed studies are presented in the Ph.D. theses [9, 73]). For a better understanding, the areas outside the caps are chosen to be uncolored. Clearly, the support of the wavelets WN^{τ_j} becomes more localized for increasing scales j. In conclusion, a calculation of an integral representation for the disturbing potential T starts with a global trend approximation using the scaling kernel at scale $j = 0$ (of course, this requires data on the whole sphere, but the data can be rather sparsely distributed since they only serve as a trend approximation). Step by step, we are able to refine this approximation by use of wavelets of increasing scale. The spatial localization of the wavelets successively allows a better spatial resolution of the disturbing potential T. Additionally, the local supports of the wavelets provide a computational advantage since the integration has to be performed on smaller and smaller spherical caps. In consequence,

the presented numerical technique becomes capable of handling heterogeneously distributed data.

All in all, keeping the space-localizing property of the regularized Neumann scaling and wavelet functions in mind, we are able to establish an approximation of the solution of the disturbing potential T from gravity disturbances D in form of a "zooming-in" multiscale method. A low-pass filtered version of the disturbing potential T at the scale j in an integral representation over the unit sphere Ω is given by (compare Eq. (2.81))

$$T^{\tau_j}(R\xi) = \frac{R}{4\pi} \int_{\Omega} D(R\eta)\; N^{\tau_j}(\xi \cdot \eta)\, d\omega(\eta), \qquad \xi \in \Omega, \tag{2.88}$$

while the j-scale band-pass filtered version of T leads to the integral representation by use of the wavelets

$$WT^{\tau_j}(R\xi) = \frac{R}{4\pi} \int\limits_{\Gamma_{\tau_j^2/(2R^2)}(\xi)} D(R\eta)\; WN^{\tau_j}(\xi \cdot \eta)\, d\omega(\eta), \qquad \xi \in \Omega. \tag{2.89}$$

Theorem 2.11. *Let $T^{\tau_{J_0}}$ be the regularized version of the disturbing potential at some arbitrary initial scale J_0 as given in* (2.88)*, and let $WT^{\tau_{J_0+j}}$, $j = 0, 1, \ldots$, be given by* (2.89)*. Then, the following reconstruction formula holds true:*

$$\lim_{N\to\infty} \sup_{\xi\in\Omega} \left| T(R\xi) - \left(T^{\tau_{J_0}}(R\xi) + \sum\nolimits_{j=0}^{N} WT^{\tau_{J_0+j}}(R\xi) \right) \right| = 0.$$

The multiscale procedure (*wavelet reconstruction*) as developed here can be illustrated by the following scheme

$$\begin{array}{ccccccccc} WT^{\tau_{J_0}} & & & & WT^{\tau_{J_0+1}} & & & & \\ & \searrow & & & & \searrow & & & \\ T^{\tau_{J_0}} & \longrightarrow & + & \longrightarrow & T^{\tau_{J_0+1}} & \longrightarrow & + & \longrightarrow & T^{\tau_{J_0+2}} \ \ldots \ . \end{array}$$

As a consequence, a tree algorithm based on the regularization in the space domain has been realized for determining the disturbing potential T from locally available data sets of gravity disturbances D. An example is shown in Figure 2.7 (following [73]).

In order to get a fully discretized solution of the Neumann boundary-value problem (ENPPG), approximate integration by use of appropriate cubature formulas is necessary (see, e.g., [16, 35] for more details about approximate integration on the (unit) sphere). The fully discretized multiscale approximations have the following representations

$$T^{\tau_j}(R\xi) \simeq \frac{R}{4\pi} \sum_{k=1}^{N_j} w_k^{N_j} D\left(R\eta_k^{N_j}\right)\; N^{\tau_j}\left(\xi \cdot \eta_k^{N_j}\right), \quad \xi \in \Omega, \tag{2.90}$$

$$WT^{\tau_j}(R\xi) \simeq \frac{R}{4\pi} \sum_{k=1}^{N_j} w_k^{N_j} D\left(R\eta_k^{N_j}\right)\; WN^{\tau_j}\left(\xi \cdot \eta_k^{N_j}\right), \quad \xi \in \Omega, \tag{2.91}$$

where $\eta_k^{N_j}$ are the N_j integration knots and $w_k^{N_j}$ the integration weights.

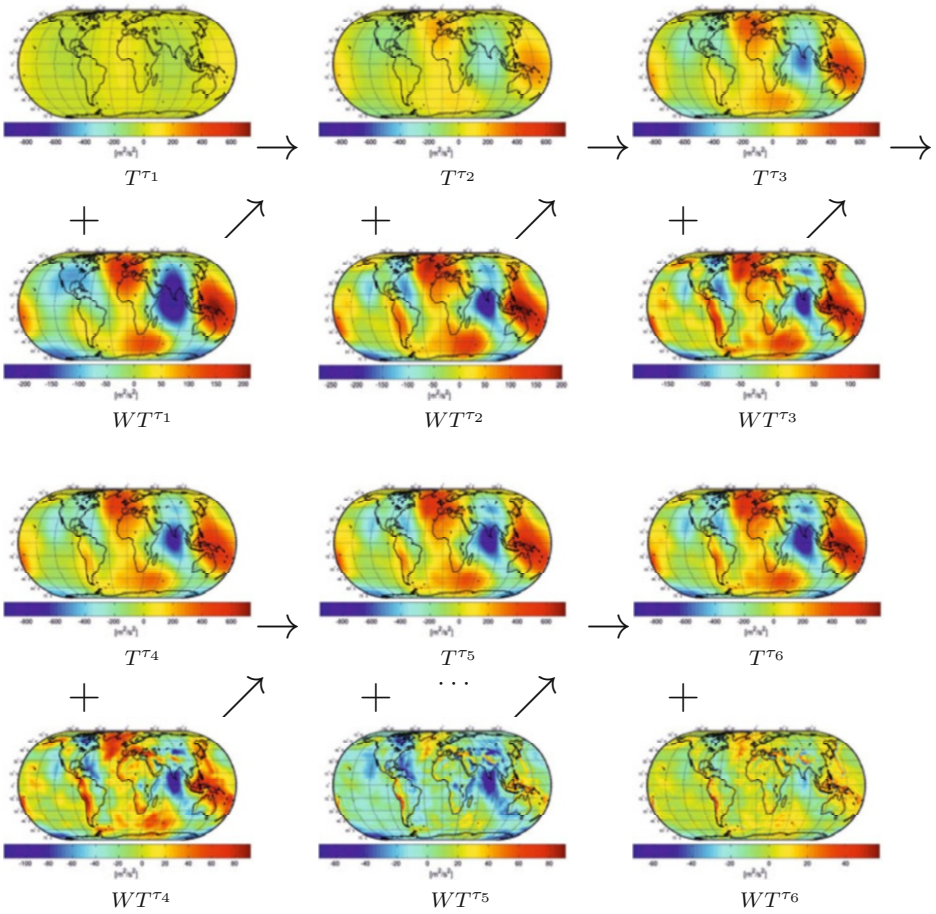
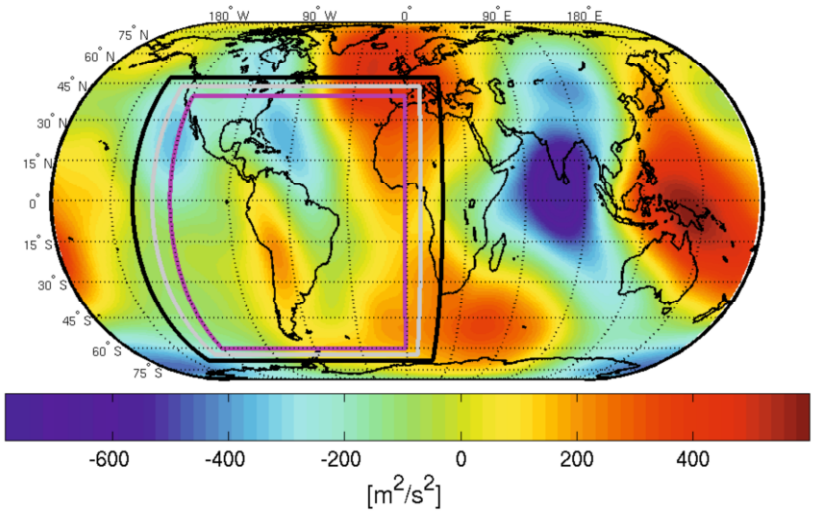


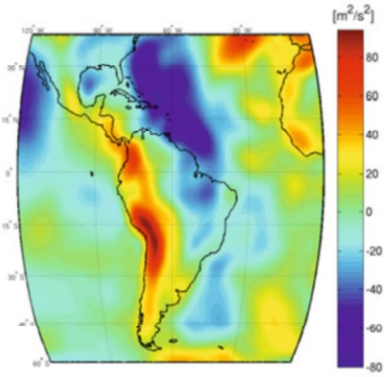
FIGURE 2.7. Illustration of a (global) multiscale approximation of the Earth's disturbing potential T in $[\frac{m^2}{s^2}]$ from gravity disturbances D, i.e., low-pass filtered versions T^{τ_j} and detail information (band-pass filtered versions) WT^{τ_j} for scales $j = 1, \ldots, 6$, by use of the linear Neumann scaling functions and wavelets computed from 4 000 000 data points distributed over the whole sphere Ω_R (from the Ph.D.-thesis [73], Geomathematics Group, University of Kaiserslautern).

Whereas the sum in (2.90) has to be calculated on the whole sphere Ω, the summation in (2.91) has to be computed only for the local supports of the wavelets (note that the symbol $\simeq$ means that the error between the right-hand and the left-hand side can be neglected).

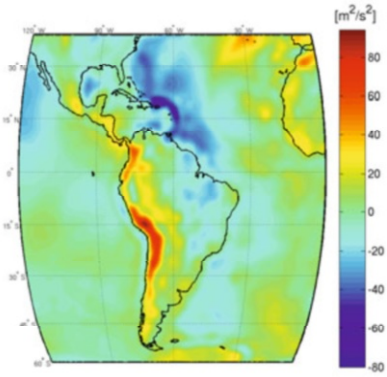
Figures 2.8 to 2.10 present a decomposition of the Earth's disturbing potential T in low-pass and band-pass filtered parts for data sets of increasing data density.



(a) Low pass part T^{τ_4} calculated from 490 000 data points distributed over the whole sphere Ω_R



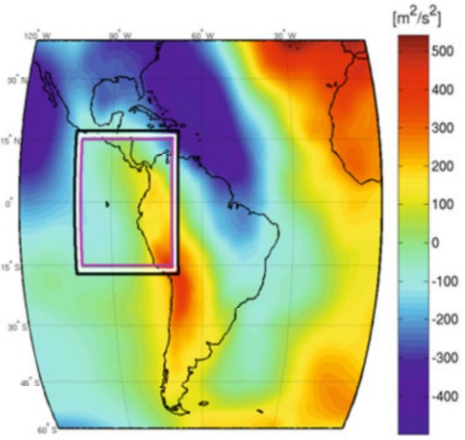
(b) Details WT^{τ_4} at scale 4 from 281 428 data points distributed within the black bordered region in Figure 2.8(a)



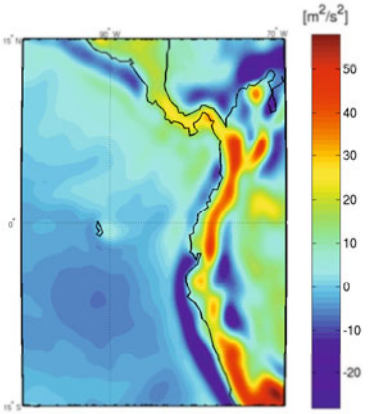
(c) Details WT^{τ_5} at scale 5 from 226 800 data points distributed within the gray bordered region in Figure 2.8(a)

FIGURE 2.8. Low-pass filtered version T^{τ_4} of the disturbing potential T in $[\frac{m^2}{s^2}]$ and the corresponding band-pass filtered versions WT^{τ_j} for scales $j = 4, 5$ of the magenta bordered region in subfigure 2.8(a) calculated from different numbers of data points (from the Ph.D.-thesis [73], Geomathematics Group, University of Kaiserslautern).

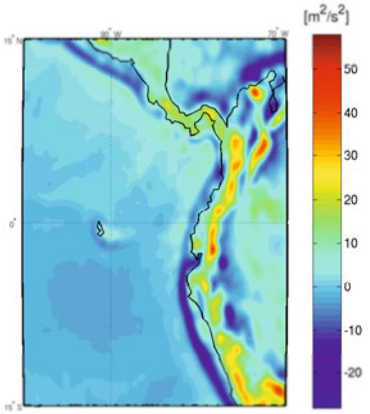
Seen from the geodetic reality, the figures are remarkable in the following sense: For getting a better accuracy in numerical integration procedures providing the



(a) Low pass part T^{τ_6} of the magenta bordered region in Figure 2.8(a) computed by the sum of T^{τ_4} (Figure 2.8(a)), WT^{τ_4} (Figure 2.8(b)), and WT^{τ_5} (Figure 2.8(c)) in this region

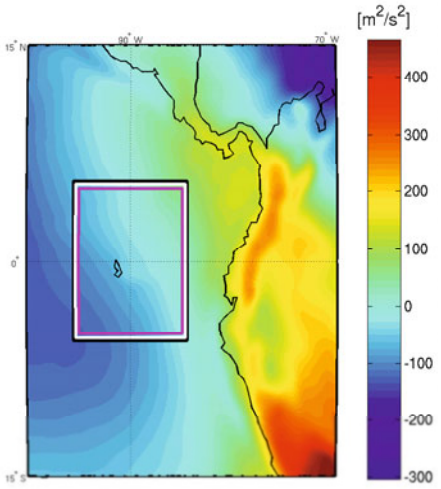


(b) Details WT^{τ_6} at scale 6 from 71 253 data points distributed within the black bordered region in Figure 2.9(a)

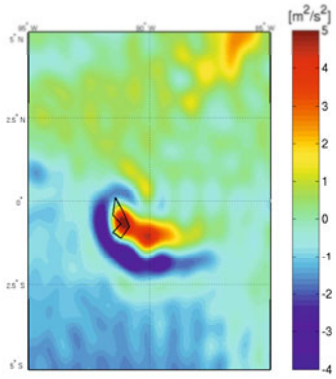


(c) Details WT^{τ_7} at scale 7 from 63 190 data points distributed within the gray bordered region in Figure 2.9(a)

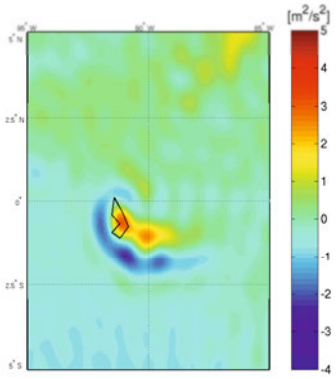
FIGURE 2.9. Low-pass filtered version T^{τ_6} of the disturbing potential T in $[\frac{m^2}{s^2}]$ of the magenta bordered region in subfigure 2.8(a) and the corresponding band-pass filtered versions WT^{τ_j} for scales $j = 6, 7$ (from the Ph.D.-thesis [73], Geomathematics Group, University of Kaiserslautern).



(a) Low pass part $T^{\tau 8}$ of the magenta bordered region in Figure 2.9(a) computed by the sum of $T^{\tau 6}$ (Figure 2.9(a)), $WT^{\tau 6}$ (Figure 2.9(b)), and $WT^{\tau 7}$ (Figure 2.9(c)) in this region



(b) Details $WT^{\tau 8}$ at scale 8 from 71 253 data points distributed within the black bordered region in Figure 2.10(a)



(c) Details $WT^{\tau 9}$ at scale 9 from 63 190 data points distributed within the gray bordered region in Figure 2.10(a)

FIGURE 2.10. Low-pass filtered version $T^{\tau 8}$ of the disturbing potential T in $[\frac{m^2}{s^2}]$ of the magenta bordered region in subfigure 2.9(a) and the corresponding band-pass filtered versions WT^{τ_j} for scales $j = 8, 9$ (from the Ph.D.-thesis [73], Geomathematics Group, University of Kaiserslautern).

(global) solution of the boundary-value problem (ENPPG) as illustrated in Figure 2.8 (a), we need denser, globally over the whole sphere Ω_R equidistributed data sets (most notably, in the sense of Weyl's Law of Equidistribution). However, in today's reality of gravitational field observation, we are confronted with the problem that terrestrial gravitational data (such as gravity disturbances, gravity anomalies) of sufficient width and quality are only available for certain parts of the Earth's surface (for more details concerning the observational aspects see, e.g., [6–8, 61, 62]). As a matter of fact, there are large gaps, particularly at sea, where no data sets of sufficient quality are available at all. This is the reason why the observational situation implies the need for specific geodetically oriented modeling techniques taking the heterogeneous data situation and the local availability of the data (usually related to latitude-longitude data grids) into consideration. In this respect, the "zooming-in" realization based on single-layer space-regularization is a suitable efficient and economic mathematical answer.

Disturbing Potential from Vertical Deflections. As already known from (2.56), the solution of the surface differential equation (see Eq. (2.36))

$$\nabla_\xi^* T(R\xi) = -\frac{\gamma M}{R}\Theta(R\xi), \qquad \xi \in \Omega, \tag{2.92}$$

determining the disturbing potential T from prescribed vertical deflections Θ under the conditions (2.40) is given by

$$T(R\xi) = \frac{R}{4\pi}\int_\Omega \Theta(R\eta)\cdot g\left(\Delta^*;\xi,\eta\right)\, d\omega(\eta), \tag{2.93}$$

where the vector kernel $g\left(\Delta^*;\xi,\eta\right),\ 1-\xi\cdot\eta>0$, reads as follows (see Eq. (2.57))

$$\begin{aligned} g\left(\Delta^*;\xi,\eta\right) &= -\frac{1}{2}\frac{\gamma M}{R^2}\frac{2}{1-\xi\cdot\eta}(\xi-(\xi\cdot\eta)\eta) \\ &= -\frac{1}{2}\frac{\gamma M}{R^2}(S(\xi\cdot\eta))^2(\xi-(\xi\cdot\eta)\eta). \end{aligned} \tag{2.94}$$

Analogously to the calculation of the disturbing potential T from known gravity disturbances D (i.e., the Neumann problem (ENPPG)), the numerical calamities of the improper integral in (2.93) can be circumvented by replacing the zonal kernel $S(\xi\cdot\eta)$ by the regularized kernel $S^\tau(\xi\cdot\eta)$. This process leads to space-regularized representations T^τ of the disturbing potential T calculated from vertical deflections Θ within a multiscale "zooming-in" procedure analogous to the approach for gravity disturbances as input data. To be more concrete, the kernel function $g(\Delta^*;\cdot,\cdot)$ is replaced by the space-regularized function using Eq. (2.63)

$$g^\tau\left(\Delta^*;\xi,\eta\right) = -\frac{\gamma M}{2R^2}\left(S^\tau(\xi\cdot\eta)\right)^2(\xi-(\xi\cdot\eta)\eta), \tag{2.95}$$

$$= \begin{cases} -\frac{\gamma M}{2R^2}\left(\frac{9R^2}{\tau^2}-\frac{12R^4}{\tau^4}(1-\xi\cdot\eta)+\frac{4R^6}{\tau^6}(1-\xi\cdot\eta)^2\right)(\xi-(\xi\cdot\eta)\eta), & 0\le 1-\xi\cdot\eta\le\frac{\tau^2}{2R^2}, \\ -\frac{\gamma M}{2R^2}\frac{2}{1-\xi\cdot\eta}(\xi-(\xi\cdot\eta)\eta), & \frac{\tau^2}{2R^2}<1-\xi\cdot\eta\le 2, \end{cases}$$

for $\tau \in (0, 2R]$. This leads to the following approximative representation of the disturbing potential T:

$$T^{\tau}(R\xi) = \frac{R}{4\pi} \int_{\Omega} \Theta(R\eta) \cdot g^{\tau}\left(\Delta^*; \xi, \eta\right) \, d\omega(\eta), \tag{2.96}$$

with $g^{\tau}(\Delta^*; \cdot, \cdot)$ given by (2.95). Using Eq. (2.70) from Lemma 2.6 we obtain

Theorem 2.12. *Suppose that T is the solution* (2.93) *of the differential equation* (2.92)*, with Θ being a member of the class of continuous vector-valued functions $c^{(0)}(\Omega_R)$. Let T^{τ}, $\tau \in (0, 2R]$, represent its regularized solution of the form* (2.96)*. Then*

$$\lim_{\tau \to 0+} \sup_{\xi \in \Omega} |T(R\xi) - T^{\tau}(R\xi)| = 0. \tag{2.97}$$

By restricting $\{g^{\tau}\left(\Delta^*; \cdot, \cdot\right)\}_{\tau \in (0, 2R]}$ to the sequence $\{g^{\tau_j}\left(\Delta^*; \cdot, \cdot\right)\}_{j \in \mathbb{N}_0}$, corresponding to a set of scaling parameters $\{\tau_j\}_{j \in \mathbb{N}_0}$ satisfying $\tau_j \in (0, 2R]$ and $\lim_{j \to \infty} \tau_j = 0$, we are canonically led to regularized vector scaling functions such that a scale-discrete solution method for the differential equation (2.92) can be formulated. The vector scaling function $g^{\tau_{j+1}}(\Delta^*; \cdot, \cdot)$ at scale $j+1$ is constituted by the sum of the vector scaling function $g^{\tau_j}(\Delta^*; \cdot, \cdot)$ and the corresponding discretized vector wavelet $wg^{\tau_j}(\Delta^*; \cdot, \cdot)$, given by

$$wg^{\tau_j}\left(\Delta^*; \xi, \eta\right) = g^{\tau_{j+1}}\left(\Delta^*; \xi, \eta\right) - g^{\tau_j}\left(\Delta^*; \xi, \eta\right). \tag{2.98}$$

Note that (cf. [15])

$$WT^{\tau_j}(R\xi) = \int_{\Omega} \Theta(R\eta) \cdot wg^{\tau_j}(\Delta^*; \xi, \eta) d\omega(\eta).$$

Application: Gravitational signatures of mantle plumes

Galapagos: "The Galapagos hotspot (Figures 2.8–2.10) is a volcanic hotspot in the East Pacific Ocean responsible for the creation of the Galapagos Islands as well as three major aseismic ridge systems, Carnegie, Cocos and Malpelso which are on two tectonic plates. The hotspot is located near the Equator on the Nazca Plate not far from the divergent plate boundary with the Cocos Plate. The tectonic setting of the hotspot is complicated by the Galapagos Triple Junction of the Nazca and Cocos plates with the Pacific Plate. The movement of the plates over the hotspot is determined not solely by the spreading along the ridge but also by the relative motion between the Pacific Plate and the Cocos and Nazca Plates.

The hotspot is believed to be over 20 million years old and in that time, there has been interaction between the hotspot, both of these plates, and the divergent plate boundary, at the Galapagos Spreading Center. Lavas from the hotspot do not exhibit the homogeneous nature of many hotspots; instead there is evidence of four major reservoirs feeding the hotspot. These mix to varying degrees at different locations on the archipelago and also within the Galapagos Spreading Center." (from [71]) (for more details the reader is referred, e.g., to [31] and the references therein).

Hawaii: [58] believe that a stationary mantle plume located beneath the Hawaiian Islands created the Hawaii-Emperor seamount chain while the oceanic lithosphere continuously passed over it. The Hawaii-Emperor chain consists of about 100 volcanic islands, atolls, and seamounts that spread nearly 6000km from the active volcanic island of Hawaii to the 75–80 million year old Emperor seamounts nearby the Aleutian trench. With moving further south east along the island chain, the geological age decreases. The interesting area is the relatively young southeastern part of the chain, situated on the Hawaiian swell, a 1200km broad anomalously shallow region of the ocean floor, extending from the island of Hawaii to the Midway atoll. Here, a distinct gravity disturbance and geoid anomaly occurs that has its maximum around the youngest island that coincides with the maximum topography and both decrease in northwestern direction. The progressive decrease in terms of the geological age is believed to result from the continuous motion of the underlying plate (cf. [50, 72]).

With seismic tomography, several features of the Hawaiian mantle plume are gained (cf. [58] and the references therein). They result in a *Low Velocity Zone* (LVZ) beneath the lithosphere, starting at a depth of about 130–140km beneath the central part of the island of Hawaii. So far, plumes have just been identified as low seismic velocity anomalies in the upper mantle and the transition zone, which is a fairly new achievement. As plumes are relatively thin with respect to their diameter, they are hard to detect in global tomography models. Hence, despite novel advances, there is still no general agreement on the fundamental questions concerning mantle plumes, like their depth of origin, their morphology, their longevity, and even their existence is still discussed controversial. This is due to the fact that many geophysical as well as geochemical observations can be explained by different plume models and even by models that do not include plumes at all (e.g., [10]). With our space-localized multiscale method of deriving gravitational signatures (more concretely, the disturbing potential) from the vertical deflections, we add a new component in specifying essential features of plumes. The vertical deflections of the plume in the region of Hawaii are visualized in Figure 2.11.

From the band-pass filtered detail approximation of the vertical deflections (Figure 2.12) and the corresponding disturbing potential (Figure 2.13), we are able to conclude that the Hawaii plume has an oblique layer structure. As can be seen in the lower scale (for which numerical evidence suggests that they reflect the higher depths), the strongest signal is located in the ocean in a westward direction of Hawaii. With increasing scale, i.e., lower depths, it moves more and more to the Big Island of Hawaii, i.e., in eastward direction.

Iceland: The plume beneath Iceland is a typical example of a ridge-centered mantle plume. An interaction between the North Atlantic ridge and the mantle plume is believed to be the reason for the existence of Iceland, resulting in melt production and crust generation since the continental break-up in the late Palaeocene and early Eocene. Nevertheless, there is still no agreement on the location of the plume before rifting started in the East. Controversial discussions, whether it was

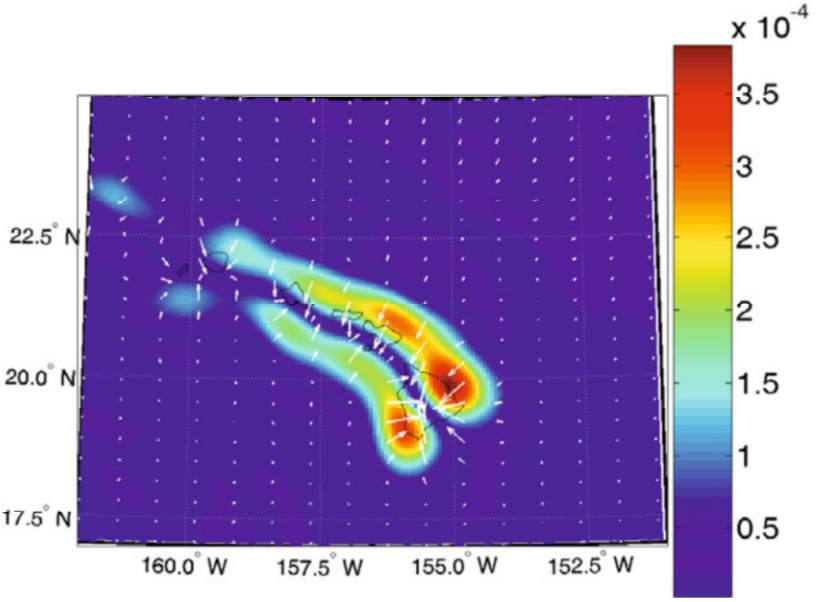


FIGURE 2.11. Illustration of the vertical deflections Θ in the region of Hawaii (from the Ph.D.-thesis [9], Geomathematics Group, University of Kaiserslautern).

located under central or eastern Greenland about 62-64 million years ago are still in progress (cf. [63] and the references therein).

Iceland itself represents the top of a nearly circular rise topography, with a maximum of about 2.8km above the surrounding seafloor in the south of the glacier "Vatnajökull". Beneath this glacier, several active volcanoes are located, which are supposed to be fed by a mantle plume. The surrounding oceanic crust consists of three different types involving a crust thickness that is more than three times as thick as average oceanic crusts. Seismic tomography provides evidence of the existence of a mantle plume beneath Iceland, resulting in low velocity zones in the upper mantle and the transition zone, but also hints for anomalies in the deeper mantle seem to exist. The low velocity anomalies have been detected in depths ranging from at least 400km up to about 150km. Above 150km, ambiguous seismic-velocity structures were obtained involving regions of low velocities covered by regions of high seismic velocities. For a deeper access into the theory of the Iceland plume, the interested reader is referred to [58] and the references therein.

From Figures 2.14 to 2.16, it can be seen that the mantle plume in lower scales, i.e., in higher depths, starts in the North of Iceland and with increasing scale, i.e., lower depths, it moves to the South. It is remarkable that from scale 13 on, the plume seems to divide into two sectors. Since it is known that the disturbing

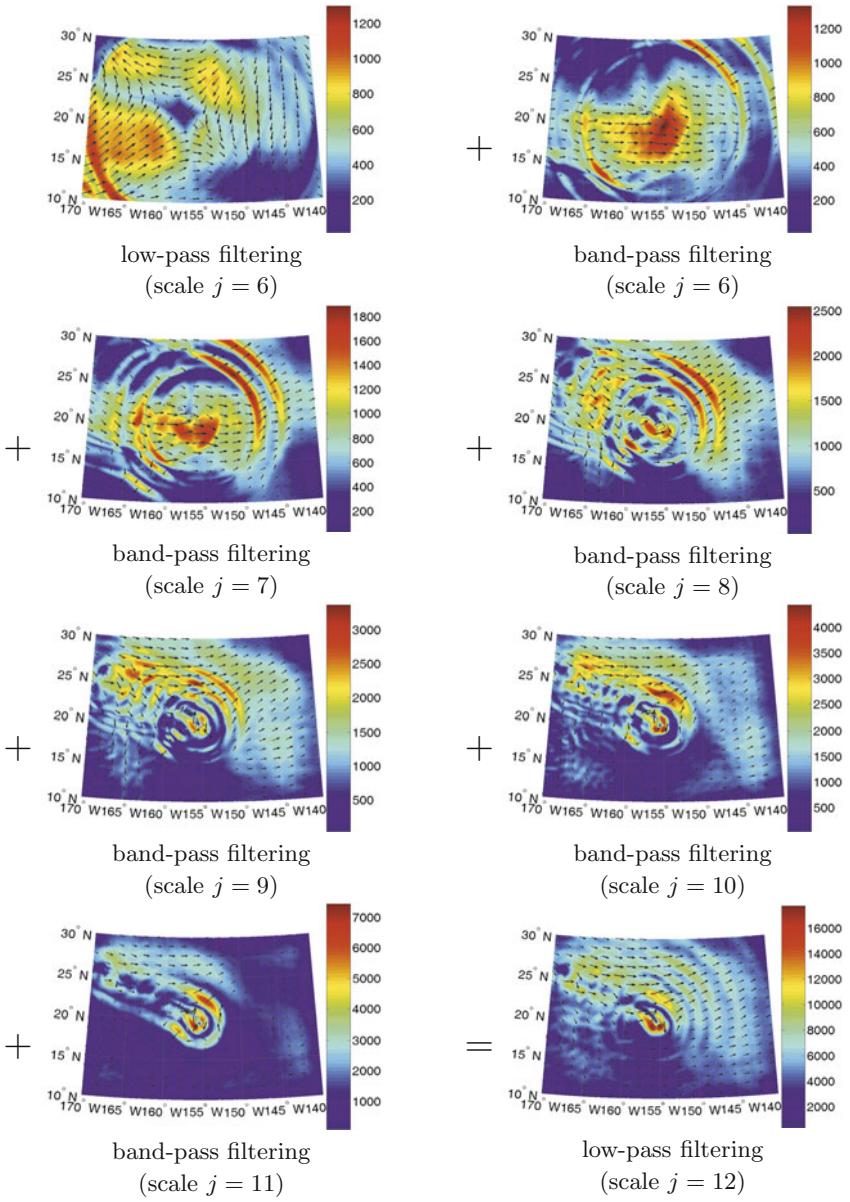


FIGURE 2.12. Approximation of the vector-valued vertical deflections Θ in $[\mathrm{ms}^{-2}]$ of the region of Hawaii (compare Fig. 2.11). A rough low-pass filtering at scale 6 is improved by several band-pass filters of scale $j = 6, \ldots, 11$, the last picture shows the multiscale approximation at scale $j = 12$, (from the Ph.D.-thesis [9], Geomathematics Group, University of Kaiserslautern).

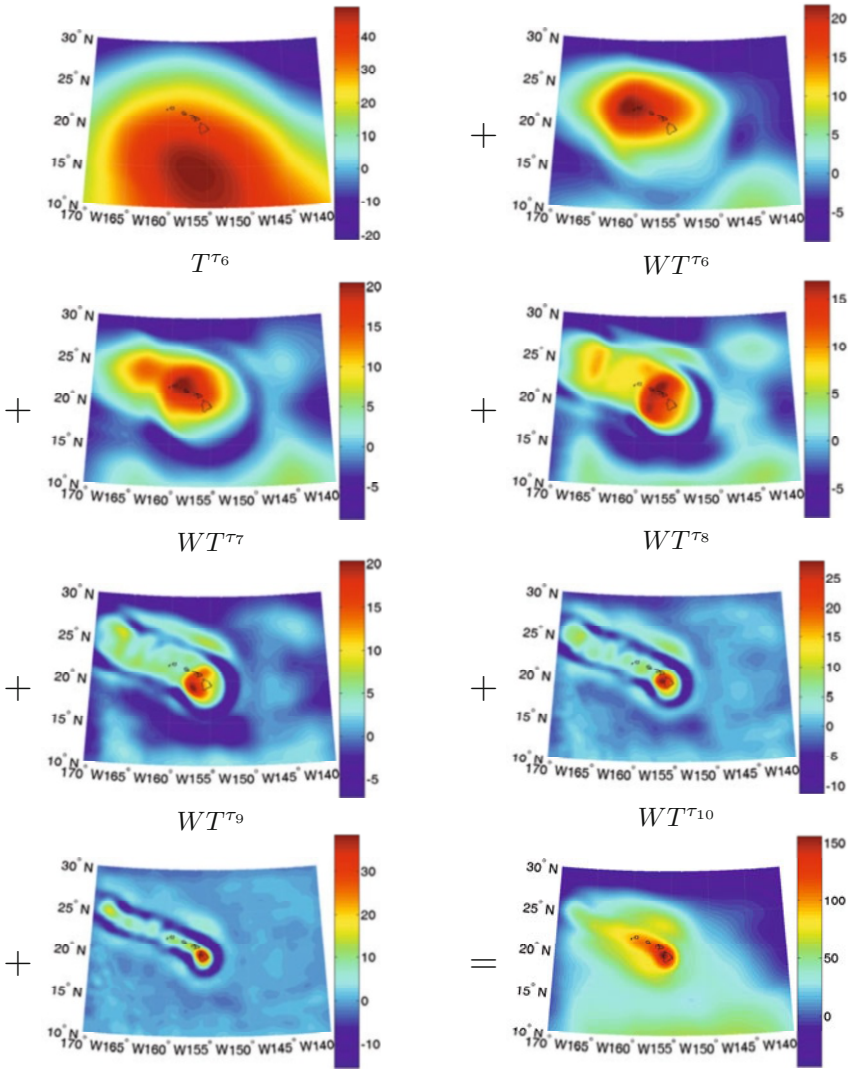


FIGURE 2.13. Multiscale reconstruction of the disturbing potential T in $[\mathrm{m}^2\mathrm{s}^{-2}]$ from vertical deflections Θ for the Hawaiian (plume) area using the scaling function g^{τ} (a rough low-pass filtering T^{τ_6} at scale $j = 6$ is improved by several band-pass filters WT^{τ_j} at scales $j = 6, \ldots, 11$, the last illustration shows the approximation $T^{\tau_{12}}$ of the disturbing potential T at scale $j = 12$, (from the Ph.D.-thesis [9], Geomathematics Group, University of Kaiserslautern).

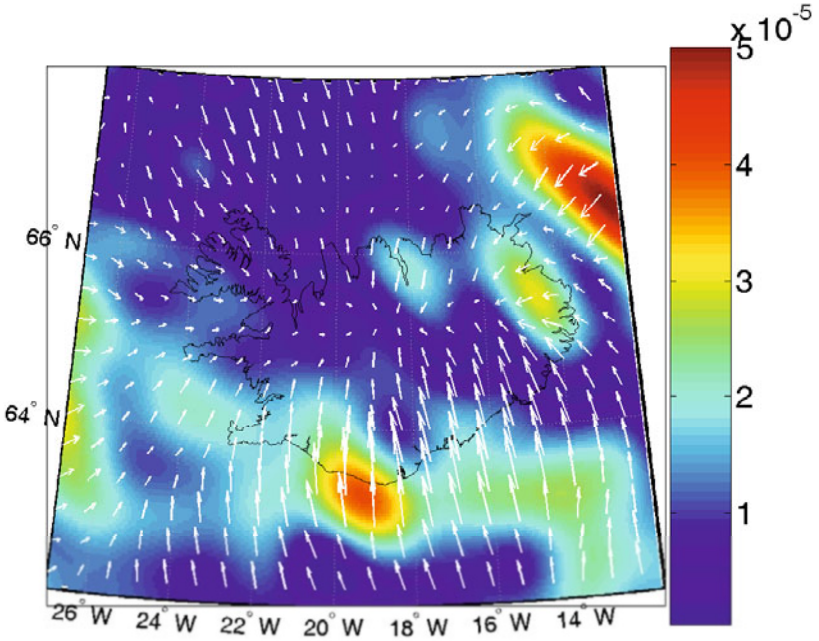


FIGURE 2.14. Illustration of the vertical deflections Θ in the region of Iceland (see [15]).

potential of the Earth is influenced by its topography, a look at a topographic map shows that the sector located more Eastern is (probably) caused by the Vatnajökull glacier (being the biggest glacier in Europe).

All in all, from our multiscale reconstruction, it can be derived that the deeper parts of the mantle plume are located in the northern part of Iceland (compare the lower scales in Figure 2.15) while shallower parts are located further south (compare the higher scales in Figure 2.15). As the North American plate moves westward and the Eurasian plate eastward, new crust is generated on both sides of the Mid-Atlantic Ridge. In the case of Iceland, which lies on the Mid-Atlantic Ridge, the neovolcanic zones are readily seen in Figure 2.16.

In Iceland, electrical production from geothermal power plants has been developed rapidly. Reflecting the geological situation, Iceland is a unique country with regard to utilization of geothermal energy, with more than 50% of its primary energy consumption coming from geothermal power plants. As shown in Figure 2.17, today's location of power plants in Iceland fits perfectly with the gravimetric investigations based on horizontal/vertical derivatives of the Earth's disturbing potential. As a matter of fact, only from these results it becomes obvious where future power plants should be placed for geothermal purposes.

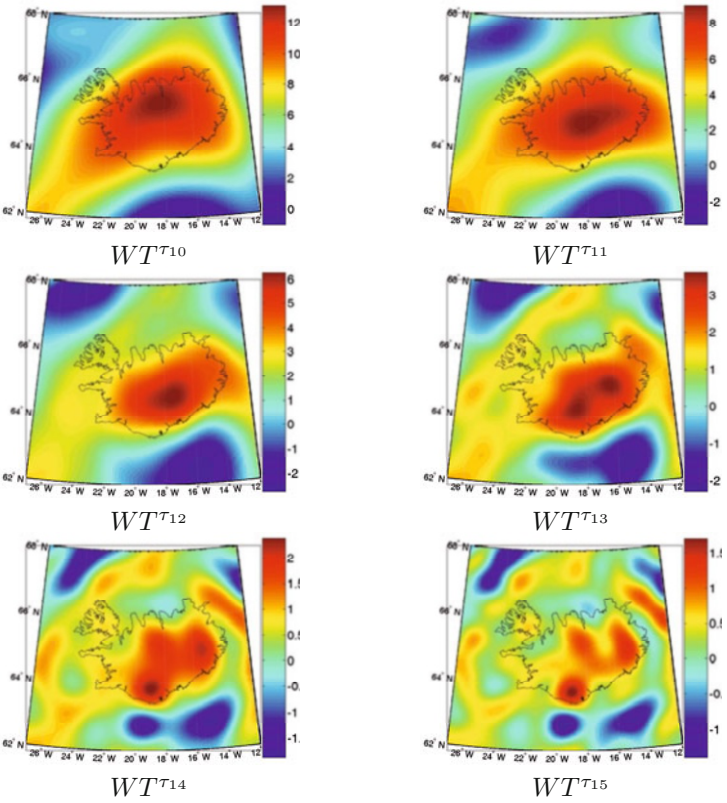


FIGURE 2.15. Band-pass filtered details WT^{τ_j} of the disturbing potential T in $[m^2 s^{-2}]$ from vertical deflections Θ in the region of Iceland with respect to the scales $j = 10, \ldots, 15$, (from [15]).

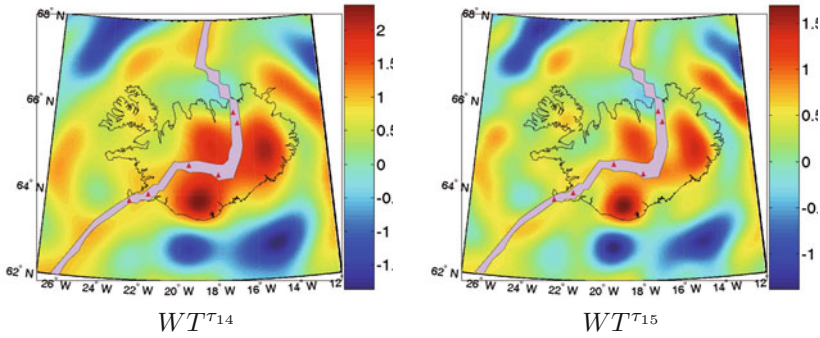


FIGURE 2.16. Band-pass filtered details WT^{τ_j} of the disturbing potential T in $\left[m^2 s^{-2}\right]$ from vertical deflections Θ in the region of Iceland for $j = 14, 15$ including the Mid-Atlantic Ridge (gray).

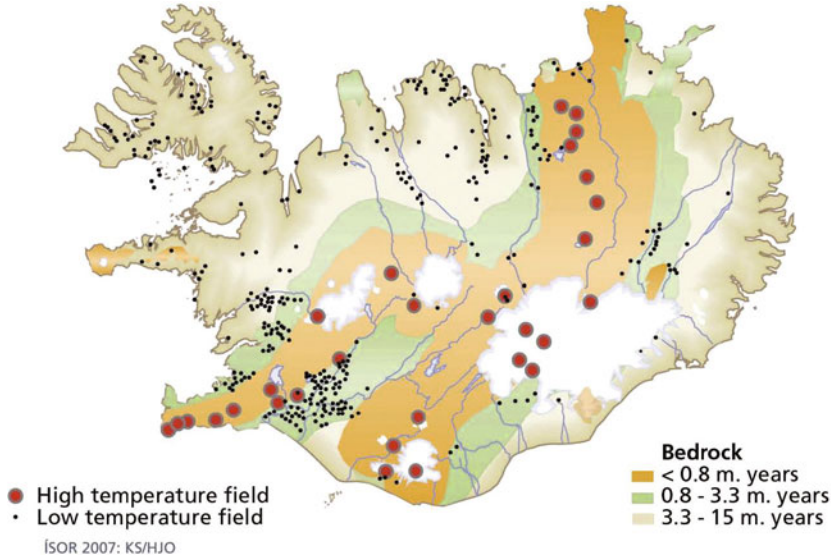


FIGURE 2.17. Geothermal power plants in Iceland [1, 37].

All in all, by the space-based multiscale techniques initiated by Freeden and Schreiner [21, 22] in gravitation we are able to come to interpretable results involving geological obligations in relation to hotspots/mantle plumes based on "surface interpretations" and just by looking at the anomalous behavior in terms of surface integrals without using the framework of Newton's volume integrals.

3. Interior gravitational potential and density distribution

3.1. Newton integral and Poisson equation

Seen from a mathematical point of view, the Earth's gravitational field v is a gradient field $v = \nabla V$, where the gravitational potential V is an infinitely often differentiable harmonic scalar field in the exterior of the Earth. As a consequence, the Earth's gravitational field v is an infinitely often differentiable vector field in the exterior of the Earth satisfying $\nabla \cdot v = 0,\ \nabla \wedge v = 0$.

According to the classical Newton Law of Gravitation (1687), knowing the density distribution of a region $\mathcal{G}$ such as the Earth, the gravitational potential (Newton potential) can be computed everywhere in $\mathbb{R}^3$. More explicitly, the gravitational potential V of the Earth's exterior $\mathcal{G}^c = \mathbb{R}^3 \backslash \overline{\mathcal{G}}$ is given by

$$V(x) = \gamma \int_{\mathcal{G}} F(y) G(\Delta; |x-y|)\, dV(y), \qquad x \in \mathbb{R}^3 \backslash \overline{\mathcal{G}}, \tag{3.1}$$

with the so-called fundamental solution $G(\Delta;\cdot)$ of the Laplace equation given by

$$G(\Delta;|x-y|) = -\frac{1}{4\pi}\frac{1}{|x-y|}, \tag{3.2}$$

and the gravitational constant γ, where F is the density function. Since γ is a constant, it has no effect on any of the following considerations. Hence, from now on, for the sake of simplicity, we neglect the gravitational constant γ in all equations, but it will be observed in numerical computations. The properties of the gravitational potential V in the Earth's exterior are easily described as follows:

(i) V is *harmonic* in $\mathbb{R}^3\backslash\overline{\mathcal{G}}$, i.e., $\Delta_x V(x) = 0,\ x \in \mathbb{R}^3\backslash\overline{\mathcal{G}}$.
(ii) V is *regular at infinity*, i.e., $|V(x)| = O\left(|x|^{-1}\right),\ |x| \to \infty$.

Let $\mathcal{G} \subset \mathbb{R}^3$ be a *regular region*, i.e., a bounded region $\mathcal{G} \subset \mathbb{R}^3$ dividing $\mathbb{R}^3$ uniquely into the inner space $\mathcal{G}$ and the outer space $\mathcal{G}^c = \mathbb{R}^3\backslash\overline{\mathcal{G}}$, $\overline{\mathcal{G}} = \mathcal{G} \cup \partial\mathcal{G}$, such that the boundary $\partial\mathcal{G}$ is an orientable smooth Lipschitzian manifold of dimension 2 (for example, ball, ellipsoid, geoid, Earth or appropriate cuboidal parts of it). It is already known, that the Newton (volume) integral over a regular region $\mathcal{G}$, corresponding to a mass density distribution F satisfies the Laplace equation in the outer space $\mathcal{G}^c = \mathbb{R}^3\backslash\overline{\mathcal{G}}$. Clearly, this property is an immediate consequence of the harmonicity of the fundamental solution for the Laplace equation (see, e.g., [40]).

Theorem 3.1. *Let $F : \overline{\mathcal{G}} \to \mathbb{R}$ be an integrable, bounded function. Then*

$$V(x) = \int_{\mathcal{G}} F(y)\ G(\Delta;|x-y|)\ dV(y), \qquad x \in \mathcal{G}^c, \tag{3.3}$$

satisfies

$$\Delta_x V(x) = \Delta_x \int_{\mathcal{G}} F(y)\ G(\Delta;|x-y|)\ dV(y) = 0, \quad x \in \mathcal{G}^c, \tag{3.4}$$

i.e., V is harmonic in $\mathcal{G}^c$.

Next, we are interested in showing that the Newton integral in the inner space satisfies the Poisson equation at least under some canonical conditions on the density function (see, e.g., [15]).

Theorem 3.2. *Let $F : \overline{\mathcal{G}} \to \mathbb{R}$ be of class $C^{(0)}(\overline{\mathcal{G}})$. Then V as defined by* (3.3) *is of class $C^{(1)}(\overline{\mathcal{G}})$. Furthermore, we have*

$$\nabla_x V(x) = \int_{\mathcal{G}} F(y)\ \nabla_x G(\Delta;|x-y|)\ dV(y), \qquad x \in \overline{\mathcal{G}}. \tag{3.5}$$

Proof. The fundamental solution

$$G(\Delta;|x-y|) = -\frac{1}{4\pi}\frac{1}{|x-y|}, \qquad |x-y| \neq 0, \tag{3.6}$$

admits a "regularization" (mollification) of the form

$$G_0^\tau(\Delta;|x-y|) = \begin{cases} -\dfrac{3\tau^2-|x-y|^2}{8\pi\tau^3}, & |x-y| \le \tau, \\ -\dfrac{1}{4\pi|x-y|}, & \tau < |x-y|. \end{cases} \tag{3.7}$$

For brevity, we set

$$V_0^\tau(x) = \int_{\mathcal{G}} F(y)\ G_0^\tau(\Delta;|x-y|)\ dV(y), \quad x \in \overline{\mathcal{G}}. \tag{3.8}$$

The integrands of V and V_0^τ only differ in the ball $\mathcal{B}_\tau(x)$ around the point x with radius τ. Moreover, the function $F : \overline{\mathcal{G}} \to \mathbb{R}$ is supposed to be continuous on $\overline{\mathcal{G}}$. Hence, it is uniformly bounded on $\overline{\mathcal{G}}$ and we derive

$$\begin{aligned} \sup_{x\in\overline{\mathcal{G}}} |V(x) - V_0^\tau(x)| &= \mathcal{O}\left(\int_{\mathcal{B}_\tau(x)} |G(\Delta;|x-y|) - G_0^\tau(\Delta;|x-y|)|\ dV(y) \right) \\ &= \mathcal{O}(\tau^2). \end{aligned} \tag{3.9}$$

Therefore, V is of class $\mathrm{C}^{(0)}(\overline{\mathcal{G}})$ as the limit of a uniformly convergent sequence of continuous functions on $\overline{\mathcal{G}}$. We let

$$v(x) = \int_{\mathcal{G}} F(y)\ \nabla_x G(\Delta;|x-y|)\ dV(y), \quad x \in \overline{\mathcal{G}}, \tag{3.10}$$

and

$$v_0^\tau(x) = \int_{\mathcal{G}} F(y)\ \nabla_x G_0^\tau(\Delta;|x-y|)\ dV(y), \quad x \in \overline{\mathcal{G}}. \tag{3.11}$$

As $|\nabla_x G(\Delta;|x-y|)| = \mathcal{O}(|x-y|^{-2})$, the integrals v and v_0^τ exist for all $x \in \overline{\mathcal{G}}$. It is not hard to see that

$$\sup_{x\in\overline{\mathcal{G}}} |v(x) - v_0^\tau(x)| = \sup_{x\in\overline{\mathcal{G}}} |v(x) - \nabla_x V_0^\tau(x)| = \mathcal{O}(\tau). \tag{3.12}$$

Consequently, v is a continuous vector field on $\overline{\mathcal{G}}$. Moreover, as the relation (3.12) holds uniformly on $\overline{\mathcal{G}}$, we obtain

$$v(x) = \nabla_x V(x) = \int_{\mathcal{G}} F(y)\ \nabla_x G(\Delta;|x-y|)\ dV(y). \tag{3.13}$$

This is the desired result. □

Remark 3.3. The proof is standard (see, e.g., [22]). Its explicit formulation, however, is helpful to understand the feature extraction method.

Next, we come to the Poisson equation under the assumption of Hölder continuity for the function F on $\overline{\mathcal{G}}$.

Theorem 3.4. *If F is of class $C^{(0,\mu)}(\overline{\mathcal{G}})$, $\mu \in (0,1]$, then the Poisson differential equation*

$$\Delta_x \int_{\mathcal{G}} F(y)\ G(\Delta; |x-y|)\ dV(y) = F(x) \tag{3.14}$$

holds true for all $x \in \mathcal{G}$.

The proof can be found in any textbook on potential theory, e.g., [15]. It is also part of Chapter 5 of this handbook.

The fundamental solution $G_0^\tau(\Delta;\cdot)$ as well as the (ordinary) *Haar function* given by

$$H_0^\tau(|x-y|) = \Delta_x G_0^\tau(\Delta; |x-y|) = \begin{cases} \frac{3}{4\pi\tau^3}, & |x-y| \le \tau, \\ 0, & |x-y| > \tau, \end{cases} \tag{3.15}$$

are depicted in Figure 3.1 for different values of τ.

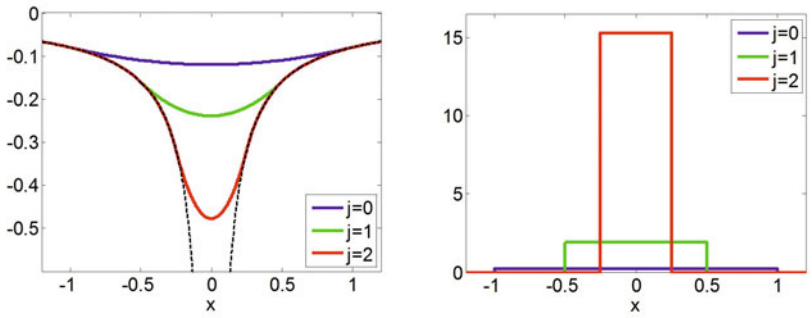


FIGURE 3.1. Sectional profile of the functions $G_0^\tau(\Delta;\cdot)$ (left) and H_0^τ (right) for the values $\tau = 2^{-j}, j = 0, 1, 2$. The black line in the left figure indicates the profile of the fundamental solution $G(\Delta;\cdot)$.

The critical point that will be expected in numerics of feature extraction by means of regularized potentials V_0^τ is the discontinuity of the Laplace derivative of $G_0^\tau(\Delta;\cdot)$, i.e., the (ordinary) Haar function H_0^τ. This is the reason why we are interested in higher-order Taylor expansions of the fundamental solution leading to a polynomial of degree $n+2$ given by $r \mapsto G_n^\tau(\Delta; r),\ r \in [0, \infty,)$ with

$$G_n^\tau(\Delta;r) = \begin{cases} -\frac{1}{4\pi r}, & \tau \le r, \\ \frac{1}{8\pi\tau^{n+3}} \sum_{l=0}^{n+1} \left((-1)^l (n+1) \binom{n+2}{l}\right) \tau^{n+2-l} r^l & \\ \quad + \frac{1}{8\pi\tau^{n+3}} \sum_{l=0}^{n+1} \left(2(-1)^{l+1} \binom{n+2}{l+1}\right) \tau^{n+2-l} r^l & \\ \quad + (-1)^{n+2}\ \frac{n+1}{8\pi\tau^{n+3}}\ r^{n+2}, & 0 \le r < \tau \end{cases} \tag{3.16}$$

instead of

$$G_0^\tau(\Delta;r) = \begin{cases} -\frac{1}{4\pi r}, & \tau \le r, \\ -\frac{3\tau^2 - r^2}{8\pi\tau^3}, & 0 \le r < \tau, \end{cases} \tag{3.17}$$

so that

$$H_n^\tau(r) = \Delta_x G_n^\tau(\Delta;r) = \begin{cases} 0, & \tau < r, \\ \frac{(n+1)(n+2)(n+3)}{8\pi} \frac{(\tau - r)^n}{\tau^{n+3}}, & 0 \le r \le \tau. \end{cases} \tag{3.18}$$

It is easy to see that $r \mapsto G_n^\tau(\Delta;r), r \in [0,\infty)$, is $(n+1)$-times continuously differentiable and $r \mapsto H_n^\tau(r), r \in [0,\infty)$, is $(n-1)$-times continuously differentiable (where, by convention in case of H_0^τ, (-1)-times continuously differentiable means piecewise continuous). Moreover, we notice that H_0^τ for $n = 0$ is the ordinary (spherically symmetric) τ-Haar function in $\mathbb{R}^3$.

As a consequence of our preparatory considerations we obtain the following statement that serves as strategic basis for our forthcoming approach to geological feature extraction.

Theorem 3.5. *For $n \in \mathbb{N}_0$, the "τ-potential functions" of order n*

$$V_n^\tau(x) = \int_{\mathcal{G}} G_n^\tau(\Delta; |x-y|) F(y)\, dV(y) \tag{3.19}$$

and the "τ-contrast functions" of order n

$$F_n^\tau(x) = \int_{\mathcal{G}} H_n^\tau(|x-y|) F(y)\, dV(y), \tag{3.20}$$

satisfy the limit relations

$$\lim_{\tau \to 0} |V(x) - V_n^\tau(x)| = 0, \quad x \in \mathcal{G} \tag{3.21}$$

and

$$\lim_{\tau \to 0} |F(x) - F_n^\tau(x)| = 0, \quad x \in \mathcal{G}, \tag{3.22}$$

provided that F is ($\mathrm{C}^{(0,\mu)}$-Hölder) continuous in the neighborhood of $x \in \mathcal{G}$.

The kernels $G_n^\tau(\Delta;\cdot)$ and H_n^τ are called *"τ-fundamental scaling function of order n"* and *"τ-Haar scaling function of order n"*, respectively. It should be remarked that $G_n^\tau(\Delta;\cdot)$ is constructed in such a way that the normalization condition

$$\int_{\mathbb{R}^3} \Delta_x G_n^\tau(\Delta;|x|)\, dV(x) = \int_{\mathbb{R}^3} H_n^\tau(|x|)\, dV(x) = 1 \tag{3.23}$$

holds true for all $\tau > 0$ and all $n \in \mathbb{N}_0$.

Unfortunately, τ-potential functions V_n^τ do not generally show a faster convergence to V than τ-potential functions V_0^τ; more concretely, we have

$$\begin{aligned}\sup_{x\in\overline{\mathcal{G}}}|V(x)-V_n^\tau(x)| &= \mathcal{O}\left(\int_{\mathcal{B}_\tau(x)}|G(\Delta;|x-y|)-G_n^\tau(\Delta;|x-y|)|\;dV(y)\right)\\ &= \mathcal{O}(\tau^2)\end{aligned} \tag{3.24}$$

for $n\in\mathbb{N}_0$. Finally it should be alluded that

$$\lim_{\tau\to 0}\sup_{x\in\overline{\mathcal{G}}}|\alpha(x)F(x)-F_n^\tau(x)|=0, \tag{3.25}$$

where $\alpha(x)$ is the solid angle subtended at $x\in\overline{\mathcal{G}}$ by the boundary surface $\partial\mathcal{G}$.

Remark 3.6. The solid angle $\alpha(x)$ in Equation (3.25) is necessary due to the fact that the support of $H_n^\tau(|\cdot-x|)$ is cut of at the boundary $\partial\mathcal{G}$ for all $\tau>0$ with $x\in\partial\mathcal{G}$.

3.2. Multiscale postprocessing of signature decorrelation

Next we deal with new mathematical mechanisms for a deeper interpretation and a better understanding of gravimetrically available pre-information inside a regular region $\mathcal{G}$. In order to make the decorrelation mechanisms transparent, our considerations start from the unrealistic assumption that the potential V is known everywhere in $\overline{\mathcal{G}}$. Our purpose is to demonstrate how the multiscale procedure for the potential canonically transfers to the density by use of "Poisson derivatives". All in all, the context of this section is meant as conceptual preparation of the Haar-type inversion process (see also [3]) discussed later on.

Suppose that $\{\tau_j\}_{j\in\mathbb{N}_0}$ is a positive, monotonously decreasing sequence with $\lim_{j\to\infty}\tau_j=0$. For $j\in\mathbb{N}_0$, we consider the differences

$$\Psi_{G_n^{\tau_j}}(\Delta;|x-y|)=G_n^{\tau_{j+1}}(\Delta;|x-y|)-G_n^{\tau_j}(\Delta;|x-y|) \tag{3.26}$$

and

$$\Psi_{H_n^{\tau_j}}(|x-y|)=H_n^{\tau_{j+1}}(|x-y|)-H_n^{\tau_j}(|x-y|). \tag{3.27}$$

$\Psi_{G_n^{\tau_j}}(\Delta;\cdot)$ and $\Psi_{H_n^{\tau_j}}$ are called *"τ_j-fundamental wavelet function of order n"* and *"τ_j-Haar wavelet function of order n"*, respectively (see Figure 3.2).

The associated *"τ_j-potential wavelet functions" of order n* and the *"τ_j-contrast wavelet functions" of order n* are given by

$$(WV)_n^{\tau_j}(x)=\int_{\mathcal{G}}\Psi_{G_n^{\tau_j}}(\Delta;|x-y|)F(y)\;dV(y) \tag{3.28}$$

and

$$(WF)_n^{\tau_j}(x)=\int_{\mathcal{G}}\Psi_{H_n^{\tau_j}}(|x-y|)F(y)\;dV(y). \tag{3.29}$$

The τ_j-potential wavelet functions of order n and the τ_j-contrast wavelet functions of order n, respectively, characterize the successive *detail information* contained in $V_n^{\tau_{j+1}}-V_n^{\tau_j}$ and $F_n^{\tau_{j+1}}-F_n^{\tau_j}$, $j\in\mathbb{N}_0$. In other words, we are able to recover the

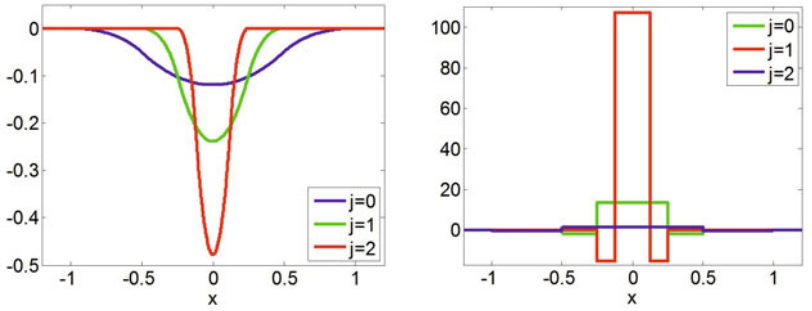


FIGURE 3.2. Sectional profile of the wavelet functions $\Psi_{G_n^{\tau_j}}(\Delta;\cdot)$ (left) and $\Psi_{H_n^{\tau_j}}(\cdot)$ (right) for $n=0$ and $\tau_j = 2^{-j}, j = 0,1,2$.

potential V and the contrast function, i.e., the "density signature" F, respectively, in form of "band structures"

$$(WV)_n^{\tau_j} = V_n^{\tau_{j+1}} - V_n^{\tau_j}, \tag{3.30}$$

and

$$(WF)_n^{\tau_j} = F_n^{\tau_{j+1}} - F_n^{\tau_j}. \tag{3.31}$$

As a consequence, the essential problem to be solved in multiscale extraction of geological features is to identify those detail information, i.e., band structures in (3.30), which contain specific geological (density) characteristics in (3.31), for example, aquifers, salt domes, etc.

Seen from a numerical point of view, it is remarkable that both wavelet functions $y \mapsto \Psi_{G_n^{\tau_j}}(\Delta;|x-y|)$ and $y \mapsto \Psi_{H_n^{\tau_j}}(|x-y|)$ vanish outside a ball around the center x due to their construction, i.e., these functions are spacelimited showing a ball as local support. Furthermore, the support becomes smaller and smaller with increasing scale parameter j, so that more and more high frequency phenomena can be highlighted without changing the features outside the balls. Explicitly written out in our nomenclature we obtain for $x \in \overline{\mathcal{G}}$

$$(WV)_n^{\tau_j}(x) = \int_{\mathcal{B}_{\tau_j}(x)\cap\mathcal{G}} \Psi_{G_n^{\tau_j}}(\Delta;|x-y|)F(y)\,dV(y), \tag{3.32}$$

and

$$(WF)_n^{\tau_j}(x) = \int_{\mathcal{B}_{\tau_j}(x)\cap\mathcal{G}} \Psi_{H_n^{\tau_j}}(|x-y|)F(y)\,dV(y). \tag{3.33}$$

Forming the sums

$$\sum_{j=0}^{J-1}(WV)_n^{\tau_j}(x) = \sum_{j=0}^{J-1}\left(V_n^{\tau_{j+1}}(x) - V_n^{\tau_j}(x)\right), \tag{3.34}$$

and

$$\sum_{j=0}^{J-1}(WF)_n^{\tau_j}(x) = \sum_{j=0}^{J-1}\left(F_n^{\tau_{j+1}}(x) - F_n^{\tau_j}(x)\right), \tag{3.35}$$

we are easily led to

$$V_n^{\tau_J}(x) = V_n^{\tau_0}(x) + \sum_{j=0}^{J-1}(WV)_n^{\tau_j}(x) \tag{3.36}$$

and

$$F_n^{\tau_J}(x) = F_n^{\tau_0}(x) + \sum_{j=0}^{J-1}(WF)_n^{\tau_j}(x). \tag{3.37}$$

Thus, we finally end up with the following multiscale relations

$$V(x) = \lim_{J\to\infty} V_n^{\tau_J}(x) = V_n^{\tau_0}(x) + \sum_{j=0}^{\infty}(WV)_n^{\tau_j}(x) \tag{3.38}$$

and

$$\alpha(x)F(x) = \lim_{J\to\infty} F_n^{\tau_J}(x) = F_n^{\tau_0}(x) + \sum_{j=0}^{\infty}(WF)_n^{\tau_j}(x) = \lim_{J\to\infty}\Delta_x V_n^{\tau_J}(x), \tag{3.39}$$

i.e.,

$$\alpha(x)F(x) = \Delta_x V_n^{\tau_0}(x) + \sum_{j=0}^{\infty}\Delta_x(WV)_n^{\tau_j}(x). \tag{3.40}$$

Altogether, the potential V as well as the contrast function, i.e., the "density signature" F can be expressed in additive way as a low-pass filtered signal $V_n^{\tau_0}$ and $F_n^{\tau_0}$ and successive band-pass filtered signals $(WV)_n^{\tau_j}$ and $(WF)_n^{\tau_j}$, $j = 0, 1, \ldots$, respectively.

It should be mentioned that our multiscale approach is constructed such that, within the spectrum of all wavebands (cf. (3.30), (3.31)), certain rock formations or aquifers, respectively, may be associated to a specific waveband characterizing typical features within the multiscale reconstruction (see Figure 3.3). Each scale parameter in the decorrelation is assigned to a data function which corresponds to the associated waveband and, thus, leads to a low-pass approximation of the data at a particular resolution.

Finally it should be noted that the key ideas of multiscale approximation as presented here lead back to evaluation methods proposed by Freeden and Schreiner [21], Freeden and Blick [13], and particularly Freeden and Gerhards [15]. For the sake of simplicity, the adaptation of this approach to the requirements of gravitational potential as well as density distribution is explained only in scale discrete form, a scale continuous formulation as presented in [21] is canonical. A variety of numerical tests and case studies of our approach are found in the Ph.D.-theses [2, 49].

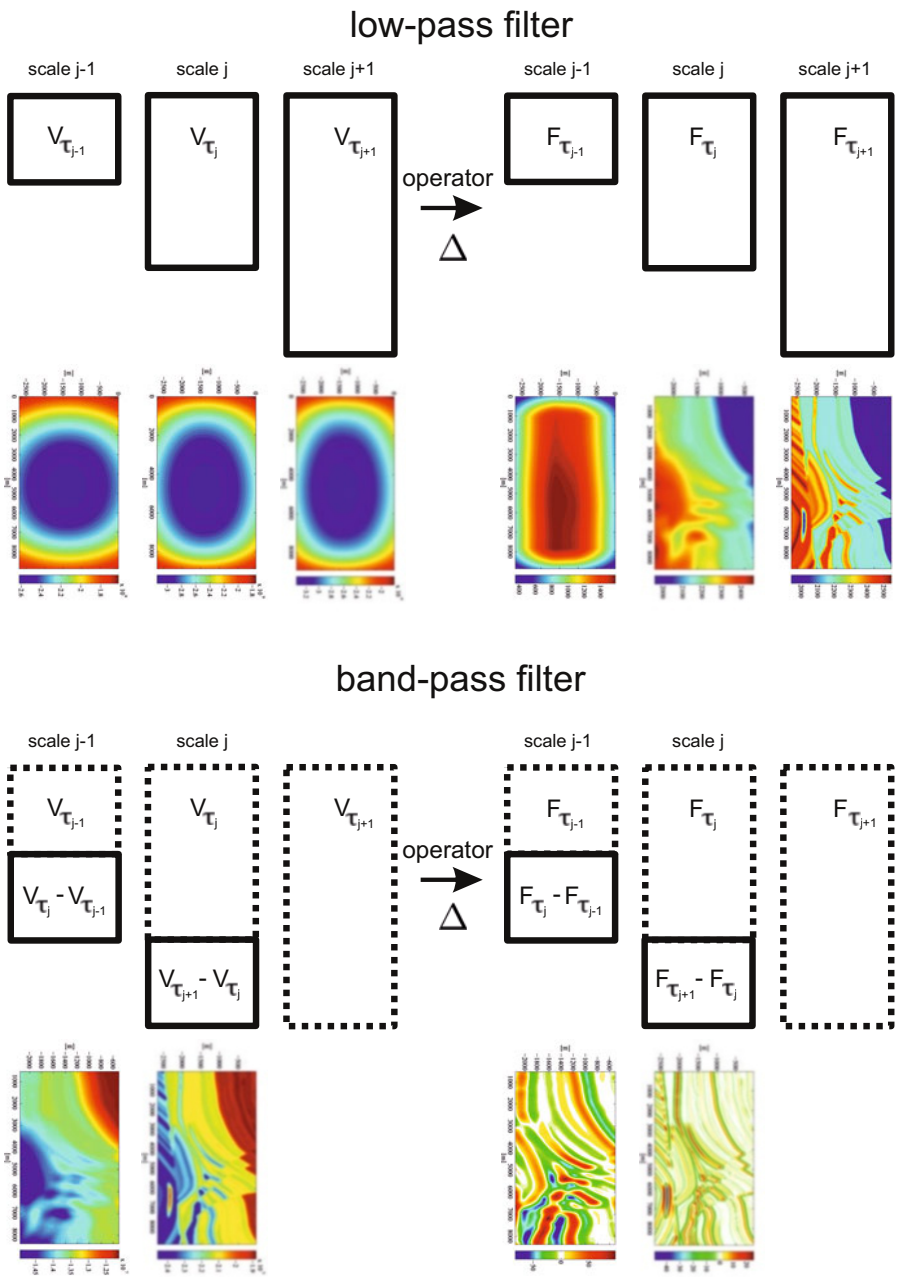


FIGURE 3.3. Schematic visualization of the multiscale decorrelation mechanism (see [3]).

3.3. Decorrelation of the Marmousi test model

Next we deal with the decorrelation of the geological signatures of a test area, namely the well-known Marmousi potential and density model (we use the canonically constructed 3D-version of the Marmousi model as proposed in the Ph.D.-thesis [2], see Figures 3.4, 3.5). In accordance with this standard test model (see also [42, 67]), the contrast function F is available as a fully interpreted 3D Marmousi density model extension (see Figure 3.6).

In order to validate the decorrelation abilities of our multiscale approach presented in the last section, we first perform a decomposition of the potential based on Eq. (3.38) (see Figure 3.7). Obviously, the low-pass filtered data, i.e., the τ_j-potential functions $V_n^{\tau_j}$ (see Eq. (3.19)) provide no essential structural information (see Figure 3.7, left column). However, for smaller scale values τ_j, by going over to finer detail information involving τ_j-wavelet potential functions, we already notice essential trends of the geological situation of the original density model (see Figure 3.7, right column).

Keeping the properties of the Newton volume integral in mind, we are not very surprised that, following the construction principles as proposed in our multiscale approach, the decomposition of the 3D Marmousi density model F based on Eq. (3.39) (Figure 3.8, left), in fact, shows a significant correlation to the decomposition of the τ_j-wavelet potential functions. Nevertheless, the τ_j-wavelet contrast functions (Figure 3.8, right) yield additional information, for example, the separation of all density transitions can be clearly detected at scale $j = 9$. This observation is of great significance in geothermal research, where the fracture transitions play a particular role for detecting areas of internal water flow.

Next we modify the original potential function V by adding three perturbations in form of mass points, i.e., fundamental solutions, at different locations obtaining the potential V_{mod}. Our purpose is to determine the locations of these three disturbances for the depth detection of geological formations. It should be noted that a decorrelation of the data with low-frequency wavelets (i.e., low values j) basically means focusing the multiscale approach on low-frequency signal components. Wavelets to higher values j allow to focus on the high-frequency interference.

As a consequence, our multiscale decorrelation mechanism shows that the low-pass filtered signals of V and V_{mod} are structurally identical (see Figures 3.7 and 3.9). However, at scale $j = 9$ (see Figure 3.10), we can identify the exact location of the centers of the introduced fundamental solutions (i.e., buried mass points) in the band-pass filtered data.

3.4. Gravimetry and Haar-type inversion

The inversion of Newton's Law of Gravitation (3.1), i.e., the determination of the internal "density function" from information of the gravitational potential is known as the *gravimetry problem*: To be more concrete, the gravimetry problem amounts to the problem of determining the "density function" F from (discrete) information of the gravitational potential V in $\mathbb{R}^3$ in accordance with the integral

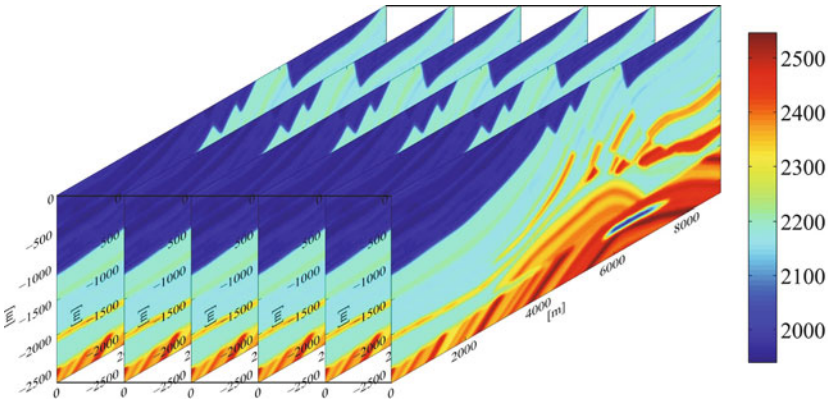


FIGURE 3.4. Artificially constructed 3D Marmousi density model.

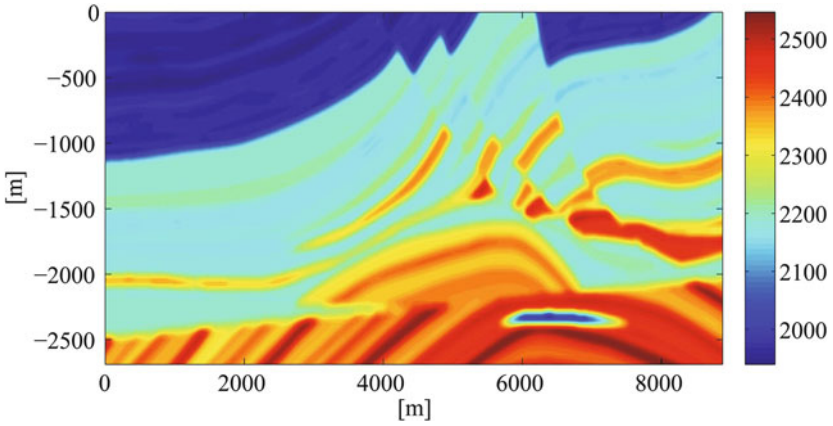


FIGURE 3.5. Cross-section of the 3D Marmousi density model (cf. [67]).

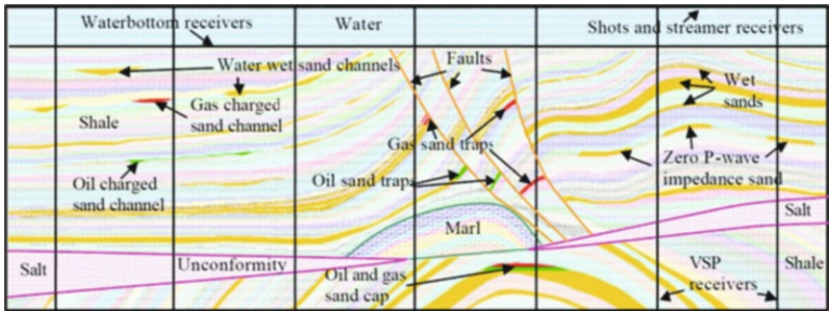


FIGURE 3.6. Marmousi density model and its geological interpretation (cf. [42]).

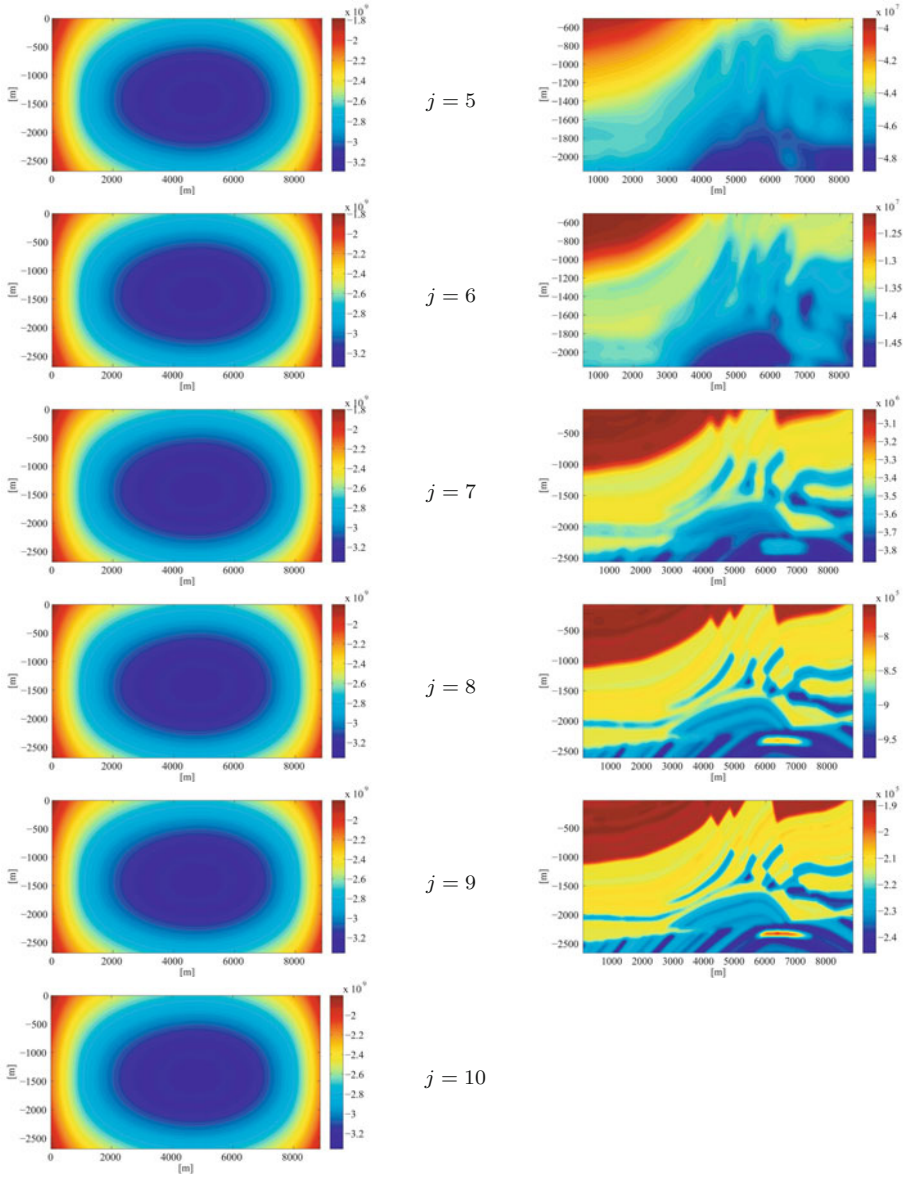


FIGURE 3.7. Decomposition of the 3D Marmousi potential in low-pass ($V_n^{\tau_j}$, left) and band-pass filtered parts ($(WV)_n^{\tau_j}$, right) for the sequence $\tau_j = 9200m \cdot 2^{-j}$ and $n = 0$ in $[kg/m]$. The choice of the sequence is adapted to the length of the density model (from the Ph.D.-thesis [2], Geomathematics Group, University of Kaiserslautern).

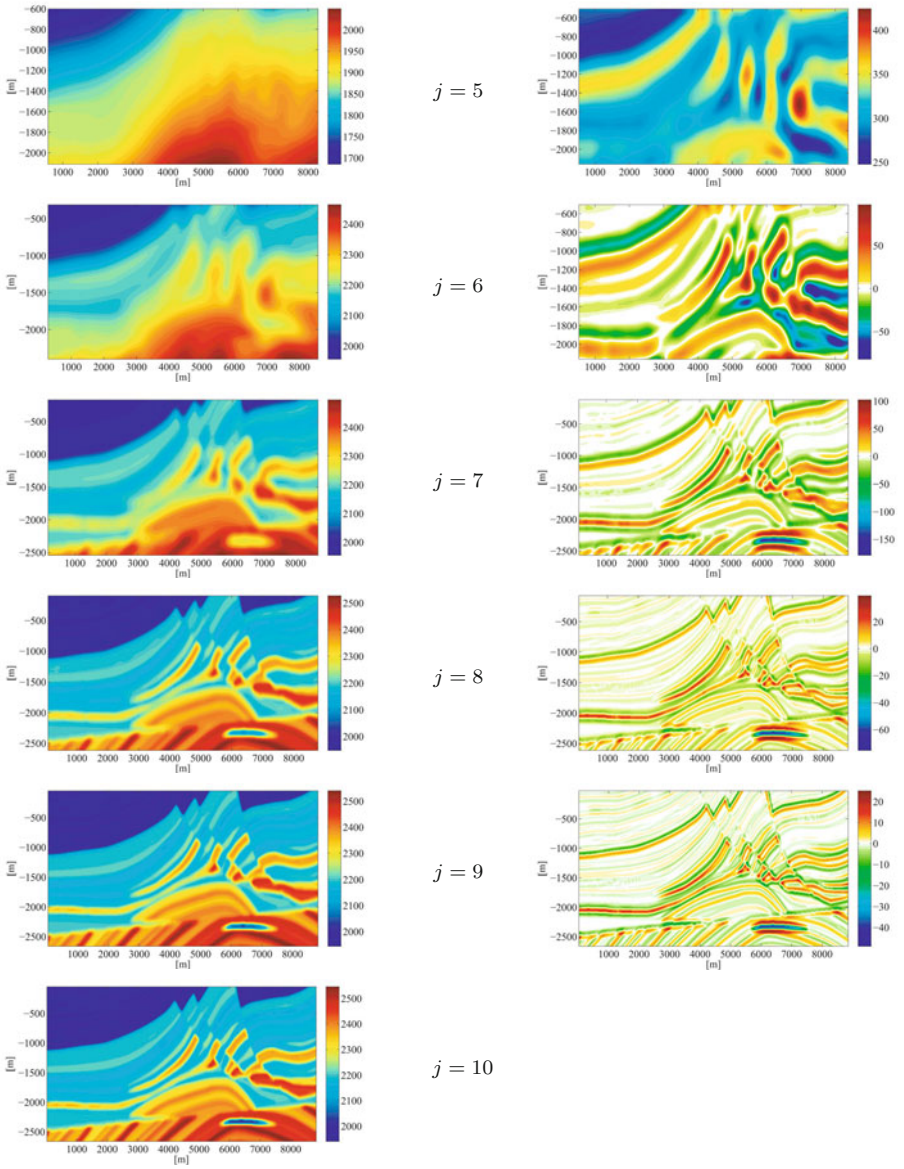


FIGURE 3.8. Decomposition of the 3D Marmousi density model into low-pass ($F_n^{\tau_j}$, left) and band-pass filtered parts ($(WF)_n^{\tau_j}$, right) for the sequence $\tau_j = 9200m \cdot 2^{-j}$ in $\left[kg/m^3\right]$ and $n = 0$. The choice of the sequence is adapted to the length of the density model (from the Ph.D.-thesis [2], Geomathematics Group, University of Kaiserslautern).

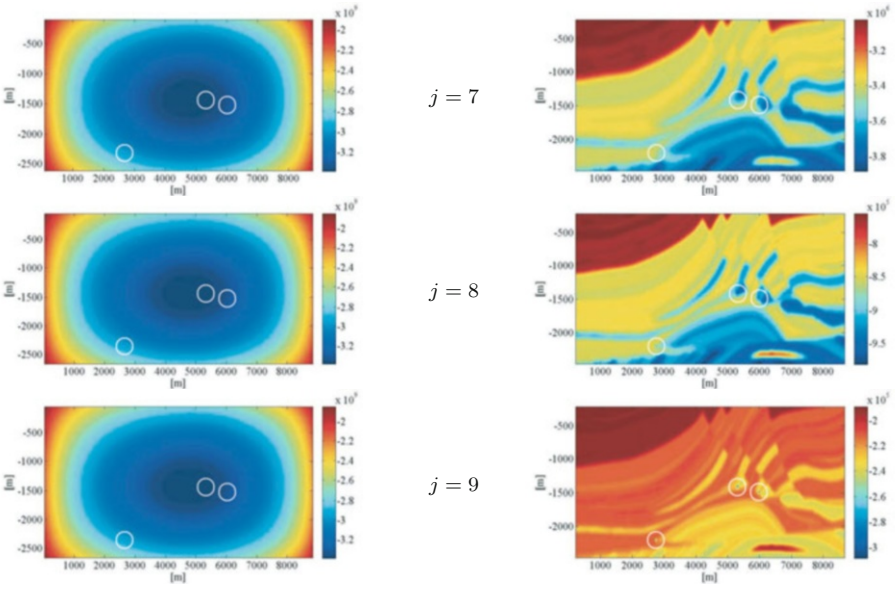


FIGURE 3.9. Decomposition of the modified 3D Marmousi potential V_{mod} in low-pass ($(V_{\text{mod}})_n^{\tau_j}$, left) and band-pass filtered parts ($(WV_{\text{mod}})_n^{\tau_j}$, right) for the sequence $\tau_j = 9200m \cdot 2^{-j}$ and $n = 0$ in $\left[\frac{kg}{m}\right]$. The choice of the sequence is adapted to the length of the density model.

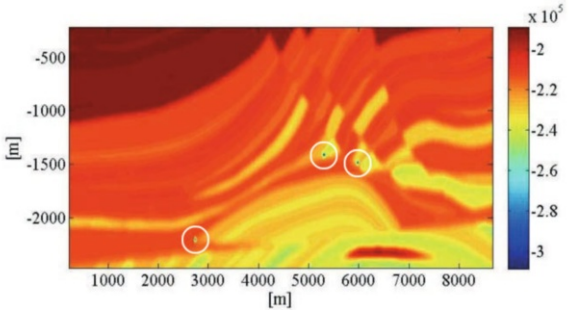


FIGURE 3.10. Illustration of the band-pass filtered signal $(WV_{\text{mod}})_n^{\tau_9}$ at scale $j = 9$ showing the locations of the three disturbing mass points, i.e., fundamental solutions and, consequently, the depth of the geological formations.

equation

$$V(x) = I[F](x) = \int_{\mathcal{G}} G(\Delta; |x-y|)F(y)\, dV(y), \quad x \in \mathbb{R}^3 \tag{3.41}$$

(note that we omit the gravitational constant γ).

In accordance with the mathematical *classification due to Hadamard*, the (classical) gravimetry problem of determining F from potential data on $\partial\mathcal{G}$, i.e., terrestrial gravitational data, violates all criteria, viz. existence, uniqueness and stability:

(i) (*Existence*) The potential V is harmonic outside $\mathcal{G}$. In accordance with the so-called Picard condition (see, e.g., [70]), a solution only exists if V belongs to (an appropriate subset in) the space of harmonic functions. However, it should be pointed out that this observation does not cause a numerical problem since, in practice, the information of V is only finite-dimensional. In particular, an approximation by an appropriate harmonic function is a natural ingredient of any practical method.

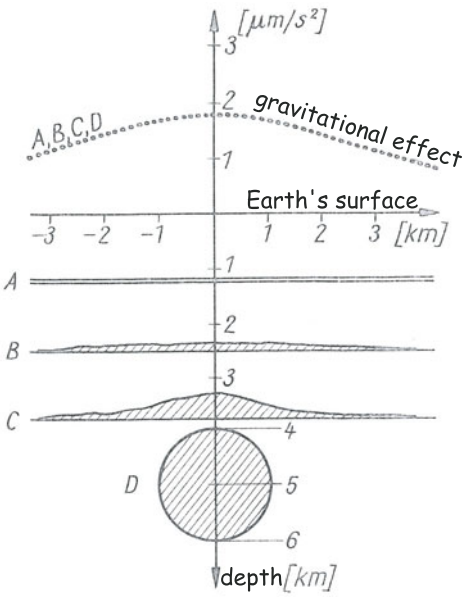


FIGURE 3.11. Equivalent gravity effect of different "sources" to generate the same gravitational potential on the Earth's surface (with kind permission of Teubner-publishing taken from [38] in modified form).

(ii) (*Uniqueness*) The most serious problem is the non-uniqueness of the solution (cf. Figure 3.11): The associated Fredholm integral operator I is of the first kind and has a kernel (null space) which is known (cf. [15, 20]) to coincide with the $L^2(\mathcal{G})$-orthogonal space of the closed linear subspace of all harmonic

functions on $\mathcal{G}$. Unfortunately, this orthogonal complement, i.e., the class of so-called anharmonic functions, is infinite-dimensional.

(iii) (*Stability*) Restricting the operator to harmonic densities leads to an injective mapping which has a discontinuous inverse implying an unstable solution.

Concerning the historical background, the question of the non-uniqueness for the classical gravimetry problem has been discussed extensively in literature, starting with a paper by Stokes [66] (for more details see, e.g., [15, 46, 47]). This calamity can be bypassed by imposing some reasonable additional condition on the density. A questionable condition, suggested by the mathematical structure of the Newton potential operator I, is to require that the density is harmonic. The approximate calculation of the harmonic density has already been implemented in several papers, whereas the problem of determining the anharmonic part seems to be still a great challenge. Due to the lack of an appropriate physical interpretation of the harmonic part of the density, various alternative variants have been discussed in the literature. In general, gravitational data yield significant information only about the uppermost part of the Earth's interior, which is not laterally homogeneous.

Seen from a mathematical point of view in constructive approximation, locally supported functions are not new, having been discussed already by Haar (1910). The importance of spacelimited (or in mathematical jargon locally supported) Haar kernels in view of a multiscale procedure is the "birth" to an entire "basis family" (scaling functions) by means of two operations, viz. dilations and translations.

In what follows, we recapitulate the already discussed "Haar philosophy" to realize an approximate determination of the mass density distribution inside $\mathcal{G}$ from discrete gravitational information. The essential tool is the regularization procedure of the Newton potential enabling to replace the integral equation (3.41) by the Fredholm integral equation of first kind

$$V_n^\tau(x) = \int_{\mathcal{G}} G_n^\tau(\Delta; |x-y|)F(y)\, dV(y), \quad x \in \mathbb{R}^3, \tag{3.42}$$

for a sufficiently large scale number J, so that the serious problem of non-uniqueness caused by the occurrence of anharmonic functions is not existent anymore for terrestrial potential data, however, at the price of non-harmonicity of the "regularizer" $y \mapsto G_n^{\tau_J}(\Delta; |x-y|)$ in a neighborhood outside $x \in \partial\mathcal{G}$.

In choosing a sufficiently large J we are aware of the fact (cf. Theorem 3.4) that

$$V(x) \simeq V_n^{\tau_J}(x) = \int_{\mathcal{G}} G_n^{\tau_J}(\Delta; |x-y|)F(y)\, dV(y), \quad x \in \mathbb{R}^3, \tag{3.43}$$

i.e., $V_n^{\tau_J}$ provides an approximation of the Newton integral (3.3) with negligible error. We remember

$$\Delta_x\, G_n^{\tau_J}(\Delta; |x-z|) = \; H_n^{\tau_J}(|x-z|) \tag{3.44}$$

for all $x, z \in \mathbb{R}^3$. From (3.44) it therefore follows that

$$\Delta_x \int_{\mathcal{G}} G_n^{\tau_J}(\Delta; |x-z|) F(z)\, dV(z) = F_n^{\tau_J}(x) \simeq \quad F(x), \quad x \in \mathcal{G}. \tag{3.45}$$

In order to realize a fully discrete approximation of F, we have to apply approximate integration formulas over $\mathcal{B}_{\tau_J}(x) \cap \overline{\mathcal{G}}$ leading to

$$V(x) \simeq V_n^{\tau_J}(x) \simeq \sum_{i=1}^{N_J} G_n^{\tau_J}(\Delta; |x - y_i^{N_J}|)\, w_i^{N_J} F(y_i^{N_J}), \tag{3.46}$$

where $w_i^{N_J}$, $y_i^{N_J} \in \mathcal{B}_{\tau_J}(x) \cap \overline{\mathcal{G}}$, $i = 1, \ldots, N_J$, are known weights and knots, respectively.

For the determination of the mass density we are confronted with the situation that all coefficients $a_i^{N_J} = w_i^{N_J} F(y_i^{N_J})$, $i = 1, \ldots, N_J$, are unknown. This, however, means that we have to solve a linear system, namely

$$V(x_k^{M_J}) = \sum_{i=1}^{N_J} G_n^{\tau_J}(\Delta; |x_k^{M_J} - y_i^{N_J}|) a_i^{N_J}, \quad k = 1, \ldots, M_J, \tag{3.47}$$

in order to determine the coefficients $a_i^{N_J}, i = 1, \ldots, N_J$, from *known gravitational values* $V(x_k^{M_J})$ at knots $x_k^{M_J} \in \mathbb{R}^3$, $k = 1, \ldots, M_J$.

Once all density values $F(y_i^{N_J})$, $i = 1, \ldots, N_J$, are available (note that the integration weights $w_i^{N_J}, i = 1, \ldots, N_J$, are known from the approximate integration rule), the density distribution F can be obtained from the formula

$$F(x) \simeq F_n^{\tau_J}(x) = \sum_{i=1}^{N_J} H_n^{\tau_J}(|x - y_i^{N_J}|)\, w_i^{N_J} F(y_i^{N_J}), \quad x \in \mathcal{G}. \tag{3.48}$$

Even more, fully discrete Haar filtered versions of F at lower scales, i.e., feature extraction, can be derived in accordance with the approximate integration rules

$$F_n^{\tau_j}(x) = \int_{\mathcal{G}} H_n^{\tau_j}(|x-y|) F(y)\, dV(y) \simeq \sum_{i=1}^{N_j} H_n^{\tau_j}(|x - y_i^{N_j}|) w_i^{N_j} F(y_i^{N_j}) \tag{3.49}$$

for $j = J_0, \ldots, J$, where $w_i^{N_j}$, $y_i^{N_j}$, $i = 1, \ldots, N_j$, are known weights and knots, respectively, such that we can take adventage of the fact that $\{y_1^{N_j}, \ldots, y_{N_j}^{N_j}\} \subset \{y_1^{N_J}, \ldots, y_{N_J}^{N_J}\} \subset \overline{\mathcal{G}}$, i.e., the sequence of knots $\{y_1^{N_j}, \ldots, y_{N_j}^{N_j}\} \subset \overline{\mathcal{G}}$ shows a hierarchical positioning.

Altogether, our approach yields Haar filtered versions (3.49) establishing a fully discrete (space-based) multiscale decomposition $F_n^{\tau_J}, \ldots, F_n^{\tau_{J_0}}$ of the density distribution F, such that an entire set of approximations is available from a single locally supported "mother function", i.e., the Haar kernel function, and this set provides useful "building block functions", which enable suitable storage and fast decorrelation of density data in consistency with geological formations.

It should be remarked that by discretizing the convolution integral by approximate integration in form of a suitable cubature formula, we arrive at a system of linear equations, which is sparse, since the utilized wavelet has a local support. The local support enables us to limit the multiscale technique to a local region, e.g., only to a relevant borehole area of interest, and guarantees that there is no change in the signature outside the support of the wavelets. All in all, our approach is given in such a way that the inversion of the equation system turns out to be numerically efficient and economical.

Remark 3.7. The linear systems occurring in this section can be handled by, e.g., use of domain decomposition techniques (see, e.g., [17, 19, 28–30, 34] and the references therein).

Remark 3.8. For simplicity, the input data of this subsection are potential data. In the same way, a linear system can be established by taking, e.g., free air gravity anomalies on the Earth's surface. In this case, however, we need Haar functions (3.18) of positive degrees n, since free air gravity anomalies are generated by a first-order derivative applied to the disturbing potential.

3.5. Improvement of in-borehole density signatures

Since both the actual potential V and the actual contrast function F, in general, cannot be measured directly inside the boundary surface $\partial\mathcal{G}$ and outside a neighborhood around the already existing boreholes without additional drilling, the a priori available potential and density information differ from the actual values and thus form only an approximation to the reality. If one associates a certain scale value τ_j within the multiscale process to the available potential data, we are given

$$V_n^{\tau_j}(x) = \int_{\mathcal{G}} G_n^{\tau_j}(\Delta; |x-y|)F(y)\ dV(y). \tag{3.50}$$

Often, in practice during borehole drilling, additional data are gathered by in-hole gravimetric measurements, so that we may assign a scale value τ_{j+1} to the improved potential data. If we now take the difference, we arrive at

$$V_n^{\tau_{j+1}}(x) - V_n^{\tau_j}(x) = \sum_{i=1}^{N_j} \Psi_{G_n^{\tau_j}}\left(\Delta; \left|x - y_i^{N_j}\right|\right) w_i^{N_j} F\left(y_i^{N_j}\right), \tag{3.51}$$

and

$$F_n^{\tau_{j+1}}(x) - F_n^{\tau_j}(x) = \sum_{i=1}^{N_j} \Psi_{H_n^{\tau_j}}\left(\left|x - y_i^{N_j}\right|\right) w_i^{N_j} F\left(y_i^{N_j}\right), \tag{3.52}$$

respectively. Once again, it should be emphasized that the linear system (3.51) can be evaluated efficiently and economically (note that the kernels $H_n^{\tau_j}$ as well as $\Psi_{G_n^{\tau_j}}$ and $\Psi_{H_n^{\tau_j}}$ have local support due to their construction and, hence, the systems of Equations (3.51) and (3.52) are sparse).

The improvement by additional data observation is shown using the example of the equation system (3.52). The input data for the inversion process are generated by smoothing of a cutout of the 3D Marmousi density model (see Figure 3.12, left for $F_n^{\tau_{j+1}}$). As a result of the inversion (see Figure 3.12, right), we obtain a sharper density model provided that the wavelet used in the inversion is sufficiently smooth, i.e., $n \geq 2$.

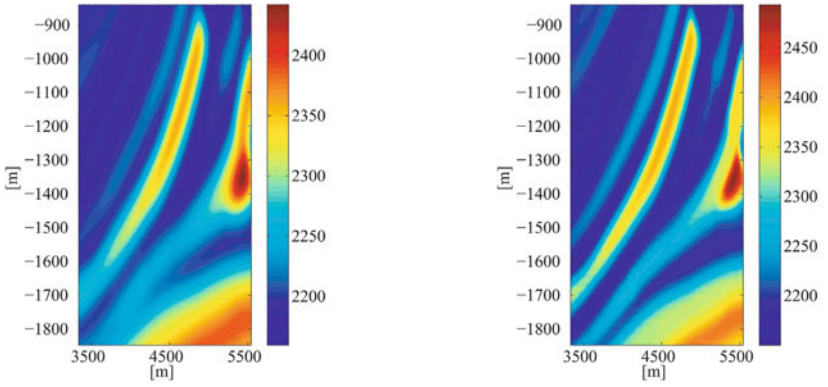


FIGURE 3.12. Illustration of the "best" data before the inversion $F_n^{\tau_{j+1}}$ (left) and the inversion result (right) for $n = 3$ and $j = 5$. The colors show the densities in [kg/m^3] (from the Ph.D.-thesis [2]).

An extensive parameter study in the Ph.D.-thesis [2] demonstrates that the inversion is numerically stable and efficient for smooth Haar-type kernels. Since the resulting relative error in the inversion depends continuously on the scale, there exists a reference interval, such that for each scale value inside this reference interval, an improvement of the data is achieved. This allows a certain tolerance in the choice of parameters.

Conclusions

Local knowledge of the gravity potential and its equipotential (level) surfaces giving information about mass distribution have become an important issue for exploration and prospecting. Indeed, the gravity field is a key component of future investigation. Seen from a numerical point of view, however, the way forward has to focus on two challenges:

(i) In reality, the distribution of geopotential data is far from being homogeneous with large gaps even in all European areas. In addition, the quality of the geopotential data under consideration is very distinct. A terrestrial data coverage now and in the foreseeable future is far from being satisfactory. For data supplementation and numerical stabilization, airborne and/or

spaceborne data are indispensable. This unfortunate situation causes particular mathematical attention for homogenization and unification to suppress undesired oscillation phenomena within the modeling process of the data.

(ii) Nowadays, the knowledge of geopotentials such as the Earth's gravitational potential and their anomalies have become an important and cost-effective issue in exploration technologies. However, it is commonly known that highly accurate sensors, when operating in an isolated manner, have their shortcomings. Combining globally available satellite data with regional airborne and/or local terrestrial observations within a physically founded and mathematically consistent multiscale process is therefore an essential step forward. In this respect, a "zooming-in" detection of specific geophysical attributes is an outstanding field of interest for validating the multiresolution method based on heterogeneous datasets and geophysically oriented multiscale "downward continuation" modeling of the different data sources starting from spaceborne data as trend solution via more accurate airborne data down to high-precision local data sets.

Geophysically relevant signatures are usually decomposed into single frequencies. Geomathematically, these techniques are well suited to resolve low and medium frequency phenomena, while their application to obtain high resolution models (such as descriptions of local orebodies, salt deposits, aquifers, etc.) is critical. Due to the quality of the data, i.e., the intrinsic scale amount of significant wave packages within the signal, spaceborne (i.e., satellite) data – continued downward to the Earth's surface – are the canonical point of departure for multiscale approximations of lower scale frequency phenomena, while the quality of airborne and/or terrestrial data can be associated to medium and/or high(er) scale frequency bands. So, the whole spectrum of spaceborne/airborne/ground data systems covers all verifiable wave packages. Actually, the advantage of satellite lower frequency band data at the ground is their availability everywhere, while (airborne) medium and (terrestrial) high(er) frequency bands usually are at the disposal for regional and local occurrence, respectively. In this respect, a helpful tool for determining the depth and size is the introduction of known artificial disturbances such as monopoles in gravitation which superpose the original wave bands of the data in an easily predictable and calculable way.

Summarizing our results, we are led to the following conclusion: The multiscale approach which is presented in this contribution breaks up a complicated signal (like the gravitational field, the geomagnetic field) into "wave band signatures" at different scales, i.e., a certain resolution. To each scale parameter, a scaling function is defined leading to an approximation of the data at that certain resolution. The difference between two successive scaling functions, i.e., the wavelets, represents the corresponding wave bands and, thus, yields the desired geological detail information. With increasing scale, the approximation is getting finer and finer starting form a low pass approximation and adding more and more wave bands. The multiscale approach guarantees that the information contained on a certain

(coarse) level is also contained in the approximations of higher scales. It is advantageous that we are able to analyze the wave bands separately (decorrelation). Thus, this multiscale concept helps to find adaptive methods to the particular structure of the data. Additionally, the resolution of the model can be adapted to the spatial structures, i.e., for areas with coarse spatial structures, the resolution of the model can be chosen to be rather low and for areas with complicated structures the resolution can be increased accordingly. Consequently, since most data show correlation both in space as in frequency, the multiscale technique is an appropriate method for a simultaneous space and frequency localization. As far as the numerical realization is concerned, fast wavelet methods (FWT) are applicable.

Considering especially the disturbing potential field approximation in gravitation, we observe – from computational point of view – two main requirements: First, the field characteristics of geological features are usually of local character such that the use of local wavelets is evident. Second, in view of physical relevance of the multiscale approach, we need wavelets which have a certain relation to the corresponding partial differential equation (here: Laplace equation). Moreover, we have to be concerned with wavelet types which are manageable from mathematical point of view and, additionally, show a close relation to the physical model. Indeed, the developed multiscale method by means of regularizing wavelets using physically motivated fundamental solutions has its origins in works of the Geomathematics Group of the University of Kaiserslautern (see, e.g., [12, 15, 18, 20–22] and the list of references therein). The main results and characteristics of our studies presented here can be summarized as follows:

- Physically based behavior and appropriate interpretability of the developed wavelets.
- Numerical efficiency and economy of the wavelets by adaptive choice of the local support and resulting fast algorithms.
- Scale dependent correlation of wavebands and geological structures in a systematic "zooming-in / zooming out" decorrelation process.
- Specific transparency of certain geological structures for an appropriate choice of parameters.
- Depth determination and localization of geological formations by artificial point source disturbances.

Furthermore seen from the point of mathematical methodology, our multiscale (postprocessing) approach is not only restricted to potential methods involving the Laplace operator. Similar approaches can be formulated, e.g., for the Helmholtz and d'Alembert operators (cf. [2]).

Regarding the signature decorrelation and Haar-Type inversion, we deal with a construction of physically relevant wavelets based on the regularization of the fundamental solution for the decomposition of gravimetric data, and analyze different examples occurring in exploration. The decomposition of the 3D Marmousi density model shows a breakdown of the signals into their constituent components. Our numerical tests have further shown that the inversion technique described for

the local improvement of records is numerically stable. In particular, the method can be restricted to the specified local region of interest without changing the remaining area due to the local support of the wavelets. The resulting linear equation systems are sparse, hence, they can thus be solved efficiently and economically.

The particular advantage of the decorrelation method proposed here is the simultaneous calculation of the potential and the contrast function (density function) without any requirement of additional mathematical and numerical effort and this while closely ensuring physical relevance and numerically acceptable effort. It is therefore expected that the method presented here, in fact, will contribute substantially to minimizing the exploration risk, for example, in geothermal obligations by providing deeper and more secure geological information.

Acknowledgment

The authors thank the "Federal Ministry for Economic Affairs and Energy, Berlin" and the "Project Management Jülich" for funding the projects "GEOFÜND" (funding reference number: 0325512A, PI Prof. Dr. W. Freeden, University of Kaiserslautern, Germany) and "SPE" (funding reference number: 0324061, PI Prof. Dr. W. Freeden, CBM – Gesellschaft für Consulting, Business und Management mbH, Bexbach, Germany, corporate manager Prof. Dr. mult. M. Bauer).

References

[1] Bauer, M., Freeden, W., Jacobi, H., and Neu, T., Eds. (2014). *Handbuch Tiefe Geothermie*. Springer, Berlin, Heidelberg.

[2] Blick, C. (2015). *Multiscale Potential Methods in Geothermal Research: Decorrelation Reflected Post-Processing and Locally Based Inversion*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern.

[3] Blick, C., Freeden, W., and Nutz, H. (2016). Feature Extraction of Geological Signatures by Multiscale Gravimetry. *GEM Int. J. Geomath.*, 8:57–83, 2017.

[4] Bruns, E.H. (1878). *Die Figur der Erde*. Königl. Preussisch. Geodätisches Institut, P. Stankiewicz Buchdruckerei, Berlin.

[5] Ernstson, K. (2018). Geologische und geophysikalische Untersuchungen. In Bauer, M., Freeden, W., Jacobi, H., and Neu, T., Eds., *Handbuch Oberflächennahe Geothermie*. Springer, Heidelberg.

[6] ESA (1996). The Nine Candidate Earth Explorer Missions. *Publications Division ESTEC, Noordwijk*, SP-1196(1).

[7] ESA (1998). European Views on Dedicated Gravity Field Missions: GRACE and GOCE. *An Earth Sciences Division Consultation Document*. ESD-MAG-REP-CON-001.

[8] ESA (1999). Gravity Field and Steady-State Ocean Circulation Mission. *Publications Division ESTEC, Noordwijk*, SP-1233(1). Report for mission selection of the four candidate earth explorer missions.

[9] Fehlinger, T. (2009). *Multiscale Formulations for the Disturbing Potential and the Deflections of the Vertical in Locally Reflected Physical Geodesy*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern.

[10] Foulger, G., Natland, J., Presnall, D., and Anderson, D., Eds. (2005). *Plates, Plumes, and Paradigms*. Geological Society of America.

[11] Freeden, W. (1980). On Integral Formulas of the (Unit) Sphere and Their Application to Numerical Computation of Integrals. *Computing*, 25:131–146.

[12] Freeden, W. (1999). *Multiscale Modelling of Spaceborne Geodata*. B.G. Teubner, Leipzig.

[13] Freeden, W. and Blick, C. (2013). Signal Decorrelation by Means of Multiscale Methods. *World of Mining*, 65(5):304–317.

[14] Freeden, W., Fehlinger, T., Mayer, C., and Schreiner, M. (2008). On the Local Multiscale Determination of the Earth's Disturbing Potential from Discrete Deflections of the Vertical. *Comput. Geosc.*, 4:473–490.

[15] Freeden, W. and Gerhards, C. (2013). *Geomathematically Oriented Potential Theory*. CRC Press, Taylor & Francis, Boca Raton.

[16] Freeden, W., Gervens, T., and Schreiner, M. (1998). *Constructive Approximation on the Sphere* (*with Applications to Geomathematics*). Oxford Science Publications, Clarendon, Oxford.

[17] Freeden, W., Glockner, O., and Schreiner, M. (1998). Spherical Panel Clustering and Its Numerical Aspects. *J. Geodesy*, 72:586–599.

[18] Freeden, W. and Gutting, M. (2013). *Special Functions of Mathematical* (*Geo-*) *Physics. Applied and Numerical Harmonic Analysis* (*ANHA*). Birkhäuser, Basel.

[19] Freeden, W. and Hesse, K. (2002). On the Multiscale Solution of Satellite Problems by use of Locally Supported Kernel Functions Corresponding to Equidistributed Data on Spherical Orbits. *Stud. Sci. Math. Hung.*, 39:37–74.

[20] Freeden, W. and Michel, V. (2004). *Multiscale Potential Theory* (*with Applications to Geoscience*). Birkhäuser, Boston.

[21] Freeden, W. and Schreiner, M. (2006). Local Multiscale Modelling of Geoid Undulations From Deflections of the Vertical. *J. Geodesy*, 79:641–651.

[22] Freeden, W. and Schreiner, M. (2009). *Spherical Functions of Mathematical Geosciences* (*A Scalar, Vectorial, and Tensorial Setup*). Springer, Heidelberg.

[23] Freeden, W. and Wolf, K. (2008). Klassische Erdschwerefeldbestimmung aus der Sicht moderner Geomathematik. *Math. Semesterb.*, 56:53–77.

[24] Grafarend, E.W., Aardalan, A.A., and Finn, G. (2006). Ellipsoidal Vertical Deflections and Ellipsoidal Gravity Disturbances: Case Studies. *Stud. Geophys. Geod.*, 50:1–57.

[25] Grafarend, E.W., Klapp, M., and Martinec, Z. (2015). Spacetime Modeling of the Earth's Gravity Field by Ellipsoidal Harmonics. In Freeden, W., Nashed, M.Z., and Sonar, T., Eds., *Handbook of Geomathematics*, volume 1, 381–496. Springer, New York, 2nd edition.

[26] Grant, E.S. and West, G.F. (1965). *Interpolation Theory in Applied Geophysics*. McGraw-Hill Book Company.

[27] Groten, E. (1979). *Geodesy and the Earth's Gravity Field I + II*. Dümmler, Bonn.

[28] Gutting, M. (2008). *Fast Multipole Methods for Oblique Derivative Problems*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern.

[29] Gutting, M. (2012). Fast Multipole Accelerated Solution of the Oblique Derivative Boundary-Value Problem. *GEM Int. J. Geomath.*, 2:223–252.

[30] Gutting, M. (2015). Fast Spherical/Harmonic Spline Modeling. In Freeden, W., Nashed, M.Z., and Sonar, T., Eds., *Handbook of Geomathematics*, volume 3,2711–2746. Springer, New York, 2nd. edition.

[31] Harpp, K.S., Fornari, D.J., Geist, D.J., and Kurz, M.D. (2003). A Manifestation of Plume-Ridge Interaction in the Northern Galapagos Islands. *Geochemistry, Geophysics, Geosystems*, 4(9):DOI: 10.1029/2003GC000531.

[32] Heiskanen, W.A. and Moritz, H. (1967). *Physical Geodesy*. Freeman, San Francisco.

[33] Helmert, F. (1884). *Die Mathematischen und Physikalischen Theorien der Höheren Geodäsie, I, II*. B.G. Teubner, Leipzig.

[34] Hesse, K. (2003). *Domain Decomposition Methods in Multiscale Geopotential Determination from SST and SGG*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern.

[35] Hesse, K., Sloan, I.H., and Womersley, R.S. (2015). Numerical Integration on the Sphere. In Freeden, W., Nashed, M.Z., and Sonar, T., Eds., *Handbook of Geomathematics*, volume 2, 2671–2710. Springer, New York, 2nd edition.

[36] Hofmann-Wellenhof, B. and Moritz, H. (2005). *Physical Geodesy*. Springer, Wien, New York.

[37] IEA (2012). Iceland Country Report 2012. Report, IEA Geothermal.

[38] Jacobs, F. and Meyer, H. (1992). *Geophysik-Signale aus der Erde*. B.G. Teubner, Leipzig, and VDF Verlag, Zürich.

[39] Keating, P. (1992). Interpretation of the Gravity Anomaly Field in the Noranda Val d'Or region, Abitibi Greenstone Belt, Canadian Shield. *Can. J. Earth Sciences*, 29(5):962–971.

[40] Kellogg, O.D. (1967). *Foundations of Potential Theory*. Springer Berlin, Heidelberg, New York.

[41] Kusche, J. (2015). Time-Variable Gravity Field and Global Deformation of the Earth. In Freeden, W., Nashed, M.Z., and Sonar, T., Eds., *Handbook of Geomathematics*, volume 1, 321–338. Springer, New York, 2nd edition.

[42] Martin, G.S., Wiley, R., and Marfurt, K.J. (2006). Marmousi2: An Elastic Upgrade for Marmousi. *The Leading Edge*, 25(2):156–166.

[43] Mathar, D. (2008). *Wavelet Variances and Their Application in Geoscientifically Relevant Feature Detection*. Diploma thesis, University of Kaiserslautern, Geomathematics Group.

[44] Meinesz, F.A.V. (1928). A Formula Expressing the Deflection of the Plumb Line in the Gravity Anomalies and Some Formulas for the Gravity Field and the Gravity Potential Outside the Geoid. *Proc. Koninklijke Akad. Wet. Amsterdam*, 31:315–322.

[45] Meissl, P.A. (1971). A Study of Covariance Functions Related to the Earth's Disturbing Potential. Scientific Report No. 151, The Ohio State University, Department of Geodetic Science, Columbus.

[46] Michel, V. (2002). *A Multiscale Approximation for Operator Equations in Separable Hilbert Spaces – Case Study: Reconstruction and Description of the Earth's Interior*. Habilitation thesis, Geomathematics Group, University of Kaiserslautern. Shaker, Aachen.

[47] Michel, V. and Fokas, A.S. (2008). A Unified Approach to Various Techniques for the Non-Uniqueness of the Inverse Gravimetric Problem and Wavelet-Based Methods. *Inverse Problems*, 24 045019 (25 pp.), DOI: 10.1088/0266-5611/24/4/045019.

[48] Militzer, H. and Weber, F. (1984). *Angewandte Geophysik, Band* 1*, Gravimetrie und Magnetik*. Springer, Wien, New York.

[49] Möhringer, S. (2014). *Decorrelation of Gravimetric Data*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern.

[50] Morgan, W.J. (1971). Convection Plumes in the Lower Mantle. *Nature*, 230:42–43.

[51] Moritz, H. (2015). Classical Physical Geodesy. In Freeden, W., Nashed, M.Z., and Sonar, T., Eds., *Handbook of Geomathematics*, volume 1, 253–289. Springer, New York, 2nd. edition.

[52] Nations, U. (2010). Current and Planned Global and Regional Navigation Satellite Systems and Satellite-Based Augmentations Systems. In *Proceedings of ICG*.

[53] Nettleton, L.L. (1971). *Elementary Gravity and Magnetics for Geologists and Seismologists*. Society Of Exploration Geophysicists, Tulsa, OK.

[54] Nettleton, L.L. (1976). *Gravity and Magnetics in Oil Prospecting*. McGraw-Hill, New York.

[55] Neumann, F. (1887). *Vorlesungen über die Theorie des Potentials und der Kugelfunktionen*. B.G. Teubner, Leipzig.

[56] Pizzetti, P. (1894). Geodesia – sulla espressione della gravità alla superficie del geoide, supposto ellissoidico. *Att. R. Acad. Lincei*, 3:331–350.

[57] Pizzetti, P. (1909). Corpi equivalenti rispetto alla attrazione newtoniana esterna. *Rom. Acc. L. Rend.*, 18:211–215.

[58] Ritter, J. and Christensen, U. (2007). *Mantle Plumes, A Multidisciplinary Approach*. Springer, Berlin, Heidelberg.

[59] Rivas, J. (2009). Gravity and Magnetic Methods. In *Short Course on Surface Exploration for Geothermal Resources*.

[60] Rummel, R. (1997). Spherical Spectral Properties of the Earth's Gravitational Potential and Its First and Second Derivatives. In Sansò, S. and Rummel, R., Eds., *Geodetic Boundary Value Problems in View of the One Centimeter Geoid*, volume 65 of *Lecture Notes in Earth Science*, 359–404. Springer, Berlin, Heidelberg.

[61] Rummel, R. (2015). GOCE: Gravitational Gradiometry in a Satellite. In Freeden, W., Nashed, M.Z., Eds., *Handbook of Geomathematics*, volume 1, 211–226. Springer, New York, 2nd edition.

[62] Rummel, R., Balmino, G., Johannessen, J., Visser, P., and Woodworth, P. (2002). Dedicated Gravity Field Missions – Principles and Aims. *J. Geodyn.*, 33:3–20.

[63] Schuber, D., Turcotte, D., and Olson, P. (2001). *Mantle Convection in the Earth and Planets*. Cambridge University Press, Cambridge.

[64] Seigel, H.O. (1995). *A Guide to High Precision Land Gravimeter Surveys*. Scintrex Limited, Concord, Ontario.

[65] Stokes, G.G. (1849). On the Variation of Gravity at the Surface of the Earth. *Trans. Cambr. Phil. Soc.*, 148:672–712.

[66] Stokes, G.G. (1867). On the internal distribution of matter which shall produce a given potential at the surface of a gravitating mass. *Proc.Royal Soc. London*, 15:482–486.

[67] Symes, W.W. (2014). T.R.I.P. the Rice inversion project, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA. URL: http://www.trip.caam.rice.edu/downloads/downloads.html, accessed 12. September 2016.

[68] Torge, W. (1989). *Gravimetry*. de Gruyter, Berlin.

[69] Torge, W. (1991). *Geodesy*. de Gruyter, Berlin.

[70] Tykhonov, A.N. (1963). Solution of Incorrectly Formulated Problems and the Regularization Method. *Sov. Math.*, 5:1035–1038. (translation from *Dokl. Akad. Nauk*, (15):501–504).

[71] Wikipedia (2016). Galápagos hotspot. URL: https://en.wikipedia.org/wiki/Galapagos_hotspot, accessed 12. May 2016.

[72] Wilson, J. (1963). A Possible Origin of the Hawaiian Island. *Can. J. Phys.*, 41:863–868.

[73] Wolf, K. (2009). *Multiscale Modeling of Classical Boundary Value Problems in Physical Geodesy by Locally Supported Wavelets*. Ph.D.-thesis, Geomathematics Group, University of Kaiserslautern.

C. Blick, W. Freeden, and H. Nutz
Geomathematics Group
University of Kaiserslautern
MPI-Gebäude, Paul-Ehrlich-Str. 26
D-67663 Kaiserslautern, Germany

Handbook of Mathematical Geodesy
Geosystems Mathematics, 753–819

Spherical Harmonics Based Special Function Systems and Constructive Approximation Methods

Willi Freeden, Volker Michel, and Frederik J. Simons

Abstract. Special function systems are reviewed that reflect particular properties of the Legendre polynomials, such as spherical harmonics, zonal kernels, and Slepian functions. The uncertainty principle is the key to their classification with respect to their localization in space and frequency/momentum. Methods of constructive approximation are outlined such as spherical harmonic and Slepian expansions, spherical spline and wavelet concepts. Regularized Functional Matching Pursuit is described as an approximation technique of combining heterogeneous systems of trial functions to a kind of a 'best basis'.

Keywords. Spherical harmonics procedures, Slepian, spline and wavelet methods, regularized functional matching pursuit.

1. Introduction

Up until the present time, modeling geoscientific data is often performed on a global scale by orthogonal expansions in terms of spherical harmonics. However, in many aspects global spherical harmonic modeling cannot keep pace with the prospects and the expectations of the 'Earth system sciences'. In particular, there is an increasing need for high-precision modeling on local areas. As we shall discuss, Slepian functions are important tools for this purpose. For their part, zonal kernel functions – in the jargon of constructive approximation: radial basis functions – have become more and more important because of their space localizing properties (even in the vectorial and tensorial context).

The addition theorem for spherical harmonics enables us to express all types of zonal kernel functions in terms of a one-dimensional function, the *Legendre polynomial*. Weighted additive clustering of Legendre polynomials generates specific classes of space localizing zonal kernel functions, i.e., Legendre series expansions,

ready for approximation within the scalar, vectorial, and tensorial framework. The closer the Legendre series expansion is to the Dirac kernel, the more localized is the zonal kernel in space, and the more economical is its role in (spatial) local computation. In addition, the Funk–Hecke formula provides the natural tool for establishing convolutions of spherical fields against zonal kernels. Consequently, by specifying Dirac families, i.e., sequences of zonal functions tending to the Dirac kernel, (space-localized) filtered versions of (square-integrable) spherical fields are obtainable by convolution, leading to 'zooming-in', multiscale approximations. Altogether, the Legendre polynomial is the keystone of any work about special functions in the mathematical geosciences. It enables the transition from spherical harmonics via zonal kernels up to the Dirac kernel. The Funk–Hecke formula and its consequences for spherical convolutions open new methodological perspectives for global as well as local approximation in scalar, vectorial and tensorial applications.

In this paper, we discuss selected systems of trial functions on the sphere with a brief excursion to basis functions on the ball. These spherical function systems are investigated with respect to their localization in space and frequency/momemtum. Moreover, we briefly summarize a method of finding a best basis by Regularized Functional Matching Pursuit.

2. Special function systems on sphere and ball

Because of the nearly spherical shape of the Earth, spherical functions and concepts play an essential part in all of the geosciences. By a spherical variant of the Weierstraß theorem, spherical polynomials, the spherical harmonics, approximate continuous functions with respect to different topologies.

2.1. Spherical harmonics

Spherical harmonics are the analogues of trigonometric functions for Fourier expansion theory on the sphere. They were introduced to study gravitational theory [61, 62]. Early publications on the theory of spherical harmonics in their original physical interpretation as 'multipoles' are by Clebsch [16], Sylvester [93], Heine [54], Neumann [77], and Maxwell [66]. Global geomagnetic data and basic spherical harmonic expansions became available in the mid 1800s [50]. Today, the use of spherical harmonics in all geosciences is well established, particularly for the representation of scalar potentials. Reference models for the Earth's gravitational or magnetic fields are distributed as tables of coefficients for the spherical harmonic expansion of their potentials. In this approach, each spherical harmonic is a polynomial 'ansatz-function', corresponding to one 'degree', or in the jargon of signal processing, to exactly one 'frequency'. Thus, orthogonal (Fourier) expansion in terms of spherical harmonics amounts to the superposition of summands with an oscillating character determined by the degree of the Legendre polynomial (see Table 1). The more spherical harmonics are involved in the expansion, the more the oscillations grow in number, but the smaller are their amplitudes.

Weierstraß approximation theorem ↓ geophysical constraint of harmonicity
spherical harmonics $Y_{n,j}$ as restrictions of homogeneous harmonic polynomials $H_{n,j}$ to the unit sphere $\Omega \subset \mathbb{R}^3$
orthonormality and invariance ↓ addition theorem
one-dimensional Legendre polynomial P_n: $P_n(\xi \cdot \eta) = \frac{4\pi}{2n+1} \sum_{j=1}^{2n+1} Y_{n,j}(\xi) Y_{n,j}(\eta), \quad \xi, \eta \in \Omega$
convolution by the Legendre kernel ↓ Funk–Hecke formula
Legendre transform of F: $(P_n * F)(\xi) = \frac{2n+1}{4\pi} \int_\Omega P_n(\xi \cdot \eta) F(\eta) \, \mathrm{d}\omega(\eta), \quad \xi \in \Omega$
superposition over frequencies ↓ orthogonal series expansion
Fourier series of $F \in \mathrm{L}^2(\Omega)$: $F(\xi) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \int_\Omega P_n(\xi \cdot \eta) F(\eta) \, \mathrm{d}\omega(\eta), \quad \xi \in \Omega$

TABLE 1. Fourier expansion of square-integrable scalar functions on the sphere.

The geosciences deal with the space $\mathrm{L}^2(\Omega)$ of square-integrable functions on the unit sphere Ω. The quantity

$$\|F\|_{\mathrm{L}^2(\Omega)} = \left(\int_\Omega (F(\xi))^2 \, \mathrm{d}\omega(\xi) \right)^{1/2} \tag{1}$$

may be understood as the energy of the 'signal' $F \in \mathrm{L}^2(\Omega)$. The appropriate representation of a finite-energy signal in terms of a countable Hilbert basis is one of the most centrally important problems in the mathematical geosciences. The spherical harmonics form a Hilbert basis in $\mathrm{L}^2(\Omega)$. Suitable systems of spherical harmonics $\{Y_{n,k}\}_{n=0,1,\ldots;\ k=1,\ldots,2n+1}$ are often defined by the restriction of homogeneous

harmonic polynomials to the sphere. The polynomial structure has tremendous advantages. First, spherical harmonics of different degrees are orthogonal. Second, the space Harm_n of spherical harmonics of degree n is finite-dimensional: $\dim(\mathrm{Harm}_n) = 2n+1$. Therefore, the basis property of $\{Y_{n,k}\}_{n=0,1,\ldots;\ k=1,\ldots,2n+1}$ is equivalently characterized by the completion of the direct sum $\bigoplus_{n=0}^{\infty} \mathrm{Harm}_n$, i.e.:

$$\mathrm{L}^2(\Omega) = \overline{\bigoplus_{n=0}^{\infty} \mathrm{Harm}_n}^{\|\cdot\|_{\mathrm{L}^2(\Omega)}} . \tag{2}$$

This is the canonical reason why spherical harmonic (multipole) expansions underlie the classical approaches to geopotentials.

Fourier transform. More explicitly, any 'signal' $F \in \mathrm{L}^2(\Omega)$ can be split into 'orthogonal contributions' involving the Fourier transforms $F^\wedge(n,k)$ defined by

$$F^\wedge(n,k) = \int_\Omega F(\xi) Y_{n,k}(\xi)\, \mathrm{d}\omega(\xi), \tag{3}$$

in terms of $\mathrm{L}^2(\Omega)$-orthonormal spherical harmonics $\{Y_{n,k}\}_{\substack{n=0,1,\ldots \\ k=1,\ldots,2n+1}}$. Parseval's identity identifies the spatial energy of a signal with the spectral energy, decomposed orthogonally into single frequency contributions

$$\|F\|^2_{\mathrm{L}^2(\Omega)} = \langle F, F\rangle_{\mathrm{L}^2(\Omega)} = \sum_{n=0}^{\infty} \sum_{k=1}^{2n+1} \left(F^\wedge(n,k)\right)^2 .$$

This explains why the (global) geosciences work more often with the '*amplitude spectrum*' $\{F^\wedge(n,k)\}_{\substack{n=0,1,\ldots \\ k=1,\ldots,2n+1}}$ than with the 'original signal' $F \in \mathrm{L}^2(\Omega)$.

Inverse Fourier transform. The 'inverse Fourier transform'

$$F = \sum_{n=0}^{\infty} \sum_{k=1}^{2n+1} F^\wedge(n,k) Y_{n,k} \tag{4}$$

allows the geoscientist to think of the function (signal) F as a sum of 'wave functions' $Y_{n,k}$ corresponding to different frequencies. One can think of measurements as operating on an 'input signal' F to produce an output signal $G = \Lambda F$, where Λ is an operator acting on $\mathrm{L}^2(\Omega)$. Fortunately, large portions of interest can be well approximated by linear rotation-invariant pseudodifferential operators (see, e.g., [33, 47, 92]). If Λ is such an operator on $\mathrm{L}^2(\Omega)$, this means that

$$\Lambda Y_{n,k} = \Lambda^\wedge(n) Y_{n,k}, n = 0, 1, \ldots; k = 1, \ldots, 2n+1, \tag{5}$$

where the 'symbol' $\{\Lambda^\wedge(n)\}_{n\in\mathbb{N}_0}$ is a sequence of real values (independent of the order k). Thus, we have the fundamental fact that the spherical harmonics are the eigenfunctions of the operator Λ. Different pseudodifferential operators Λ are characterized by their eigenvalues $\Lambda^\wedge(n)$. All eigenvalues $\{\Lambda^\wedge(n)\}_{n\in\mathbb{N}_0}$ are collected in the so-called symbol of Λ. The 'amplitude spectrum' $\{G^\wedge(n,k)\}$ of the response of Λ is described in terms of the amplitude spectrum of functions (signals) by a simple multiplication by the 'transfer' $\Lambda^\wedge(n)$.

Bandlimited/spacelimited functions. Physical devices do not transmit spherical harmonics of arbitrarily high frequency without severe attenuation. The 'transfer' $\Lambda^\wedge(n)$ usually tends to zero with increasing n. It follows that the amplitude spectra of the responses (observations) to functions (signals) of finite energy are also negligibly small beyond some finite frequency. Thus, both because of the frequency limiting nature of the devices used, and because of the nature of the 'transmitted signals', the geoscientist is soon led to consider bandlimited functions. These are the functions $F \in \mathrm{L}^2(\Omega)$ whose 'amplitude spectra' vanish for all $n > N$ ($N \in \mathbb{N}$ fixed). In other words, each bandlimited function $F \in \mathrm{L}^2(\Omega)$ can be written as a finite Fourier series. So, any function F of the form $F = \sum_{n=0}^{N} \sum_{k=1}^{2n+1} F^\wedge(n,k) Y_{n,k}$ is said to be *bandlimited with the band* N, if $F^\wedge(N,k) \neq 0$ for at least one k. In analogous manner, $F \in \mathrm{L}^2(\Omega)$ is said to be *locally supported* (*spacelimited*) *with spacewidth* ρ around an axis $\eta \in \Omega$, if for some $\rho \in (-1,1)$ the function F vanishes on the set of all $\xi \in \Omega$ with $-1 \leq \xi \cdot \eta \leq \rho$ (where ρ is the largest number for which this is the case). Bandlimited functions are infinitely often differentiable everywhere. Moreover, it is clear that any bandlimited function F is an analytic function. From the analyticity, it follows immediately that a non-trivial bandlimited function cannot vanish on any (non-degenerate) subset of Ω. The only function that is both bandlimited and spacelimited is the zero function.

In addition to bandlimited but non-spacelimited functions, numerical analysis would like to deal with spacelimited functions. However, as we have seen, such a function (signal) of finite (space) support cannot be bandlimited, it must contain spherical harmonics of arbitrarily large frequencies. Thus, there is a dilemma of seeking functions that are somehow concentrated in both space and frequency (more accurately, angular momentum domain). There is a way of mathematically expressing the impossibility of simultaneous confinement of a function to space and angular momentum, namely the *uncertainty principle*.

2.2. Zonal kernel functions

To understand the *transition from the theory of spherical harmonics through zonal kernel functions to the Dirac kernel*, we have to realize the relative advantages of the classical Fourier expansion method by means of spherical harmonics, and this not only in the frequency domain, but also in the space domain. It is characteristic for Fourier techniques that the spherical harmonics as polynomial trial functions admit no localization in space domain, while in the frequency domain (or: angular momentum domain), they always correspond to exactly one degree, i.e., frequency, and therefore, are said to show ideal frequency localization. Because of the ideal frequency localization and the simultaneous absence of space localization, in fact, local changes of fields (signals) in the space domain affect the whole table of orthogonal (Fourier) coefficients. This, in turn, causes global changes of the corresponding (truncated) Fourier series in the space domain. Nevertheless, ideal frequency localization is often helpful for meaningful physical interpretations by relating the different observables of a geopotential to each other at a fixed frequency.

Taking these aspects on spherical harmonic modeling by Fourier series into account, trial functions which simultaneously show ideal frequency localization as well as ideal space localization would be a desirable choice. In fact, such an ideal system of trial functions would admit models of highest spatial resolution which were expressible in terms of single frequencies. However, from the uncertainty principle – the connection between space and frequency localization – we will see that both characteristics are mutually exclusive.

In conclusion, Fourier expansion methods are well suited to resolve low and medium frequency phenomena, i.e., the 'trends' of a signal, while their application to obtain high resolution in global or local models is critical. This difficulty is also well known to theoretical physics, e.g., when describing monochromatic electromagnetic waves or considering the quantum-mechanical treatment of free particles. There, plane waves with fixed frequencies (ideal frequency localization, no space localization) are the solutions of the corresponding differential equations, but they do certainly not reflect the physical reality. As a remedy, plane waves of different frequencies are superposed into 'wave-packages' that gain a certain amount of space localization, while losing their ideal spectral localization. In a similar way, a suitable superposition of polynomial functions leads to so-called zonal kernel functions, in particular to kernel functions with a reduced frequency, but increased space localization.

More concretely, any kernel function $K : \Omega \times \Omega \to \mathbb{R}$ that is characterized by the property that there exists a function $\tilde{K} : [0, 2] \to \mathbb{R}$ such that

$$K(\xi, \eta) = \tilde{K}(|\xi - \eta|) = \tilde{K}\left(\sqrt{2 - 2\xi \cdot \eta}\right) = \hat{K}(\xi \cdot \eta), \quad \xi, \eta \in \Omega, \tag{6}$$

is called a (spherical) *radial basis function* (at least in the theory of constructive approximation).

Zonal kernels. The application of a rotation (i.e., a 3×3 'orthogonal' matrix $\mathbf{t}$ with $\mathbf{t}^{\mathrm{T}} = \mathbf{t}^{-1}$) leads to $K(\mathbf{t}\xi, \mathbf{t}\eta) = \hat{K}((\mathbf{t}\xi) \cdot (\mathbf{t}\eta)) = \hat{K}(\xi \cdot (\mathbf{t}^{\mathrm{T}}\mathbf{t}\eta)) = \hat{K}(\xi \cdot \eta) = K(\xi, \eta)$. In particular, a rotation around the axis $\xi \in \Omega$ (i.e., $\mathbf{t}\xi = \xi$) yields $K(\xi, \eta) = K(\xi, \mathbf{t}\eta)$ for all $\eta \in \Omega$. Hence, $K(\xi, \cdot)$ possesses a rotational symmetry with respect to the axis ξ. In the theory of special functions of mathematical physics, a kernel $\hat{K} : \Omega \times \Omega \to \mathbb{R}$ satisfying $\hat{K}(\xi \cdot \eta) = \hat{K}(\mathbf{t}\xi \cdot \mathbf{t}\eta)$, $\xi, \eta \in \Omega$, for all orthogonal transformations $\mathbf{t}$ is known as a *zonal kernel function*. To highlight the reducibility of $\hat{K}$ to a function defined on the interval $[-1, 1]$, the notation $(\xi, \eta) \mapsto \hat{K}(\xi \cdot \eta)$, $(\xi, \eta) \in \Omega \times \Omega$, is used throughout this chapter (see also (6)).

From the theory of spherical harmonics we get a representation of any $\mathrm{L}^2(\Omega)$-zonal kernel function K in terms of a Legendre expansion

$$K(\xi \cdot) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} K^{\wedge}(n) P_n(\xi \cdot) \tag{7}$$

(in the $\|\cdot\|_{\mathrm{L}^2(\Omega)}$-sense), where the sequence $\{K^\wedge(n)\}_{n\in\mathbb{N}_0}$ given by

$$K^\wedge(n) = 2\pi \int_{-1}^{1} K(t) P_n(t)\,\mathrm{d}t \tag{8}$$

is called the *Legendre symbol* of the zonal kernel $K(\xi\cdot)$. A simple but extreme example (with optimal frequency localization and no space localization) is the Legendre kernel where $K^\wedge(n) = 1$ for one particular n and $K^\wedge(m) = 0$ for $m \neq n$, i.e., the Legendre kernel is given by

$$\Omega \times \Omega \ni (\xi, \eta) \mapsto \frac{2n+1}{4\pi} P_n(\xi \cdot \eta).$$

In other words, additive clustering of weighted Legendre kernels generates zonal kernel functions. It is of importance to distinguish bandlimited kernels (i.e., $K^\wedge(n) = 0$ for all $n \geq N$) and non-bandlimited ones, for which infinitely many numbers $K^\wedge(n)$ do not vanish. Non-bandlimited kernels show a much stronger space localization than their bandlimited counterparts. Empirically, if $K^\wedge(n) \approx K^\wedge(n+1) \approx 1$ for many successive large integers n, then the support of the series (7) in the space domain is small, i.e., the kernel is spacelimited (i.e., in the jargon of approximation theory 'locally supported'). This leads to the other extremal kernel (in contrast to the Legendre kernel) which is the Dirac kernel with optimal space localization but no frequency localization and $K^\wedge(n) = 1$ for all n, where, however, the Dirac kernel does not exist as a classical function in the mathematical sense. Nevertheless, it is well known that, if we have a family of kernels $\{K_J\}_{J=0,1,\ldots}$ where $\lim_{J\to\infty} K_J^\wedge(n) = 1$ for each n and an additional (technical) condition holds, then $K_J * F$ tends to F in the sense of $\mathrm{L}^2(\Omega)$ for all $F \in \mathrm{L}^2(\Omega)$.

Assuming $\lim_{n\to\infty} K^\wedge(n) = 0$, necessary to get a 'proper' function, the slower the sequence $\{K^\wedge(n)\}_{n=0,1,\ldots}$ converges to zero, the lower the frequency localization, and the higher the space localization. A unified scheme is found in Table 2. Zonal kernel function theory relies on the following principles:

(i) Weighted Legendre kernels are the summands of zonal kernel functions.
(ii) The Legendre kernel is ideally localized in frequency. The Dirac kernel is ideally localized in space.
(iii) The only frequency- *and* spacelimited zonal kernel is the zero function.

Legendre kernels	zonal kernels general case		Dirac kernel
	bandlimited	spacelimited	

TABLE 2. From Legendre kernels via zonal kernels to the Dirac kernel

2.3. Slepian functions

As we have seen, $\|F\|^2_{\mathrm{L}^2(\Omega)}$ is the energy of a certain function $F \in \mathrm{L}^2(\Omega)$. Suppose now that there is a particular region $C \subset \Omega$, and let us define the 'local' energy of that function as $\|F\|^2_{\mathrm{L}^2(C)}$. Functions F that are *band*limited,

$$F = \sum_{n=0}^{N} \sum_{k=1}^{2n+1} F^{\wedge}(n,k) Y_{n,k} \tag{9}$$

cannot also be *space*limited, but they can be space*concentrated*.

Bandlimited/spaceconcentrated Slepian functions. By maximizing the spatial energy ratio

$$\lambda_C(F) = \frac{\|F\|^2_{\mathrm{L}^2(C)}}{\|F\|^2_{\mathrm{L}^2(\Omega)}}, \quad 0 < \lambda_C(F) < 1, \tag{10}$$

we obtain bandlimited spherical 'Slepian functions' [56, 67, 87], named in analogy with the prolate spheroidal wave functions of Slepian [90]. They are not, in general, zonal functions.

The Fourier coefficients of the Slepian functions are the $(N+1)^2$ orthogonal eigenvectors of the symmetric concentration matrix whose elements are the limited-domain inner-product terms $\langle Y_{m,j}, Y_{n,k} \rangle_{\mathrm{L}^2(C)}$, $0 \le m, n \le N$, i.e.,

$$\sum_{n=0}^{N} \sum_{k=1}^{2n+1} \langle Y_{m,j}, Y_{n,k} \rangle_{\mathrm{L}^2(C)} F^{\wedge}(n,k) = \lambda_C(F) F^{\wedge}(m,j). \tag{11}$$

We will give their associated eigenvalues superscripted labels and rank them in decreasing order of concentration, $1 > \lambda_C^{(1)}(F) \ge \lambda_C^{(\alpha)}(F) \ge \lambda_C^{((N+1)^2)}(F) > 0$. The bandlimited Slepian functions can alternatively be obtained by solving a Fredholm integral equation with a 'Shannon' concentration kernel:

$$\int_C \sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(\xi \cdot \eta) F(\eta) \, \mathrm{d}\omega(\eta) = \lambda_C(F) F(\xi), \quad \xi \in \Omega. \tag{12}$$

Spacelimited/bandconcentrated Slepian functions. We can define space*limited* Slepian functions which are band*concentrated*. They are obtained by the restriction of the bandlimited Slepian functions F to the region of interest C, or, equivalently, their Fourier coefficients are

$$\sum_{n=0}^{N} \sum_{k=1}^{2n+1} \langle Y_{m,j}, Y_{n,k} \rangle_{\mathrm{L}^2(C)} F^{\wedge}(n,k), \tag{13}$$

extending the $F^{\wedge}(n,k)$ to *all* degrees $m = 0, 1, \ldots, \infty$ and order indices $j = 1, \ldots, 2m+1$.

A central concept is the *effective* dimension of functions that are 'essentially' space- and bandlimited. The *Shannon number* is the trace of the concentration operators in (11)–(12), given by (using the addition theoremin 14b–14c),

$$N_C = \sum_{\alpha=1}^{(N+1)^2} \lambda_C^{(\alpha)}(F) \tag{14a}$$

$$= \sum_{n=0}^{N} \sum_{k=1}^{2n+1} \langle Y_{n,k}, Y_{n,k} \rangle_{\mathrm{L}^2(C)} \tag{14b}$$

$$= \int_C \sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(\eta \cdot \eta)\, \mathrm{d}\omega(\eta) \tag{14c}$$

$$= (N+1)^2 \frac{\int_C \mathrm{d}\omega(\eta)}{4\pi}. \tag{14d}$$

The eigenvalue spectrum $\lambda_C^{(\alpha)}(F)$ has a characteristic step-like shape, with the property $\sum_{\alpha=1}^{(N+1)^2} \lambda_C^{(\alpha)}(F) \approx \sum_{\alpha=1}^{N_C} \lambda_C^{(\alpha)}(F)$ revealing that N_C will be close to the number of Slepian functions that usefully contribute to the approximation of arbitrary target functions on domains $C \subset \Omega$.

While computation can be carried out via either (11) or (12), when the region of interest C is a spherical cap (one whose boundary ∂C is a circle and whose half-opening angle is Θ), the integral equation (12) commutes with a Sturm–Liouville differential equation whose spectral-domain representation has an extremely simple analytical form, rendering the computation of Slepian functions of domains essentially trivial [51]. In that case, the Slepian functions degenerate to being the solutions of fixed-order (j) versions of equation (11), with a partial Shannon number given in terms of products of the associated Legendre functions and their derivatives (primed), namely

$$N_j = \frac{(N-j+1)!}{2(N+j)!} \int_{\cos\Theta}^{1} \left[P'_{N+1,j}(t) P_{N,j}(t) - P'_{N,j}(t) P_{N+1,j}(t) \right] \mathrm{d}t. \tag{15}$$

Only on circularly symmetric domains and when the spherical-harmonic order $j = 0$ are the Slepian functions zonal, and in that case, the fixed-order partial Shannon number is well approximated by Wieczorek and Simons [104] as

$$N_0 \approx 2\frac{\sqrt{N_C}}{\pi} \approx (N+1)\frac{\Theta}{\pi}. \tag{16}$$

Figure 1 shows examples of spherical-cap Slepian functions, their power spectra, and their eigenvalue spectra.

When the concentration domain is a spherical cap, the best-concentrated (highest-eigenvalue) bandlimited Slepian function is a zonal function that is close to optimally localized under the uncertainty principle (see Section 3). All the lower-eigenvalue zonal Slepian functions, and finally, all the non-zonal Slepian functions, together form a complete orthonormal basis for the space of functions on the unit

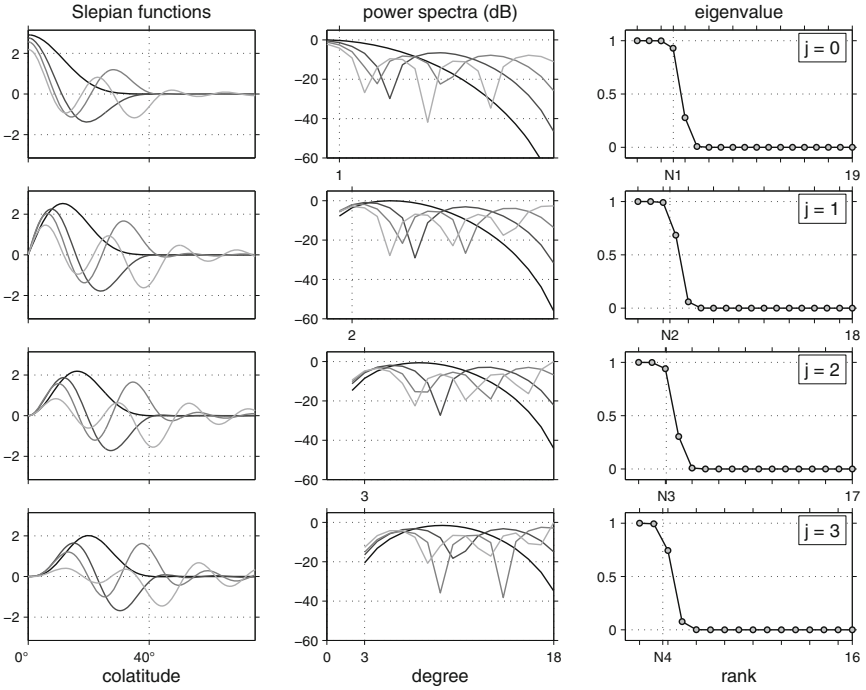


FIGURE 1. The four best-concentrated (in decreasing gray shading) fixed-order (top to bottom, $j = 0, 1, 2, 3$) Slepian functions and their power spectra, for a common bandwidth $N = 18$, with the domain C a 40° spherical cap. Also shown are the complete eigenvalue spectra with the fixed-order (partial) Shannon numbers indicated. Compare to [87] (their Figs. 5.1, 5.2 and 5.3.).

sphere Ω that are bandlimited to N. The partial Shannon numbers N_j sum to the full Shannon number N_C via

$$N_C = \sum_{j=0}^{2N+1} N_j. \tag{17}$$

2.4. From the scalar to the vector and tensor context

In the second half of the last century, a physically motivated approach for the decomposition of spherical vector and tensor fields was presented based on a spherical variant of the Helmholtz theorem, e.g., [6–8, 75]. Following this concept, the tangential part of a spherical vector field is split up into a curl-free and a divergence-free field by use of two differential operators, viz. the surface gradient and the surface curl gradient. Of course, an analogous splitting is valid in tensor theory.

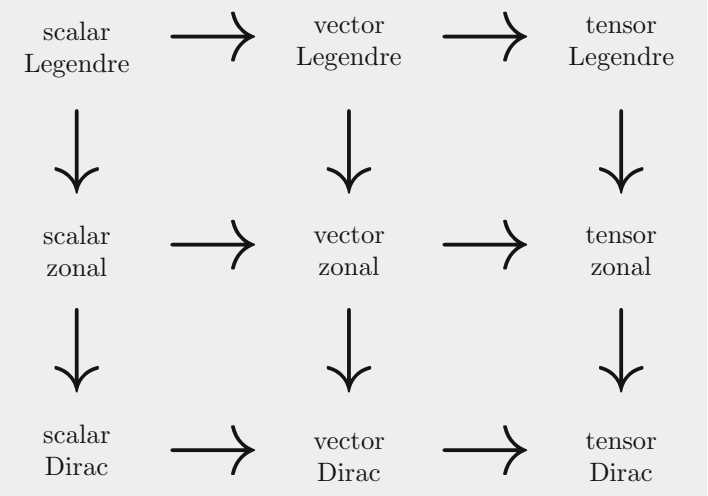


TABLE 3. From scalar via vectorial to tensorial kernels

In subsequent publications during the second half of the last century, however, the vector spherical harmonic theory was usually written in local coordinate expressions that make mathematical formulations lengthy and hard to read. Tensor spherical harmonic settings were even more difficult to understand. In addition, when using local coordinates within a global spherical concept, differential geometry tells us that there is no representation of vector and tensor spherical harmonics that is free of singularities. As a consequence, vector and tensor spherical harmonics have suffered from an inadequately complex and inconsistent literature. Absent coordinate-free explicit formulas, the orthogonal invariance based on specific vector/tensor extensions of the Legendre polynomials was not worked out suitably in a unifying scalar/vector/tensor framework, nor was the concept of zonal (kernel) functions adequately generalized to the spherical vector/tensor case.

All new structures for spherical functions in mathematical (geo)physics were developed by Freeden and Schreiner [43] and Freeden and Gutting [32]. Two fundamental transitions underlie their approach: one from spherical harmonics via zonal kernel functions to the Dirac kernels, and the other one from scalar to vector and tensor theory (see Table 3).

Helmholtz decomposition of spherical vector/tensor fields. To explain the *transition from the theory of scalar spherical harmonics to its vectorial and tensorial extensions*, Freeden and Schreiner [43] start from physically motivated dual pairs of operators (the reference space being always the space of signals with finite energy, i.e., the space of square-integrable fields). The pair $o^{(i)}, O^{(i)}, i \in \{1,2,3\}$, originates in the ingredients of the Helmholtz decomposition of a vector field, while

$o^{(i,k)}, O^{(i,k)}, i, k \in \{1,2,3\}$, take the analogous role for the Helmholtz decomposition of tensor fields (see, e.g., [7, 43, 47]).

For example, in vector theory, $o^{(1)}F$ is the normal field

$$\xi \mapsto o_\xi^{(1)} F(\xi) = F(\xi)\xi, \quad \xi \in \Omega,$$

whereas $o^{(2)}F$ is the surface gradient field

$$\xi \mapsto o_\xi^{(2)} F(\xi) = \nabla_\xi^* F(\xi), \quad \xi \in \Omega,$$

and $o^{(3)}F$ is the surface curl gradient field

$$\xi \mapsto o_\xi^{(3)} F(\xi) = \mathrm{L}_\xi^* F(\xi), \quad \mathrm{L}_\xi^* = \xi \wedge \nabla_\xi^*, \quad \xi \in \Omega,$$

applied to a scalar function F. In addition, $O^{(1)}f$ is the normal component

$$\xi \mapsto O_\xi^{(1)} f(\xi) = f(\xi) \cdot \xi, \quad \xi \in \Omega,$$

while $O^{(2)}f$ is the negative surface divergence

$$\xi \mapsto O_\xi^{(2)} f(\xi) = -\nabla_\xi^* \cdot f(\xi), \quad \xi \in \Omega,$$

and $O^{(3)}f$ is the negative surface curl

$$\xi \mapsto O_\xi^{(3)} f(\xi) = -\mathrm{L}_\xi^* \cdot f(\xi), \quad \xi \in \Omega,$$

taken over a vector-valued function f.

Clearly, the operators $o^{(i,k)}, O^{(i,k)}$ are also definable in orientation to the tensor Helmholtz decomposition theorem (for reasons of simplicity, however, their explicit description is omitted here). The pairs $o^{(i)}, O^{(i)}$ and $o^{(i,k)}, O^{(i,k)}$ of dual operators lead us to an associated palette of Legendre kernel functions, all of them generated by the classical one-dimensional Legendre polynomial P_n of degree n. To be more specific, three types of Legendre kernels occur in the vectorial as well as tensorial context (see Table 4).

The Legendre kernels $o^{(i)}P_n$ and $o^{(i)}o^{(i)}P_n$ pertain to the vector approach for spherical harmonics, whereas $o^{(i,k)}P_n$ and $o^{(i,k)}o^{(i,k)}P_n$, $i, k = 1,2,3$, form the analogues in tensorial theory. Corresponding to each Legendre kernel, we are led to two variants for representing square-integrable fields by orthogonal (Fourier) expansion, where the reconstruction – as in the scalar case – is undertaken by superposition over all frequencies.

In a unified notation, the formalism for vector/tensor spherical harmonic theory is based on the following principles (cf. [43]):

(i) The vector/tensor spherical harmonics involving the $o^{(i)}, o^{(i,k)}$-operators, respectively, are obtainable as restrictions of three-dimensional homogeneous harmonic vector/tensor polynomials, respectively.

(ii) The vector/tensor Legendre kernels are obtainable as the outcome of sums extended over a maximal orthonormal system of vector/tensor spherical harmonics of degree (frequency) n, respectively.

Scalar Legendre polynomial: $P_n = \frac{O^{(i)}O^{(i)}\mathbf{p}_n^{(i,i)}}{\mu_n^{(i)}} = \frac{O^{(i,k)}O^{(i,k)}\mathbf{P}_n^{(i,k)}}{\mu_n^{(i,k)}}$	
application of $o^{(i)}$ ↓↑ application of $O^{(i)}$	application of $o^{(i,k)}$ ↓↑ application of $O^{(i,k)}$
vector Legendre kernel	tensor Legendre kernel (order 2)
$p_n^{(i)} = \frac{o^{(i)}P_n}{\left(\mu_n^{(i)}\right)^{1/2}} = \frac{O^{(i)}\mathbf{p}_n^{(i,i)}}{\left(\mu_n^{(i)}\right)^{1/2}}$	$\mathbf{p}_n^{(i,k)} = \frac{o^{(i,k)}P_n}{\left(\mu_n^{(i,k)}\right)^{1/2}} = \frac{O^{(i,k)}\mathbf{P}_n^{(i,k)}}{\left(\mu_n^{(i,k)}\right)^{1/2}}$
application of $o^{(i)}$ ↓↑ application of $O^{(i)}$	application of $o^{(i,k)}$ ↓↑ application of $O^{(i,k)}$
tensor Legendre kernel (order 2)	tensor Legendre kernel (order 4)
$\mathbf{p}_n^{(i,i)} = \frac{o^{(i)}p_n^{(i)}}{\left(\mu_n^{(i)}\right)^{1/2}} = \frac{o^{(i)}o^{(i)}P_n}{\mu_n^{(i)}}$	$\mathbf{P}_n^{(i,k,i,k)} = \frac{o^{(i,k)}\mathbf{p}_n^{(i,k)}}{\left(\mu_n^{(i,k)}\right)^{1/2}} = \frac{o^{(i,k)}o^{(i,k)}P_n}{\mu_n^{(i,k)}}$
vectorial context	*tensorial context*

TABLE 4. Legendre scalar, vectorial, and tensorial kernel functions.

(iii) The vector/tensor Legendre kernels are zonal kernel functions, rotation-invariant (in vector/tensor sense, respectively) with respect to orthogonal transformations (leaving one point of the unit sphere Ω fixed).

(iv) Spherical harmonics of degree (frequency) n form an irreducible subspace of the reference space of (square-integrable) fields on Ω.

(v) Each Legendre kernel implies an associated Funk–Hecke formula that determines the constituting features of the convolution (filtering) of a square-integrable field against the Legendre kernel.

(vi) The orthogonal Fourier expansion of a square-integrable field is the sum of the convolutions of the field against the Legendre kernels being extended over all frequencies.

To summarize, the theory of spherical harmonics provides us with a framework to unify, review and supplement the different approaches in real scalar, vector, and tensor theory. The essential tools are the Legendre functions, used in orthogonal Fourier expansions and endowed with rotational invariance. The coordinate-free construction yields a number of formulas and theorems that previously were derived only in coordinate (e.g., polar) representations. Consequently, any kind of singularities is avoided at the poles. Finally, our transition from the scalar to

the vectorial as well as the tensorial case opens new promising perspectives of constructing important zonal classes of spherical trial functions by summing up Legendre kernel expressions, thereby providing (geo-)physical relevance and increasing local applicability [43]. Similar considerations apply to the construction of vector/tensor Slepian functions, e.g., [21, 79].

2.5. From the sphere to the ball

The modeling of structures inside the Earth requires basis functions on the ball $\mathcal{B} = \{x \in \mathbb{R}^3 : |x| \leq \beta\}$ with $\beta > 0$. Several approaches for the construction of such basis systems exist. Of course, from the mathematical point of view, one could easily take a basis $\{B_k\}_{k=0,1,\ldots}$ on the Cartesian domain $[-\beta, \beta]$ to construct a basis on the cube $[-\beta, \beta]^3$ by simply taking the tensor product basis $(x_1, x_2, x_3) \mapsto B_{k_1}(x_1)B_{k_2}(x_2)B_{k_3}(x_3)$, $k_1, k_2, k_3 = 0, 1, \ldots$ However, the Earth's interior is usually subdivided into structural layers that are approximately bounded by spheres. In view of this fact, the use of cartesian-coordinate-based trial functions appears to be inappropriate and the spherical harmonics also here play an essential role.

An intuitive approach is to look for basis functions of the form

$$G_{m,n,k}(r\xi) = F_{m,n}(r)Y_{n,k}(\xi), \quad \xi \in \Omega,\ r \in [0, \beta],$$

for $m, n = 0, 1, \ldots$ and $k = 1, \ldots, 2n + 1$. Also here, orthogonality appears to be useful, which leads to the requirement that

$$\begin{aligned}
&\int_{\mathcal{B}} G_{m_1,n_1,k_1}(x)\, G_{m_2,n_2,k_2}(x)\, \mathrm{d}x \\
&\quad = \int_0^\beta r^2 F_{m_1,n_1}(r)\, F_{m_2,n_2}(r)\, \mathrm{d}r \int_\Omega Y_{n_1,k_1}(\xi)\, Y_{n_2,k_2}(\xi)\, \mathrm{d}\omega(\xi) \\
&\quad = \int_0^\beta r^2 F_{m_1,n_1}(r)\, F_{m_2,n_1}(r)\, \mathrm{d}r = 0,
\end{aligned} \tag{18}$$

if $m_1 \neq m_2$ or $n_1 \neq n_2$ or $k_1 \neq k_2$. The weight function r^2 in the radial integral in (18) suggests the use of the Jacobi polynomials as building blocks for $F_{m,n}$. However, there is a notable degree of freedom in the choice of (e.g., polynomial) functions for $F_{m,n}$. This degree of freedom can be used to construct the $G_{m,n,k}$ in a manner such that they characterize the non-uniqueness of solutions of tomographic inverse problems in the geosciences or medical imaging. For further details, see [9, 63, 67, 68, 70, 97] and the contribution by Leweke, Michel, and Telschow (this book, pp. 883–919). Note that some of the obtained systems become discontinuous or even singular at the origin $0 \in \mathcal{B}$ but in a way such that they are still elements of $\mathrm{L}^2(\mathcal{B})$.

The fact that such orthonormal basis functions on the ball arise from the spherical harmonics as orthonormal basis functions on the sphere yields a way to formulate analogies regarding the methodologies and the associated properties – though often further difficulties occur due to the additional radial coordinate.

Particular analogies exist with respect to the space and 'frequency' localization of kernels

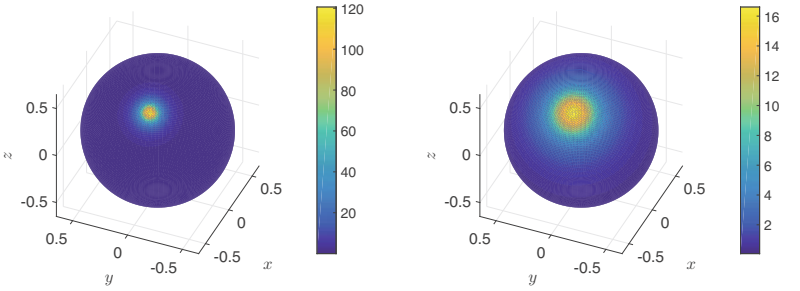
$$K(x,y) = \sum_{m,n=0}^{\infty} \sum_{k=1}^{2n+1} K^\wedge(m,n)\, G_{m,n,k}(x)\, G_{m,n,k}(y)$$
$$= \sum_{m,n=0}^{\infty} K^\wedge(m,n) F_{m,n}(|x|) F_{m,n}(|y|) \frac{2n+1}{4\pi} P_n\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right), \qquad (19)$$

$x, y \in \mathcal{B} \setminus \{0\}$. It should be noted, however, that most of the choices of $F_{m,n}$ do *not* lead to functions $G_{m,n,k}$ which are algebraic polynomials in x_1, x_2, x_3. Nevertheless, the properties of the Jacobi polynomials and the spherical harmonics imply that the $G_{m,n,k}$ also show an increasing oscillatory behavior for increasing m or n. Furthermore, the Dirac kernel can also here be associated to the case where all coefficients satisfy $K^\wedge(m,n) = 1$.

One of the advantages of this approach – in contrast to a cartesian setup – is that the obtained kernels inherit the rotational invariance from the spherical kernels in the sense that

$$K(r\mathbf{t}\xi, s\mathbf{t}\eta) = \sum_{n=0}^{\infty} K^\wedge(m,n) F_{m,n}(r) F_{m,n}(s) \frac{2n+1}{4\pi} P_n((\mathbf{t}\xi)\cdot(\mathbf{t}\eta))$$
$$= K(r\xi, s\eta)$$

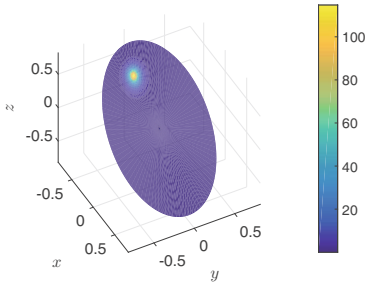
for all 3×3-orthogonal matrices (i.e., rotations) $\mathbf{t}$, see Figure 2.



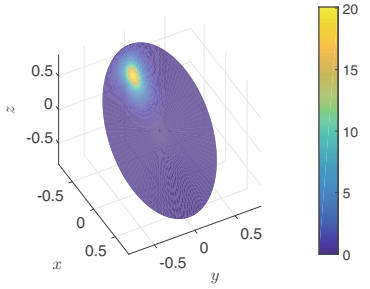
(A) Kernel with $K^\wedge(m,n) = 0.8^m 0.9^n$ (B) Kernel with $K^\wedge(m,n) = 0.8^m 0.7^n$

FIGURE 2. The figures show localized trial functions $\mathcal{B} \ni y \mapsto K(x,y)$ based on a kernel of the kind in (19) with a fixed point $x = (-0.4, -0.1, 0.5)^{\mathrm{T}}$. The functions are plotted on the sphere with radius $|x|$. Each function is a hat function concentrated around x. Its restriction to a sphere around 0 is a rotationally symmetric function, as it is known for the case of spherical kernels. Note that the series representations were truncated at $n = m = 400$ in the numerical implementation.

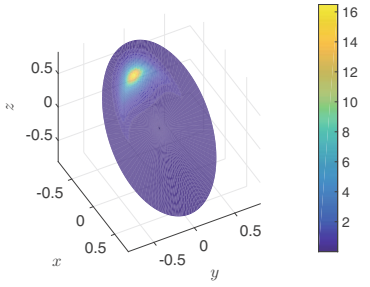
Moreover, the localization with respect to the radius (or the 'depth') can be separated from the localization with respect to the angular coordinates by taking, for example, symbols of the form $K^\wedge(m,n) = A_m B_n$ like $K^\wedge(m,n) = h_{\mathrm{r}}^m h_{\mathrm{ang}}^n$ for parameters $h_{\mathrm{r}}, h_{\mathrm{ang}} \in (0,1)$, see Figure 3. This is useful, e.g., for tomographic problems where it is known that the solution has a finer structure in the angular domain than in the radial domain (or vice versa).



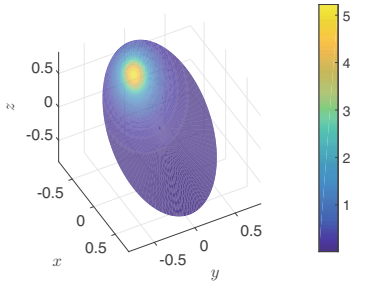
(A) Kernel with $h_{\mathrm{r}} = 0.8$ and $h_{\mathrm{ang}} = 0.9$



(B) Kernel with $h_{\mathrm{r}} = 0.5$ and $h_{\mathrm{ang}} = 0.9$



(C) Kernel with $h_{\mathrm{r}} = 0.8$ and $h_{\mathrm{ang}} = 0.7$



(D) Kernel with $h_{\mathrm{r}} = 0.5$ and $h_{\mathrm{ang}} = 0.7$

FIGURE 3. The figures show localized trial functions $\mathcal{B} \ni y \mapsto K(x,y)$ based on a kernel of the kind in (19) with a fixed point $x = (-0.4, -0.1, 0.5)^{\mathrm{T}}$. The functions are plotted on the planar cross section with normal vector $(1,1,1)^{\mathrm{T}}$. By choosing a symbol $K^\wedge(m,n) = h_{\mathrm{r}}^m h_{\mathrm{ang}}^n$, the localization in radial and in angular domain can be controlled separately. Note that the series representations were truncated at $n = m = 400$ in the numerical implementation.

3. Spherical uncertainty principle

As pointed out in Section 2, four classes of zonal kernel functions can be distinguished, namely bandlimited and non-bandlimited, spacelimited and non-spacelimited ones. In addition, Slepian functions exist in bandlimited and spacelimited varieties. What is the right kernel function for the purpose of local approximation? Of course, the user of a mathematical method is interested in knowing the trial system which fits 'adequately' to the problem. When several choices are possible or an optimal choice cannot be found it is necessary to choose the trial systems in close adaptation to the data extent or density, and the required smoothness of the field to be approximated. This, however, is often a local property, i.e., the data density can be high in one area and low in another. In addition, the field to be approximated can have a high-detail structure in some parts of the sphere (e.g., over mountainous regions) and a low-detail structure elsewhere (e.g., over the oceans). This makes the selection of appropriate trial functions even more challenging.

3.1. Derivation and basic theory

An uncertainty principle that specifies the degree of space and frequency localization is helpful to serve as a decisive criterion. The essential outcome is a better understanding of the classification of zonal kernel functions, and Slepian functions, based on the development of suitable bounds for their quantification with respect to space and frequency localization.

Localization in space. Assume F is of class $\mathrm{L}^2(\Omega)$ with energy

$$\|F\|_{\mathrm{L}^2(\Omega)} = \left(\int_\Omega \left(F(\eta)\right)^2 \, \mathrm{d}\omega(\eta) \right)^{1/2} = 1.$$

We associate to F the normal (radial) field $\eta \mapsto \eta F(\eta) = o_\eta^{(1)} F(\eta)$, $\eta \in \Omega$. This function maps $\mathrm{L}^2(\Omega)$ into the associated set of normal fields on Ω. The 'center of gravity' of F is the *expectation of the normal operator* $o^{(1)}$ *on* Ω,

$$g_F^{o^{(1)}} = \int_\Omega \left(o_\eta^{(1)} F(\eta) \right) F(\eta) \, \mathrm{d}\omega(\eta) = \int_\Omega \eta (F(\eta))^2 \, \mathrm{d}\omega(\eta) \in \mathbb{R}^3 \tag{20}$$

thereby interpreting $(F(\eta))^2 \, \mathrm{d}\omega(\eta)$ as surface mass distribution over the sphere Ω embedded in Cartesian space $\mathbb{R}^3$. It is clear that $g_F^{o^{(1)}}$ lies in the closed inner space $\overline{\Omega^{\mathrm{int}}}$ of Ω: $|g_F^{o^{(1)}}| \leq 1$. The variance of the operator $o^{(1)}$ is understood in the canonical sense as the *variance in the space domain*,

$$\begin{aligned} \left(\sigma_F^{o^{(1)}} \right)^2 &= \int_\Omega \left(\left(o_\eta^{(1)} - g_F^{o^{(1)}} \right) F(\eta) \right)^2 \, \mathrm{d}\omega(\eta) \\ &= \int_\Omega \left(\eta - g_F^{o^{(1)}} \right)^2 \left(F(\eta)\right)^2 \, \mathrm{d}\omega(\eta) \in \mathbb{R}. \end{aligned} \tag{21}$$

Observing the identity $(\eta - g_F^{o^{(1)}})^2 = 1 + (g_F^{o^{(1)}})^2 - 2\eta \cdot g_F^{o^{(1)}}$, $\eta \in \Omega$, it follows immediately that $(\sigma_F^{o^{(1)}})^2 = 1 - (g_F^{o^{(1)}})^2$. Naturally, $0 \leq (\sigma_F^{o^{(1)}})^2 \leq 1$.

Since we are particularly interested in zonal functions, some simplifications can be made. Let K be of class $\mathrm{L}^2[-1,1]$ and $\|K\|_{\mathrm{L}^2[-1,1]} = 1$, where $\|F\|_{\mathrm{L}^2[-1,1]} = (2\pi \int_{-1}^{1} (F(t))^2 \,\mathrm{d}t)^{1/2}$ for $F \in \mathrm{L}^2[-1,1]$. Then the corresponding center of gravity can be computed readily as follows ($\varepsilon^3 = (0,0,1)^{\mathrm{T}}$):

$$g^{o^{(1)}}_{K(\cdot\varepsilon^3)} = \int_\Omega \eta \left(K\left(\eta \cdot \varepsilon^3\right)\right)^2 \,\mathrm{d}\omega(\eta) = \left(2\pi \int_{-1}^{1} t\,(K(t))^2 \,\mathrm{d}t\right) \varepsilon^3. \tag{22}$$

Letting $t^{o^{(1)}}_K = \left| g^{o^{(1)}}_{K(\cdot\varepsilon^3)} \right| = 2\pi \left| \int_{-1}^{1} t\,(K(t))^2 \,\mathrm{d}t \right| \in \mathbb{R}$ we find for the variance

$$\begin{aligned} \left(\sigma^{o^{(1)}}_K\right)^2 &= \int_\Omega \left(\eta - g^{o^{(1)}}_{K(\cdot\varepsilon^3)}\right)^2 \left(K\left(\eta \cdot \varepsilon^3\right)\right)^2 \,\mathrm{d}\omega(\eta) \\ &= 1 - \left(t^{o^{(1)}}_K\right)^2 = 1 - \left(g^{o^{(1)}}_{K(\cdot\varepsilon^3)}\right)^2 \in \mathbb{R}. \end{aligned} \tag{23}$$

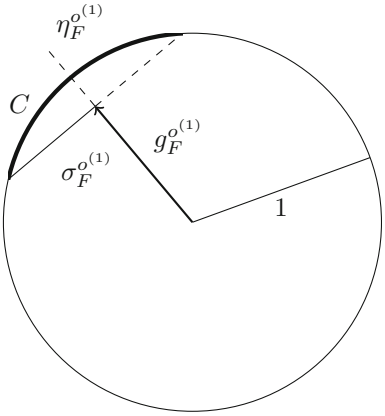


FIGURE 4. Localization in a spherical cap.

Figure 4 gives a geometric interpretation of $g^{o^{(1)}}_F$ and $\sigma^{o^{(1)}}_F$. We associate to $g^{o^{(1)}}_F$, $g^{o^{(1)}}_F \neq 0$, and its projection $\eta^{o^{(1)}}_F$ onto the sphere Ω the spherical cap $C = \{\eta \in \Omega \mid 1 - \eta \cdot \eta^{o^{(1)}}_F \leq 1 - |g^{o^{(1)}}_F|\}$. Then the boundary ∂C is a circle with radius $\sigma^{o^{(1)}}_F$. Thinking of a zonal function F as a 'spherical window function' on Ω, the window is determined by C, and its width is given by $\sigma^{o^{(1)}}_F$.

Localization in frequency ('momentum space'). The 'expectation in the frequency domain' is introduced as the *expectation of the surface curl operator* $o^{(3)}$ on Ω. Then, for $F \in \mathrm{H}^{(2l)}(\Omega)$, $l \in \mathbb{N}$, i.e., for all $F \in \mathrm{L}^2(\Omega)$ such that there exists a function $G \in \mathrm{L}^2(\Omega)$ with $G^\wedge(n,k) = (-n(n+1))^l F^\wedge(n,k)$ for all $n = 0,1,\ldots$; $k = 1,\ldots,2n+1$, we have

$$g^{o^{(3)}}_F = \int_\Omega \left(o^{(3)}_\eta F(\eta)\right) F(\eta) \,\mathrm{d}\omega(\eta) = 0 \in \mathbb{R}^3. \tag{24}$$

	operator	expectation value
space	$o^{(1)}$	$g_F^{o^{(1)}} = \int_\Omega \left(o_\eta^{(1)} F(\eta)\right) F(\eta)\, \mathrm{d}\omega(\eta)$
frequency	$o^{(3)}$	$g_F^{o^{(3)}} = \int_\Omega \left(o_\eta^{(3)} F(\eta)\right) F(\eta)\, \mathrm{d}\omega(\eta)$

	operator	variance
space	$o^{(1)}$	$\left(\sigma_F^{o^{(1)}}\right)^2 = \int_\Omega \left(\left(o_\eta^{(1)} - g_F^{o^{(1)}}\right) F(\eta)\right)^2 \mathrm{d}\omega(\eta)$
frequency	$o^{(3)}$	$\left(\sigma_F^{o^{(3)}}\right)^2 = \int_\Omega \left(\left(o_\eta^{(3)} - g_F^{o^{(3)}}\right) F(\eta)\right)^2 \mathrm{d}\omega(\eta)$

TABLE 5. Localization in terms of the normal and curl operators $o^{(1)}$ and $o^{(3)}$.

Correspondingly, the *variance in the frequency domain* is given by

$$\left(\sigma_F^{o^{(3)}}\right)^2 = \int_\Omega \left(\left(o_\eta^{(3)} - g_F^{o^{(3)}}\right) F(\eta)\right)^2 \mathrm{d}\omega(\eta) \in \mathbb{R}. \tag{25}$$

The surface theorem of Stokes shows us that

$$\begin{aligned}\left(\sigma_F^{o^{(3)}}\right)^2 &= \int_\Omega \left(o_\eta^{(3)} F(\eta)\right) \cdot \left(o_\eta^{(3)} F(\eta)\right) \mathrm{d}\omega(\eta) \\ &= \int_\Omega \left(-\Delta_\eta^* F(\eta)\right) F(\eta)\, \mathrm{d}\omega(\eta) \quad = \quad g_F^{-\Delta^*}.\end{aligned} \tag{26}$$

Expressed in terms of spherical harmonics we get via the Parseval identity

$$\left(\sigma_F^{o^{(3)}}\right)^2 = \sum_{n=0}^{\infty} \sum_{k=1}^{2n+1} n(n+1) \left(F^\wedge(n,k)\right)^2. \tag{27}$$

Note that we require $\|F\|^2_{\mathrm{L}^2(\Omega)} = \sum_{n=0}^{\infty} \sum_{k=1}^{2n+1} \left(F^\wedge(n,k)\right)^2 = 1$. The meaning of $\sigma_F^{o^{(3)}}$ as measure of 'frequency localization' is as follows: the range of $\sigma_F^{o^{(3)}}$ is the interval $[0, \infty]$; a large value of $\sigma_F^{o^{(3)}}$ occurs if many Fourier coefficients contribute to $\sigma_F^{o^{(3)}}$. In conclusion, relating any spherical harmonic to a 'single wavelength', a large value $\sigma_F^{o^{(3)}}$ informs us that F is spread out widely in 'frequency domain'. In contrast, a small value $\sigma_F^{o^{(3)}}$ indicates that only a few number of Fourier coefficients is significant (see Table 5).

Again we reformulate our quantities in the specific context of zonal functions. Let $K(\cdot\varepsilon^3)$ be of class $\mathrm{H}^{(2)}(\Omega)$ satisfying $\|K(\cdot\varepsilon^3)\|_{\mathrm{L}^2(\Omega)} = 1$, then

$$\begin{aligned}\left(\sigma^{o^{(3)}}_{K(\cdot\varepsilon^3)}\right)^2 &= -\int_\Omega \Delta^*_\eta K\left(\eta\cdot\varepsilon^3\right) K\left(\eta\cdot\varepsilon^3\right)\,\mathrm{d}\omega(\eta)\\ &= -2\pi\int_{-1}^{1} K(t)\mathrm{L}_t K(t)\,\mathrm{d}t \end{aligned} \tag{28}$$

where L_t denotes the Legendre operator as given by $\mathrm{L}_t = \frac{\mathrm{d}}{\mathrm{d}t}(1-t^2)\frac{\mathrm{d}}{\mathrm{d}t}$.

Uncertainties and the uncertainty principle. The square roots of the variances, i.e., $\sigma^{o^{(1)}}$ and $\sigma^{o^{(3)}}$, are called the *uncertainties* in $o^{(1)}$ and $o^{(3)}$, respectively. For these quantities we get (see [43, 76]) an estimate given by $(\sigma^{o^{(1)}}_F)^2(\sigma^{o^{(3)}}_F)^2 \geq |g^{o^{(1)}}_F|^2$. We summarize our results in Theorem 1. For details on the proof the reader is referred to [29].

Theorem 1. *Let $F \in \mathrm{H}^{(2)}(\Omega)$ satisfy $\|F\|_{\mathrm{L}^2(\Omega)} = 1$. Then*

$$\left(\sigma^{o^{(1)}}_F\right)^2\left(\sigma^{o^{(3)}}_F\right)^2 \geq \left|g^{o^{(1)}}_F\right|^2. \tag{29}$$

If $g^{o^{(1)}}_F$ is non-vanishing, then

$$\Delta^{o^{(1)}}_F \Delta^{o^{(3)}}_F \geq 1, \tag{30}$$

where we have used the abbreviations

$$\Delta^{o^{(1)}}_F = \frac{\sigma^{o^{(1)}}_F}{\left|g^{o^{(1)}}_F\right|}, \quad \Delta^{o^{(3)}}_F = \sigma^{o^{(3)}}_F. \tag{31}$$

The *uncertainty relation* measures the tradeoff between 'space localization' and 'frequency localization' ('spread in frequency'). It states that *sharp localization in space and frequency are mutually exclusive.*

An immediate consequence of Theorem 1 is its reformulation for zonal functions $K(\varepsilon^3\cdot) : \eta \mapsto K(\varepsilon^3\cdot\eta)$, $\eta \in \Omega$.

Corollary 2. *Let $K(\varepsilon^3\cdot) \in \mathrm{H}^{(2)}(\Omega)$ satisfy $\|K\|_{\mathrm{L}^2[-1,1]} = 1$. If $t^{o^{(1)}}_K$ is non-vanishing, then*

$$\Delta^{o^{(1)}}_K \Delta^{o^{(3)}}_K \geq 1, \tag{32}$$

where

$$\Delta^{o^{(1)}}_K = \frac{\sigma^{o^{(1)}}_K}{t^{o^{(1)}}_K}, \quad \Delta^{o^{(3)}}_K = \sigma^{o^{(3)}}_K. \tag{33}$$

The interpretation of $(\sigma^{o^{(3)}}_K)^2$ as variance in 'total angular momentum' helped us to prove Theorem 1. However, this interpretation shows two essential drawbacks: first, the expectation of the surface curl gradient is a vector which seems to be inadequate in 'momentum localization' in terms of scalar spherical harmonics, and secondly the value of $g^{o^{(3)}}_F$ vanishes for all candidates F. This means that the

'center of gravity of the spherical window' in 'momentum domain' is independent of the function F under consideration. Therefore, we are finally interested in the variance of the operator $-\Delta^*$

$$\left(\sigma_F^{-\Delta^*}\right)^2 = \int_\Omega \left| \left((-\Delta_\eta^*) - g_F^{-\Delta^*} \right) F(\eta) \right|^2 \, \mathrm{d}\omega(\eta) \tag{34}$$

which is a measure for the 'spread in momentum'. Now the corresponding expectation value $g_F^{-\Delta^*}$ is scalar-valued and non-vanishing. It can be easily seen that

$$\left(\sigma_F^{-\Delta^*}\right)^2 = g_F^{(-\Delta^*)^2} - \left(g_F^{-\Delta^*}\right)^2. \tag{35}$$

In connection with Theorem 1 this leads to the following result.

Theorem 3. *Let F be of class $\mathrm{H}^{(4)}(\Omega)$ such that $\|F\|_{\mathrm{L}^2(\Omega)} = 1$. Then*

$$\left(\sigma_F^{o^{(1)}}\right)^2 \left(\sigma_F^{-\Delta^*}\right)^2 \geq \left| g_F^{o^{(1)}} \right| \, \frac{g_F^{(-\Delta^*)^2} - \left(g_F^{-\Delta^*}\right)^2}{g_F^{-\Delta^*}} \tag{36}$$

provided that $g_F^{-\Delta^*} \neq 0$. If the right-hand side of (36) *is non-vanishing, then*

$$\Delta_F^{o^{(1)}} \Delta_F^{-\Delta^*} \geq 1, \tag{37}$$

where

$$\Delta_F^{-\Delta^*} = \left(\frac{\left(\sigma_F^{-\Delta^*}\right)^2}{\frac{g_F^{(-\Delta^*)^2} - \left(g_F^{-\Delta^*}\right)^2}{g_F^{-\Delta^*}}} \right)^{1/2} = \left(g_F^{-\Delta^*}\right)^{1/2} = \Delta_F^{o^{(3)}}. \tag{38}$$

3.2. Classification of examples

We continue with some examples of particular interest for geoscientific research.

Localization of the spherical harmonics. We know that

$$\int_\Omega (Y_{n,k}(\xi))^2 \, \mathrm{d}\omega(\xi) = 1 \ . \tag{39}$$

One can prove that

$$g_{Y_{n,k}}^{o^{(1)}} = 0, \qquad \sigma_{Y_{n,k}}^{o^{(1)}} = 1. \tag{40}$$

Moreover, we have

$$g_{Y_{n,k}}^{-\Delta^*} = n(n+1), \qquad \sigma_{Y_{n,k}}^{-\Delta^*} = 0. \tag{41}$$

In other words, *spherical harmonics show an ideal frequency localization, but no space localization* (see Figure 5 for an illustration of space and frequency localization for the Legendre polynomials).

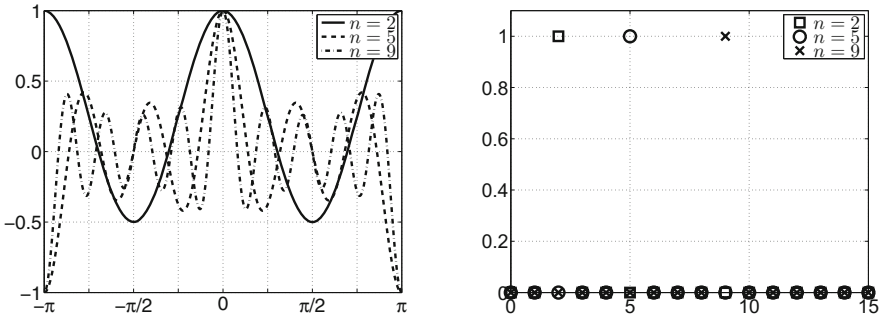


FIGURE 5. The Legendre kernel P_n for $n = 2, 5, 9$, space representation for $\vartheta \mapsto P_n(\cos(\vartheta))$ (left), and frequency representation $m \mapsto (P_n)^\wedge(m)$ (right).

Localization of the ideally bandlimited Legendre kernel. We have, with $P_n^* = \sqrt{\frac{2n+1}{4\pi}} P_n$,

$$\int_\Omega (P_n^*(\xi \cdot \zeta))^2 \, \mathrm{d}\omega(\zeta) = 1 \tag{42}$$

for all $\xi \in \Omega$, such that

$$g^{o^{(1)}}_{P_n^*(\xi\cdot)} = 0, \qquad \sigma^{o^{(1)}}_{P_n^*(\xi\cdot)} = 1, \tag{43}$$

$$g^{-\Delta^*}_{P_n^*(\xi\cdot)} = n(n+1), \quad \sigma^{-\Delta^*}_{P_n^*(\xi\cdot)} = 0. \tag{44}$$

Localization of the bandlimited Shannon kernel. The Shannon kernel Φ_ρ, $\rho > 0$, given by

$$\Phi_\rho(\xi \cdot \eta) = \sum_{n \le \rho^{-1}} \frac{2n+1}{4\pi} P_n(\xi \cdot \eta), \quad \xi, \eta \in \Omega, \tag{45}$$

may be interpreted as a truncated Dirac kernel. It is not surprising that the Shannon kernel as a 'finite polynomial kernel' shows strong oscillations in space. This is the price to be paid for the sharp separation in frequency space.

The investigation of the uncertainty properties of the Shannon kernel starts from (cf. [43])

$$\|\Phi_\rho\|^2_{\mathrm{L}^2(\Omega)} = \sum_{n=0}^{\lfloor \rho^{-1} \rfloor} \frac{2n+1}{4\pi} = \frac{1}{4\pi} \left((\lfloor \rho^{-1} \rfloor + 1) + \lfloor \rho^{-1} \rfloor \lfloor \rho^{-1} + 1 \rfloor \right), \tag{46}$$

where, as usual, $\lfloor \rho^{-1} \rfloor$ is the largest integer which is less or equal ρ^{-1}. Observing this result, we introduce the normalized Shannon kernel by

$$\tilde{\Phi}_\rho = \frac{1}{\|\Phi_\rho\|_{\mathrm{L}^2(\Omega)}} \Phi_\rho. \tag{47}$$

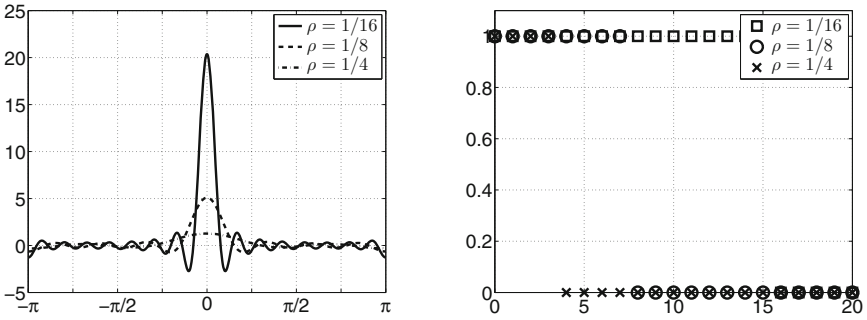


FIGURE 6. The Shannon scaling function Φ_ρ for $\rho = 1/16,\ 1/8,\ 1/4$. Space representation $\vartheta \mapsto \Phi_\rho(\cos(\vartheta))$ (left) and frequency representation $n \mapsto (\Phi_\rho)^\wedge(n)$ (right).

Its localization in space satisfies

$$\begin{aligned}\left(\sigma^{o(1)}_{\tilde{\Phi}_\rho}\right)^2 &= 1 - \frac{1}{\|\Phi_\rho\|^2}\left(\sum_{n=1}^{\lfloor \rho^{-1}-1\rfloor} \frac{2n+2}{4\pi}\right)^2 \\ &= 1 - \left(\frac{2\lfloor \rho^{-1}-1\rfloor + \lfloor \rho^{-1}\rfloor\lfloor \rho^{-1}-1\rfloor}{\lfloor \rho^{-1}+1\rfloor + \lfloor \rho^{-1}\rfloor\lfloor \rho^{-1}+1\rfloor}\right)^2, \end{aligned} \tag{48}$$

so that

$$\Delta^{o(1)}_{\tilde{\Phi}_\rho} = \sqrt{\frac{1 - \left(\frac{2\lfloor \rho^{-1}-1\rfloor + \lfloor \rho^{-1}\rfloor\lfloor \rho^{-1}-1\rfloor}{\lfloor \rho^{-1}+1\rfloor + \lfloor \rho^{-1}\rfloor\lfloor \rho^{-1}+1\rfloor}\right)^2}{\frac{2\lfloor \rho^{-1}-1\rfloor + \lfloor \rho^{-1}\rfloor\lfloor \rho^{-1}-1\rfloor}{\lfloor \rho^{-1}+1\rfloor + \lfloor \rho^{-1}\rfloor\lfloor \rho^{-1}+1\rfloor}}}. \tag{49}$$

Moreover, we find

$$\begin{aligned}\left(\sigma^{o(3)}_{\tilde{\Phi}_\rho}\right)^2 &= \frac{4\pi}{\lfloor \rho^{-1}\rfloor + 1 + \lfloor \rho^{-1}\rfloor\lfloor \rho^{-1}+1\rfloor} \sum_{n=0}^{\lfloor \rho^{-1}\rfloor} \frac{2n+1}{4\pi}\, n(n+1) \\ &= \frac{1}{2}\frac{\lfloor \rho^{-1}\rfloor(1+\lfloor \rho^{-1}\rfloor)^2(2+\lfloor \rho^{-1}\rfloor)}{\lfloor \rho^{-1}\rfloor + 1 + \lfloor \rho^{-1}\rfloor\lfloor \rho^{-1}+1\rfloor} \end{aligned} \tag{50}$$

such that

$$\Delta^{o(3)}_{\tilde{\Phi}_\rho} = \sqrt{\frac{1}{2}\frac{\lfloor \rho^{-1}\rfloor(1+\lfloor \rho^{-1}\rfloor)^2(2+\lfloor \rho^{-1}\rfloor)}{\lfloor \rho^{-1}\rfloor + 1 + \lfloor \rho^{-1}\rfloor\lfloor \rho^{-1}+1\rfloor}}. \tag{51}$$

The results are graphically illustrated in Figure 7.

Localization of the non-bandlimited/non-spacelimited Abel–Poisson kernel. Let us consider the function $Q_h : [-1,1] \to \mathbb{R}$, $h < 1$, given by

$$Q_h(t) = \frac{1}{4\pi}\frac{1-h^2}{(1+h^2-2ht)^{3/2}} = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi}\, h^n P_n(t). \tag{52}$$

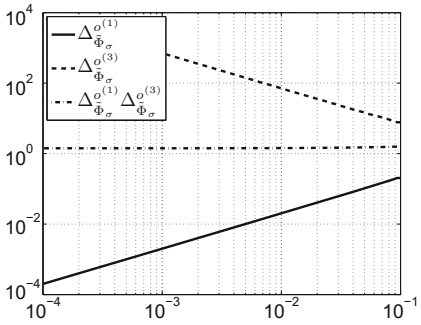


FIGURE 7. Uncertainty classification of the normalized Shannon Dirac family $\tilde{\Phi}_\rho$. Shown are $\Delta^{o^{(1)}}_{\tilde{\Phi}_\rho}$, $\Delta^{o^{(3)}}_{\tilde{\Phi}_\rho}$, and the product $\Delta^{o^{(1)}}_{\tilde{\Phi}_\rho}\Delta^{o^{(3)}}_{\tilde{\Phi}_\rho}$ as functions of ρ in a double logarithmic setting.

An easy calculation gives us

$$\|Q_h\|_{\mathrm{L}^2[-1,1]} = (Q_{h^2}(1))^{1/2} = \left(\frac{1+h^2}{4\pi}\right)^{1/2} \frac{1}{1-h^2}. \tag{53}$$

Furthermore, for $\tilde{Q}_h(t) = \|Q_h\|^{-1}_{\mathrm{L}^2[-1,1]}\ Q_h(t)$, $t \in [-1,1]$, we obtain after an elementary calculation (see also Figure 8)

$$\Delta^{o^{(1)}}_{\tilde{Q}_h} = \frac{1-h^2}{2h}, \quad \Delta^{-\Delta^*}_{\tilde{Q}_h} = \frac{\sqrt{6}h}{1-h^2}. \tag{54}$$

Thus, we finally obtain

$$\Delta^{o^{(1)}}_{\tilde{Q}_h}\Delta^{-\Delta^*}_{\tilde{Q}_h} = \frac{\sqrt{6}}{2} = \sqrt{\frac{3}{2}} > 1. \tag{55}$$

Here, the value $\Delta^{o^{(1)}}_{\tilde{Q}_h}\Delta^{-\Delta^*}_{\tilde{Q}_h}$ is independent of h. All intermediate cases of 'space-frequency localization' are realized by the Abel–Poisson kernel, but the Abel–Poisson kernel does not satisfy a minimum uncertainty state.

Localization of the spacelimited Haar kernel. Let k be a non-negative integer, i.e., $k \in \mathbb{N}_0$. The (smoothed) Haar kernel $\{B^{(k)}_h\}_{h\in(0,1)} \subset \mathrm{C}^{(k-1)}[-1,1]$ is defined by

$$B^{(k)}_h(t) = \begin{cases} 0 & , \quad t \in [-1,h) \\ \dfrac{(t-h)^k}{(1-h)^k} & , \quad t \in [h,1]. \end{cases} \tag{56}$$

By definition, $B^{(k)}_h$ is non-negative and has the support $[h,1]$. Obviously, the function $B^{(0)}_h$, $h \in (-1,1)$, represents the (classical) Haar function (cf. [53]). The

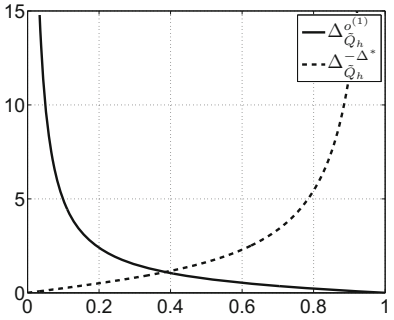


FIGURE 8. Abel–Poisson kernel uncertainty classification. The curves graphically illustrate the functions $h \mapsto \Delta^{o^{(1)}}_{\tilde{Q}_h}$ and $h \mapsto \Delta^{-\Delta^*}_{\tilde{Q}_h}$.

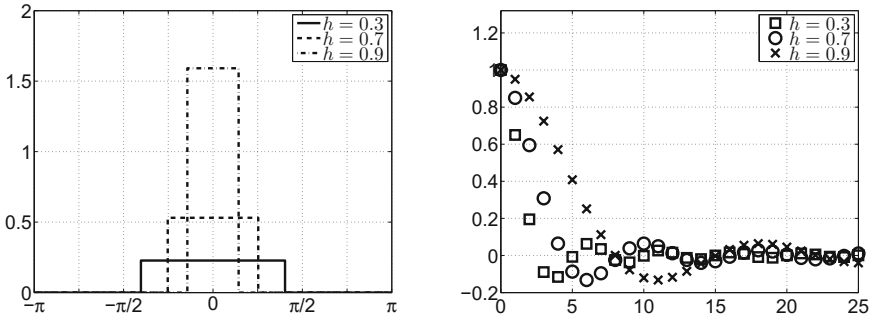


FIGURE 9. The Haar kernel$B^{(0)}_h$ for $h = 0.3,\ 0.7,\ 0.9$. Space representation $\vartheta \mapsto B^{(0)}_h(\cos(\vartheta)), \vartheta \in [-\pi, \pi]$, (left) and frequency representation $n \mapsto (B^{(0)}_h)^\wedge(n)$ (right).

Legendre coefficients of $B^{(k)}_h$, $h \in (-1, 1)$, $k \in \mathbb{N}_0$, can be calculated recursively (cf. [47]):

$$\left(B^{(k)}_h\right)^\wedge(0) = 2\pi \frac{1-h}{k+1} \neq 0, \tag{57}$$

$$\left(B^{(k)}_h\right)^\wedge(1) = 2\pi \frac{1-h}{k+1}\left(1 - \frac{1-h}{k+2}\right), \tag{58}$$

$$\left(B^{(k)}_h\right)^\wedge(n+1) = \frac{2n+1}{n+k+2}\, h\, \left(B^{(k)}_h\right)^\wedge(n) + \frac{k+1-n}{n+k+2}\, \left(B^{(k)}_h\right)^\wedge(n-1). \tag{59}$$

An elementary calculation shows

$$\begin{aligned} \left\| B^{(k)}_h \right\|^2_{\mathrm{L}^2(\Omega)} &= 2\pi \int_{-1}^{1} \left[B^{(k)}_h(t)\right]^2 \mathrm{d}t \\ &= 2\pi \frac{1-h}{2k+1}. \end{aligned} \tag{60}$$

We define the kernel

$$\tilde{B}_h^{(k)} = \sqrt{\frac{2k+1}{2\pi(1-h)}}\, B_h^{(k)}, \tag{61}$$

since the uncertainty properties are normally defined for kernels with norm one. We find

$$g^{o^{(1)}}_{\tilde{B}_h^{(k)}(\,\cdot\varepsilon^3)} = 2\pi \int_{-1}^{1} t \left(\tilde{B}_h^{(k)}(t)\right)^2 \mathrm{d}t\, \varepsilon^3 = \frac{1+h+2k}{2+2k}\, \varepsilon^3. \tag{62}$$

Consequently,

$$\left(\sigma^{o^{(1)}}_{\tilde{B}_h^{(k)}}\right)^2 = 1 - \left(\frac{1+h+2k}{2+2k}\right)^2 = \frac{(1-h)(h+4k+3)}{(2k+2)^2}. \tag{63}$$

Using (31), we finally arrive at

$$\Delta^{o^{(1)}}_{\tilde{B}_h^{(k)}} = \frac{1}{1+h+2k}\sqrt{(1-h)(h+4k+3)}. \tag{64}$$

For the localization in frequency, we assume $k \geq 2$. We have

$$\begin{aligned}\left(\sigma^{o^{(3)}}_{\tilde{B}_h^{(k)}(\,\cdot\varepsilon^3)}\right)^2 &= -2\pi \int_{-1}^{1} \tilde{B}_h^{(k)}(t)\, \mathrm{L}_t \tilde{B}_h^{(k)}(t)\, \mathrm{d}t \\ &= \frac{2k+1}{2\pi(1-h)} \frac{-2\pi}{(1-h)^{2k}} \int_h^1 (t-h)^k \mathrm{L}_t (t-h)^k\, \mathrm{d}t \\ &= \frac{k(h+2k)}{(1-h)(2k-1)},\end{aligned} \tag{65}$$

so that

$$\Delta^{o^{(3)}}_{\tilde{B}_h^{(k)}} = \sqrt{\frac{k(h+2k)}{(1-h)(2k-1)}}. \tag{66}$$

The application of L_t requires that the kernel is twice differentiable. However, using integration by parts, the results immediately carry over to the case $k = 1$. Figure 10 gives a graphical impression of these results for the particular cases $k = 1$ and $k = 3$.

Localization of the ideally spacelimited Dirac kernel. Letting h formally tend to 1 in the results provided by the uncertainty principle for the Abel–Poisson kernel function we are able to interpret the localization properties of the Dirac kernel on Ω satisfying $\delta^\wedge(n) = 1$ for all $n \in \mathbb{N}_0$:

$$\delta(\xi \cdot \eta) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} P_n(\xi \cdot \eta), \quad \xi, \eta \in \Omega, \tag{67}$$

where the convergence is understood in distributional sense. As a matter of fact, letting h tend to 1 shows us that the variances in the space domain take the constant value 0. On the other hand, the variances in the frequency domain converge to ∞. Hence, the Dirac kernel shows ideal space localization, but no frequency localization.

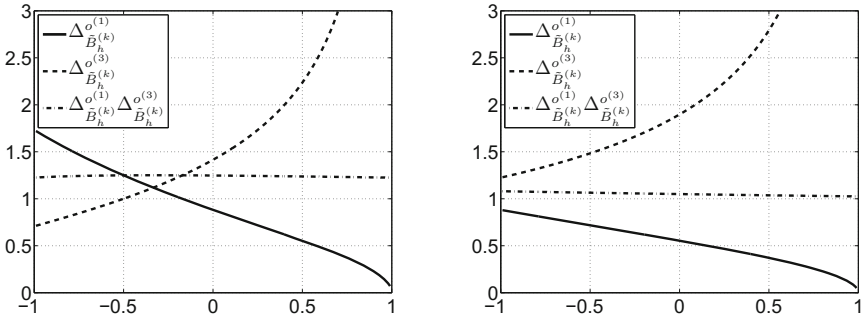


FIGURE 10. Uncertainty classification of the normalized smoothed Haar scaling function $\tilde{B}^{(k)}_h$ ($k = 1$, left; $k = 3$ right). $\Delta^{o^{(1)}}_{\tilde{B}^{(k)}_h}$, $\Delta^{o^{(3)}}_{\tilde{B}^{(k)}_h}$ and the product $\Delta^{o^{(1)}}_{\tilde{B}^{(k)}_h}\Delta^{o^{(3)}}_{\tilde{B}^{(k)}_h}$ are shown as functions of h.

Localization of the non-bandlimited/non-spacelimited Gaussian function. The minimum uncertainty state within the uncertainty relation is provided by the Gaussian probability density function (see [47, 59]). Consider the function G_λ given by

$$G_\lambda(t) = e^{-(\lambda/2)(1-t)}, \qquad t \in [-1, 1], \quad \lambda > 0. \tag{68}$$

An elementary calculation shows us that

$$\tilde{G}_\lambda(t) = \gamma(\lambda) e^{-(\lambda/2)(1-t)}, \tag{69}$$

with

$$\gamma(\lambda) = \left(1/\sqrt{4\pi}\right)\left(\frac{1}{2\lambda}\left(1 - e^{-2\lambda}\right)\right)^{-1/2}, \tag{70}$$

satisfies $\|\tilde{G}_\lambda\|_{\mathrm{L}^2[-1,1]} = 1$. It is not difficult to deduce that $\Delta^{o^{(1)}}_{\tilde{G}_\lambda}\Delta^{-\Delta^*}_{\tilde{G}_\lambda} \to 1$ as $\lambda \to \infty$: the best value of the uncertainty principle (Theorem 3) is 1.

Localization of Slepian functions. The bandlimited Slepian functions solve the concentration criterion (10) on general domains $C \subset \Omega$. If we restrict our attention to spherical caps as in Figures 1 and 4, the solutions degenerate and equations (11) and (12) can be solved for fixed spherical-harmonic orders j, with twice-repeated eigenvalues for the nonzonal functions at the same nonzero absolute orders.

While the Slepian functions do not formally optimize the uncertainty relation (38), calculations by Wieczorek and Simons [104] reveal that, again on spherical caps C of various opening half-angles Θ, the values attained by the largest-eigenvalue ($\lambda^{(\alpha)}_C \approx 1$ for $\alpha = 1$) zonal Slepian functions of varying bandwidths N are very close to satisfying the bounds (38) for Shannon numbers $N_0 = (N+1)\Theta/\pi$ (see (16)) greater than about 2. Furthermore, for increasing Shannon numbers, the uncertainty products for the αth best-concentrated Slepian function, when $N_0 \geq \alpha + 1$, tend to $2\alpha - 1$. This favorable behavior was illustrated by Wieczorek

and Simons [104], see their Figures 5 and 6b. Subsequent work by, among others, Guilloux et al. [52] and Khalid et al. [57], has substantiated and elaborated on these early analyses.

Slepian functions vs. the Gaussian. Another way by which the spatiospectral localization properties of the Slepian functions may be appreciated is by comparing how close they are to the family of minimum-uncertainty 'squeezed' *coherent states* (e.g., [15, 58]), a common root for many later developments in spline, Slepian function, and wavelet analysis [20]. This is of importance because in practical problems in the geosciences (e.g., [17, 84]), as in cosmology (e.g., [94]), we place as much value on the precise bandwidth, or bandwidth resolution, of our observations as on the spatial domain of interest. The Gaussian (68) may satisfy the uncertainty lower bound exactly, but it is not a bandlimited kernel. In contrast, the Slepian functions (11–12) can be bandlimited and spaceconcentrated at the same time. Formally, they are the optimizers of (10), though not of (38).

That they get close is shown in Figure 11. Inspired by Bluhm et al. [14] we determine the squeeze factor, s that renders the suitably normalized function

$$G_s(\cos\theta) = \gamma(s) e^{s\cos\theta}, \quad 0 \le \theta \le \pi, \tag{71}$$

as close as possible, in the mean-squared sense, to the best-concentrated bandlimited zonal Slepian function, concentrated to a spherical cap of a certain radius Θ, and whereby the tradeoff between spatial (the area of the spherical cap) and spectral concentration (the bandwidth N) is parameterized via the partial Shannon number $N_0 = (N+1)\Theta/\pi$.

3.3. Closing remarks

The uncertainty principle represents a trade-off between two 'spreads', in position and in frequency. *Sharp localization in space and in frequency are mutually exclusive.* The reason for the validity of the uncertainty relation (Theorem 1) is that the normal and curl operators $o^{(1)}$ and $o^{(3)}$ do not commute, hence, they cannot be sharply defined simultaneously. Extremal members of the uncertainty relation are polynomials (spherical harmonics) and Dirac function(al)s. An asymptotically optimal kernel is the Gaussian function.

Corollary 2 allows a quantitative classification and a hierarchy of the space and frequency localization properties of kernel functions of the form

$$K(t) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} K^{\wedge}(n) P_n(t), \qquad t = \xi \cdot \eta, \quad (\xi, \eta) \in \Omega \times \Omega. \tag{72}$$

In view of their space/frequency localization, it is also important to distinguish bandlimited kernels (i.e., $K^{\wedge}(n) = 0$ for all $n \ge N \in \mathbb{N}_0$) and non-bandlimited ones ($K^{\wedge}(n) \ne 0$ for an infinite number of integers n). Non-bandlimited kernels show a much stronger space localization than their bandlimited counterparts. It is

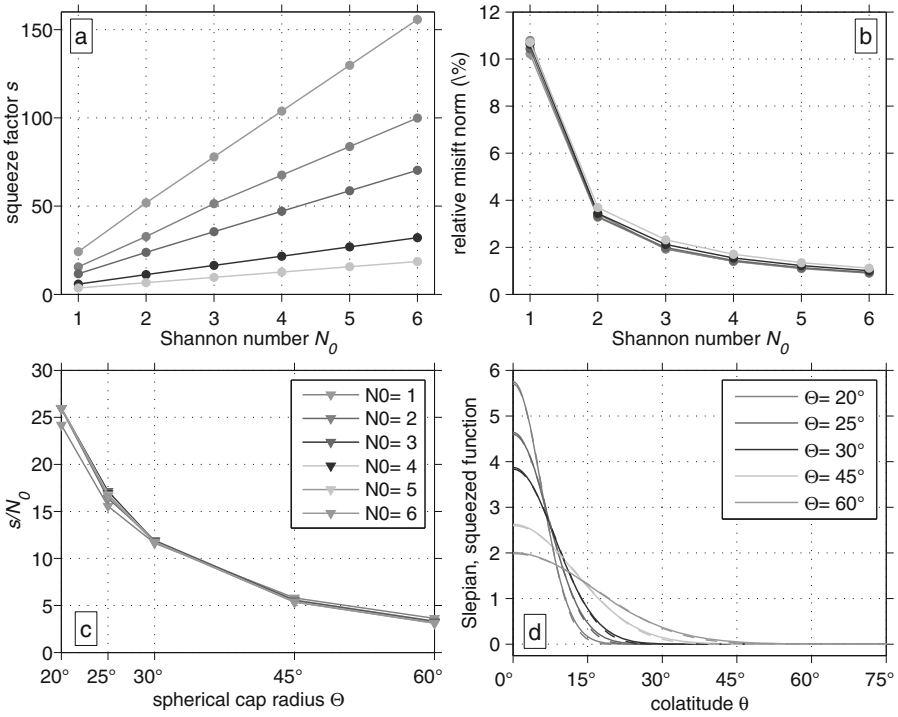


FIGURE 11. Slepian functions compared to Gaussian 'squeezed coherent states'. The chosen squeeze factors render functions of the type (71) as close as possible to the best-concentrated zonal Slepian functions in the relative mean-squared sense, for a variety of spherical cap sizes Θ and Slepian-function bandwidths N, linked through the Shannon number N_0.

not difficult to prove that, if $K \in \mathrm{L}^2[-1,1]$ with $\|K(\xi\cdot\,)\|_{\mathrm{L}^2(\Omega)} = 1$,

$$\left(\sigma^{o\,(1)}_{K(\xi\cdot\,)}\right)^2 = 1 - \left(\sum_{n=1}^{\infty} \frac{2n+1}{4\pi}\, K^\wedge(n) K^\wedge(n+1)\right)^2. \tag{73}$$

If $K^\wedge(n) \approx K^\wedge(n+1) \approx 1$ for many successive integers n, the space-domain support of $K(t)$ in (72) is small.

Space/frequency localization on the sphere can also be illustrated directly from (72). Choosing $K^\wedge(n) = \delta_{nk}$ we obtain a Legendre kernel of degree k, on the left in our scheme (Table 6). Setting $K^\wedge(n) = 1$ for $n = 0, 1, \ldots$, we obtain the Dirac kernel. The slower the sequence $\{K^\wedge(n)\}_{n=0,1,\ldots}$ converges to zero, the lower the frequency localization, but the higher the space localization.

Altogether, Table 6 gives a qualitative illustration of the consequences of the uncertainty principle in the theory of zonal kernel functions on the sphere: on the

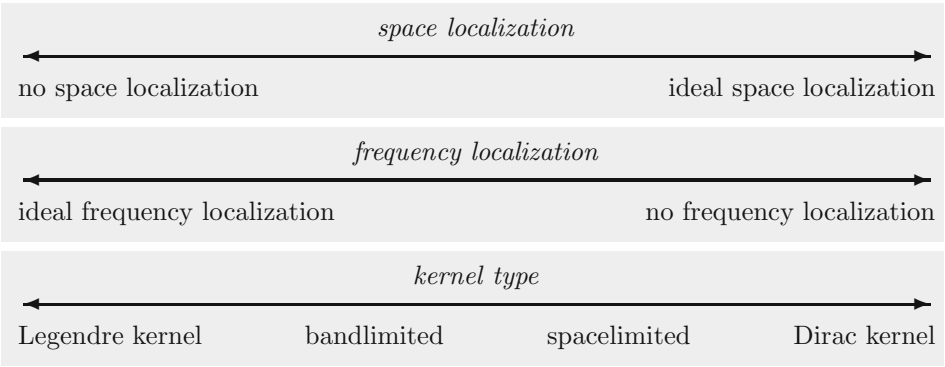


TABLE 6. The uncertainty principle and its consequences.

left end of this scheme, we have the Legendre kernels with their ideal frequency (momentum) localization. However, they show no space localization, as they are of polynomial nature. Thus, the present standard way in applications of increasing the accuracy in spherical harmonic (Fourier) expansions is to increase the maximum degree of the spherical harmonics expansions under consideration. On the right end of the scheme, there is the Dirac kernel which maps a function to its value at a certain point. Hence, this (generalized) function has an ideal space localization but no frequency localization. Consequently, it can be used in a finite pointset approximation.

4. Constructive approximation on the sphere

In Section 4.1, we discuss an approach using Slepian functions, Section 4.2 is an approach based on splines, and Section 4.3 treats the case of wavelets. Section 4.4 helps combine benefits of various approaches.

4.1. Approximation by Slepian functions

Given a certain region of interest C on the unit sphere Ω and a certain bandwidth N (a limiting spherical-harmonic degree in the sense of (9)), optimization of a *concentration criterion* yields linear combinations of spherical harmonics that we call *Slepian functions*. In Section 2.3, we gave their formulation in terms of band*limited* functions that are space*concentrated*. We shall denote these functions from now on as $G_N^C(\xi)$. Of course, we can equally well ask for space*limited* functions that are band*concentrated* – see [84, 87] for details. We shall denote those functions from now on as $H_C^N(\xi)$. The Fourier coefficients of the H_C^N can be calculated from those of the G_N^C by extension as in (13). We refer to [60] for an extensive discussion on the properties of what are, essentially, cases *intermediate* between these two endmembers, for functions defined on the real line.

If we introduce the space of all square-integrable scalar spherical functions that are spacelimited to the region C as $\mathcal{S}_C$, and the space of all square-integrable spherical functions that are bandlimited to the spherical-harmonic degree N as $\mathcal{S}_N$, then it is implied that $H_C^N \in \mathcal{S}_C$ and $G_N^C \in \mathcal{S}_N$.

Reproducing properties. We can show that the spectral-domain kernel that we first encountered in bandlimited form in (11), and which we now extend to $0 \leq m, n < \infty$,

$$d^C_{(m,j),(n,k)} = \langle Y_{m,j}, Y_{n,k} \rangle_{\mathrm{L}^2(C)} = \int_C Y_{m,j}(\xi) Y_{n,k}(\xi) \, \mathrm{d}\omega(\xi), \tag{74}$$

is a reproducing kernel in the space $\mathcal{S}_C$. Indeed, for any function $F \in \mathcal{S}_C$,

$$\begin{aligned} \sum_{n=0}^{\infty} \sum_{k=1}^{2n+1} d^C_{(m,j),(n,k)} F^{\wedge}(n,k) &= \int_C Y_{m,j}(\xi) \left(\sum_{n=0}^{\infty} \sum_{k=1}^{2n+1} Y_{n,k}(\xi) F^{\wedge}(n,k) \right) \mathrm{d}\omega(\xi) \\ &= \int_C Y_{m,j}(\xi) F(\xi) \, \mathrm{d}\omega(\xi) \\ &= F^{\wedge}(m,j). \end{aligned} \tag{75}$$

At the same time, the spatial-domain Shannon kernel that we encountered in (12), and which we rebaptize

$$D_N(\xi, \eta) = \sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(\xi \cdot \eta) = \sum_{n=0}^{N} \sum_{k=1}^{2n+1} Y_{n,k}(\xi) Y_{n,k}(\eta), \tag{76}$$

is a reproducing kernel in the space $\mathcal{S}_N$, since, for any function $F \in \mathcal{S}_N$,

$$\begin{aligned} &\int_{\Omega} \sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(\xi \cdot \eta) F(\eta) \, \mathrm{d}\omega(\eta) \\ &\quad = \sum_{n=0}^{N} \sum_{k=1}^{2n+1} Y_{n,k}(\xi) \int_{\Omega} Y_{n,k}(\eta) F(\eta) \, \mathrm{d}\omega(\eta) \\ &\quad = \sum_{n=0}^{N} \sum_{k=1}^{2n+1} Y_{n,k}(\xi) F^{\wedge}(n,k) \\ &\quad = F(\xi). \end{aligned} \tag{77}$$

Equations (75) and (77) hold the key to the approximation properties of the Slepian functions, since they imply that the spacelimited Slepian functions H_C^N provide a complete basis for all spacelimited functions in $\mathcal{S}_C \subset \mathrm{L}^2(\Omega)$, whereas the bandlimited Slepian functions G_N^C are a complete basis for all bandlimited functions in $\mathcal{S}_N \subset \mathrm{L}^2(\Omega)$.

Mercer's theorem. A second set of properties that solidifies these notions is established through an identity known as *Mercer's theorem*, which in this context takes the form

$$D_N(\xi,\eta) = \sum_{\alpha=1}^{(N+1)^2} \left(G_N^C\right)_\alpha(\xi)\left(G_N^C\right)_\alpha(\eta), \tag{78}$$

for all the α-indexed bandlimited Slepian functions G_N^C, with eigenvalues $\lambda_C^{(\alpha)}$, from which we establish, using (76) and as in (14d), that

$$D_N(\xi,\xi) = \sum_{\alpha=1}^{(N+1)^2} \left(G_N^C\right)_\alpha^2(\xi) = \frac{(N+1)^2}{4\pi} = \frac{N_C}{\int_C \mathrm{d}\omega(\eta)}, \tag{79}$$

recovering the spherical Shannon number N_C and the area of the domain of interest, $\int_C \mathrm{d}\omega(\eta)$.

A useful corollary is that the eigenvalue-weighted sum of squares of the bandlimited Slepian eigenfunctions closely *approximates* the value $N_C/\int_C \mathrm{d}\omega(\eta)$ when $\xi \in C$, and vanishes otherwise,

$$\sum_{\alpha=1}^{(N+1)^2} \lambda_C^{(\alpha)}\left(G_N^C\right)_\alpha^2 \approx \sum_{\alpha=1}^{N_C} \lambda_C^{(\alpha)}\left(G_N^C\right)_\alpha^2 \approx \begin{cases} N_C/\int_C \mathrm{d}\omega(\eta) & \text{if } \xi \in C \\ 0 & \text{otherwise,} \end{cases}$$

which is a consequence of the step-shaped eigenvalue spectrum that we saw in Figure 1. Eq. (80) testifies to the fact that the *effective* dimension of the space $\mathcal{S}_N$ of bandlimited functions that are *also* spaceconcentrated to C, is reduced from the canonical $(N+1)^2$ to the Shannon number N_C. It is our first clue to the approximation qualities of the Slepian functions, e.g., for (linear) signal estimation from regionally available data [85].

Power spectrum. If we furthermore define the power spectrum or degree variance of the bandlimited Slepian functions as

$$\mathscr{P}_n^{(\alpha)} = \frac{1}{2n+1}\sum_{k=1}^{2n+1}\left|\left(G_N^C\right)_\alpha^\wedge(n,k)\right|^2 \tag{80}$$

we get, via the spectral theorem, equation (74), and the addition theorem a spectral-domain equation equivalent to (79)–(80), namely,

$$\begin{aligned}\sum_{\alpha=1}^{(N+1)^2} \lambda_C^{(\alpha)}\mathscr{P}_n^{(\alpha)} &= \frac{1}{2n+1}\sum_{k=1}^{2n+1} d^C_{(n,k),(n,k)} = \frac{\int_C \mathrm{d}\omega(\eta)}{4\pi} \\ &\approx \sum_{\alpha=1}^{N_C} \lambda_C^{(\alpha)}\mathscr{P}_n^{(\alpha)},\end{aligned} \tag{81}$$

which is suggestive of the *spectral*-domain approximation properties of the Slepian functions, as arises, e.g., in the theory of (quadratic) power-spectral estimation from regionally available data [17].

Equations (80) and (81) together, show that the set of $N_C < (N+1)^2$ Slepian functions provide essentially uniform coverage over the spatial domain C and spectral bandwidth N. This is of interest when estimating (interpolating, approximating) functions from observations, as is common to a large number of research fields, not limited to the geosciences.

Alternative Mercer theorem. We note for completeness, and since the relevant identities have not been published before, that an alternative version of Mercer's theoremwould have transformed (12) and (76) from

$$\int_C D_N(\xi,\eta)F(\eta)\,\mathrm{d}\omega(\eta) = \lambda_C(F)F(\xi) \tag{82}$$

into the full-domain

$$\int_\Omega D^N(\xi,\eta)F(\eta)\,\mathrm{d}\omega(\eta) = \lambda_C(F)F(\xi), \tag{83}$$

which have the same eigenfunctions, but where we have defined

$$D^N(\xi,\eta) = \sum_{m=0}^{N}\sum_{j=1}^{2m+1}\sum_{n=0}^{N}\sum_{k=1}^{2n+1} d^C_{(m,j),(n,k)}Y_{m,j}(\xi)Y_{n,k}(\eta). \tag{84}$$

In that case, the equivalent to (78) is the to some more familiar expression

$$D^N(\xi,\eta) = \sum_{\alpha=1}^{(N+1)^2} \lambda_C^{(\alpha)} \left(G_N^C\right)_\alpha(\xi)\left(G_N^C\right)_\alpha(\eta). \tag{85}$$

Approximation 0: Noiseless data (interpolation). Imagine a certain function is 'known' as a spherical-harmonic expansion. Clearly, considering such a situation is merely postponing the problem of how to estimate an *unknown* function from observations. However, it is a common occurrence in the geosciences that, for example, space agencies perform exhaustive satellite data reductions that end up in the official release of spherical harmonic 'models' (typically of gravity or magnetic fields) that are then available for further research [103]. Another situation is where spectral forward-modeling codes deliver 'simulations' that are subsequently in need of interpretation and evaluation [102].

Whatever the source, and however large the bandlimit, the key property of the Slepian function basis is that the function expansion coefficients can be obtained by a simple transformation. If indeed the known function is F, then it is immaterial whether it is expressed in the spherical-harmonic basis, or in a bandlimited Slepian basis designed for *whichever* region C of interest, *as long as its bandwidth N matches the original*:

$$F = \sum_{n=0}^{N}\sum_{k=1}^{2n+1} F^\wedge(n,k)Y_{n,k} = \sum_{\alpha=1}^{(N+1)^2} F^\wedge(\alpha)\left(G_N^C\right)_\alpha. \tag{86}$$

The Slepian-function expansion coefficients $F^\wedge(\alpha)$, $\alpha = 1,\ldots,(N+1)^2$ are simply obtained from the spherical-harmonic expansion coefficients $F^\wedge(n,k)$,

$n = 0, 1, \ldots, N$ and order indices $k = 1, \ldots, 2n+1$, by the (orthogonal) transformation [85]

$$F^{\wedge}(\alpha) = \sum_{n=0}^{N} \sum_{k=1}^{2n+1} \left(G_N^C\right)_{\alpha}^{\wedge}(n,k) F^{\wedge}(n,k). \tag{87}$$

A linear basis transformation (87) is exact and thus, strictly speaking 'uninteresting'. However, the properties of the Slepian functions designed for a region C are such that after a *partial* Slepian expansion to $J < (N+1)^2$ terms, denoted $F_J(\xi)$, equation (86) will hold *approximately* in the region of interest:

$$F(\xi) = \sum_{n=0}^{N} \sum_{k=1}^{2n+1} F^{\wedge}(n,k) Y_{n,k}(\xi) \approx \sum_{\alpha=1}^{J} F^{\wedge}(\alpha) \left(G_N^C\right)_{\alpha}(\xi), \quad \xi \in C. \tag{88}$$

Clearly, a truncation of the spherical-harmonic series to its first J terms, however ordered, would generally result in poor approximations, *precisely* because of the non-localized spatial behavior of the basis functions. The eigenvalue-ranked Slepian transformation (87), on the other hand, has reordered the basis such that its first J functions increasingly uniformly 'cover' the spatial region of interest while providing an increasingly complete coverage over the entire spectral band, see (80) and (81). As a measure of approximation quality we take the area-weighted relative mean-squared error. It can be easily shown to depend on the truncation level in the manner

$$\frac{\|F - F_J\|^2_{\mathrm{L}^2(C)}}{\|F\|^2_{\mathrm{L}^2(C)}} = \frac{\sum_{\alpha>J}^{(N+1)^2} \left(F^{\wedge}(\alpha)\right)^2 \lambda_C^{(\alpha)}}{\sum_{\alpha=0}^{(N+1)^2} \left(F^{\wedge}(\alpha)\right)^2 \lambda_C^{(\alpha)}}. \tag{89}$$

Given the universally favorable decay of the eigenvalue spectrum of the spatio-spectral concentration problem (11), in this noiseless case, the Shannon number N_C is an obvious practical first choice for the truncation level J, although (89) of course shows the role played by the spectrum of the signal itself. An illustrative numerical example is given by Simons et al. [88], their Figure 3.

Approximation 1: Noisy data. We finally turn to the approximation problem that is most familiar in geophysical inverse theory, namely that of the *estimation* of a certain unknown signal from noisily observed data. We will briefly discuss the traditional spherical-harmonics based approach, and then clarify the beneficial role that localized basis functions (here: *Slepian functions*) may play in this context. We adhere to the continuous viewpoint for notational convenience and to lay bare the structure of the solutions. In practice, all datasets will be sample values at discrete geographic locations. As a consequence, the properties derived for constructive approximation by Slepian functions will themselves hold only approximately – to the degree by which continuous integrals are (hopefully, well) approximated by their Riemann sums [17, 84].

However, therein lies the crux of the Slepian-function method: if the data are regionally (in some region C) and densely (warranting a certain 'Nyquist' bandlimit at spherical-harmonic degree N) available, computing the Slepian basis for the idealized acquisition geometry *ahead of time* is what will lead to manageably sized inverse problems (on the order of the Shannon number N_C, and $N_C \ll (N+1)^2$ when $|C| \ll |\Omega|$) that solve for the unknown signal from which we assume the data to have been sampled.

Such a viewpoint, in a sense, embodies a strict *geographical prior*, and is very different from the splines and wavelets that will be discussed in the remaining Sections 4.2 and 4.3. Indeed, in contrast to Slepian functions, splines and wavelets made from zonal kernel functions do not *strictly* select for *particular* regions of interest, although of course, when particular combinations of any of those constructions are sought by optimization, as they are in Section 4.4, *effectively*, they do. Simons et al. [89] discuss a hybrid situation termed 'Slepian trees', as well as an alternative spherical wavelet transform obtained via a simple 'cubed-sphere' mapping of the 'usual' separable Cartesian discrete wavelet transforms [20].

The most detailed and up-to-date discussion of approximation by Slepian functions (both scalar and vector-valued, and for geomathematics problems involving measurements made by satellites at altitude) is found in the works by Simons and Plattner [80, 81, 86]. From these references, we retain and present a few essential points.

Suppose that we have 'data', M, consisting of a superposition of 'signal', F, and 'noise', E. What is F? The measurements are only available over some closed region C of the unit sphere Ω, i.e.,

$$M(\xi) = \begin{cases} F(\xi) + E(\xi) & \text{if } \xi \in C \\ \text{unknown/undesired} & \text{if } \xi \in \Omega \setminus C. \end{cases} \tag{90}$$

We assume that both signal and noise can be represented via an infinite spherical harmonic expansion as in (4), and we furthermore assume that they are uncorrelated realizations of zero-mean Gaussian random processes. Paying no heed to the structure of the noise (i.e., without explicit prior information that could be weighted into the norms in the form of a noise covariance) we elect to seek solutions to the optimization problem that results in a regularized bandlimited (to N, which remains to be determined) *estimate* of the signal, $\hat{F}$, in the form of equation (9), and which solves

$$\|\hat{F} - M\|^2_{\mathrm{L}^2(C)} + \lambda \|\hat{F}\|^2_{\mathrm{L}^2(\Omega\setminus C)} = \text{minimum}, \tag{91}$$

where $\lambda \geq 0$ is a regularization (damping) parameter forcing the solution to vanish outside of the observation domain. In the following two paragraphs, we distinguish solutions $\hat{F}_N$ and $\hat{F}_J$, both bandlimited.

Approximation 2: Regularized spherical-harmonic expansions. Simons and Dahlen [85] give the Fourier coefficients that solve equation (91) as

$$\hat{F}_N^{\wedge}(m,j) = \sum_{n=0}^{N} \sum_{k=1}^{2n+1} \left(d^{C}_{(m,j),(n,k)} + \lambda\, d^{(\Omega\setminus C)}_{(m,j),(n,k)} \right)^{-1} \langle M, Y_{n,k} \rangle_{\mathrm{L}^2(C)} \,. \tag{92}$$

We note from equation (74) that $d^{C}_{(m,j),(n,k)} + d^{(\Omega\setminus C)}_{(m,j),(n,k)}$ is the identity matrix. Regularization is unavoidable: as we have seen, the eigenvalues of $d^{C}_{(m,j),(n,k)}$ trail off quickly to nearly zero, see Figure 1. Restricted-region data availability is the prime reason for our inverse problem to be *ill posed* – even if no downward continuation from satellite height is required and if no internal density distributions (in the case of gravimetry) are being sought.

How well are we doing when accepting (92) as our solution? Rewriting the inverse Slepian eigenvalues $\lambda_C^{(\alpha)}$ with the damping parameter λ as

$$\left(\lambda_C^{(\alpha)}\right)^{*}(\lambda) = \left[\lambda_C^{(\alpha)} + \lambda\left(1-\lambda_C^{(\alpha)}\right)\right]^{-1}, \tag{93}$$

[85] derive the regional relative mean-squared error, the *expected value* of the ratio of approximation-error to signal norms as

$$\frac{E\left\{\left\|\hat{F}_N - F\right\|^2_{\mathrm{L}^2(C)}\right\}}{E\left\{\|F\|^2_{\mathrm{L}^2(C)}\right\}} = \sum_{\alpha=1}^{(N+1)^2} \frac{\lambda_C^{(\alpha)}}{N_C} \left[\left(\lambda_C^{(\alpha)}\right)^{*}(\lambda)\right]^2 \left[\mathscr{R}^{-1}\lambda_C^{(\alpha)} + \lambda^2\left(1-\lambda_C^{(\alpha)}\right)^2\right]. \tag{94}$$

In the expression above, both signal and noise were assumed to be characterized by a white (flat) power spectrum (defined in (80)), and we introduced $\mathscr{R}$, the signal-to-noise ratio. Valid only for this admittedly idealized case, (94) nevertheless contains all the elements by which the quality of the approximation can be appreciated: the bandwidth N and the size and shape of the region C enter through the eigenvalues $\lambda_C^{(\alpha)}$ and the Shannon number N_C, and of course the dependence on the signal-to-noise ratio $\mathscr{R}$ and the damping parameter λ are important controlling factors. Minimization of the relative error norm provides an implicit criterion for the regularization parameter:

$$\lambda_{\mathrm{opt}} = \mathscr{R}^{-1} \frac{\sum_{\alpha=1}^{(N+1)^2} \left[\left(\lambda_C^{(\alpha)}\right)^{*}(\lambda)\right]^3 \left(\lambda_C^{(\alpha)}\right)^2 \left(1-\lambda_C^{(\alpha)}\right)}{\sum_{\alpha=1}^{(N+1)^2} \left[\left(\lambda_C^{(\alpha)}\right)^{*}(\lambda)\right]^3 \left(\lambda_C^{(\alpha)}\right)^2 \left(1-\lambda_C^{(\alpha)}\right)^2}. \tag{95}$$

At high signal-to-noise ratios, (95) is well approximated by $\lambda_{\mathrm{opt}} \approx \mathscr{R}^{-1}$.

Approximation 3: Truncated Slepian expansions. Where did the Slepian functions go? We solved (91) using spherical harmonics, but we discussed the statistics of the solution (92) in terms of the eigenvalues of the Slepian concentration problem. The link, of course, is that the spherical-harmonic solution is derived via the

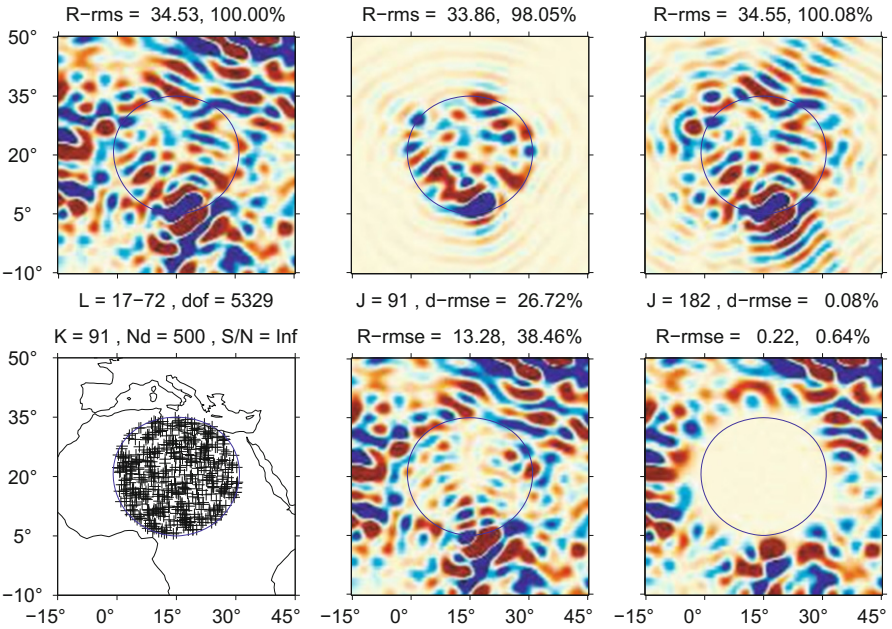


FIGURE 12. Example of Slepian-basis (Shannon number K = 91) approximation of a non-white bandlimited (bandpass, spherical-harmonic degrees L = 17 − 72) geomagnetic field from Nd = 500 noiseless data, for two truncation levels, J = 91 and J = 182, over a circular domain R. Top: the field, and the two reconstructions. Bottom: the location of the data points, and the difference between the truth and the approximation. The relative regional root-mean square signal, reconstruction and error strengths are indicated.

intermediary of the inverse of the Slepian localization matrix $d^{C}_{(m,j),(n,k)}$ and, with regularization, its complement, $d^{(\Omega\setminus C)}_{(m,j),(n,k)}$. Both of these are large, full (though banded) matrices whose inverses (especially at large spherical-harmonic degrees N) are computed at significant cost. We have previously seen how a partial set of Slepian functions provides excellent regional approximations in noiseless cases. To conclude this section, we thus propose an estimator for the situation of the form (90), where we attempt to reconstruct the unknown signal F from a regionally observed set of noisy measurements, M.

This time, our estimator does not take the form of a spherical-harmonic expansion that needs to be *regularized* (sometimes at great computational cost), but rather of a Slepian-function expansion which can be *truncated* (usually without any difficulty at all). In the context of equation (88): we prefer the approximate identity over the equality which may well furnish us with a 'complete' expansion, but whose coefficients we can only calculate approximately, after regularization.

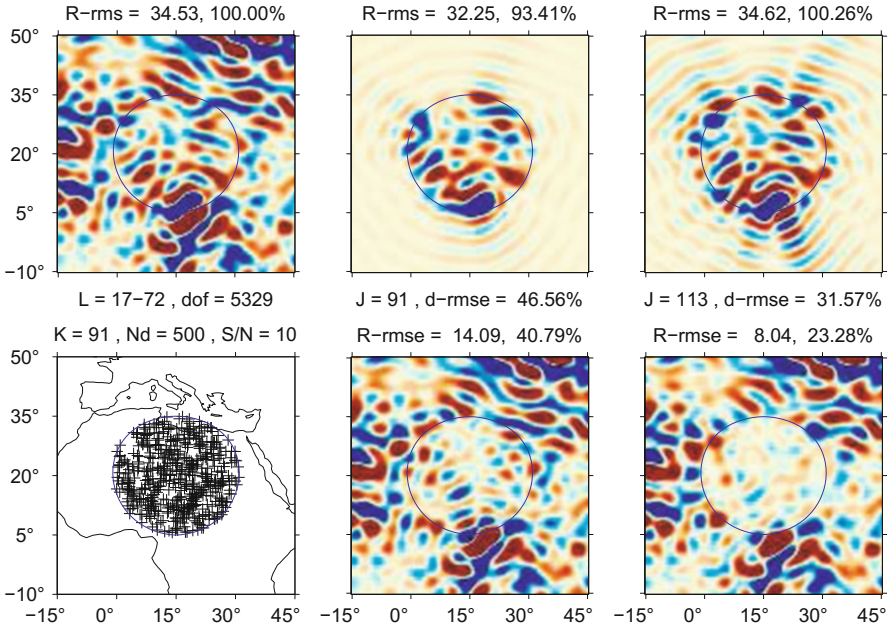


FIGURE 13. Example of Slepian-basis (Shannon number 91) approximation of a bandlimited field from data with a signal-to-noise ratio of 10, for two truncation levels, 91 and 113. Layout as in Figure 12.

In the framework of Slepian-function estimation, truncation *is* our regularization. The Slepian-basis solutions to the 'unregularized' ($\lambda = 0$) problem (91) are, quite simply,

$$\hat{F}^{\wedge}(\alpha) = \left(\lambda_C^{(\alpha)}\right)^{-1} \left\langle M, \left(G_N^C\right)_\alpha \right\rangle_{\mathrm{L}^2(C)} . \tag{96}$$

Truncation means that we only compute J of them, which gives us the freedom to avoid the blowup of the inverse eigenvalues, i.e., the estimate in the Slepian basis is given by

$$\hat{F}_J = \sum_{\alpha=1}^{J} \hat{F}^{\wedge}(\alpha) \left(G_N^C\right)_\alpha . \tag{97}$$

By the same metric of (94), we evaluate the quality of this solution as

$$\frac{E\left\{ \left\| \hat{F}_J - F \right\|^2_{\mathrm{L}^2(C)} \right\}}{E\left\{ \|F\|^2_{\mathrm{L}^2(C)} \right\}} = \mathscr{R}^{-1} \frac{J}{N_C} + \frac{1}{N_C} \sum_{\alpha>J}^{(N+1)^2} \lambda_C^{(\alpha)} . \tag{98}$$

As (94), but unlike (89), again (98) is only applicable in the case of white noise and white signal with a signal-to-noise power ratio $\mathscr{R}$. Of course, the signal contained in the neglected terms of what should be a complete Slepian expansion exerts

a controlling factor on the mean squared error behavior. [85] show how, in the Slepian basis, the neglected terms positively affect the variance of the estimate, but negatively the bias; the mean-squared error being the combination of the two. Minimization of (98) to determine the optimal truncation level for these circumstances yields it in terms of the Slepian eigenvalue and the signal-to-noise level, namely

$$\lambda_C^{(J_{\mathrm{opt}})} \approx \mathscr{R}^{-1}. \tag{99}$$

In other words, we include Slepian functions in the expansion until their ranked eigenvalues drop below the noise-to-signal ratio.

We have ignored that in order to 'solve' data-driven approximation problems, we need to determine an optimal bandwidth N *and* an optimal truncation level J for data situations that are more involved than just being given by white noise and white signal. Such vital practical matters are discussed by Slobbe et al. [91] and Plattner and Simons [81]. The solution procedures involved are always cumbersome – but the computational complexity, and the overall size of the numerical problem, of the truncated Slepian-function approach is always smaller than via regularized spherical-harmonics. Slepian functions lend themselves well to solving approximation problems involving noisy and partially observed data on the sphere.

Two realistic examples of truncated Slepian-basis approximation problems are given in Figures 12 and 13.

4.2. Approximation by splines

Only relatively recently have zonal kernel function techniques such as spline interpolation/approximation and wavelet analysis been playing a fundamental role in numerical analysis on the sphere. *Spherical splines* (independently introduced by Freeden [28] andWahba [99] in 1981) are canonical generalizations of 'spherical polynomials' (spherical harmonics) which have desirable characteristics as interpolating, smoothing, and best approximating functions (see also [100]). By spline interpolation we mean the variational problem of minimizing an 'energy'-norm of a suitable Sobolev space. Depending on the chosen norm, bandlimited and non-bandlimited splines are distinguished. Spherical splines have been successfully applied to many areas of application in particular in geodesy for gravitational field determination, radio occultation, ocean flow, etc. (for more details see [29, 36, 47] and the references therein). Spherical splines, especially their counterparts on the ball, have been applied to tomographic inverse problems in geophysics ([1–3, 10–12, 22, 74]) and in medical imaging ([27]).

To understand spherical splines, we adopt the idea of one-dimensional cubic splines to the sphere. Cubic splines in one-dimension are well known for having minimal 'bending energy' (roughly, minimal 'curvature energy' understood in a linearized sense). More concretely, among all interpolating functions of the Sobolev space $\mathrm{H}^{(2)}([a,b])$, the integral $\int_a^b |F''(x)|\,\mathrm{d}x$ becomes minimal, where F may be physically interpreted as the deflection normal to the rest position which is supposed to be horizontal. The physical model is suggested by the classical inter-

pretation of the potential energy of a statically deflected thin beam which indeed is proportional to the integral taken over the square of the linearized curvature of the elastic beam. Analogously, the concept can be applied to the sphere by choosing $\int_\Omega |\Delta_\xi^* F(\xi)|^2 \, \mathrm{d}\omega(\xi)$, where F now denotes the deflection of a thin membrane normal to the rest position supposed to be spherical. In other words, the second derivative canonically takes on the form of the Beltrami operator Δ^*. Indeed, our interest now is to state that the interpolating spline to a given dataset has minimum 'bending energy' for all interpolants within the Sobolev space $\mathrm{H}^{(2)}(\Omega)$. Furthermore, the spline functions defined in this section are able to simultaneously interpolate and smooth the data. Hence, we can decide in our spline application, which knots of the input data should be strictly interpolated and which ones should be 'near' the interpolating function, i.e., the points subjected to smoothing.

Reproducing kernel Hilbert reference space. As usual (see, e.g., [47]), we introduce the Sobolev space $\mathrm{H}^{(2)}(\Omega)$ as the completion of $\mathrm{C}^{(2)}(\Omega)$ with respect to a specific scalar product thereby specifying $\mathrm{H}^{(2)}(\Omega)$ as a certain reproducing kernel space. In more detail, the inner product $\langle \cdot, \cdot \rangle_{\mathrm{H}^{(2)}(\Omega)}$ is defined by

$$\begin{aligned}\langle F, G \rangle_{\mathrm{H}^{(2)}(\Omega)} = & \underbrace{\int_\Omega F(\eta) Y_{0,1}(\eta) \, \mathrm{d}\omega(\eta) \int_\Omega G(\eta) Y_{0,1}(\eta) \, \mathrm{d}\omega(\eta)}_{= \langle F, G \rangle_{\mathrm{H}_0}} \\ & + \underbrace{\sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} (n(n+1))^2 \int_\Omega F(\eta) Y_{n,j}(\eta) \, \mathrm{d}\omega(\eta) \int_\Omega G(\eta) Y_{n,j}(\eta) \, \mathrm{d}\omega(\eta)}_{= \langle F, G \rangle_{\mathrm{H}_0^\perp}}, \end{aligned} \tag{100}$$

which is equivalent in accordance with Parseval's identity to

$$\begin{aligned}\langle F, G \rangle_{\mathrm{H}^{(2)}(\Omega)} = & \underbrace{\int_\Omega F(\eta) Y_{0,1}(\eta) \, \mathrm{d}\omega(\eta) \int_\Omega G(\eta) Y_{0,1}(\eta) \, \mathrm{d}\omega(\eta)}_{= \langle F, G \rangle_{H_0}} \\ & + \underbrace{\int_\Omega (\Delta_\eta^* F(\eta))(\Delta_\eta^* G(\eta)) \, \mathrm{d}\omega(\eta)}_{= \langle F, G \rangle_{\mathrm{H}_0^\perp}} \\ = & \langle F, G \rangle_{\mathrm{H}_0} + \langle F, G \rangle_{\mathrm{H}_0^\perp} \end{aligned} \tag{101}$$

for all $F, G \in \mathrm{C}^{(2)}(\Omega)$. The Sobolev space $\mathrm{H}^{(2)}(\Omega)$ as defined in Section 3.1 is the completion of $\mathrm{C}^{(2)}(\Omega)$ under the norm $\| \cdot \|_{\mathrm{H}^{(2)}(\Omega)}$, i.e.,

$$\mathrm{H}^{(2)}(\Omega) = \overline{\mathrm{C}^{(2)}(\Omega)}^{\| \cdot \|_{\mathrm{H}^{(2)}(\Omega)}}, \tag{102}$$

where $\|F\|_{\mathrm{H}^{(2)}(\Omega)} = \sqrt{\langle F, F \rangle_{\mathrm{H}^{(2)}(\Omega)}}$.

Consider the kernel $K : (\xi, \eta) \mapsto K(\xi, \eta)$, $(\xi, \eta) \in \Omega^2 = \Omega \times \Omega$ given in the form

$$K(\xi, \eta) = Y_{0,1}(\xi) Y_{0,1}(\eta) + \sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} \frac{1}{(n(n+1))^2} Y_{n,j}(\xi) Y_{n,j}(\eta). \tag{103}$$

Then $K(\xi, \cdot)$, $\xi \in \Omega$ fixed, is a member of $\mathrm{H}^{(2)}(\Omega)$. Inserting K into the inner product, we see via the orthogonal properties of the spherical harmonics that

$$\langle F, K(\xi, \cdot) \rangle_{\mathrm{H}^{(2)}(\Omega)} = F(\xi), \qquad \xi \in \Omega, \tag{104}$$

for all $F \in \mathrm{H}^{(2)}(\Omega)$. Hence, $K(\cdot, \cdot)$ is the unique reproducing kernel of the Hilbert space $\mathrm{H}^{(2)}(\Omega)$. The reproducing kernel $K(\cdot, \cdot)$ can be decomposed into the reproducing kernels of the spaces H_0 and $\mathrm{H}_0^{\perp}$, respectively, via

$$K(\xi, \eta) = \underbrace{Y_{0,1}(\xi) Y_{0,1}(\eta)}_{= K_0(\xi, \eta) = \frac{1}{4\pi}} + \underbrace{\sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} \frac{1}{(n(n+1))^2} Y_{n,j}(\xi) Y_{n,j}(\eta)}_{= K_0^{\perp}(\xi, \eta)}. \tag{105}$$

Applying the spherical-harmonic addition theorem and comparing with (105), we get

$$\begin{aligned} K_0^{\perp}(\xi, \eta) &= \sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} \frac{1}{(n(n+1))^2} Y_{n,j}(\xi) Y_{n,j}(\eta) \\ &= \frac{1}{4\pi} \sum_{n=1}^{\infty} \frac{2n+1}{(n(n+1))^2} P_n(\xi \cdot \eta) \\ &= G((\Delta^*)^2; \xi, \eta) \end{aligned} \tag{106}$$

where $G((\Delta^*)^2, \cdot, \cdot)$ is the Green function with respect to the iterated Beltrami operator $(\Delta^*)^2 = \Delta^* \Delta^*$ (see [28]). Summarizing our results we, therefore, see that

$$K(\xi, \eta) = \underbrace{Y_{0,1}(\xi) Y_{0,1}(\eta)}_{= K_0(\xi, \eta) = \frac{1}{4\pi}} + \underbrace{G\left((\Delta^*)^2; \xi, \eta\right)}_{= K_0^{\perp}(\xi, \eta)}, \qquad \xi, \eta \in \Omega, \tag{107}$$

is the uniquely determined reprokernel of the space $(\mathrm{H}^{(2)}(\Omega), \langle \cdot, \cdot \rangle_{\mathrm{H}^{(2)}(\Omega)})$, i.e.:

(i) For each fixed $\xi \in \Omega$, $K(\xi, \eta)$, a function of η, is an element of $\mathrm{H}^{(2)}(\Omega)$.
(ii) For every function $F \in \mathrm{H}^{(2)}(\Omega)$ and for every point $\xi \in \Omega$, the reproducing property holds:

$$F(\xi) = \langle F, K(\xi, \cdot) \rangle_{\mathrm{H}^{(2)}(\Omega)}. \tag{108}$$

Explicit representation of the reproducing kernel. Keeping the reprostructure of $\mathrm{H}^{(2)}(\Omega)$ in mind, we are able to handle our announced spline interpolation and smoothing problem. To this end, we follow the concept presented by Freeden [28] and observe, in addition, the explicit representation of $G((\Delta^*)^2; \cdot, \cdot)$ known from

[41]. In fact, Green's function corresponding to the iterated Beltrami operator $(\Delta^*)^2$ is continuous on $\Omega \times \Omega$ and admits the explicit formulation:

$$G((\Delta^*)^2;\xi,\eta) = \begin{cases} \frac{1}{4\pi}, & 1-\xi\cdot\eta = 0 \\ \frac{1}{4\pi}(1-\ln(1-\xi\cdot\eta)(\ln(1+\xi\cdot\eta)-\ln(2)) & \\ -\mathfrak{L}_2(\frac{1-t}{2}) - (\ln(2))^2 + \ln(2)\ln(1+\xi\cdot\eta)), & 1\pm\xi\cdot\eta \neq 0 \\ \frac{1}{4\pi} - \frac{\pi}{24}, & 1+\xi\cdot\eta = 0, \end{cases} \tag{109}$$

where the function $\mathfrak{L}_2(x)$ is the dilogarithm given by

$$\mathfrak{L}_2(x) = -\int_0^x \frac{\ln(1-t)}{t}\,\mathrm{d}t = \sum_{k=1}^{\infty}\frac{x^k}{k^2}. \tag{110}$$

Spline concept. We have come to the definition of spherical splines corresponding to one-dimensional cubic splines (a more general concept involving pseudo-differential operators is known from [47]). Let $M_1,\ldots,M_n$ be a linearly independent system of bounded linear functionals on $\mathrm{H}^{(2)}(\Omega)$. Any function $S \in \mathrm{H}^{(2)}(\Omega)$ of the form

$$S(\eta) = c_0 Y_{0,1}(\eta) + \sum_{i=1}^{n} a_i M_i G((\Delta^*)^2;\eta,\cdot), \quad a_i \in \mathbb{R}, \quad \eta \in \Omega \tag{111}$$

with

$$\sum_{i=1}^{n} a_i M_i(Y_{0,1}) = 0 \tag{112}$$

is called a *spherical spline* in $\mathrm{H}^{(2)}(\Omega)$ relative to $M_1,\ldots,M_n$.

By virtue of (111) and (112), we are now prepared to formulate the following uniqueness result: let $M_1,\ldots,M_n$ be a system of bounded linear functionals on the Sobolev space $\mathrm{H}^{(2)}(\Omega)$ such that the $((n+1)\times(n+1))$-matrix

$$\begin{pmatrix} \alpha & \kappa \\ \kappa^{\mathrm{T}} & 0 \end{pmatrix} \tag{113}$$

is non-singular, where the matrix α and the vector κ are given as follows:

$$\alpha = \left(M_i M_j G((\Delta^*)^2;\cdot,\cdot)\right)_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}}, \tag{114}$$

$$\kappa = \left(M_i Y_{0,1}(\cdot)\right)_{i=1,\ldots,n}. \tag{115}$$

Then, there exists a unique spline in $\mathrm{H}^{(2)}(\Omega)$ relative to $M_1,\ldots,M_n$ that solves the interpolation problem $M_i S = \mu_i$, $i=1,\ldots,n$. This spline is called the *interpolating spline*. The proof easily follows by inserting the representation (111) into $M_i S = \mu_i$, $i=1,\ldots,n$, resulting in a linear system for the coefficients a_i, c_0, whose coefficient matrix is given by (113). Since the matrix is assumed to be non-singular, the coefficients are uniquely determined.

The key to spline approximation is the so-called *spline integration formula*

$$\int_\Omega \Delta_\eta^* S(\eta)\ \Delta_\eta^* F(\eta)\,\mathrm{d}\omega(\eta) = \sum_{k=1}^{n} a_k M_k F, \tag{116}$$

valid for the uniquely determined interpolating spline S and all members $F \in \mathrm{H}^{(2)}(\Omega)$, provided that the constraint $\kappa^{\mathrm{T}} a = 0$ is fulfilled. The proof is a direct conclusion of the reproducing kernel property. Its idea is to inspect the sum $\sum_{k=1}^{n} a_k M_k F$ and substitute F by the reproducing kernel property

$$\sum_{k=1}^{n} a_k M_k F = \sum_{k=1}^{n} a_k M_k \langle F(\cdot), K(\eta,\cdot)\rangle_{\mathrm{H}^{(2)}(\Omega)}. \tag{117}$$

Evaluating the inner product by inserting the reproducing kernel function leads to the desired result.

Next, we turn to dealing with the *'minimum energy property' of strict spline interpolation.*

Theorem 4. *Let $M_1, \ldots, M_n$ be a linearly independent system of bounded linear functionals on $\mathrm{H}^{(2)}(\Omega)$. Let S be the unique spline which solves the interpolation problem $M_i S = \mu_i$, $i = 1, \ldots, n$. Then, for all twice continuously differentiable functions F on Ω, which interpolate the given data, i.e., $M_i F = \mu_i$, $i = 1, \ldots, n$, the following inequality*

$$\int_\Omega (\Delta_\eta^* S(\eta))^2\,\mathrm{d}\omega(\eta) \le \int_\Omega (\Delta_\eta^* F(\eta))^2\,\mathrm{d}\omega(\eta) \tag{118}$$

holds true with equality if and only if $S = F$.

The proof easily follows from arguments given by Freeden [28]. Theorem 4 tells us that the 'bending energy' (the integral over the second derivative) of the spline is minimal among all functions in $\mathrm{H}^{(2)}(\Omega)$ interpolating the data.

Combined spline interpolation and smoothing. Theorem 4 allows an extension to include smoothing at predefined points while interpolating the remaining pointset (in accordance with [46]). This technique was used by Blick and Freeden [13] to visualize radio occultation data collected by the satellite CHAMP.

Given $n = p + q$ data points, where the data points μ_i, $i = 1, \ldots, p$, are subjected to smoothing and the points ν_i, $i = 1, \ldots, q$, are subjected to strict interpolation, we are lead to the following result.

Theorem 5. *Suppose that δ and $\beta_1^2, \ldots, \beta_p^2$ are prescribed positive weights and that μ_i, $i = 1, \ldots, p$; ν_j, $j = 1, \ldots, q$ are given data points. Let $M_1, \ldots, M_p$ and $N_1, \ldots, N_q$ be systems of bounded linear functionals on $\mathrm{H}^{(2)}(\Omega)$ such that the $((p+q)+1) \times ((p+q)+1)$-matrix*

$$\begin{pmatrix} \alpha & \beta & \kappa \\ \beta^{\mathrm{T}} & \gamma & \zeta \\ \kappa^{\mathrm{T}} & \zeta^{\mathrm{T}} & 0 \end{pmatrix} \tag{119}$$

is non-singular, where the matrices $\alpha,\ \beta,\ \gamma,\ \kappa,\ \zeta$ *are given as follows*

$$\alpha = \left(M_i M_j G\left(\left(\Delta^*\right)^2;\cdot,\cdot\right) + \delta\beta_i^2\delta_{ij}\right)_{\substack{i=1,\ldots,p\\ j=1,\ldots,p}}, (\textit{Kronecker } \delta_{ij}) \tag{120}$$

$$\beta = \left(M_i N_j G\left(\left(\Delta^*\right)^2;\cdot,\cdot\right)\right)_{\substack{i=1,\ldots,p\\ j=1,\ldots,q}}, \tag{121}$$

$$\gamma = \left(N_i N_j G\left(\left(\Delta^*\right)^2;\cdot,\cdot\right)\right)_{\substack{i=1,\ldots,q\\ j=1,\ldots,q}}, \tag{122}$$

$$\kappa = \left(M_i Y_{0,1}(\cdot)\right)_{i=1,\ldots,p}, \tag{123}$$

$$\zeta = \left(N_j Y_{0,1}(\cdot)\right)_{j=1,\ldots,q}. \tag{124}$$

Then the smoothing spline function S *of the form*

$$S(\zeta) = c_0 Y_{0,1}(\xi) + \sum_{i=1}^{p} a_i M_i G\left(\left(\Delta^*\right)^2;\xi,\cdot\right) + \sum_{j=1}^{q} b_j N_j G\left(\left(\Delta^*\right)^2;\xi,\cdot\right), \xi\in\Omega, \tag{125}$$

with coefficients $a\in\mathbb{R}^p$, $a^{\mathrm{T}} = (a_1,\ldots,a_p)$; $b\in\mathbb{R}^q$, $b^{\mathrm{T}} = (b_1,\ldots,b_q)$ *and* $c_0\in\mathbb{R}$ *subjected to the constraint*

$$\sum_{i=1}^{p} a_i M_i(Y_{0,1}) + \sum_{j=1}^{q} b_j N_j(Y_{0,1}) = 0 \tag{126}$$

is the unique solution of the interpolation and smoothing problem given by

$$\begin{aligned} M_i S + \delta\beta_i^2 a_i &= \mu_i, \quad i = 1,\ldots,p,\\ N_j S &= \nu_j, \quad j = 1,\ldots,q, \end{aligned}$$

corresponding to the data points μ_i, $i = 1,\ldots,p$; ν_j, $j = 1,\ldots,q$ *and represents the only element of* $\mathrm{H}^{(2)}(\Omega)$ *satisfying*

$$\sum_{i=1}^{p}\left(\frac{M_i S - \mu_i}{\beta}\right)^2 + \delta\langle S,S\rangle_{H_0^\perp} \le \sum_{i=1}^{p}\left(\frac{M_i F - \mu_i}{\beta_i}\right)^2 + \delta\langle F,F\rangle_{\mathrm{H}_0^\perp} \tag{127}$$

for all $F\in\mathrm{H}^{(2)}(\Omega)$ *with* $N_j F = \nu_j,\ j = 1,\ldots,q$.

As already mentioned, the proof can be given in parallel to the arguments stated by Freeden and Witte [46]. Moreover, Theorem 4 leads us to the following comments:

(i) The values $\mu_1,\ldots,\mu_p,\ \nu_1,\ldots,\nu_q$ are regarded as the observed quantities, e.g., geodetic observations and measurements.

(ii) The spline function $S\in\mathrm{H}^{(2)}(\Omega)$ satisfies that $M_i S$ is 'near' μ_i, $i = 1,\ldots,p$ and $N_j S$ is equal to ν_j, $j = 1,\ldots,q$. The 'nearness' of the values $M_i S$ to μ_i, $i = 1,\ldots,p$ can be controlled by choosing the constant δ in a suitable way. A small value of δ emphasizes fidelity to the observed data at the expense of smoothness, while a large value does the opposite.

(iii) Taking $\delta = 0$ yields $M_i S = \mu_i$, $i = 1,\ldots,p$, i.e., the combined smoothing and interpolation procedure leads back to strict interpolation.

(iv) For numerical purposes, it is advantageous to adapt the quantities $\beta_1^2,\ldots,\beta_p^2$ to the standard deviations of the measured values.

4.3. Approximation by wavelets

As already pointed out, the context of the spectral representation of a square-integrable function by means of spherical harmonics is essential to solving many problems in today's applications. In future research, however, orthogonal (Fourier) expansions in terms of spherical harmonics $\{Y_{n,j}\}$ will not be the only way of representing a square-integrable function. In order to explain this in more detail, we think of a square-integrable function as a signal in which the spectrum evolves over space in significant way. We imagine that, at each point on the sphere Ω, the function refers to a certain combination of frequencies, and that these frequencies are continuously changing. This space-evolution of the frequencies, however, is not reflected in the Fourier expansion in terms of non-space localizing spherical harmonics, at least not directly. Therefore, in theory, any member F of the space $\mathrm{L}^2(\Omega)$ can be reconstructed from its Fourier transforms, i.e., the 'amplitude spectrum' $\{F^\wedge(n,j)\}_{\substack{n=0,1,\dots,\\ j=1,\dots,2n+1}}$, but the Fourier transform contains information about the frequencies of the function over all positions instead of showing how the frequencies vary in space.

Dirac families. In what follows, we present a two-parameter, i.e., scale- and space-dependent method of achieving a reconstruction of a function $F \in \mathrm{L}^2(\Omega)$ involving (scalar) zonal kernel functions which we refer to as a *Dirac family* $\{\Phi_\rho\}_{\rho\in(0,\infty)}$ converging to the (zonal) Dirac kernel δ. In other words, a Dirac family is a set of zonal kernels $\Phi_\rho : [-1,1] \to \mathbb{R}$, $\rho \in (0,\infty)$, of the form

$$\Phi_\rho(\xi\cdot\eta) = \sum_{n=0}^{\infty} \Phi_\rho^\wedge(n)\,\frac{2n+1}{4\pi}\,P_n(\xi\cdot\eta), \quad \xi,\eta\in\Omega, \tag{128}$$

converging to the 'Dirac kernel' δ as $\rho\to 0$, $\rho>0$. Consequently, if $\{\Phi_\rho\}_{\rho\in(0,\infty)}$ is a Dirac family, its 'symbol' $\{\Phi_\rho^\wedge(n))\}_{n=0,1,\dots}$ constitutes a sequence satisfying the limit relation

$$\lim_{\rho\to 0,\,\rho>0} \Phi_\rho^\wedge(n) = 1, \quad n = 0,1,\dots\ . \tag{129}$$

Accordingly, if $\{\Phi_\rho\}_{\rho\in(0,\infty)}$ is a scaling kernel function, the convolution integrals

$$(\Phi_\rho * F)(\xi) = \int_\Omega \Phi_\rho(\xi\cdot\eta)F(\eta)\,\mathrm{d}\omega(\eta), \quad \xi\in\Omega, \tag{130}$$

converge (in a certain topology) to the limit

$$F(\xi) = (\delta * F)(\xi) = \int_\Omega \delta(\xi\cdot\eta)F(\eta)\,\mathrm{d}\omega(\eta), \quad \xi\in\Omega, \tag{131}$$

for all $\xi\in\Omega$ as ρ tends to 0 (from the positive side). In more detail, if F is a function of class $\mathrm{L}^2(\Omega)$ and $\{\Phi_\rho\}$ is a (suitable) Dirac family (tending to the Dirac kernel), then the following limit relation holds true:

$$\lim_{\rho\to 0,\,\rho>0} \|F - \Phi_\rho * F\|_{\mathrm{L}^2(\Omega)} = 0. \tag{132}$$

There is a large number of Dirac families that is of interest for geoscientific application (for more details, the reader is referred to, e.g., [39, 44] and the

references therein). Only three prototypes of Dirac families should be mentioned here: the bandlimited Shannon family, the neither bandlimited nor spacelimited Abel–Poisson and Gauss–Weierstraß families, and the spacelimited Haar family.

It should be noted that an approximate convolution identity (132) acts as a space and frequency localization procedure in the following way. As $\{\Phi_\rho\}_{\rho\in(0,\infty)}$ is a Dirac family of zonal scalar kernel functions tending to the Dirac kernel, the function $\Phi_\rho(\eta\cdot)$, is highly concentrated around the point $\eta \in \Omega$, if the 'scale parameter' is a small positive value. Moreover, as ρ tends to infinity, $\Phi_\rho(\eta\cdot)$ becomes more and more localized in frequency. Correspondingly, the uncertainty principle states that the space localization of $\Phi_\rho(\eta\cdot)$ becomes more and more decreasing. In conclusion, the products $\eta \mapsto \Phi_\rho(\xi\cdot\eta)F(\eta)$, $\eta \in \Omega$, $\xi \in \Omega$, for each fixed value ρ, display information in $F \in \mathrm{L}^2(\Omega)$ at various levels of spatial resolution or frequency bands. Consequently, as ρ approaches ∞, the convolution integrals $\Phi_\rho * F = \int_\Omega \Phi_\rho(\cdot\eta)F(\eta)\,\mathrm{d}\omega(\eta)$ display coarser, lower-frequency features. As ρ approaches 0, the integrals give sharper and sharper spatial resolution. Thus, the convolution integrals can measure the space-frequency variations of spectral components, but they have a different space-frequency resolution.

Scaling and wavelet functions. Next we come to the bilinear theory of scaling and wavelet functions (note that we only deal with the bilinear theory, for basic aspects of the linear case the reader is referred to, e.g., [39, 44]).

The point of departure for our multi-scale approach is a particular type of a Dirac family: a *scaling* (*kernel*) *function* $\{\Phi_\rho^{(2)}\}_{\rho\in(0,\infty)}$ is a set of zonal kernels $\Phi_\rho^{(2)} = \Phi_\rho * \Phi_\rho : [-1,1] \to \mathbb{R}$, $\rho \in (0,\infty)$, of the form

$$\Phi_\rho^{(2)}(\xi\cdot\eta) = \sum_{n=0}^{\infty} \Phi_\rho^{(2)\wedge}(n)\,\frac{2n+1}{4\pi}\,P_n(\xi\cdot\eta), \quad \xi,\eta\in\Omega, \tag{133}$$

with

$$\lim_{\rho\to 0,\,\rho>0} \Phi_\rho^{(2)\wedge}(n) = \lim_{\rho\to 0,\,\rho>0} (\Phi_\rho^\wedge(n))^2 = 1, \quad n = 0,1,\dots\ . \tag{134}$$

and

$$\Phi_\rho^{(2)\wedge}(0) = 1. \tag{135}$$

Accordingly, the convolution integrals

$$\left(\Phi_\rho^{(2)} * F\right)(\xi) = \int_\Omega \Phi_\rho^{(2)}(\xi\cdot\eta)F(\eta)\,\mathrm{d}\omega(\eta), \quad \xi\in\Omega, \tag{136}$$

converge (in a certain topology) to the limit

$$F(\xi) = (\delta * F)(\xi) = \int_\Omega \delta(\xi\cdot\eta)F(\eta)\,\mathrm{d}\omega(\eta), \quad \xi\in\Omega, \tag{137}$$

for all $\xi \in \Omega$ as ρ tends to 0 (from the positive side). In other words, if F is a function of class $\mathrm{L}^2(\Omega)$ and $\{\Phi_\rho^{(2)}\}$ is a certain Dirac family (tending to the Dirac

kernel), then the approximate identity

$$\lim_{\rho\to 0,\,\rho>0} \left\| F - \Phi_\rho^{(2)} * F \right\|_{\mathrm{L}^2(\Omega)} = 0 \tag{138}$$

holds true.

Each scale approximation $\Phi_\rho^{(2)} * F$ of a function $F \in \mathrm{L}^2(\Omega)$ must be made directly by computing the relevant convolution integrals. In doing so, however, it is inefficient to use no information from the approximation $\Phi_\rho^{(2)} * F$ within the computation of $\Phi_{\rho'}^{(2)} * F$ provided that $\rho' < \rho$. In fact, the efficient construction of multiscale approximation based on Dirac families usually begins by a *multiresolution analysis* in terms of wavelets, i.e., a recursive method which is efficient for computation, but not all economic multiscale approaches constitute multiresolution procedures (see, e.g., [35, 36, 38, 40–43, 47] and the references therein).

Let $\Psi_\rho(\xi,\eta)$, $(\xi,\eta) \in \Omega \times \Omega$, be defined via the series expansion

$$\Psi_\rho(\xi,\eta) = \sum_{n=0}^{\infty} {\Psi_\rho}^\wedge(n)\, \frac{2n+1}{4\pi}\, P_n(\xi\cdot\eta), \quad (\xi,\eta) \in \Omega\times\Omega, \tag{139}$$

such that the symbol $\{{\Psi_\rho^{(2)}}^\wedge(n)\}_{n=0,1,\ldots}$ of $\Psi_\rho^{(2)} = \Psi_\rho * \Psi_\rho$ is derived from ${\Phi_\rho^{(2)}}^\wedge(n)$ via the differential equation ('scale equation')

$${\Psi_\rho^{(2)}}^\wedge(n) = -\rho\, \frac{\mathrm{d}}{\mathrm{d}\rho} {\Phi_\rho^{(2)}}^\wedge(n). \tag{140}$$

As immediate consequences, we obtain from (135) the properties

$${\Psi_\rho}^\wedge(0) = 0 \tag{141}$$

and

$$\lim_{\rho\to 0,\,\rho>0} \Psi_\rho^\wedge(n) = 0$$

for $n = 1, 2, \ldots$ As in classical one-dimensional theory, the condition (135), therefore, justifies the notion *wavelet* of order 0.

Typically, within wavelet nomenclature, we may write

$$\Psi_{\rho;\eta} : \xi \mapsto \Psi_{\rho;\eta}(\xi) = \Psi_\rho(\xi\cdot\eta) = R_\eta D_\rho \Psi(\cdot\xi), \quad \xi \in \Omega,$$

to indicate $\Psi_{\rho;\eta}$ as generated by two parameters, namely the '*η-rotation operator*' R_η and the '*ρ-dilation operator*' D_ρ, respectively, given by

$$R_\eta : \Psi(\cdot\xi) \mapsto R_\eta \Psi(\cdot\xi) = \Psi(\eta\cdot\xi), \tag{142}$$

$$D_\rho : \Psi(\cdot\xi) \mapsto D_\rho \Psi(\cdot\xi) = \Psi_\rho(\cdot\xi). \tag{143}$$

The function $\Psi = \Psi_1$ (i.e., $\rho = 1$) is called the *mother wavelet*.

The *wavelet transform* WT is defined as the $\mathrm{L}^2(\Omega)$-inner product (convolution) of $F \in \mathrm{L}^2(\Omega)$ with the set of 'rotations' and 'dilations' of F

$$(\mathrm{WT})(F)(\rho;\eta) = (\Psi_{\rho;\eta}, F)_{\mathrm{L}^2(\Omega)} = \int_\Omega \Psi_{\rho;\eta}(\xi) F(\xi)\, \mathrm{d}\omega(\xi), \tag{144}$$

i.e., the wavelet transform acts as a space and frequency localization operator. The wavelet transform (WT) is invertible on the space of functions $F \in \mathrm{L}^2(\Omega)$ satisfying $F^\wedge(0,1) = 0$, i.e.,

$$F = \int_\Omega \int_0^\infty (\mathrm{WT})(F)(\rho;\eta)\Psi_{\rho;\eta}(\cdot)\,\frac{\mathrm{d}\rho}{\rho}\,\mathrm{d}\omega(\eta) \tag{145}$$

holds true (in the sense of $\|\cdot\|_{\mathrm{L}^2(\Omega)}$) for all $F \in \mathrm{L}^2(\Omega)$ satisfying $F^\wedge(0,1) = 0$.

The reconstruction formula (145), in fact, is based on the simple idea of dilation and rotation of the mother wavelet.

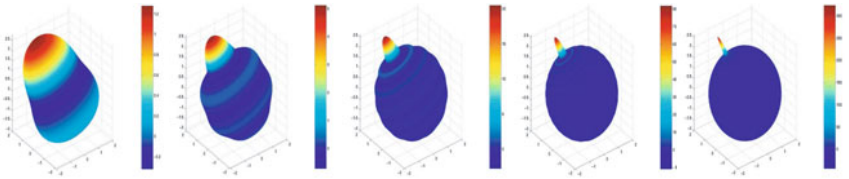


FIGURE 14. Shannon scaling (kernel) functions for decreasing scales ρ.

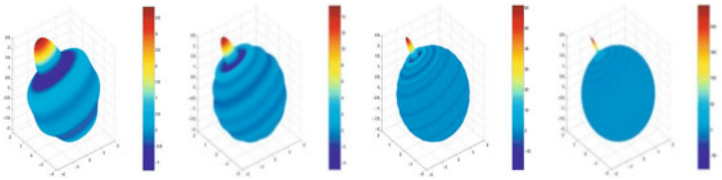


FIGURE 15. Shannon wavelet (kernel) functions for decreasing scales ρ.

Spectral interrelation between Fourier and wavelet transform. In terms of filtering, $\{\Phi_\rho\}_{\rho\in(0,\infty)}$ and $\{\Psi_\rho\}_{\rho\in(0,\infty)}$ may be interpreted (cf. Figures 14 and 15) as lowpass filter and bandpass filter, respectively. Correspondingly, the convolution operators are given by

$$\Phi_\rho * F, \quad F \in \mathrm{L}^2(\Omega), \tag{146}$$

$$\Psi_\rho * F, \quad F \in \mathrm{L}^2(\Omega). \tag{147}$$

The Fourier transforms read as follows:

$$(\Phi_\rho * F)^\wedge(n,j) = F^\wedge(n,j)\Phi_\rho^\wedge(n), \tag{148}$$

$$(\Psi_\rho * F)^\wedge(n,j) = F^\wedge(n,j)\Psi_\rho^\wedge(n). \tag{149}$$

These formulas provide the transition from the wavelet transform to the Fourier transform. Since all scales ρ are used, the reconstruction is highly redundant.

If $F, G \in \mathrm{L}^2(\Omega)$ have vanishing moments of order 0, i.e., if the property $F^\wedge(0,1) = G^\wedge(0,1) = 0$ is satisfied, then it follows from

$$\int_0^\infty (\Psi_\rho^\wedge(n))^2\,\frac{\mathrm{d}\rho}{\rho} = 1 \tag{150}$$

and the Parseval identity of the theory of spherical harmonics that

$$\begin{aligned}
&\int_\Omega \int_0^\infty \langle F, \Psi_{\rho;\eta}\rangle_{\mathrm{L}^2(\Omega)} \langle G, \Psi_{\rho;\eta}\rangle_{\mathrm{L}^2(\Omega)} \frac{\mathrm{d}\rho}{\rho}\,\mathrm{d}\omega(\eta) \\
&\quad = \int_0^\infty \sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} F^\wedge(n,j) G^\wedge(n,j) (\Psi_\rho^\wedge(n))^2 \frac{\mathrm{d}\rho}{\rho} \\
&\quad = \sum_{n=1}^{\infty} \sum_{j=1}^{2n+1} F^\wedge(n,j) G^\wedge(n,j) \\
&\quad = \langle F, G\rangle_{\mathrm{L}^2(\Omega)}. \qquad (151)
\end{aligned}$$

Denote by $\mathrm{L}^2((0,\infty)\times\Omega)$ the space of all integrable functions $H : (0,\infty)\times\Omega \to \mathbb{R}$ such that

$$\int_\Omega \int_0^\infty |H(\rho;\eta)|^2 \frac{\mathrm{d}\rho}{\rho}\,\mathrm{d}\omega(\eta) < \infty. \qquad (152)$$

On the space $\mathrm{L}^2((0,\infty)\times\Omega)$, an inner product $\langle\cdot,\cdot\rangle_{\mathrm{L}^2((0,\infty)\times\Omega)}$ can be imposed corresponding to the norm

$$\|H\|_{\mathrm{L}^2((0,\infty)\times\Omega)} = \left(\int_\Omega \int_0^\infty |H(\rho;\eta)|^2 \frac{\mathrm{d}\rho}{\rho}\,\mathrm{d}\omega(\eta)\right)^{1/2}. \qquad (153)$$

From (151), it follows that

$$\left\langle \langle F, \Psi_{\cdot,\cdot}\rangle_{\mathrm{L}^2(\Omega)}, \langle G, \Psi_{\cdot,\cdot}\rangle_{\mathrm{L}^2(\Omega)} \right\rangle_{\mathrm{L}^2((0,\infty)\times\Omega)} = \langle F, G\rangle_{\mathrm{L}^2(\Omega)} \qquad (154)$$

and

$$\| \langle F, \Psi_{\cdot,\cdot}\rangle_{\mathrm{L}^2(\Omega)} \|^2_{\mathrm{L}^2((0,\infty)\times\Omega)} = \|F\|^2_{\mathrm{L}^2(\Omega)}. \qquad (155)$$

In other words, the total energy of a signal can be continuously distributed by the wavelet transform into scale and spatially dependent 'signal subenergy'.

Least energy representation. WT is a transformation from the one-parameter space $\mathrm{L}^2(\Omega)$ into the two-parameter space $\mathrm{L}^2((0,\infty)\times\Omega)$. Thus, it is clear that (WT) is not surjective on $\mathrm{L}^2((0,\infty)\times\Omega)$. That means that $\mathcal{W} = (\mathrm{WT})(\mathrm{L}^2(\Omega))$ is a proper subspace of $\mathrm{L}^2((0,\infty)\times\Omega)$:

$$\mathcal{W} \subsetneqq \mathrm{L}^2((0,\infty)\times\Omega). \qquad (156)$$

Thus, the problem is to characterize $\mathcal{W}$ within the framework of $\mathrm{L}^2((0,\infty)\times\Omega)$. For that purpose, we consider the operator $P : \mathrm{L}^2((0,\infty)\times\Omega) \to \mathcal{W}$ given by

$$(PH)(\rho';\eta') = \int_0^\infty \int_\Omega K(\rho';\eta' \mid \rho;\eta) H(\rho;\eta)\,\mathrm{d}\omega(\eta)\,\frac{\mathrm{d}\rho}{\rho}, \qquad (157)$$

where

$$K(\rho';\eta' \mid \rho;\eta) = \int_\Omega \Psi_{\rho';\eta'}(\xi)\Psi_{\rho;\eta}(\xi)\,\mathrm{d}\omega(\xi).$$

$\mathcal{W}$ is characterized as follows: $H \in \mathcal{W}$ if and only if

$$H(\rho';\eta') = \int_0^\infty \int_\Omega K(\rho';\eta' \mid \rho;\eta) H(\rho;\eta)\, \mathrm{d}\omega(\eta)\, \frac{\mathrm{d}\rho}{\rho}. \tag{158}$$

It can easily be seen that $K(\rho';\eta' \mid \cdot;\cdot) \in \mathcal{W}$ and $K(\cdot;\cdot \mid \rho;\eta) \in \mathcal{W}$. The kernel $K(\rho';\eta' \mid \rho;\eta)$ is the reproducing kernel in $\mathcal{W}$. The reproducing property (158) can also be understood in such a way that $H \in \mathcal{W}$ is calculable by superpositions of itself. This shows that there is a kind of linear dependence, which can be interpreted as redundancy. Although it might seem inefficient, such redundancy has certain advantages. Unlike a non-redundant expansion, errors can be detected and corrected.

The tendency for correcting errors is expressed in the next result (see [35, 47]):

– Let H be an arbitrary element of $\mathrm{L}^2((0,\infty)\times\Omega)$. Then the unique function $F_H \in \mathrm{L}^2(\Omega)$ which satisfies the property

$$\left\| H - \tilde{F}_H \right\|_{\mathrm{L}^2((0,\infty)\times\Omega)} = \inf_{F\in\mathrm{L}^2(\Omega)} \left\| H - \tilde{F} \right\|_{\mathrm{L}^2((0,\infty)\times\Omega)}$$

(with $\tilde{F}_H = (\mathrm{WT})(F_H)$) is given by

$$F_H(\xi) = \int_0^\infty \int_\Omega H(\rho;\eta) \Psi_{\rho;\eta}(\xi)\, \mathrm{d}\omega(\eta)\, \frac{\mathrm{d}\rho}{\rho}.$$

Indeed, $\tilde{F}_H$ is the orthogonal projection of H onto $\mathcal{W}$, which explains the aforementioned statement.

The linear dependence of $\tilde{F} \in \mathcal{W}$ leads to the effect that the coefficients in $\mathrm{L}^2((0,\infty)\times\Omega)$ for reconstructing a function $F \in \mathrm{L}^2(\Omega)$ are not unique. This can be easily seen from the following identity:

$$F(\xi) = \int_0^\infty \int_\Omega \left(\tilde{F}(\rho;\eta) + \tilde{F}^\perp(\rho;\eta) \right) \Psi_{\rho;\eta}(\xi)\, \mathrm{d}\omega(\eta)\, \frac{\mathrm{d}\rho}{\rho}$$

where $\tilde{F} = (\mathrm{WT})(F)$ and $\tilde{F}^\perp$ is an arbitrary member of $\mathcal{W}^\perp$. Nevertheless, we are able to deal with the following question: given an arbitrary $H(\rho;\xi) = (\mathrm{WT})(F)(\rho;\xi)$, $\rho \in (0,\infty)$, and $\xi \in \Omega$, for some $F \in \mathrm{L}^2(\Omega)$, how can we reconstruct F? The answer (see [35, 47]) is provided by the so-called *least-energy representation*:

– Of all possible functions $H \in \mathrm{L}^2((0,\infty)\times\Omega)$ for $F \in \mathrm{L}^2(\Omega)$, the function $H = (\mathrm{WT})(F)$ is unique in that it minimizes the 'energy' $\|H\|^2_{\mathrm{L}^2((0,\infty)\times\Omega)}$. More explicitly,

$$\|(\mathrm{WT})(F)\|_{\mathrm{L}^2((0,\infty)\times\Omega)} = \inf_{\substack{H\in\mathrm{L}^2((0,\infty)\times\Omega) \\ (\mathrm{WT})^{-1}(H)=F}} \|H\|_{\mathrm{L}^2((0,\infty)\times\Omega)}.$$

Wavelet variants. The construction of *spherical wavelets* has seen an enormous increase of activity in the last few years. Three features are essential in the thinking about georelevant wavelets: basis property, decorrelation, and fast computation.

First, wavelets are building blocks for general datasets derived from functions. By virtue of the basis property, each element of a general class of functions (e.g., a geopotential seen as a member of a set of potentials within a Sobolev space framework) can be expressed in stable way as a linear combination of dilated and shifted copies of a 'mother function' (see [29, 31, 35, 36] and the references therein). The role of the wavelet transform as a mapping from the class of functions into an associated two-parameter family of space and scale dependent functions is properly characterized by least squares properties.

Second, wavelets have the power to decorrelate. In other words, the representation of data in terms of wavelets is somehow 'more compact' than the original representation. We search for an accurate approximation by only using a small fraction of the original information of a function. Typically, the decorrelation is achieved by building wavelets which have a compact support (localization in space), which are smooth (decay towards high frequencies), and which have vanishing moments (decay towards low frequencies). Different types of wavelets can be found from certain constructions of space/momentum localization. The uncertainty principle tells us that sharp localization in 'space and momentum' are mutually exclusive. Nevertheless, decay towards long and short wavelengths (i.e., bandpass filtering) can be assured without any difficulty. Moreover, vanishing moments of wavelets (see, e.g., [45, 47]) enable us to combine (polynomial) outer harmonic expansions (responsible for the long-wavelength part of a function) with wavelet multiscale expansions (responsible for the medium-to-short-wavelengths contributions).

Third, the main question of recovering a function on the sphere, e.g., the Earth's gravitational potential, is how to decompose the function into wavelet coefficients, and how to reconstruct efficiently the potential from the coefficients. There is a 'tree algorithm' or 'pyramid algorithm' (cf. [29, 47]) that makes these steps simple and fast. In this respect, it is desirable to switch between the original representation of the data and its wavelet representation in a time proportional to the size of the data. In fact, the fast decorrelation power of wavelets is the key to applications such as data compression, fast data transmission, noise cancelation, signal recovery, etc.

In the last years, wavelets on the sphere have been the focus of several research groups which led to different wavelet approaches. Common to all these proposals is a multiresolution analysis which enables a balanced amount of both frequency (more accurately, angular momentum) and space localization (see, e.g., [18, 64, 82, 83, 101]).

A group theoretical approach to a continuous wavelet transform on the sphere is followed by Antoine and Vandergheynst [5], Antoine et al. [4], and Holschneider [55]. The parameter choice of their continuous wavelet transform is the product of SO(3) (for the motion on the sphere) and $\mathbb{R}^+$ (for the dilations). A continuous

wavelet transform approach for analyzing functions on the sphere is presented by Dahlke and Maass [19].

The Kaiserslautern constructions (see, e.g., [32, 39, 43, 47]) are intrinsically based on the specific properties concerning the theory of spherical harmonics. Wavelet regularization and multiresolution techniques are applied to 'downward continuation' of spaceborne (satellite) observations (see, e.g., [31, 37, 38, 48] and the references therein). Multiscale signal-to noise ratio modeling is done by signal and noise decorrelation Freeden and Maier [33, 34]. Freeden and Schreiner [42] are interested in a compromise connecting zonal function expressions and structured grids on the sphere to obtain fast algorithms. Freeden et al. [49] and Freeden and Gerhards [31] generate locally supported wavelets by regularizing fundamental solutions to pseudodifferential operators.

Finally, much of the material presented in this paper within a spherical framework can be readily formulated for non-spherical reference surfaces, even for vector and tensor data. Nevertheless, work remains to be done for more realistic geometries such as (the actual) Earth's surface, real satellite orbits, etc. These are challenges for future research.

4.4. Regularized functional matching pursuit and its variants

The Regularized Functional Matching Pursuit (RFMP) and its variants were developed by Fischer [23], Fischer and Michel [24], Michel [69], Michel and Telschow [72, 73], and Telschow [95]. They are based on the Matching Pursuit (MP) and its enhancements as described by Mallat and Zhang [65] and Vincent and Bengio [98], where the problem consisted of finding a greedy algorithm for the approximation of an unknown signal F based on given samples $F(x_j)$, $j = 1, \ldots, N$, usually on Euclidean domains.

For the RFMP, matching pursuit had to be extended to the inverse problem

$$\mathcal{F}F = y$$

for a linear and continuous operator $\mathcal{F} : \mathcal{H}(D) \to \mathbb{R}^l$, a Hilbert space $\mathcal{H}(D)$ of (some) functions on $D \subset \mathbb{R}^d$ (e.g., $\mathrm{L}^2(D)$ or, more generally, a Sobolev space), a given data vector $y \in \mathbb{R}^l$ and an unknown function $F \in \mathcal{H}(D)$. Many inverse problems of this kind, such as the downward continuation (F is the gravitational potential at the surface $D(= \Omega)$ and y is a vector of samples at satellite height) or the inverse gravimetric problem (F is a volume or a surface mass distribution and y is a vector of samples of the gravitational potential), are ill posed. For this reason, a regularization technique also had to be included into the RFMP.

The different algorithms are summarized here starting with a short introduction of the MP. All algorithms have in common that a set of possibly useful trial functions, the 'dictionary' $\mathcal{D} \subset \mathcal{H}(D)$, is chosen in advance. These trial functions need not originate from one single basis system: $\mathcal{D}$ may be (and is often chosen on purpose as) overcomplete. If $\mathcal{D}$ is heterogeneous, it may contain different kinds of basis systems (in particular, with different frequency and space localization). For instance, in several numerical applications of the RFMP, an approximate solution

F was combined from spherical harmonics (for a coarse global approximation) and radial basis functions with different levels of localization (locally improving the result). Without loss of generality, one can assume that $\|d\|_{\mathcal{H}(D)} = 1$ for all $d \in \mathcal{D}$.

Matching pursuit. Assume that a function (signal) $F \in \mathcal{H}(D)$ is to be approximated by m elements of $\mathcal{D}$. In this context, the expression of the best-m-term approximation (see, e.g., [96]) occurs. It means that one looks for m elements $d_1, \ldots, d_m \in \mathcal{D}$ and associated coefficients $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ such that the approximation error

$$\left\| F - \sum_{k=1}^{m} \alpha_k d_k \right\|_{\mathcal{H}(D)}$$

becomes minimal in comparison to all other choices of d_k and α_k. In formal language, the objective is

$$\sigma_m(F, \mathcal{D}) = \inf_{d_j \in \mathcal{D},\, \alpha_j \in \mathbb{R};\; j=1,\ldots,m} \left\| F - \sum_{k=1}^{m} \alpha_k d_k \right\|_{\mathcal{H}(D)} .$$

For large m, it is often numerically too expensive to find an exact minimizer. However, this concept can be a guideline for the construction of a less expensive algorithm with still 'good' results.

The first idea is to construct an iterative algorithm, i.e., to find the pairs $(\alpha_1, d_1), \ldots, (\alpha_m, d_m)$ consecutively. The initial problem is to find $\alpha_1 \in \mathbb{R}$ and $d_1 \in \mathcal{D}$ such that

$$J(\alpha_1, d_1) = \|F - \alpha_1 d_1\|^2_{\mathcal{H}(D)} = \|F\|^2_{\mathcal{H}(D)} - 2\alpha_1 \langle F, d_1 \rangle_{\mathcal{H}(D)} + \alpha_1^2 \tag{159}$$

is minimal. With $\frac{\partial}{\partial \alpha_1} J(\alpha_1, d_1) = 0$, one obtains

$$-2 \langle F, d_1 \rangle_{\mathcal{H}(D)} + 2\alpha_1 = 0, \qquad \text{i.e.,} \qquad \alpha_1 = \langle F, d_1 \rangle_{\mathcal{H}(D)} .$$

Inserting this result in (159), one gets

$$J(\alpha_1, d_1) = \|F\|^2_{\mathcal{H}(D)} - \langle F, d_1 \rangle^2_{\mathcal{H}(D)} .$$

Consequently, this dictionary element $d_1 \in \mathcal{D}$ for which F has the largest projection, i.e., the dictionary element which is most collinear to F, is the optimal choice in the first step. The first approximation is, therefore,

$$F_1 = \langle F, d_1 \rangle_{\mathcal{H}(D)} \, d_1,$$

where $d_1 \in \mathcal{D}$ is a maximizer of $\langle F, d_1 \rangle^2_{\mathcal{H}(D)}$, i.e.,

$$d_1 = \arg\max_{d \in \mathcal{D}} \langle F, d \rangle^2_{\mathcal{H}(D)} .$$

With the residual $R^1 = F - F_1$, one can analogously proceed. In general, if R^n is given, then one has to find $d_{n+1} \in \mathcal{D}$ such that $\langle R^n, d_{n+1} \rangle^2_{\mathcal{H}(D)}$ is maximal and then sets

$$F_{n+1} = F_n + \langle R^n, d_{n+1} \rangle_{\mathcal{H}(D)} \, d_{n+1}.$$

Functional matching pursuit. In the case of an inverse problem $\mathcal{F}F = y$, one minimizes the data misfit

$$\begin{aligned} J(\alpha_1, d_1) &= \|y - \mathcal{F}(\alpha_1 d_1)\|^2_{\mathbb{R}^l} \\ &= \|y\|^2_{\mathbb{R}^l} - 2\alpha_1 \langle y, \mathcal{F}d_1\rangle_{\mathbb{R}^l} + \alpha_1^2 \|\mathcal{F}d_1\|^2_{\mathbb{R}^l}, \end{aligned}$$

which implies that, again by assuming that $\frac{\partial}{\partial\alpha_1} J(\alpha_1, d_1) = 0$,

$$\alpha_1 = \frac{\langle y, \mathcal{F}d_1\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d_1\|^2_{\mathbb{R}^l}}.$$

Consequently,

$$J(\alpha_1, d_1) = \|y\|^2_{\mathbb{R}^l} - \frac{\langle y, \mathcal{F}d_1\rangle^2_{\mathbb{R}^l}}{\|\mathcal{F}d_1\|^2_{\mathbb{R}^l}}$$

shows that d_1 has to be chosen such that

$$\frac{\langle y, \mathcal{F}d_1\rangle^2_{\mathbb{R}^l}}{\|\mathcal{F}d_1\|^2_{\mathbb{R}^l}}$$

is maximal. Then,

$$F_1 = \frac{\langle y, \mathcal{F}d_1\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d_1\|^2_{\mathbb{R}^l}} d_1$$

is the first approximation. With the residual $R^1 = y - \mathcal{F}F_1$, one proceeds again analogously. Hence, for a given residual R^n, one chooses d_{n+1} such that

$$\frac{\langle R^n, \mathcal{F}d_{n+1}\rangle^2_{\mathbb{R}^l}}{\|\mathcal{F}d_{n+1}\|^2_{\mathbb{R}^l}}$$

is maximal and we set

$$F_{n+1} = F_n + \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l}}{\|\mathcal{F}d_{n+1}\|^2_{\mathbb{R}^l}} d_{n+1}.$$

Regularized functional matching pursuit. For the handling of ill-posed inverse problems, the Regularized Functional Matching Pursuit (RFMP) includes a Tikhonov-type regularization term

$$\lambda \|F_n\|^2_{\mathcal{H}(D)},$$

where $\lambda \in \mathbb{R}^+$ is a regularization parameter. Note that the choice of the (Sobolev) space $\mathcal{H}(D)$ influences the obtained result by requiring a particular kind of 'smoothness'. For instance, the Sobolev space $\mathrm{H}^{(2)}(\Omega)$ yields a regularization term which is not equal but similar to the norm which occurs in the minimum principle of spherical spline interpolation (see Theorem 4).

In analogy to the above, let $F_n \in \mathcal{H}(D)$ be the approximation after iteration n and $R^n = y - \mathcal{F}F_n$ be the residual, the error on the right-hand side of the inverse problem $\mathcal{F}F = y$. We find $d_{n+1} \in \mathcal{D}$ and $\alpha_{n+1} \in \mathbb{R}$ such that

$$J_\lambda(\alpha_{n+1}, d_{n+1}) = \|R^n - \mathcal{F}(\alpha_{n+1} d_{n+1})\|^2_{\mathbb{R}^l} + \lambda \|F_n + \alpha_{n+1} d_{n+1}\|^2_{\mathcal{H}(D)}$$

is minimal. Treating the functional J_λ like J above, one obtains

$$J_\lambda(\alpha_{n+1}, d_{n+1}) = \|R^n\|^2_{\mathbb{R}^l} - 2\alpha_{n+1}\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} + \alpha^2_{n+1}\|\mathcal{F}d_{n+1}\|^2_{\mathbb{R}^l} + \lambda\left(\|F_n\|^2_{\mathcal{H}(D)} + 2\alpha_{n+1}\langle F_n, d_{n+1}\rangle_{\mathcal{H}(D)} + \alpha^2_{n+1}\right), \quad (160)$$

where the necessary condition $\frac{\partial}{\partial\alpha_{n+1}} J_\lambda(\alpha_{n+1}, d_{n+1}) = 0$ yields

$$\alpha_{n+1} = \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d_{n+1}\rangle_{\mathcal{H}(D)}}{\|\mathcal{F}d_{n+1}\|^2_{\mathbb{R}^l} + \lambda}. \quad (161)$$

If one inserts (161) into (160), one gets

$$J_\lambda(\alpha_{n+1}, d_{n+1}) = \|R^n\|^2_{\mathbb{R}^l} + \lambda\|F_n\|^2_{\mathcal{H}(D)} - \frac{\left(\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d_{n+1}\rangle_{\mathcal{H}(D)}\right)^2}{\|\mathcal{F}d_{n+1}\|^2_{\mathbb{R}^l} + \lambda}$$

such that $d_{n+1} \in \mathcal{D}$ has to be chosen as a maximizer of

$$\frac{\left(\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d_{n+1}\rangle_{\mathcal{H}(D)}\right)^2}{\|\mathcal{F}d_{n+1}\|^2_{\mathbb{R}^l} + \lambda}.$$

This yields the following algorithm (where the Functional Matching Pursuit is a particular case for $\lambda = 0$).

Algorithm 6 (RFMP). *Let a data vector $y \in \mathbb{R}^l$, a linear and continuous operator $\mathcal{F} : \mathcal{H}(D) \to \mathbb{R}^l$, a dictionary $\mathcal{D} \subset \{d \in \mathcal{H}(D) \,|\, \|d\|_{\mathcal{H}(D)} = 1\}$ and an initial approximation $F_0 \in \mathcal{H}(D)$ be given.*

(i) *Initialize the iteration with $n = 0$ and $R^0 = y - \mathcal{F}F_0$ and select a stopping criterion (data-misfit-based, i.e., choose $\varepsilon > 0$ to require $\|R^{n+1}\| < \varepsilon$, or iteration-based, i.e., choose $N \in \mathbb{N}$ to require $n + 1 \leq N$) as well as a regularization parameter $\lambda \in \mathbb{R}^+_0$.*

(ii) *Determine*

$$d_{n+1} = \arg\max_{d\in\mathcal{D}} \frac{\left(\langle R^n, \mathcal{F}d\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d\rangle_{\mathcal{H}(D)}\right)^2}{\|\mathcal{F}d\|^2_{\mathbb{R}^l} + \lambda}, \quad (162)$$

$$\alpha_{n+1} = \frac{\langle R^n, \mathcal{F}d_{n+1}\rangle_{\mathbb{R}^l} - \lambda\langle F_n, d_{n+1}\rangle_{\mathcal{H}(D)}}{\|\mathcal{F}d_{n+1}\|^2_{\mathbb{R}^l} + \lambda} \quad (163)$$

and set $F_{n+1} = F_n + \alpha_{n+1}d_{n+1}$ and $R^{n+1} = R^n - \alpha_{n+1}\mathcal{F}d_{n+1}$.

(iii) *If the stopping criterion is satisfied, then use F_{n+1} as an approximate solution to $\mathcal{F}F = y$. Otherwise, increase n by 1 and go to step* (ii).

The algorithm is accelerated if one implements the following procedures.

- *Normalize the dictionary:* use the assumption above and choose all $d \in \mathcal{D}$ such that $\|d\|_{\mathcal{H}(D)} = 1$, otherwise the norm of the dictionary elements occurs in (162) and (163) (see, e.g., [24]).

- *Move as much as possible to the preprocessing:* calculate $\|\mathcal{F}d\|_{\mathbb{R}^l}$ for all $d \in \mathcal{D}$ and the (symmetric) matrices with the components $\langle d, \tilde{d} \rangle_{\mathcal{H}(D)}$ and $\langle \mathcal{F}d, \mathcal{F}\tilde{d} \rangle_{\mathbb{R}^l}$, respectively, (with $d, \tilde{d} \in \mathcal{D}$) once and store them.
- *Use preprocessing for finding d_{n+1} and α_{n+1}:* note, in particular, that

$$\langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l} = \langle R^{n-1}, \mathcal{F}d \rangle_{\mathbb{R}^l} - \alpha_n \langle \mathcal{F}d_n, \mathcal{F}d \rangle_{\mathbb{R}^l} ,$$
$$\langle F_n, d \rangle_{\mathcal{H}(D)} = \langle F_{n-1}, d \rangle_{\mathcal{H}(D)} + \alpha_n \langle d_n, d \rangle_{\mathcal{H}(D)} ,$$

  where, in both cases, the first summands on the right-hand side are already known from the previous iteration step (i.e., step $n-1$).

It should be mentioned that, in (162), the maximizer need not be uniquely determined. In this case, no particular strategy for choosing between several maximizers has been applied yet.

One essential result is the following convergence theorem.

Theorem 7 (Convergence Theorem). *Let the dictionary $\mathcal{D}$ satisfy:*

(i) *'semi-frame condition': There exist a constant $c > 0$ and an integer N such that, for all expansions $H = \sum_{k=1}^{\infty} \beta_k d_k$ with $\beta_k \in \mathbb{R}$ and $d_k \in \mathcal{D}$, where the d_k are not necessarily pairwise distinct but $\{j \in \mathbb{N} \,|\, d_j = d_k\}$ is a finite set with at most N elements for each $k \in \mathbb{N}$,*

$$c\|H\|^2_{\mathcal{H}(D)} \leq \sum\nolimits_{k=1}^{\infty} \beta_k^2 .$$

(ii) *$\|d\|_{\mathcal{H}(D)} = 1$ for all $d \in \mathcal{D}$ and, if $\lambda = 0$, then $\inf_{d \in \mathcal{D}} \|\mathcal{F}d\|_{\mathbb{R}^l} > 0$ is required additionally.*

If the sequence $(F_n)_n$ is produced by the RFMP and no dictionary element is chosen more than N times, then $(F_n)_n$ converges in $\mathcal{H}(D)$ to $F_\infty = F_0 + \sum_{n=1}^{\infty} \alpha_n d_n \in \mathcal{H}(D)$. Moreover, the following holds true:

(a) *If $\overline{\operatorname{span} \mathcal{D}}^{\|\cdot\|_{\mathcal{H}(D)}} = \mathcal{H}(D)$ and $\lambda \in \mathbb{R}_0^+$ is an arbitrary parameter, then F_∞ solves*

$$(\mathcal{F}^*\mathcal{F} + \lambda \mathcal{I}) F_\infty = \mathcal{F}^* y ,$$

where $\mathcal{F}^*$ is the adjoint operator corresponding to $\mathcal{F}$ and $\mathcal{I}$ is the identity operator on $\mathcal{H}(D)$. In other words,

$$\|y - \mathcal{F}F_\infty\|^2_{\mathbb{R}^l} + \lambda \|F_\infty\|^2_{\mathcal{H}(D)} = \min_{F \in \mathcal{H}(D)} \left(\|y - \mathcal{F}F\|^2_{\mathbb{R}^l} + \lambda \|F\|^2_{\mathcal{H}(D)} \right) ,$$

where the minimizer is unique, if $\lambda > 0$.

(b) *If $\operatorname{span} \{\mathcal{F}d \,|\, d \in \mathcal{D}\} = \mathbb{R}^l$ and $\lambda = 0$, then F_∞ solves $\mathcal{F}F_\infty = y$.*

Note that the semi-frame condition has been changed (including the requirement on repeated choices of dictionary elements) in comparison to earlier publications on the RFMP by Michel [69] and Michel and Telschow [72], since an unlimited number of equally chosen dictionary elements would allow a counterexample for which the semi-frame condition could not be achieved, as it was pointed out in [73]. For a proof of the convergence theorem and additional properties, see [71]. For numerical examples of RFMP applied to geodetic problems, see [23–26, 72].

Regularized orthogonal functional matching pursuit. Numerical experiments show that the RFMP chooses some dictionary elements several times, which actually means that some of the previously calculated coefficients $\alpha_1, \ldots, \alpha_n$ are corrected. The reason for this phenomenon is that the dictionary elements (or their images in the data space) are typically non-orthogonal. In the case of the Matching Pursuit (MP), this effect is compensated for by introducing a particular orthogonal projection procedure in the Orthogonal Matching Pursuit (OMP, see [78]) and by using 'prefitting' (see [98]). However, the OMP requires that the data and the solution are in the same space for performing the projection and it also does not contain a regularization.

In [95] and [73], the idea behind OMP and 'prefitting' was used to enhance RFMP to Regularized Orthogonal Functional Matching Pursuit (ROFMP). It is now possible to update the coefficients α_i in every iteration. For this reason, the approximation after step n is represented by

$$F_n = \sum_{i=1}^{n} \alpha_i^{(n)} d_i.$$

If one measures the quality of an approximate solution in the data space, i.e., in the sense of the data misfit, then the best approximation (without a regularization) in terms of (fixed) $d_1, \ldots, d_n$ would be given by requiring that $\mathcal{F}F_n$ equals the orthogonal projection of y onto

$$\mathcal{V}_n = \text{span}\,\{\mathcal{F}d_1, \ldots, \mathcal{F}d_n\},$$

i.e., $\mathcal{F}F_n = \mathcal{P}_{\mathcal{V}_n} y$. This is equivalent to requiring that the residual $R^n = y - \mathcal{F}F_n$ is orthogonal to $\mathcal{V}_n$. Geometrically speaking, $\mathcal{F}F_n$ is the projection of y onto the hyperplane $\mathcal{V}_n$ and R^n is the associated plumbline, see Figure 16.

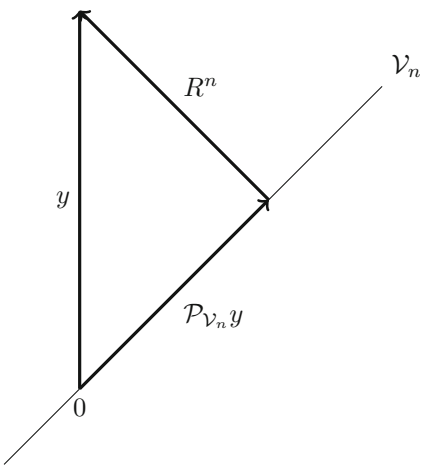


FIGURE 16. Illustration of the orthogonal projection $\mathcal{P}_{\mathcal{V}_n} y$ in $\mathbb{R}^l$.

Consequently, the next summand $\alpha_{n+1}d_{n+1}$ should complement the previous approximation F_n such that $\mathcal{F}F_{n+1} = \mathcal{P}_{\mathcal{V}_{n+1}}y$. However, in general, $\mathcal{P}_{\mathcal{V}_n}(\mathcal{F}d_{n+1}) \neq 0$. This projection would, however, deteriorate the previously exact approximation of $\mathcal{P}_{\mathcal{V}_n}y$ by $\mathcal{F}F_n$. For this reason, this redundant part is subtracted, i.e., one is interested in

$$\mathcal{F}F_{n+1} = \mathcal{F}F_n + \alpha_{n+1}\left[\mathcal{F}d_{n+1} - \mathcal{P}_{\mathcal{V}_n}\left(\mathcal{F}d_{n+1}\right)\right].$$

If one sets $\mathcal{P}_{\mathcal{V}_n}(\mathcal{F}d) = \sum_{i=1}^{n}\beta_i^{(n)}(d)\,\mathcal{F}d_i$, then

$$\begin{aligned}\mathcal{F}F_{n+1} &= \sum_{i=1}^{n}\alpha_i^{(n)}\,\mathcal{F}d_i - \alpha_{n+1}\sum_{i=1}^{n}\beta_i^{(n)}\left(d_{n+1}\right)\,\mathcal{F}d_i + \alpha_{n+1}\mathcal{F}d_{n+1}\\ &= \sum_{i=1}^{n}\left(\alpha_i^{(n)} - \alpha_{n+1}\beta_i^{(n)}\left(d_{n+1}\right)\right)\mathcal{F}d_i + \alpha_{n+1}\mathcal{F}d_{n+1}.\end{aligned}$$

Hence, the task is now (in step $n+1$) to find $\alpha \in \mathbb{R}$ and $d \in \mathcal{D}$ such that

$$\left\| y - \sum_{i=1}^{n}\left(\alpha_i^{(n)} - \alpha\beta_i^{(n)}\left(d\right)\right)\mathcal{F}d_i - \alpha\mathcal{F}d\right\|_{\mathbb{R}^l}$$

is minimized. As an approximation at step $n+1$, one uses then

$$F_{n+1} = \sum_{i=1}^{n}\left(\alpha_i^{(n)} - \alpha_{n+1}\beta_i^{(n)}\left(d_{n+1}\right)\right)d_i + \alpha_{n+1}d_{n+1},$$

$$\alpha_i^{(n+1)} = \alpha_i^{(n)} - \alpha_{n+1}\beta_i^{(n)}\left(d_{n+1}\right) \quad \text{for } i = 1,\ldots,n$$

and $\alpha_{n+1}^{(n+1)} = \alpha_{n+1}$. With the regularization, the functional to minimize is

$$\begin{aligned}&\left\| y - \sum_{i=1}^{n}\left(\alpha_i^{(n)} - \alpha\beta_i^{(n)}\left(d\right)\right)\mathcal{F}d_i - \alpha\mathcal{F}d\right\|_{\mathbb{R}^l}^2\\ &+ \lambda\left\|\sum_{i=1}^{n}\left(\alpha_i^{(n)} - \alpha\beta_i^{(n)}\left(d\right)\right)d_i + \alpha d\right\|_{\mathcal{H}(D)}^2.\end{aligned}$$

This is the principle of the ROFMP. We now introduce some abbreviations.

- The orthogonal complement of $\mathcal{V}_n$ in $\mathbb{R}^l$ is denoted by $\mathcal{W}_n$, i.e., $\mathcal{V}_n \oplus \mathcal{W}_n = \mathbb{R}^l$, and the projection of $\mathcal{F}d$ onto $\mathcal{W}_n$ is

$$\mathcal{P}_{\mathcal{W}_n}(\mathcal{F}d) = \mathcal{F}d - \sum_{i=1}^{n}\beta_i^{(n)}\left(d\right)\,\mathcal{F}d_i.$$

- The function associated to $\mathcal{P}_{\mathcal{V}_n}(\mathcal{F}d)$ in $\mathcal{H}(D)$ is denoted by

$$B_n(d) = \sum_{i=1}^{n}\beta_i^{(n)}\left(d\right)\,d_i.$$

Similar derivations as in the cases above finally yield the following algorithm.

Algorithm 8 (ROFMP). *Let a data vector $y \in \mathbb{R}^l$, a linear and continuous operator $\mathcal{F} : \mathcal{H}(D) \to \mathbb{R}^l$ and a dictionary $\mathcal{D} \subset \mathcal{H}(D) \setminus \{0\}$ be given.*

(i) *Initialize the iteration with $n = 0$, $F_0 = 0$ and $R^0 = y$ and select a stopping criterion (data-misfit-based, i.e., choose $\varepsilon > 0$ to require $\|R^{n+1}\| < \varepsilon$, or iteration-based, i.e., choose $N \in \mathbb{N}$ to require $n + 1 \leq N$) as well as a regularization parameter $\lambda \in \mathbb{R}_0^+$.*

(ii) *Determine*

$$d_{n+1} = \arg\max_{d \in \mathcal{D}} \frac{\left(\langle R^n, \mathcal{P}_{\mathcal{W}_n}(\mathcal{F}d) \rangle_{\mathbb{R}^l} + \lambda \langle F_n, B_n(d) - d \rangle_{\mathcal{H}(D)} \right)^2}{\|\mathcal{P}_{\mathcal{W}_n}(\mathcal{F}d)\|_{\mathbb{R}^l}^2 + \lambda \|B_n(d) - d\|_{\mathcal{H}(D)}^2},$$

$$\alpha_{n+1} = \frac{\langle R^n, \mathcal{P}_{\mathcal{W}_n}(\mathcal{F}d_{n+1}) \rangle_{\mathbb{R}^l} + \lambda \langle F_n, B_n(d_{n+1}) - d_{n+1} \rangle_{\mathcal{H}(D)}}{\|\mathcal{P}_{\mathcal{W}_n}(\mathcal{F}d_{n+1})\|_{\mathbb{R}^l}^2 + \lambda \|B_n(d_{n+1}) - d_{n+1}\|_{\mathcal{H}(D)}^2}.$$

(iii) *Update the coefficients as follows:*

$$\alpha_i^{(n+1)} = \alpha_i^{(n)} - \alpha_{n+1} \beta_i^{(n)}(d_{n+1}) \quad \text{for } i = 1, \ldots, n,$$
$$\alpha_{n+1}^{(n+1)} = \alpha_{n+1}$$

and set $F_{n+1} = \sum_{i=1}^{n+1} \alpha_i^{(n+1)} d_i$ as well as $R^{n+1} = y - \mathcal{F}F_{n+1}$.

(iv) *If the stopping criterion is satisfied, then use F_{n+1} as an approximate solution to $\mathcal{F}F = y$. Otherwise, increase n by 1 and go to step* (ii).

Obviously, a normalization of the dictionary elements to $\|d\|_{\mathcal{H}(D)} = 1$ does not yield an improvement for the implementation in the case of the ROFMP.

Note that the orthogonal projection becomes more and more expensive with an increasing number n. For this reason, it is advisable to restart the algorithm after a certain number of steps N by using $y - \mathcal{F}F_N$ as the new data vector to be approximated and recounting from $n = 0$. Due to the linearity of $\mathcal{F}$, the consecutively produced approximations can be summed up in the end to obtain an approximation of the solution F of $\mathcal{F}F = y$. It turned out to be useful to keep, after each restart, the previous approximation $F_N = \tilde{F}$ in the regularization term and to regularize with $\|\tilde{F} + F_n\|_{\mathcal{H}(D)}^2$, where (F_n) is the approximating sequence after the restart.

For details of the implementation, see [95]. For numerical experiments and theoretical results, see [73, 95]. Note that, in the non-regularized case ($\lambda = 0$), the algorithm is able to produce an exact solution of $\mathcal{F}F = y$ in at most l steps, where $y \in \mathbb{R}^l$.

5. Conclusion

For the last decades, the possibilities and challenges which have presented themselves to geodesists have changed dramatically. Due to tremendously increased precisions in measurement technologies and the availability of satellite missions, huge

amounts of highly accurate data related to the Earth have become available. This has opened previously unexpected options for observing, analyzing and predicting the processes of the Earth system. Such progresses can be seen in manifold ways, for example when the ocean dynamics can be understood better, when the mass transports due to climate change or seasonal climatic phenomena can be better quantified and localized, when static and dynamic models of the Earth's interior can be validated and improved by a more precise model or when unprecedented ways of determining heights become available to geodesists.

Since mathematics plays a central role in the processes of, e.g., denoising, analyzing or inverting geoscientific data, the changes in the data situation can be mapped to changes in the requirements on the methodologies in mathematical geodesy (see also [30]). In this paper, we focussed on the uncertainty principle of spherical signal analysis which tells us that precise localization in space and in frequency/momentum are mutually exclusive. Moreover, we can interpret the uncertainty principle as a fundamental property of a spectrum ranging from ideal frequency localization (i.e., no space localization) to ideal space localization (i.e., no frequency localization). The former is associated to the use of spherical harmonics, which have been a common choice as basis system in geodesy. Away from this extremal case, in order that trial functions possess a space localization, they need to be sums of several spherical harmonics. The closer we come to the latter end of the spectrum with ideal space localization, the more spherical harmonics degrees have to be summed up in a trial function leading, as a limit, finally to the (only as a theoretical concept existing) Dirac functional which includes all degrees.

The aforementioned new challenges due to today's data situation can be reflected in this spectrum. In former days, when only a few data were available which allowed a very coarse global modeling only, spherical harmonics were the ultimate and reasonable choice. Today, the demands on highly accurate models which are, in particular, provided with a very high resolution in space define the limits of the use of spherical harmonics. These models can be better constructed with trial functions which combine certain extents of space and frequency localization.

As we have shown, there are many facets of localized trial functions which can be positioned in the spectrum of space and momentum localization. They include basis functions generated from (reproducing) kernels of particular function spaces. Such tools have successfully been used for spline and wavelet approximations in the geosciences. They leave sufficient degrees of freedom to control their variance in space and momentum. Furthermore, also Slepian functions provide another equally valuable tool for regionally approximating or analyzing a signal. They provide us with an orthonormal basis which is, in contrast to spherical harmonics, spaceconcentrated (to a region which can be arbitrarily chosen). Moreover, the Slepian functions are also orthogonal in the L^2-space of the chosen region, which is essentially useful for the modeling of a signal which is only regionally available. Furthermore, Slepian functions can also be not only spaceconcentrated but even spacelimited with the price (due to the uncertainty principle) that they become nonbandlimited, i.e., they sum up an infinite number of spherical harmonics degrees.

Certainly, there exist many other systems of trial functions on the sphere but also the ball, which have their own characteristics regarding space- and frequency localization. We added some references to other methods in appropriate paragraphs but do not claim to have provided a complete overview. In general, a wide range of special functions systems is available for the analysis of geoscientific data. However, it appears that, still, the main focus of (too) many research projects in geodesy and other disciplines of Earth sciences lies on the data alone but not on the choice of the methodology for their handling.

In this paper, we have tried to break new synoptical ground in dealing with spherical harmonics based special function systems and their role in constructive approximation methods of mathematical geodesy. We have presented a short insight and guide for the zoo of spherical trial functions to encourage geoscientists to question the mathematical basis functions which they use for their models and not to use mathematical tools as 'blackboxes'. We have also summarized briefly the possibility that regularized functional matching pursuit and its variants yield as algorithms for generating a kind of a best basis out of a selection of different basis systems.

Further research on finding the 'optimal' basis system for particular problems in mathematical geodesy has to be done. However, the present state-of-the-art shows that there is a high potential in improving (not only) geodetic models by using sophisticated mathematical methodologies. Obviously, our work as presented here is selective, but not only with respect to the choice of discussed basis functions. Also, not all details on the treated topics could be discussed up to an appropriate extent. For example, most of the proofs have been left out completely, so that the interested reader is referred to the attached list of literature. Nonetheless, we believe that we have provided a deeper insight on how geoscientific and, particularly, geodetic problems can be attacked in a mathematically systematic and rigorous way.

References

[1] Amirbekyan, A. (2007) The Application of Reproducing Kernel Based Spline Approximation to Seismic Surface and Body Wave Tomography: Theoretical Aspects and Numerical Results. Ph.D.-thesis, Geomathematics Group, Dept. Mathematics, University of Kaiserslautern

[2] Amirbekyan, A., Michel, V. (2008) Splines on the Three-dimensional Ball and Their Application to Seismic Body Wave Tomography, Inverse Probl., 24, 015022 (25pp)

[3] Amirbekyan, A., Michel, V., Simons, F.J. (2008) Parameterizing Surface-wave Tomographic Models with Harmonic Spherical Splines, Geophys. J. Int., 174, pp. 617–628

[4] Antoine, J.P., Demanet, L., Jaques, L., Vandergheynst, P. (2002) Wavelets on the Sphere: Implementations and Approximations, Appl. Comput. Harmon. Anal., 13, pp. 177–200

[5] Antoine, J.P., Vandergheynst, P. (1999) Wavelets on the 2-Sphere: A Group-theoretic Approach, Appl. Comput. Harmon. Anal., 7, pp. 1–30

[6] Backus, G.E. (1966) Potentials for Tangent Tensor Fields on Spheroids, Arch. Ration. Mech. Anal., 22, pp. 210–252

[7] Backus, G.E. (1967) Converting Vector and Tensor Equations to Scalar Equations in Spherical Coordinates, Geophys. J.R. Astron. Soc., 13, pp. 61–101

[8] Backus, G.E. (1986) Poloidal and Toroidal Fields in Geomagnetic Field Modelling, Reviews of Geophysics, 24, pp. 75–109

[9] Ballani, L., Engels, J., Grafarend, E.W. (1993) Global Base Functions for the Mass Density in the Interior of a Massive Body (Earth), Manuscr. Geod., 18, pp. 99–114

[10] Berkel, P. (2009) Multiscale Methods for the Combined Inversion of Normal Mode and Gravity Variations. Ph.D.-thesis, Geomathematics Group, Dept. Mathematics, University of Kaiserslautern, Shaker, Aachen

[11] Berkel, P., Michel, V. (2010) On Mathematical Aspects of a Combined Inversion of Gravity and Normal Mode Variations by a Spline Method, Math. Geosci., 42, pp. 795–816

[12] Berkel, P., Fischer, D., Michel, V. (2011) Spline Multiresolution and Numerical Results for Joint Gravitation and Normal Mode Inversion with an Outlook on Sparse Regularisation, Int. J. Geomath., 1, pp. 167–204

[13] Blick, C., Freeden, W. (2011) Spherical Spline Application to Radio Occultation Data, J. Geodetic Science, 1, 379–396

[14] Bluhm, R., Kostelecký, V.A., Tudose, B. (1995) Elliptical Squeezed States and Rydberg Wave Packets, Phys. Rev. A, 52, pp. 2234–2244

[15] Carruthers, P., Nieto, M.M. (1968) Phase and Angle Variables in Quantum Mechanics, Rev. Mod. Phys., 40, pp. 411–440

[16] Clebsch, R.F.A. (1862) Ueber eine Eigenschaft der Kugelfunctionen. J. Reine Angew. Math., 1862(60), pp. 343–350

[17] Dahlen, F.A., Simons, F.J. (2008) Spectral Estimation on a Sphere in Geophysics and Cosmology, Geoph. J. Int., 174, pp. 774–807

[18] Dahlke, S., Dahmen, W., Schmitt, W., Weinreich, I. (1995) Multiresolution Analysis and Wavelets on S^2 and S^3, Numer. Funct. Anal. Optimiz., 16, pp. 19–41

[19] Dahlke, S., Maass, P. (1996) Continuous Wavelet Transforms with Applications to Analyzing Functions on Spheres, J. Fourier Anal. Appl., 2, pp. 379–396

[20] Daubechies, I. (1992) Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, 61, SIAM, Philadelphia

[21] Eshagh, M. (2009) Spatially Restricted Integrals in Gradiometric Boundary Value Problems, Artif. Sat., 44, pp. 131–148

[22] Fengler, M., Michel, D., Michel, V. (2006) Harmonic Spline-wavelets on the 3-dimensional Ball and Their Application to the Reconstruction of the Earth's Density Distribution From Gravitational Data at Arbitrarily Shaped Satellite Orbits, ZAMM-Z. Angew. Math. Me., 86, pp. 856–873

[23] Fischer, D. (2011) Sparse Regularization of a Joint Inversion of Gravitational Data and Normal Mode Anomalies. Ph.D.-thesis, Geomathematics Group, Dept. Mathematics, University of Siegen, Verlag Dr. Hut, Munich

[24] Fischer, D., Michel, V. (2012) Sparse Regularization of Inverse Gravimetry – Case Study: Spatial and Temporal Mass Variations in South America. Inverse Probl., 28, 065012 (34pp.)

[25] Fischer, D., Michel, V. (2013a) Automatic Best-basis Selection for Geophysical Tomographic Inverse Problems. Geophys. J. Int., 193, pp. 1291–1299

[26] Fischer, D., Michel, V. (2013b) Inverting GRACE Gravity Data for Local Climate Effects, J. Geodetic Science, 3, pp. 151–162

[27] Fokas, A.S., Hauk, O., Michel, V. (2012) Electro-Magneto-Encephalography for the Three-shell Model: Numerical Implementation via Splines for Distributed Current in Spherical Geometry, Inverse Probl., 28, 035009 (28pp.)

[28] Freeden, W. (1981) On Spherical Spline Interpolation and Approximation, Math. Meth. in the Appl. Sci., 3, pp. 551–575

[29] Freeden, W. (1999) Multiscale Modelling of Spaceborne Geodata, B.G. Teubner, Leipzig

[30] Freeden, W. (2015) Geomathematics: Its Role, Its Aim, and Its Potential, in: Handbook of Geomathematics (W. Freeden, M.Z. Nashed, and T. Sonar, Eds.), 2nd ed., Springer, Berlin, Heidelberg, pp. 3–79

[31] Freeden, W., Gerhards, C. (2013) Geomathematically Oriented Potential Theory. Chapman and Hall, CRC Press, Boca Raton, London, New York

[32] Freeden, W., Gutting, M. (2013) Special Functions of Mathematical (Geo)Physics. Birkhäuser, Basel

[33] Freeden, W., Maier, T. (2002) On Multiscale Denoising of Spherical Functions: Basic Theory and Numerical Aspects, Electr. Transact. on Numer. Anal. (ETNA), 14, pp. 40–62

[34] Freeden, W., Maier, T. (2003) Spectral and Multiscale Signal-to-Noise Thresholding of Spherical Vector Fields, Comput. Geosciences, 7, pp. 215–250

[35] Freeden, W., Michel, V. (2000) Least-Squares Geopotential Approximation by Windowed Fourier Transform and Wavelet Transform, in: Wavelets in the Geosciences (R. Klees and R. Haagmans, Eds.), Lecture Notes in Earth Sciences, Springer, 90, pp. 189–241

[36] Freeden, W., Michel, V. (2004) Multiscale Potential Theory (with Applications to Geoscience), Birkhäuser Verlag, Boston, Basel, Berlin

[37] Freeden, W., Nutz, H. (2011) Satellite Gravity Gradiometry as Tensorial Inverse Problem, Int. J. Geomath., 2, pp. 177–218

[38] Freeden, W., Schneider, F. (1998) Regularization Wavelets and Multiresolution, Inverse Probl., 14, pp. 225–243

[39] Freeden, W., Schreiner, M. (1995) Non-Orthogonal Expansions on the Sphere, Math. Meth. Appl. Sci., 18, pp. 83–120

[40] Freeden, W., Schreiner, M. (1998) Orthogonal and Non-Orthogonal Multiresolution Analysis, Scale Discrete and Exact Fully Discrete Wavelet Transform on the Sphere. Constr. Approx., 14, pp. 493–515

[41] Freeden, W., Schreiner, M. (2006), Local Multiscale Modelling of Geoid Undulations from Deflections of the Vertical, J. Geodesy, 79, 641–651

[42] Freeden, W., Schreiner, M. (2007) Biorthogonal Locally Supported Wavelets on the Sphere Based on Zonal Kernel Functions, J. Fourier Anal. Appl., 13, pp. 693–709

[43] Freeden, W., Schreiner, M. (2009) Spherical Functions of Mathematical Geosciences – A Scalar, Vectorial, and Tensorial Setup, Springer, Berlin

[44] Freeden, W., Windheuser, U. (1996) Spherical Wavelet Transform and its Discretization, Adv. Comput. Math., 5, pp. 51–94

[45] Freeden, W., Windheuser, U. (1997) Combined Spherical Harmonic and Wavelet Expansion. Appl. Comput. Harmon. Anal., 4, pp. 1–37

[46] Freeden, W., Witte, B. (1982) A Combined (Spline-) Interpolation and Smoothing Method for the Determination of the External Gravitational Potential from Heterogeneous Data, Bull. Geod., 56, pp. 53–62

[47] Freeden, W., Gervens, T., Schreiner, M. (1998) Constructive Approximation on the Sphere (with Applications to Geomathematics), Oxford Science Publications, Clarendon, Oxford

[48] Freeden, W., Michel, V., Nutz, H. (2002) Satellite-to-Satellite Tracking and Satellite Gravity Gradiometry (Advanced Techniques for High-Resolution Geopotential Field Determination), J. Eng. Math., 43, pp. 19–56

[49] Freeden, W., Fehlinger, T., Klug, M., Mathar, D., and Wolf, K. (2009) Classical Globally Reflected Gravity Field Determination in Modern Locally Oriented Multiscale Framework, J. Geodesy, 83, pp. 1171–1191

[50] Gauß, C.F. (1838) Allgemeine Theorie des Erdmagnetismus, Resultate aus den Beobachtungen des magnetischen Vereins, Göttingen

[51] Grünbaum, F.A., Longhi, L., Perlstadt, M. (1982) Differential Operators Commuting with Finite Convolution Integral Operators: Some Non-Abelian Examples, SIAM J. Appl. Math., 42, pp. 941–955

[52] Guilloux, F., Faÿ, G., Cardoso, J.F. (2009) Practical Wavelet Design on the Sphere, Appl. Comput. Harmon. Anal., 26, pp. 143–160

[53] Haar, A. (1910) Zur Theorie der orthogonalen Funktionensysteme, Math. Ann., 69, pp. 331–371

[54] Heine, E. (1878) Handbuch der Kugelfunktionen, Reimer, Berlin

[55] Holschneider M (1996) Continuous Wavelet Transforms on the Sphere, J. Math. Phys., 37, pp. 4156–4165

[56] Kennedy, R.A. and Sadeghi, P. (2013) Hilbert Space Methods in Signal Processing, Cambridge University Press, Cambridge, UK

[57] Khalid, Z., Durrani, S., Sadeghi, P., Kennedy, R.A. (2012) Spatio-spectral Analysis on the Sphere Using Spatially Localized Spherical Harmonics Transform, IEEE Trans. Signal Process., 60, pp. 1487–1492

[58] Kowalski, K., Rembieliński, J. (2000) Quantum Mechanics on a Sphere and Coherent States, J. Phys. A: Math. Gen., 33, pp. 6035–6048

[59] Laín Fernández, N. (2003) Polynomial Bases on the Sphere, Ph.D.-thesis, University of Lübeck

[60] Landau, H.J. and Pollak, H.O. (1961) Prolate spheroidal wave functions, Fourier analysis and uncertainty – II, Bell Syst. Tech. J., 40, pp. 65–84

[61] Laplace, P.S. de (1785) Théorie des Attractions des Sphéroides et de la Figure des Planètes, Mém. de l'Acad., Paris

[62] Legendre, A.M. (1785) Recherches sur l'Attraction des Sphéroides Homogènes, Mém. math. phys. prés. à l'Acad. Aci. par divers savants, 10, pp. 411–434

[63] Leistedt, B., McEwen, J.D. (2012) Exact Wavelets on the Ball, IEEE Trans. Signal Process., 60, pp. 6257–6269

[64] Lyche, T., Schumaker, L. (2000) A Multiresolution Tensor Spline Method for Fitting Functions on the Sphere, SIAM J. Sci. Comput., 22, pp. 724–746

[65] Mallat, S.G., Zhang, Z. (1993) Matching Pursuits with Time-Frequency Dictionaries, IEEE Trans. Signal Process., 41, pp. 3397–3415

[66] Maxwell, J.C. (1891) A Treatise on Electricity and Magnetism (1873, 1881, 1891) Bd. 1 und 2 Ungekürzter Nachdruck der letzten Auflage (1954), Dover, New York

[67] Michel, V. (2013) Lectures on Constructive Approximation – Fourier, Spline, and Wavelet Methods on the Real Line, the Sphere, and the Ball, Birkhäuser, Boston

[68] Michel, V. (2015a) Tomography – Problems and Multiscale Solutions, in: Handbook of Geomathematics (W. Freeden, M.Z. Nashed, and T. Sonar, Eds.), 2nd ed., Springer, Berlin, Heidelberg, pp. 2087–2119

[69] Michel, V. (2015b) RFMP – An Iterative Best Basis Algorithm for Inverse Problems in the Geosciences, in: Handbook of Geomathematics (W. Freeden, M.Z. Nashed, and T. Sonar, Eds.), 2nd ed., Springer, Berlin, Heidelberg, pp. 2121–2147

[70] Michel, V., Orzlowski, S. (2016) On the Null Space of a Class of Fredholm Integral Equations of the First Kind, J. Inverse and Ill-Posed Problems, 24, pp. 687–710

[71] Michel, V., Orzlowski, S. (2017) On the Convergence Theorem for the Regularized Functional Matching Pursuit (RFMP) Algorithm, Int. J. Geomath., 8, pp. 183–190

[72] Michel, V., Telschow, R. (2014) A Non-linear Approximation Method on the Sphere, Int. J. Geomath., 5, pp. 195–224

[73] Michel, V., Telschow, R. (2016) The Regularized Orthogonal Functional Matching Pursuit for Ill-posed Inverse Problems, SIAM J. Numer. Anal., 54, pp. 262–287

[74] Michel, V., Wolf, K. (2008) Numerical Aspects of a Spline-based Multiresolution Recovery of the Harmonic Mass Density out of Gravity Functionals, Geophys. J. Int., 173, pp. 1–16

[75] Morse, P.M., Feshbach, H. (1953) Methods of Theoretical Physics, McGraw-Hill, New York

[76] Narcowich, F.J., Ward, J.D. (1996) Nonstationary Wavelets on the m-Sphere for Scattered Data, Appl. Comput. Harmon. Anal., 3, pp. 324–336

[77] Neumann, F. (1887) Vorlesungen über die Theorie des Potentials und der Kugelfunktionen, Teubner, Leipzig, pp. 135–154

[78] Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.S. (1993) Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition, in: Asilomar Conference on Signals, Systems and Computers, IEEE Conf. Pub., 1, pp. 40–44

[79] Plattner, A., Simons, F.J. (2014) Spatiospectral Concentration of Vector Fields on a Sphere, Appl. Comput. Harmon. Anal., 36, pp. 1–22

[80] Plattner, A., Simons, F.J. (2015a) Potential-Field Estimation Using Scalar and Vector Slepian Functions at Satellite Altitude, in: Handbook of Geomathematics (W. Freeden, M.Z. Nashed, and T. Sonar, Eds.), 2nd ed., Springer, Berlin, Heidelberg, pp. 2003–2055

[81] Plattner, A., Simons, F.J. (2015b) High-Resolution Local Magnetic Field Models for the Martian South Pole From Mars Global Surveyor Data, J. Geophys. Res., pp. 1543–1566

[82] Potts, D., Tasche, M. (1995) Interpolatory Wavelets on the Sphere, in: Approximation Theory VIII (C.K. Chui, L.L. Schumaker, Eds.), World Scientific, Singapore, pp. 335–342

[83] Schröder, P., Sweldens, W. (1995) Spherical Wavelets: Efficiently Representing Functions on the Sphere, in: Computer Graphics Proceedings (SIGGRAPH95), pp. 161–175

[84] Simons, F.J. (2010) Slepian Functions and Their Use in Signal Estimation and Spectral Analysis, in: Handbook of Geomathematics (W. Freeden, M.Z. Nashed, and T. Sonar, Eds.), 1st ed., Springer, Heidelberg, pp. 891–923

[85] Simons, F.J., Dahlen, F.A. (2006) Spherical Slepian Functions and the Polar Gap in Geodesy, Geoph. J. Int., 166, pp. 1039–1061

[86] Simons, F.J., Plattner, A. (2015) Scalar and Vector Slepian Functions, Spherical Signal Estimation and Spectral Analysis, in: Handbook of Geomathematics (W. Freeden, M.Z. Nashed, and T. Sonar, Eds.), 2nd ed., Springer, Berlin, Heidelberg, pp. 2563–2608

[87] Simons, F.J., Dahlen, F.A., Wieczorek, M. (2006) Spatiospectral Concentration on a Sphere, SIAM Rev., 48, pp. 504–536

[88] Simons, F.J., Hawthorne, J.C., Beggan, C.D. (2009) Efficient Analysis and Representation of Geophysical Processes Using Localized Spherical Basis Functions, in: Wavelets XIII, Proc. SPIE, (V. Goyal, M. Papadakis and D. Van de Ville, Eds.), 7446(74460G)

[89] Simons, F.J., Loris, I., Brevdo, E., Daubechies, I.C. (2011) Wavelets and Wavelet-like Transforms on the Sphere and Their Application to Geophysical Data Inversion, in: Wavelets and Sparsity XIV, Proc. SPIE, (M. Papadakis, V. Goyal, and D. Van de Ville, Eds.), 8138(81380X)

[90] Slepian, D. (1983) Some Comments on Fourier Analysis, Uncertainty and Modeling, SIAM Rev., 25, pp. 379–393

[91] Slobbe, D.C., Simons, F.J., Klees, R. (2012) The Spherical Slepian Basis as a Means to Obtain Spectral Consistency Between Mean Sea Level and the Geoid, J. Geod., 86, pp. 609–628

[92] Svensson, S.L. (1983) Pseudodifferential Operators – A New Approach to the Boundary Value Problems of Physical Geodesy, Manuscr. Geod., 8, pp. 1–40

[93] Sylvester, T. (1876) Note on Spherical Harmonics, Phil. Mag., II, 291 and 400

[94] Tegmark, M. (1996) A Method for Extracting Maximum Resolution Power Spectra from Microwave Sky Maps, Mon. Not. Roy. Astron. Soc., 280, pp. 299–308

[95] Telschow, R. (2015) An Orthogonal Matching Pursuit for the Regularization of Spherical Inverse Problems, Ph.D.-thesis, Geomathematics Group, Dept. Mathematics, University of Siegen, Verlag Dr. Hut, Munich

[96] Temlyakov, V.N. (2003) Nonlinear Methods of Approximation, Found. Comput. Math. 3, pp. 33–107

[97] Tscherning, C.C. (1996) Isotropic Reproducing Kernels for the Inner of a Sphere or Spherical Shell and Their Use as Density Covariance Functions. Math. Geol. 28, pp. 161–168

[98] Vincent, P., Bengio, Y. (2002) Kernel Matching Pursuit, Mach. Learn., 48, pp. 169–191

[99] Wahba, G. (1981) Spline Interpolation and Smoothing on the Sphere, SIAM J. Sci. Stat. Comput., 2, pp. 5–16, (also errata: SIAM J. Sci. Stat. Comput., 3, pp. 385–386)

[100] Wahba, G. (1990) Spline Models for Observational Data, CBMS-NSF Regional Conference Series in Applied Mathematics, 59, SIAM, Philadelphia

[101] Weinreich, I. (2001) A Construction of $C^{(1)}$-Wavelets on the Two-dimensional Sphere, Appl. Comput. Harmon. Anal., 10, pp. 1–26

[102] Wicht, J., Stellmach, S., Harder, H. (2010) Numerical Dynamo Simulations: From Basic Concepts to Realistic Models, in: Handbook of Geomathematics (W. Freeden, M.Z. Nashed, and T. Sonar, Eds.), 1st ed., pp. 459–502, Springer, Heidelberg

[103] Wieczorek, M.A. (2015) The Gravity and Topography of the Terrestrial Planets, in: Treatise on Geophysics (T. Spohn, ed.), 2nd ed., vol. 10, pp. 153–2193, Elsevier, Amsterdam

[104] Wieczorek, M.A., Simons, F.J. (2005) Localized Spectral Analysis on the Sphere, Geophys. J. Int., 162, pp. 655–675

Willi Freeden
Geomathematics Group
University of Kaiserslautern
MPI-Gebäude, Paul-Ehrlich-Str. 26
D-67663 Kaiserslautern, Germany

Volker Michel
Geomathematics Group
University of Siegen
Walter-Flex-Str. 3
D-57068 Siegen, Germany

Frederik J. Simons
Princeton University
Department of Geosciences
Princeton, NJ, USA

Handbook of Mathematical Geodesy
Geosystems Mathematics, 821–853

Spherical Potential Theory: Tools and Applications

Christian Gerhards

Abstract. In the current chapter, we transfer classical potential theoretic concepts from the Euclidean space $\mathbb{R}^3$ to a setting intrinsic on the sphere. We present uniqueness results for the Poisson equation on the sphere, explicitly construct Green functions for spherical caps and complete function systems for harmonic approximation, and elaborate on decompositions of vector fields on the sphere. Among the intended applications are problems from oceanography, geodesy, and geomagnetism. Some examples are presented at the end.

Keywords. Green's functions, Helmholtz decomposition, Hardy–Hodge decomposition, fundamental solutions, complete function systems, boundary value problems, surface potentials, curve potentials,limit- and jump-relations, Green's formulas, vertical deflections, disturbing potential, multiscale approximation, geostrophic flow, mean dynamic topography, vortex motion.

Classical potential theoretic concepts in the Euclidean space $\mathbb{R}^3$ have been described in an earlier chapter of this handbook. They appear frequently in geodesy when treating the harmonic gravitational potential in the exterior of the spherical Earth. The sphere $\Omega_R = \{x \in \mathbb{R}^3 : |x| = R\}$ occurs as the boundary surface of a subdomain in $\mathbb{R}^3$. Opposed to this, in the present chapter, the sphere is not representing a boundary surface, it is rather regarded as the underlying domain on which a problem is formulated. Examples for this are the spherical Navier–Stokes equations and shallow water equations in meteorology and ocean modeling (see, e.g., [5, 12, 19, 29, 42, 44]). But also simpler spherical differential equations occur in geodesy and geomagnetism (see, e.g., [2, 10, 11, 14, 16, 23, 25]) and vortex dynamics (see, e.g., [32, 33, 42–44]), more precisely, those based on the Beltrami operator Δ^* (the spherical counterpart to the Laplace operator Δ). Latter is going to be the focus of this chapter. In particular, we are interested in the Beltrami equation on subdomains $\Gamma_R \subset \Omega_R$ of the sphere, which eventually leads to potential theoretic concepts analogous to those of the Euclidean case. Subdomains appear naturally, e.g., due to only regionally available data or coastal/continental boundaries. The problems we take a closer look at are the following (note that

Ω and Γ are simply abbreviations for the unit sphere Ω_1 and a corresponding subdomain Γ_1):

Poisson Problem (PP): Let H be of class $\mathrm{C}^{(1)}(\overline{\Gamma})$. We are looking for a function U of class $\mathrm{C}^{(2)}(\Gamma)$ such that

$$\Delta^* U(\xi) = H(\xi), \quad \xi \in \Gamma. \tag{1}$$

Dirichlet Problem (DP): Let F be of class $\mathrm{C}^{(0)}(\partial\Gamma)$. We are looking for a function U of class $\mathrm{C}^{(2)}(\Gamma) \cap \mathrm{C}^{(0)}(\overline{\Gamma})$ such that

$$\Delta^* U(\xi) = 0, \qquad \xi \in \Gamma, \tag{2}$$

$$U^-(\xi) = F(\xi), \quad \xi \in \partial\Gamma. \tag{3}$$

Neumann Problem (NP): Let F be of class $\mathrm{C}^{(0)}(\partial\Gamma)$. We are looking for a function U of class $\mathrm{C}^{(2)}(\Gamma) \cap \mathrm{C}^{(0)}(\overline{\Gamma})$, with a well-defined normal derivative $\frac{\partial}{\partial\nu}U^-$ on $\partial\Gamma$, such that

$$\Delta^* U(\xi) = 0, \qquad \xi \in \Gamma, \tag{4}$$

$$\frac{\partial}{\partial\nu}U^-(\xi) = F(\xi), \quad \xi \in \partial\Gamma. \tag{5}$$

In the setting above, $\partial\Gamma$ denotes the boundary curve of Γ, $\nu(\xi)$ the outward directed unit normal vector at $\xi \in \partial\Gamma$, and $\frac{\partial}{\partial\nu}$ the corresponding normal derivative. The minus of U^- simply indicates that we are approaching the boundary $\partial\Gamma$ from within Γ.

Certainly, the problems above and its potential theoretic consequences can be and have been treated on more general manifolds than the sphere (e.g., in [8, 37, 38]). However, we focus on the geophysically relevant case of the sphere where explicit representations of the fundamental solution and some Green's functions are known. In large parts, we follow the course of [15] and emphasize similarities and differences to the Euclidean case.

The first section supplies the reader with necessary notations and several mathematical tools related to spherical potential theory. In Section 2, we treat the problems (PP), (DP), and (NP). In particular, we are interested in integral representations of their solutions. In Section 3, we turn towards spherical differential operators of order one, namely the surface gradient ∇^* (the spherical counterpart to the gradient ∇) and the surface curl gradient L^*. We investigate the corresponding differential equations on Γ as well as the so-called spherical Helmholtz decomposition and the spherical Hardy–Hodge decomposition. Section 4 comments briefly on complete function systems and approximation methods on the sphere. Finally, in Section 5, applications of the previous concepts to some geophysical problems are discussed, namely, vertical deflections, (geostrophic) ocean flow, and a toy problem for point vortex motion.

1. Fundamental tools

Of fundamental importance to us is the Beltrami operator Δ^* which denotes the tangential contribution to the Euclidean Laplace operator Δ. More precisely,

$$\Delta_x = \frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2}\Delta_\xi^*, \tag{6}$$

where Δ_x acts on $x \in \mathbb{R}^3$ while Δ_ξ^* acts on $\xi = \frac{x}{|x|} \in \Omega$. The length $|x|$ is usually denoted by r. Furthermore, ∇^* stands for the (spherical) surface gradient, which denotes the tangential contribution to the gradient ∇:

$$\nabla_x = \xi \frac{\partial}{\partial r} + \frac{1}{r}\nabla_\xi^*. \tag{7}$$

The occasionally occurring (spherical) surface curl gradient L^* acts via $\xi \wedge \nabla_\xi^*$ at a point $\xi \in \Omega$ ("$\wedge$" denotes the vector product). It should be noted that $\Delta^* = \nabla^* \cdot \nabla^* = \mathrm{L}^* \cdot \mathrm{L}^*$ ("$\cdot$" denotes the Euclidean inner product). If it is clear on which variables the operators act, we usually omit the subindices ξ and x. For convenience, we typically use Greek letters ξ, η to indicate unit vectors in Ω while we use latin letters x, y for general vectors in $\mathbb{R}^3$. Upper case letters F, G denote scalar-valued functions mapping $\Gamma \subset \Omega$ into $\mathbb{R}$ while lower case letters f, g denote vector-valued functions mapping $\Gamma \subset \Omega$ into $\mathbb{R}^3$. Correspondingly, the set of k-times continuously differentiable scalar-valued functions on Γ is designated by $\mathrm{C}^{(k)}(\Gamma)$ and the set of k-times continuously differentiable vector-valued functions on Γ by $c^{(k)}(\Gamma)$. The closure of Γ is denoted by $\overline{\Gamma}$ and the open complement by $\Gamma^c = \Omega \setminus \overline{\Gamma}$.

Whenever we talk about subdomains $\Gamma \subset \Omega$ in this chapter, we mean, without further mention, regular regions, i.e., subdomains with a sufficiently smooth boundary curve $\partial\Gamma$ (for details, the reader is referred to [15]; an exemplary illustration is supplied in Figure 1). For such regular regions, the positively oriented unit tangential vector $\tau(\xi)$ at a point $\xi \in \partial\Gamma$ is well defined. The unit normal vector $\nu(\xi)$ at $\xi \in \partial\Gamma$ points into the exterior of Γ and is perpendicular to $\tau(\xi)$

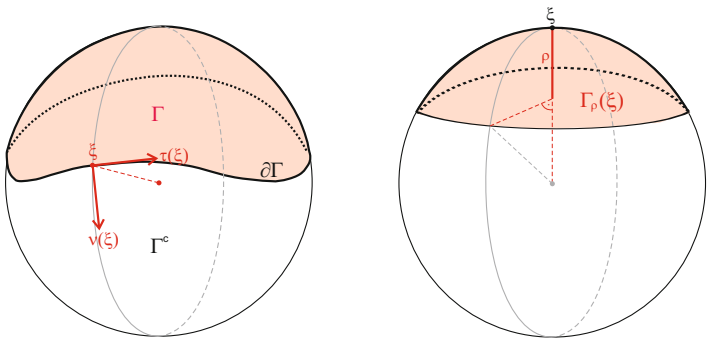


FIGURE 1. Examples for a general regular regionΓ (left) and a spherical cap $\Gamma_\rho(\xi)$ with center ξ and radius ρ (right).

and ξ (i.e., $\nu(\xi)$ is perpendicular to the boundary curve $\partial\Gamma$ but tangential to the unit sphere Ω). The normal derivative of a scalar-valued function F at $\xi \in \partial\Gamma$ is defined as

$$\frac{\partial}{\partial\nu}F(\xi) = \nu(\xi) \cdot \nabla^*_\xi F(\xi). \tag{8}$$

1.1. Green's formulas

We frequently need integral expressions that describe the shifting of differential operators from one integrand to another, so-called Green formulas. Some spherical versions are stated in the next theorem.

Theorem 1.1 (Spherical Green Formulas I).
(a) *If f is of class $\mathrm{c}^{(1)}(\overline{\Gamma})$ and tangential, i.e., $\xi \cdot f(\xi) = 0$ for $\xi \in \overline{\Gamma}$, then*

$$\int_\Gamma \nabla^* \cdot f(\eta) d\omega(\eta) = \int_{\partial\Gamma} \nu(\eta) \cdot f(\eta) d\sigma(\eta), \tag{9}$$

$$\int_\Gamma \mathrm{L}^* \cdot f(\eta) d\omega(\eta) = \int_{\partial\Gamma} \tau(\eta) \cdot f(\eta) d\sigma(\eta). \tag{10}$$

(b) *If F is of class $\mathrm{C}^{(1)}(\overline{\Gamma})$ and f of class $\mathrm{c}^{(1)}(\overline{\Gamma})$, then*

$$\begin{aligned}\int_\Gamma f(\eta) \cdot \nabla^* F(\eta) d\omega(\eta) &+ \int_\Gamma F(\eta) \nabla^* \cdot f(\eta) d\omega(\eta) \\ &= \int_{\partial\Gamma} \nu(\eta) \cdot (F(\eta) f(\eta))\, d\sigma(\eta) + 2\int_\Gamma \eta \cdot (F(\eta) f(\eta))\, d\omega(\eta),\end{aligned} \tag{11}$$

$$\begin{aligned}\int_\Gamma f(\eta) \cdot \mathrm{L}^* F(\eta) d\omega(\eta) &+ \int_\Gamma F(\eta) \mathrm{L}^* \cdot f(\eta) d\omega(\eta) \\ &= \int_{\partial\Gamma} \tau(\eta) \cdot (F(\eta) f(\eta))\, d\sigma(\eta).\end{aligned} \tag{12}$$

(c) *If F, H are functions of class $\mathrm{C}^{(2)}(\overline{\Gamma})$, then*

$$\begin{aligned}\int_\Gamma F(\eta) \Delta^* H(\eta) d\omega(\eta) &- \int_\Gamma H(\eta) \Delta^* F(\eta) d\omega(\eta) \\ &= \int_{\partial\Gamma} F(\eta) \frac{\partial}{\partial\nu} H(\eta) d\sigma(\eta) - \int_{\partial\Gamma} H(\eta) \frac{\partial}{\partial\nu} F(\eta) d\sigma(\eta).\end{aligned} \tag{13}$$

Generally, '$d\omega$' denotes the surface element in $\Gamma \subset \Omega$ and '$d\sigma$' the line element on $\partial\Gamma$.

Remark 1.2. The formulas (11)–(13) are direct consequences of (9) and (10). Dropping the boundary terms $\int_{\partial\Gamma} \ldots d\sigma$, all of these formulas also hold true for the choice $\Gamma = \Omega$.

A crucial step for later considerations is the combination of Green's formulas with the *fundamental solution for the Beltrami operator* $G(\Delta^*; \cdot) : [-1, 1) \to \mathbb{R}$, which is uniquely determined by the following properties:

1. For any fixed $\xi \in \Omega$, the function $\eta \mapsto G(\Delta^*; \xi \cdot \eta)$ is twice continuously differentiable on $\Omega \setminus \{\xi\}$ and
$$\Delta^*_\eta G(\Delta^*; \xi \cdot \eta) = -\frac{1}{4\pi}, \quad \eta \in \Omega \setminus \{\xi\}. \tag{14}$$
2. For any fixed $\xi \in \Omega$, the function $\eta \mapsto G(\Delta^*; \xi \cdot \eta) - \frac{1}{4\pi} \ln(1 - \xi \cdot \eta)$ is continuously differentiable on Ω.
3. For any fixed $\xi \in \Omega$, it holds $\frac{1}{4\pi} \int_\Omega G(\Delta^*; \xi \cdot \eta) d\omega(\eta) = 0$.

Some basic calculations show that the function given by
$$G(\Delta^*; t) = \frac{1}{4\pi} \ln(1 - t) + \frac{1}{4\pi}(1 - \ln(2)), \quad t \in [-1, 1), \tag{15}$$
satisfies the properties (i)–(iii). The property (i) denotes the major difference between the fundamental solution for the Laplace operator $G(\Delta; \cdot)$ and its spherical counterpart. While $G(\Delta; \cdot)$ generates a 'true' Dirac distribution in the sense that $\Delta_y G(\Delta; |x - y|) = 0$, $y \in \mathbb{R}^3 \setminus \{x\}$, the fundamental solution $G(\Delta^*; \cdot)$ only generates a Dirac distribution up to an additive constant (reflecting the null space of the Beltrami operator Δ^*). Eventually, applying Green's formulas from Theorem 1.1, the properties of $G(\Delta^*; \cdot)$ lead to the following integral representations.

Theorem 1.3 (Spherical Green Formulas II).

(a) *If F is of class $\mathrm{C}^{(2)}(\overline{\Gamma})$, then we have for $\xi \in \Omega$,*
$$\begin{aligned}\frac{\alpha(\xi)}{2\pi} F(\xi) = \frac{1}{4\pi} \int_\Gamma F(\eta) d\omega(\eta) + \int_\Gamma G(\Delta^*; \xi \cdot \eta) \Delta^*_\eta F(\eta) d\omega(\eta) \\ + \int_{\partial\Gamma} F(\eta) \frac{\partial}{\partial \nu(\eta)} G(\Delta^*; \xi \cdot \eta) d\sigma(\eta) - \int_{\partial\Gamma} G(\Delta^*; \xi \cdot \eta) \frac{\partial}{\partial \nu(\eta)} F(\eta) d\sigma(\eta).\end{aligned} \tag{16}$$

(b) *If F is of class $\mathrm{C}^{(1)}(\overline{\Gamma})$, then we have for $\xi \in \Omega$,*
$$\begin{aligned}\frac{\alpha(\xi)}{2\pi} F(\xi) &= \frac{1}{4\pi} \int_\Gamma F(\eta) d\omega(\eta) - \int_\Gamma \nabla^*_\eta G(\Delta^*; \xi \cdot \eta) \cdot \nabla^*_\eta F(\eta) d\omega(\eta) \\ &\quad + \int_{\partial\Gamma} F(\eta) \frac{\partial}{\partial \nu(\eta)} G(\Delta^*; \xi \cdot \eta) d\sigma(\eta), \\ &= \frac{1}{4\pi} \int_\Gamma F(\eta) d\omega(\eta) - \int_\Gamma \mathrm{L}^*_\eta G(\Delta^*; \xi \cdot \eta) \cdot \mathrm{L}^*_\eta F(\eta) d\omega(\eta) \\ &\quad + \int_{\partial\Gamma} F(\eta) \frac{\partial}{\partial \nu(\eta)} G(\Delta^*; \xi \cdot \eta) d\sigma(\eta).\end{aligned} \tag{17}$$

The solid angle α of a regular region Γ is defined such that $\alpha(\xi) = 2\pi$ for $\xi \in \Gamma$, $\alpha(\xi) = \pi$ for $\xi \in \partial\Gamma$, and $\alpha(\xi) = 0$ for $\xi \in \Gamma^c$.

Remark 1.4. Again, dropping the boundary terms $\int_{\partial\Gamma} \ldots d\sigma$ in the expressions of Theorem 1.3 leads to results that hold true for the global choice $\Gamma = \Omega$.

Remark 1.5. Choosing $F \equiv 1$ in any of the formulas in Theorem 1.3 implies

$$\int_{\partial\Gamma} \frac{\partial}{\partial\nu(\eta)} G(\Delta^*; \xi\cdot\eta) d\sigma(\eta) = \begin{cases} 1 - \frac{\|\Gamma\|}{4\pi}, & \xi \in \Gamma, \\ \frac{1}{2} - \frac{\|\Gamma\|}{4\pi}, & \xi \in \partial\Gamma, \\ -\frac{\|\Gamma\|}{4\pi}, & \xi \in \Gamma^c, \end{cases} \tag{18}$$

where $\|\Gamma\|$ denotes the surface area of Γ. The behaviour of (18) across the boundary $\partial\Gamma$ states a first hint at the limit and jump relations of the layer potentials in Section 1.4. Apart from the additive constant $\frac{\|\Gamma\|}{4\pi}$, they are identical to the Euclidean setting.

1.2. Harmonic functions

In this subsection, we turn towards functions that are *harmonic* (*with respect to the Beltrami operator*) in Γ, i.e., functions U of class $\mathrm{C}^{(2)}(\Gamma)$ that satisfy

$$\Delta^* U = 0 \quad \text{in } \Gamma. \tag{19}$$

If no confusion with the Euclidean case is likely to arise, we just say that U is harmonic. Plugging such functions into (16), together with the choice of Γ being a spherical cap $\Gamma_\rho(\xi) = \{\eta \in \Omega : 1 - \xi\cdot\eta < \rho\}$ with center $\xi \in \Omega$ and radius $\rho \in (0, 2)$, we end up with the mean value property for harmonic functions.

Theorem 1.6 (Mean Value Property I). *A function U of class $\mathrm{C}^{(0)}(\Gamma)$ is harmonic if and only if*

$$U(\xi) = \frac{1}{4\pi} \int_{\Gamma_\rho(\xi)} U(\eta) d\omega(\eta) + \frac{\sqrt{2-\rho}}{4\pi\sqrt{\rho}} \int_{\partial\Gamma_\rho(\xi)} U(\eta) d\sigma(\eta), \quad \xi \in \Gamma, \tag{20}$$

for any spherical cap $\overline{\Gamma_\rho(\xi)} \subset \Gamma$.

The mean value property above contains the typical additive term for spherical problems. However, we can get rid of this additive constant when using Green's functions for spherical caps as described later on in Section 2.3. We are led to the following representation which resembles a Mean Value Property that is more closely related to the Euclidean case of functions that are harmonic with respect to the Laplace operator.

Theorem 1.7 (Mean Value Property II). *A function U of class $\mathrm{C}^{(0)}(\Gamma)$ is harmonic if and only if*

$$U(\xi) = \frac{1}{2\pi\sqrt{\rho(2-\rho)}} \int_{\partial\Gamma_\rho(\xi)} U(\eta) d\sigma(\eta), \quad \xi \in \Gamma, \tag{21}$$

for any spherical cap $\overline{\Gamma_\rho(\xi)} \subset \Gamma$.

Once a Mean Value Property is established, it can be used to derive a Maximum Principle. For details, we refer to [15] or, in the Euclidean case, any book on classical potential theory such as [17, 26, 30, 45].

Theorem 1.8 (Maximum Principle). *If U of class $\mathrm{C}^{(2)}(\Gamma) \cap \mathrm{C}^{(0)}(\overline{\Gamma})$ is harmonic, then*

$$\sup_{\xi \in \Gamma} |U(\xi)| \leq \sup_{\xi \in \partial\Gamma} |U(\xi)|. \tag{22}$$

1.3. Surface potentials

Analogous to the Euclidean setting, we can define a Newton potential and layer potentials for the spherical setting which take over the corresponding roles. The obvious difference is that now the Newton potential is a surface potential and the layer potentials represent curve potentials. Throughout this section, we take a closer look at the surface potential

$$U(\xi) = \int_\Gamma G(\Delta^*; \xi \cdot \eta) H(\eta) d\omega(\eta), \quad \xi \in \Omega, \tag{23}$$

From the properties of the fundamental solution for the Beltrami operator it becomes directly clear that U is of class $\mathrm{C}^{(2)}(\Gamma^c)$ and that

$$\Delta_\xi^* \int_\Gamma G(\Delta^*; \xi \cdot \eta) H(\eta) d\omega(\eta) = -\frac{1}{4\pi} \int_\Gamma H(\eta) d\omega(\eta), \quad \xi \in \Gamma^c. \tag{24}$$

Yet, the interesting question is what happens if $\xi \in \Gamma$, i.e., when the integration region contains the singularity of $G(\Delta^*; \cdot)$.

Theorem 1.9. *If H is of class $\mathrm{C}^{(0)}(\overline{\Gamma})$ and U is given by* (23)*, then U is of class $\mathrm{C}^{(1)}(\Omega)$ and*

$$\nabla_\xi^* \int_\Gamma G(\Delta^*; \xi \cdot \eta) H(\eta) d\omega(\eta) = \int_\Gamma \nabla_\xi^* G(\Delta^*; \xi \cdot \eta) H(\eta) d\omega(\eta), \quad \xi \in \Omega. \tag{25}$$

The proof of the theorem above can be based on a regularization of the fundamental solution $G(\Delta^*; \cdot)$. This approach also works for the application of the Beltrami operator to U. However, in connection with Theorem 1.3, we find that $\Delta^* U$ is not continuous across $\partial\Gamma$ anymore (compare equation (24) and Theorem 1.10). For brevity, we do not supply the proofs at this point but refer the reader, e.g., to [15]. A related regularized Green function plays an important role in the applications in Section 5 and is explained in more detail later on.

Theorem 1.10. *If H is of class $\mathrm{C}^{(1)}(\overline{\Gamma})$ and U is given by* (23)*, then U is of class $\mathrm{C}^{(2)}(\Gamma)$ and satisfies*

$$\Delta_\xi^* \int_\Gamma G(\Delta^*; \xi \cdot \eta) H(\eta) d\omega(\eta) = H(\xi) - \frac{1}{4\pi} \int_\Gamma H(\eta) d\omega(\eta), \quad \xi \in \Gamma. \tag{26}$$

1.4. Curve potentials

While surface potentials are useful to deal with the Poisson problem(PP), curve potentials are particularly useful when dealing with functions that are harmonic (with respect to the Beltrami operator). More precisely, we take a closer look at the two layer potentials

$$U_1[\tilde{Q}](\xi) = \int_{\partial\Gamma} G(\Delta^*; \xi \cdot \eta) \tilde{Q}(\eta)\, d\sigma(\eta), \quad \xi \in \Gamma, \tag{27}$$

and

$$U_2[Q](\xi) = \int_{\partial\Gamma} \left(\frac{\partial}{\partial\nu(\eta)} G(\Delta^*; \xi \cdot \eta) \right) Q(\eta)\, d\sigma(\eta), \quad \xi \in \Gamma. \tag{28}$$

From the properties of the fundamental solution $G(\Delta^*; \cdot)$ it can be seen that the so-called *double-layer potential* $U_2[Q]$ is harmonic in Γ for any Q of class $\mathrm{C}^{(0)}(\partial\Gamma)$. The *single-layer potential* $U_1[\tilde{Q}]$ is harmonic in Γ if $\tilde{Q}$ is of class $\mathrm{C}^{(0)}(\partial\Gamma)$ and if the integral over $\partial\Gamma$ vanishes, i.e., if $\int_{\partial\Gamma} \tilde{Q}(\eta) d\sigma(\eta) = 0$ (we say that $\tilde{Q}$ is of class $\mathrm{C}_0^{(0)}(\partial\Gamma)$). Therefore, these two potentials represent good candidates for solutions to the boundary value problems (DP) and (NP). The aim of the present section is to investigate the behaviour of the single- and double-layer potentials $U_1[Q]$ and $U_2[\tilde{Q}]$, respectively, when they approach the boundary $\partial\Gamma$. The essential behaviour of the double-layer potential $U_2[Q]$ is already reflected by the relation (18). Based on this relation and a set of several more technical estimates, one can prove the following set of limit- and jump-relations at the boundary $\partial\Gamma$.

Theorem 1.11 (Limit- and Jump-Relations). *Let Q, $\tilde{Q}$ be of class $\mathrm{C}^{(0)}(\partial\Gamma)$ and U_1, U_2 be given as in* (27) *and* (28)*, respectively. Furthermore, let $\xi \in \partial\Gamma$.*

(a) *For the single-layer potential, we have the limit-relations*

$$\lim_{\tau\to 0+} U_1[\tilde{Q}] \left(\frac{\xi \pm \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) - U_1[\tilde{Q}](\xi) = 0, \tag{29}$$

$$\lim_{\tau\to 0+} \left(\frac{\partial}{\partial\nu} U_1[\tilde{Q}] \right) \left(\frac{\xi \pm \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) - \left(\frac{\partial}{\partial\nu} U_1[\tilde{Q}] \right)(\xi) = \pm\frac{1}{2}\tilde{Q}(\xi). \tag{30}$$

For the double-layer potential, we have

$$\lim_{\tau\to 0+} U_2[Q] \left(\frac{\xi \pm \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) - U_2[Q](\xi) = \mp\frac{1}{2}Q(\xi). \tag{31}$$

(b) *For the single-layer potential, we have the jump-relations*

$$\lim_{\tau\to 0+} \left(U_1[\tilde{Q}] \left(\frac{\xi + \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) - U_1[\tilde{Q}] \left(\frac{\xi - \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) \right) = 0, \tag{32}$$

$$\lim_{\tau\to 0+} \left(\left(\frac{\partial}{\partial\nu} U_1[\tilde{Q}] \right) \left(\frac{\xi + \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) - \left(\frac{\partial}{\partial\nu} U_1[\tilde{Q}] \right) \left(\frac{\xi - \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) \right) = \tilde{Q}(\xi). \tag{33}$$

For the double-layer potential, we have

$$\lim_{\tau\to 0+} \left(U_2[Q] \left(\frac{\xi + \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) - U_2[Q] \left(\frac{\xi - \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) \right) = -Q(\xi), \tag{34}$$

$$\lim_{\tau\to 0+} \left(\left(\frac{\partial}{\partial\nu} U_2[Q] \right) \left(\frac{\xi + \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) - \left(\frac{\partial}{\partial\nu} U_2[Q] \right) \left(\frac{\xi - \tau\nu(\xi)}{\sqrt{1+\tau^2}} \right) \right) = 0. \tag{35}$$

All of the relations above hold uniformly with respect to $\xi \in \partial\Gamma$.

Remark 1.12. Theorem 1.11 essentially tells us that the single-layer potential $U_1[\tilde{Q}]$ and the normal derivative of the double-layer potential $\frac{\partial}{\partial\nu} U_2[Q]$ are continuous across the boundary $\partial\Gamma$ while the double-layer potential $U_2[Q]$ and the normal

derivative of the single-layer potential $\frac{\partial}{\partial\nu}U_1[\tilde{Q}]$ are not. However, one has to be careful about $\frac{\partial}{\partial\nu}U_2[Q]$: it is only well-defined on $\partial\Gamma$ under higher smoothness assumptions on Q than just $\mathrm{C}^{(0)}(\partial\Gamma)$. Therefore, we only supplied the jump relation for this particular case but not the limit relation, which is sufficient for most theoretical considerations.

Remark 1.13. The relations in Theorem 1.11 were formulated with respect to the uniform topology for $Q, \tilde{Q} \in \mathrm{C}^{(0)}(\Omega)$. However, they can also be formulated with respect to the $\mathrm{L}^2(\Omega)$-topology for $Q, \tilde{Q} \in \mathrm{L}^2(\Omega)$. For details, the reader is again referred to [15, 17] and earlier references therein.

2. Boundary value problems for the Beltrami operator

In this section, we investigate the problems (PP), (DP), and (NP) and try to obtain integral representations of their solutions.

2.1. Poisson problem

We remember the Poisson problem (PP) from the beginning of this chapter: Let H be of class $\mathrm{C}^{(1)}(\overline{\Gamma})$, then we are looking for a function U of class $\mathrm{C}^{(2)}(\Gamma)$ such that

$$\Delta^* U(\xi) = H(\xi), \quad \xi \in \Gamma. \tag{36}$$

If we choose $\bar{H} = H - \frac{1}{\|\Gamma\|}\int_\Gamma H(\eta)d(\eta)$, we find that $\int_\Gamma \bar{H}(\eta)d\omega(\eta) = 0$ and, by Theorem 1.10, that

$$\bar{U}(\xi) = \int_\Gamma G(\Delta^*; \xi\cdot\eta)\bar{H}(\eta)d\omega(\eta), \quad \xi \in \Gamma, \tag{37}$$

satisfies $\Delta^*\bar{U}(\xi) = \bar{H}(\xi)$, for $\xi \in \Gamma$.

Setting $U(\xi) = \bar{U}(\xi) - \frac{1}{\|\Gamma\|}\ln(1-\xi\cdot\bar{\xi})\int_\Gamma H(\eta)d\omega(\eta)$, for some fixed $\bar{\xi} \in \Gamma^c$, we eventually obtain the desired solution satisfying

$$\Delta^* U(\xi) = \bar{H}(\xi) + \frac{1}{\|\Gamma\|}\int_\Gamma H(\eta)d\omega(\eta) = H(\xi), \quad \xi \in \Gamma. \tag{38}$$

The solution of (36), however, is not unique. Subscribing further boundary values on U, e.g., Dirichlet boundary values $U^-(\xi) = F(\xi)$, for $\xi \in \partial\Gamma$, it is possible to obtain uniqueness. Letting $\tilde{U}$ denote the function U from (38) that we constructed before, we can formulate the boundary value problem of finding a function $\tilde{\tilde{U}}$ that solves

$$\Delta^*\tilde{\tilde{U}}(\xi) = 0, \quad \xi \in \Gamma, \tag{39}$$

$$\tilde{\tilde{U}}^-(\xi) = F(\xi) - \tilde{U}^-(\xi), \quad \xi \in \partial\Gamma. \tag{40}$$

The newly obtained function $U = \tilde{U} + \tilde{\tilde{U}}$ would then satisfy the desired differential equation (36) and the desired Dirichlet boundary values. Boundary value problems such as (39), (40) are studied in more detail in the upcoming section.

2.2. Dirichlet and Neumann problem

We take a closer look at the following boundary value problems that have already been mentioned in the introduction:

Dirichlet Problem (DP): Let F be of class $\mathrm{C}^{(0)}(\partial\Gamma)$. We are looking for a function U of class $\mathrm{C}^{(2)}(\Gamma) \cap \mathrm{C}^{(0)}(\overline{\Gamma})$ such that

$$\Delta^* U(\xi) = 0, \qquad \xi \in \Gamma, \tag{41}$$

$$U^-(\xi) = F(\xi), \quad \xi \in \partial\Gamma. \tag{42}$$

Neumann Problem (NP): Let F be of class $\mathrm{C}^{(0)}(\partial\Gamma)$. We are looking for a function U of class $\mathrm{C}^{(2)}(\Gamma) \cap \mathrm{C}^{(0)}(\overline{\Gamma})$, with a well-defined normal derivative $\frac{\partial}{\partial\nu}U^-$ on $\partial\Gamma$, such that

$$\Delta^* U(\xi) = 0, \qquad \xi \in \Gamma, \tag{43}$$

$$\frac{\partial}{\partial\nu}U^-(\xi) = F(\xi), \quad \xi \in \partial\Gamma. \tag{44}$$

First, we formalize the term $U^-(\xi)$. For $\xi \in \partial\Gamma$, it is meant in the sense

$$U^-(\xi) = \lim_{\tau\to 0+} U\left(\frac{\xi - \tau\nu(\xi)}{\sqrt{1+\tau^2}}\right), \tag{45}$$

i.e., we approach the boundary $\partial\Gamma$ in normal direction from within Γ. The term $U^+(\xi)$ is meant in the sense

$$U^+(\xi) = \lim_{\tau\to 0+} U\left(\frac{\xi + \tau\nu(\xi)}{\sqrt{1+\tau^2}}\right), \tag{46}$$

i.e., we approach the boundary $\partial\Gamma$ in normal direction from the outside of Γ (or, in other words, from within Γ^c). The expressions $\frac{\partial}{\partial\nu}U^\pm(\xi)$ are meant analogously. We can already see the connection to the limit- and jump-relations from Theorem 1.11. More precisely, making the ansatz $U = U_2[Q]$ for the Dirichlet problem (DP) and $U = U_1[\tilde{Q}]$ for the Neumann problem (NP), Theorem 1.11 yields the following closely related problems:

Integral Dirichlet Problem (IDP): Let F be of class $\mathrm{C}^{(0)}(\partial\Gamma)$. We are looking for some Q of class $\mathrm{C}^{(0)}(\partial\Gamma)$ that satisfies

$$F(\xi) = U_2[Q](\xi) + \frac{1}{2}Q(\xi), \quad \xi \in \partial\Gamma. \tag{47}$$

Integral Neumann Problem (INP): Let F be of class $\mathrm{C}^{(0)}(\partial\Gamma)$. We are looking for some $\tilde{Q}$ of class $\mathrm{C}_0^{(0)}(\partial\Gamma)$ that satisfies

$$F(\xi) = U_1[\tilde{Q}](\xi) - \frac{1}{2}\tilde{Q}(\xi), \quad \xi \in \partial\Gamma. \tag{48}$$

In other words, the Dirichlet problem (DP) and the Neumann problem (NP) have been reduced to the Fredholm equations (IDP) and (INP). These boundary integral formulations have been used, e.g., in [20, 34] to numerically solve the original boundary value problems for the Beltrami operator. In this section, however, we

are mainly interested in (IDP) and (INP) as tools to guarantee the existence of solutions to (DP) and (NP) via the Fredholm alternative. Uniqueness of the solutions can be obtained via the application of the maximum principle from Theorem 1.8 and the Green formulas from Theorem 1.1.

Remark 2.1. There are two noteworthy differences in comparison to the Euclidean case. First, considerations on the sphere do not require a clear distinction between interior and exterior problems since the open complement Γ^c of a bounded regular region $\Gamma \subset \Omega$ is again a bounded regular region. Second, the single-layer potential $U_1[\tilde{Q}]$ is only harmonic if $\tilde{Q} \in \mathrm{C}_0^{(0)}(\partial\Gamma)$. A solution of (48) in $\mathrm{C}_0^{(0)}(\partial\Gamma)$ exists if and only if F is of class $\mathrm{C}_0^{(0)}(\partial\Gamma)$, which suits the general necessary condition for the existence of a solution to (NP) that can be obtained from Green's formulas. However, it should be mentioned that the integral equation (48) additionally has a unique solution $\tilde{Q} \in \mathrm{C}^{(0)}(\partial\Gamma)$ if F is of class $\mathrm{C}^{(0)}(\partial\Gamma)$. This is not true for the Euclidean counterpart.

Summarizing, and including the considerations from Section 2.1, we obtain the following results. For details, the reader is again referred to [15] and, for the Euclidean counterparts, to [17, 26, 30, 45].

Theorem 2.2 (Uniqueness).

(a) *A solution of* (DP) *is uniquely determined.*
(b) *A solution of* (NP) *is uniquely determined up to an additive constant.*

Theorem 2.3 (Existence for Generalized (DP)). *Let F be of class $\mathrm{C}^{(0)}(\partial\Gamma)$ and H of class $\mathrm{C}^{(1)}(\overline{\Gamma})$. Then there exists a unique solution U of class $\mathrm{C}^{(2)}(\Gamma) \cap \mathrm{C}^{(0)}(\overline{\Gamma})$ of the Dirichlet problem*

$$\Delta^* U(\xi) = H(\xi), \quad \xi \in \Gamma, \tag{49}$$

$$U^-(\xi) = F(\xi), \quad \xi \in \partial\Gamma. \tag{50}$$

Theorem 2.4 (Existence for Generalized (NP)). *Let F be of class $\mathrm{C}^{(0)}(\partial\Gamma)$ and H of class $\mathrm{C}^{(1)}(\overline{\Gamma})$. Then there exists an up to an additive constant uniquely determined solution U of class $\mathrm{C}^{(2)}(\Gamma) \cap \mathrm{C}^{(0)}(\overline{\Gamma})$, with a well-defined normal derivative $\frac{\partial}{\partial\nu}U^-$ on $\partial\Gamma$, to the Neumann problem*

$$\Delta^* U(\xi) = H(\xi), \quad \xi \in \Gamma, \tag{51}$$

$$\frac{\partial}{\partial\nu} U^-(\xi) = F(\xi), \quad \xi \in \partial\Gamma, \tag{52}$$

if and only if

$$\int_{\partial\Gamma} F(\eta) d\sigma(\eta) - \int_{\Gamma} H(\eta) d\omega(\eta) = 0. \tag{53}$$

Proof. The condition (53) is a simple consequence from

$$\int_{\Gamma} H(\eta) d\omega(\eta) = \int_{\Gamma} \Delta^* U(\eta) d\omega(\eta) = \int_{\partial\Gamma} \frac{\partial}{\partial\nu} U(\eta) d\sigma(\eta) = \int_{\partial\Gamma} F(\eta) d\sigma(\eta), \tag{54}$$

where Green's formulas have been used for the second equation. The general existence follows from the application of the Fredholm alternative to (INP). □

2.3. Green's functions

Next, we are interested in the representation of a solution to (DP) and (NP). A possibility is indicated in Theorem 1.3(a). However, this representation requires the simultaneous knowledge of U and $\frac{\partial}{\partial \nu}U$ on the boundary $\partial\Gamma$, which is not necessary and can be problematic since the two quantities are not independent from each other. As a remedy, Green's functions for Dirichlet and Neumann boundary values can be used.

More precisely, a function $G_D(\Delta^*;\cdot,\cdot)$ is called a *Dirichlet Green function* (*with respect to the Beltrami operator*) if it can be decomposed in the form

$$G_D(\Delta^*;\xi,\eta) = G(\Delta^*;\xi\cdot\eta) - \Phi_D(\xi,\eta), \quad \eta\in\overline{\Gamma},\xi\in\Gamma,\xi\neq\eta, \tag{55}$$

where $\Phi_D(\xi,\cdot)$ is of class $\mathrm{C}^{(2)}(\Gamma)\cap\mathrm{C}^{(1)}(\overline{\Gamma})$ and satisfies

$$\Delta_\eta^*\Phi_D(\xi,\eta) = -\frac{1}{4\pi}, \quad \eta\in\Gamma, \tag{56}$$

$$\Phi_D^-(\xi,\eta) = G(\Delta^*;\xi\cdot\eta), \quad \eta\in\partial\Gamma, \tag{57}$$

for every $\xi\in\Gamma$. Analogously, a function $G_N(\Delta^*;\cdot,\cdot)$ is called a *Neumann Green function* (*with respect to the Beltrami operator*) if it can be decomposed in the form

$$G_N(\Delta^*;\xi,\eta) = G(\Delta^*;\xi\cdot\eta) - \Phi_N(\xi,\eta), \quad \eta\in\overline{\Gamma},\xi\in\Gamma,\xi\neq\eta, \tag{58}$$

where $\Phi_N(\xi,\cdot)$ is of class $\mathrm{C}^{(2)}(\Gamma)\cap\mathrm{C}^{(1)}(\overline{\Gamma})$ and satisfies the conditions

$$\Delta_\eta^*\Phi_N(\xi,\eta) = \frac{1}{\|\Gamma\|} - \frac{1}{4\pi}, \quad \eta\in\Gamma, \tag{59}$$

$$\frac{\partial}{\partial\nu(\eta)}\Phi_N^-(\xi,\eta) = \frac{\partial}{\partial\nu(\eta)}G(\Delta^*;\xi\cdot\eta), \quad \eta\in\partial\Gamma, \tag{60}$$

for every $\xi\in\Gamma$. Using Theorem 1.3(a) and Theorem 1.1(c) for Φ_D and Φ_N, we eventually achieve the representations

$$U(\xi) = \int_\Gamma G_D(\Delta^*;\xi,\eta)\Delta_\eta^*U(\eta)d\omega(\eta) + \int_{\partial\Gamma}U(\eta)\frac{\partial}{\partial\nu(\eta)}G_D(\Delta^*;\xi,\eta)d\sigma(\eta) \tag{61}$$

and

$$\begin{aligned} U(\xi) =& \frac{1}{\|\Gamma\|}\int_\Gamma U(\eta)d\omega(\eta) + \int_\Gamma G_N(\Delta^*;\xi,\eta)\Delta_\eta^*U(\eta)d\omega(\eta) \\ &- \int_{\partial\Gamma}G_N(\Delta^*;\xi,\eta)\frac{\partial}{\partial\nu(\eta)}U(\eta)d\sigma(\eta), \end{aligned} \tag{62}$$

which yield integral representations for solutions to (DP) and (NP), respectively, under the condition that U is of class $\mathrm{C}^{(2)}(\overline{\Gamma})$. It remains to construct the auxiliary functions Φ_D and Φ_N. Some general construction principles on the sphere can be found, e.g., in [24, 32]. In this chapter, we focus on spherical caps $\Gamma_\rho(\zeta)$. The

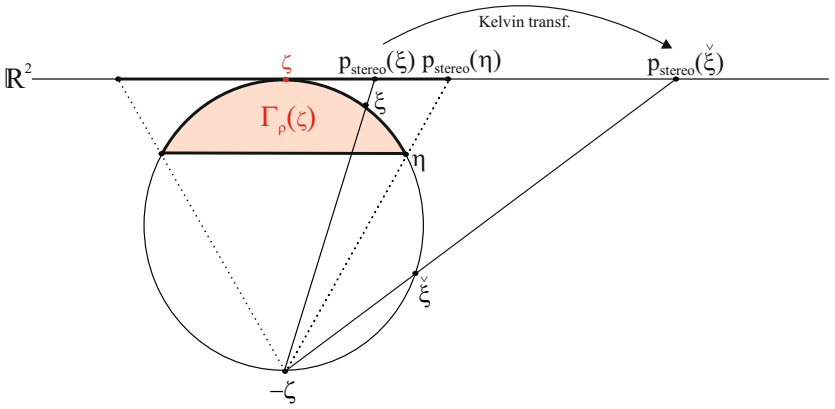


FIGURE 2. Schematic description of the construction of the reflection point $\check{\xi}$.

procedure is similar to the construction of a Dirichlet Green function for a disc in $\mathbb{R}^2$. For $\xi \in \Gamma_\rho(\zeta)$, we need to find a reflection point $\check{\xi} \in (\Gamma_\rho(\zeta))^c$ and a scaling factor $\check{r} \in \mathbb{R}$ such that

$$1 - \xi \cdot \eta = \check{r}\left(1 - \check{\xi} \cdot \eta\right), \quad \eta \in \partial\Gamma_\rho(\zeta), \xi \in \Gamma_\rho(\zeta). \tag{63}$$

Indeed, under this assumption, it is clear that

$$\Phi_D(\xi, \eta) = \frac{1}{4\pi} \ln(\check{r}(1 - \check{\xi} \cdot \eta)) + \frac{1}{4\pi}(1 - \ln(2)) \tag{64}$$

satisfies the desired conditions (56) and (57). The reflection point $\check{\xi}$ can be obtained by a stereographic projection of ξ onto $\mathbb{R}^2$, then applying a Kelvin transform to the projection point, and eventually projecting it back to the sphere (cf. Figure 2 for an illustration). $\check{\xi}$ represents the spherical Kelvin transformation of ξ. The scaling factor $\check{r}$ is obtained by solving (63). Alternatively, the entire Dirichlet Green function $G_D(\Delta^*; \cdot, \cdot)$ can be obtained from a stereographic projection of the Dirichlet Green function for the Laplace operator on a disc in $\mathbb{R}^2$. But this route would not supply us with a spherical counterpart to the Kelvin transform. We can conclude our considerations with the following theorem.

Theorem 2.5. *Let $\Gamma = \Gamma_\rho(\zeta)$ be a spherical cap with center $\zeta \in \Omega$ and radius $\rho \in (0, 2)$. Furthermore, for $\xi \in \Gamma_\rho(\zeta)$ we set*

$$\check{\xi} = \frac{1}{\check{r}}\xi - \frac{\check{r} - 1}{\check{r}(\rho - 1)}\zeta, \tag{65}$$

$$\check{r} = -\frac{1 + 2\xi \cdot \zeta(\rho - 1) + (\rho - 1)^2}{\rho(\rho - 2)}. \tag{66}$$

Then

$$G_D(\Delta^*; \xi, \eta) = \frac{1}{4\pi} \ln(1 - \xi \cdot \eta) - \frac{1}{4\pi} \ln(\check{r}(1 - \check{\xi} \cdot \eta)), \tag{67}$$

and a solution $U \in \mathrm{C}^{(2)}(\overline{\Gamma})$ *of the Dirichlet problem* (DP) *can be represented by*

$$U(\xi) = \frac{1}{2\pi} \frac{\xi \cdot \zeta + \rho - 1}{\sqrt{\rho(2-\rho)}} \int_{\partial\Gamma_\rho(\zeta)} \frac{1}{1 - \xi \cdot \eta} F(\eta)\, d\sigma(\eta), \quad \xi \in \Gamma_\rho(\zeta). \tag{68}$$

Remark 2.6. Applying Theorem 2.5 for $\zeta = \xi$ leads to the Mean Value Property II from Theorem 1.7.

A Neumann Green function for the Beltrami operator cannot be obtained by a simple stereographic projection of the Neumann Green function for the Laplace operator on a disc in $\mathbb{R}^2$. But some computations based on the previously obtained auxiliary function Φ_D yield the following theorem.

Theorem 2.7. *Let* $\Gamma = \Gamma_\rho(\zeta)$ *be a spherical cap with center* $\zeta \in \Omega$ *and radius* $\rho \in (0, 2)$. *Furthermore, let* $\check{\xi}$ *and* $\check{r}$ *be given as in Theorem* 2.5. *Then, a Neumann Green function is given by*

$$G_N(\Delta^*; \xi, \eta) = \frac{1}{4\pi} \ln(1 - \xi \cdot \eta) + \frac{1}{4\pi} \ln(\check{r}(1 - \check{\xi} \cdot \eta)) + \frac{1-\rho}{2\pi\rho} \ln(1 + \zeta \cdot \eta). \tag{69}$$

A solution $U \in \mathrm{C}^{(2)}(\overline{\Gamma})$ *of the Neumann problem* (NP) *can be represented by*

$$\begin{aligned} U(\xi) = & \frac{1}{2\pi\rho} \int_{\Gamma_\rho(\zeta)} U(\eta) d\omega(\eta) \\ & - \int_{\partial\Gamma_\rho(\zeta)} \left(\frac{1}{2\pi} \ln(1 - \xi \cdot \eta) + \frac{1-\rho}{2\pi\rho} \ln(2-\rho) \right) F(\eta)\, d\sigma(\eta), \quad \xi \in \Gamma_\rho(\zeta). \end{aligned} \tag{70}$$

3. Spherical decompositions and first-order differential equations

In this section we treat differential equations for the surface gradient ∇^* and the surface curl gradient L^*. They come up, e.g., when dealing with vertical deflections and geostrophic ocean flow. Additionally, we take a look at some spherical decompositions of vector fields that are particularly useful in geosciences.

3.1. Surface gradient and surface curl gradient

Different from the Poisson equation, solutions of the differential equations with respect to the surface gradient and the surface curl gradient are uniquely determined up to an additive constant on regular regions $\Gamma \subset \Omega$, without the necessity of boundary values. Also the existence of a solution can be easily guaranteed. This is summarized in the following two lemmas.

Lemma 3.1 (Uniqueness). *Let* U *be of class* $\mathrm{C}^{(1)}(\Gamma)$. *Then*

$$\nabla^* U(\xi) = 0, \quad \xi \in \Gamma, \tag{71}$$

if and only if U *is constant on* Γ. *The same holds true for* $\mathrm{L}^* U(\xi) = 0$, $\xi \in \Gamma$.

Theorem 3.2 (Existence).

(a) *Let* $f \in \mathrm{c}^{(1)}(\Gamma)$ *be a tangential vector field satisfying*

$$\mathrm{L}^* \cdot f(\xi) = 0, \quad \xi \in \Gamma. \tag{72}$$

Then there exists a function U of class $\mathrm{C}^{(2)}(\Gamma)$, which is uniquely determined up to an additive constant, such that

$$f(\xi) = \nabla^* U(\xi), \quad \xi \in \Gamma. \tag{73}$$

(b) *Let $f \in \mathrm{c}^{(1)}(\Gamma)$ be a tangential vector field satisfying*

$$\nabla^* \cdot f(\xi) = 0, \quad \xi \in \Gamma. \tag{74}$$

Then there exists a function U of class $\mathrm{C}^{(2)}(\Gamma)$, which is uniquely determined up to an additive constant, such that

$$f(\xi) = \mathrm{L}^* U(\xi), \quad \xi \in \Gamma. \tag{75}$$

From Theorem 1.3(b), we know a possible expression of the solutions to the differential equations for the surface gradient and the surface curl gradient. However, this representation requires the knowledge of U on the boundary $\partial\Gamma$, which is actually not necessary according to Theorem 3.2. Using a Neumann Green function together with the identities in Theorem 1.3(b) directly implies the following results.

Theorem 3.3.

(a) *Let f of class $\mathrm{c}^{(1)}(\overline{\Gamma})$ be a tangential vector field satisfying $\mathrm{L}^* \cdot f(\xi) = 0$, $\xi \in \Gamma$. Then a solution of*

$$f(\xi) = \nabla^* U(\xi), \quad \xi \in \Gamma, \tag{76}$$

is given by

$$U(\xi) = \frac{1}{\|\Gamma\|} \int_\Gamma U(\eta) d\omega(\eta) - \int_\Gamma \left(\nabla^*_\eta G_N(\Delta^*; \xi, \eta)\right) \cdot f(\eta) d\omega(\eta). \tag{77}$$

(b) *Let f of class $\mathrm{c}^{(1)}(\overline{\Gamma})$ be a tangential vector field satisfying $\nabla^* \cdot f(\xi) = 0$, $\xi \in \Gamma$. Then a solution of*

$$f(\xi) = \mathrm{L}^* U(\xi), \quad \xi \in \Gamma, \tag{78}$$

is given by

$$U(\xi) = \frac{1}{\|\Gamma\|} \int_\Gamma U(\eta) d\omega(\eta) - \int_\Gamma \left(\mathrm{L}^*_\eta G_N(\Delta^*; \xi, \eta)\right) \cdot f(\eta) d\omega(\eta). \tag{79}$$

Remark 3.4. If we deal with the entire sphere $\Gamma = \Omega$, the same results as in the preceding theorem hold true. For the integral representations, one simply has to substitute the Neumann Green function by the fundamental solution $G(\Delta^*; \cdot)$.

3.2. Helmholtz and Hardy–Hodge decomposition

We begin with the spherical Helmholtz decomposition of a vector field f. It essentially describes the split-up of the vector field into a radial and two tangential components, of which one is surface curl-free and the other one surface divergence-free. In geomagnetism, this has applications, e.g., in the separation of polar ionospheric current systems into field-aligned currents (which are nearly radial in polar regions) and Pedersen and Hall currents (see, e.g., [1, 2, 21, 39]). In other areas, the

spherical Helmholtz decomposition has a natural connection as well: geostrophic ocean flow, e.g., is purely tangential and surface divergence-free while the vertical deflection of the geoidal normal vector is approximately purely tangential and surface curl-free. For convenience, we use the following notations for the Helmholtz operators acting on a scalar function F at a point $\xi \in \Omega$:

$$o^{(1)}F(\xi) = \xi F(\xi), \qquad o^{(2)}F(\xi) = \nabla^* F(\xi), \qquad o^{(3)}F(\xi) = \mathrm{L}^* F(\xi). \tag{80}$$

Writing $f = o^{(1)}F_1 + o^{(2)}F_2 + o^{(3)}F_3$ on a subdomain Γ and using the orthogonality of the three operators, we obtain $\Delta^* F_3(\xi) = \mathrm{L}^* \cdot f(\xi)$, $\xi \in \Gamma$. Latter can be solved by the methods of the previous section. We need to prescribe boundary data on F_3 in order to obtain uniqueness of the scalar function F_3. All in all, we can formulate Decomposition Theorem 3.5. More details can be found, e.g., in [15, 21].

Theorem 3.5 (Spherical Helmholtz Decomposition). *Let f be of class $\mathrm{c}^{(2)}(\overline{\Gamma})$. Then there exist scalar fields F_1 of class $\mathrm{C}^{(2)}(\overline{\Gamma})$ and F_2, F_3 of class $\mathrm{C}^{(2)}(\Gamma)$ such that*

$$f(\xi) = o^{(1)}F_1(\xi) + o^{(2)}F_2(\xi) + o^{(3)}F_3(\xi), \quad \xi \in \Gamma. \tag{81}$$

Uniqueness of F_1, F_2, F_3 is guaranteed by the properties

$$\int_\Gamma F_2(\eta) d\omega(\eta) = 0 \tag{82}$$

and

$$F_3^-(\xi) = F(\xi), \quad \xi \in \partial\Gamma, \tag{83}$$

for a fixed function F of class $\mathrm{C}^{(0)}(\partial\Gamma)$. The Helmholtz scalars F_1, F_2, and F_3 can be then represented by

$$\begin{aligned} F_2(\xi) = &- \int_\Gamma \big(\nabla_\eta^* G_N(\Delta^*; \xi, \eta)\big) \cdot f(\eta) d\omega(\eta) \\ &+ \int_{\partial\Gamma} F(\eta)\, \tau_\eta \cdot \nabla_\eta^* G_N(\Delta^*; \xi, \eta) d\sigma(\eta), \quad \xi \in \Gamma \end{aligned} \tag{84}$$

$$\begin{aligned} F_3(\xi) = &- \int_\Gamma \big(\mathrm{L}_\eta^* G_D(\Delta^*; \xi, \eta)\big) \cdot f(\eta) d\omega(\eta) \\ &+ \int_{\partial\Gamma} G_D(\Delta^*; \xi, \eta) \tau_\eta \cdot f(\eta) d\sigma(\eta) \\ &+ \int_{\partial\Gamma} F(\eta) \frac{\partial}{\partial \nu_\eta} G_D(\Delta^*; \xi, \eta) d\sigma(\eta), \quad \xi \in \Gamma \end{aligned} \tag{85}$$

for $\xi \in \Gamma$. Additionally, if $\int_\Gamma F_1(\eta) d\omega(\eta) = 0$, then

$$F_1(\xi) = \xi \cdot f(\xi) = \Delta_\xi^* \int_\Gamma G(\Delta^*; \xi \cdot \eta)\, \eta \cdot f(\eta) d\omega(\eta), \quad \xi \in \Gamma. \tag{86}$$

Remark 3.6. Clearly, the type of boundary conditions that have to be prescribed to obtain uniqueness of the Helmholtz decomposition can be varied. They can be imposed on F_2 instead of F_3, or the Dirichlet boundary conditions can be substituted

by Neumann boundary conditions. Neumann boundary conditions are occasionally more advantageous as they allow the imposition of boundary information on the normal and tangential direction of the vectorial quantities $o^{(2)}F_2$ and $o^{(3)}F_3$, respectively, which are in some cases better accessible from the given data than the scalars F_2 or F_3. Representations analogous to Theorem 3.5 can be derived by Green's formulas and the results from Section 2.

Remark 3.7. For the particular case $\Gamma = \Omega$, the results from Theorem 3.5 hold true as well if the boundary integrals $\int_{\partial\Gamma} \ldots d\sigma$ are dropped and the Neumann and Dirichlet Green functions are substituted by the fundamental solution $G(\Delta^*;\cdot)$. For the uniqueness, condition (83) has to be substituted by $\int_\Omega F_3(\eta)d\omega(\eta) = 0$. We then obtain

$$F_2(\xi) = -\int_\Omega \left(\nabla^*_\eta G(\Delta^*;\xi\cdot\eta)\right)\cdot f(\eta)d\omega(\eta), \quad \xi\in\Omega, \tag{87}$$

$$F_3(\xi) = -\int_\Omega \left(\mathrm{L}^*_\eta G(\Delta^*;\xi\cdot\eta)\right)\cdot f(\eta)d\omega(\eta), \quad \xi\in\Omega. \tag{88}$$

Thus, in the global case $\Gamma = \Omega$, the Helmholtz scalars F_2 and F_3 are determined uniquely up to an additive constant without further constraints. The vectorial quantities $o^{(2)}F_2$ and $o^{(3)}F_3$ are actually uniquely determined. This is not true for general subdomains $\Gamma \subset \Omega$.

Next, we turn to a different spherical decomposition, the so-called spherical Hardy–Hodge decomposition (the name is adopted from the Euclidean decomposition presented in [3], although its spherical version is known and used significantly longer, e.g., in [2, 18, 21, 36, 40] and references therein). It is based on the set of operators

$$\tilde{o}^{(1)} = o^{(1)}\left(\mathrm{D}+\frac{1}{2}\right) - o^{(2)}, \qquad \tilde{o}^{(2)} = o^{(1)}\left(\mathrm{D}-\frac{1}{2}\right) + o^{(2)}, \qquad \tilde{o}^{(3)} = o^{(3)}, \tag{89}$$

where the operator D is given by $\mathrm{D} = \left(-\Delta^* + \frac{1}{4}\right)^{\frac{1}{2}}$. A decomposition in terms of these operators can be interpreted as a decomposition of a spherical vectorial signal with respect to sources lying inside a given sphere (reflected by the $\tilde{o}^{(1)}$-contributions), sources lying in the exterior of the sphere ($\tilde{o}^{(2)}$-contributions), and sources on the sphere ($\tilde{o}^{(3)}$-contributions). For the gravitational field measured at satellite altitude, e.g., only the $\tilde{o}^{(1)}$-contribution is of relevance. Concerning the Earth's crustal magnetization, only the $\tilde{o}^{(2)}$-contribution of the magnetization generates a magnetic effect at satellite altitude. The generated magnetic field itself, however, only consists of $\tilde{o}^{(1)}$-contributions since its source (i.e., the magnetization) is located inside the satellite's orbit. The decomposition and the integral representation of its scalar functions can be closely related to the spherical Helmholtz decomposition. For details, we refer the reader to [15, 18, 21, 22]. Yet, the non-local structure of the operator D makes it very difficult to obtain results on subdomains $\Gamma \subset \Omega$. Therefore, the following theorem only treats the decomposition for the case $\Gamma = \Omega$.

Theorem 3.8 (Spherical Hardy–Hodge Decomposition). *Let f be of class $\mathrm{c}^{(1)}(\Omega)$. Then there exist scalar fields $\tilde{F}_1$, $\tilde{F}_2$, $\tilde{F}_3$ of class $\mathrm{C}^{(2)}(\Omega)$ such that*

$$f(\xi) = \tilde{o}^{(1)}\tilde{F}_1(\xi) + \tilde{o}^{(2)}\tilde{F}_2(\xi) + \tilde{o}^{(3)}\tilde{F}_3(\xi), \quad \xi \in \Omega. \tag{90}$$

Uniqueness of $\tilde{F}_1, \tilde{F}_2, \tilde{F}_3$ is guaranteed by the properties

$$\int_\Omega \tilde{F}_3(\eta)d\omega(\eta) = 0, \tag{91}$$

$$\int_\Omega \tilde{F}_1(\eta) - \tilde{F}_2(\eta)d\omega(\eta) = 0. \tag{92}$$

The Hardy–Hodge scalars $\tilde{F}_1$, $\tilde{F}_2$, and $\tilde{F}_3$ can then be represented by

$$\tilde{F}_1 = \frac{1}{2}\,\mathrm{D}^{-1}\,F_1 + \frac{1}{4}\,\mathrm{D}^{-1}\,F_2 - \frac{1}{2}F_2, \tag{93}$$

$$\tilde{F}_2 = \frac{1}{2}\,\mathrm{D}^{-1}\,F_1 + \frac{1}{4}\,\mathrm{D}^{-1}\,F_2 + \frac{1}{2}F_2, \tag{94}$$

$$\tilde{F}_3 = F_3, \tag{95}$$

where F_1, F_2, F_3 are the Helmholtz scalars from Theorem 3.5 *and Remark* 3.7.

Remark 3.9. The operator D^{-1} can be represented as the convolution operator

$$\mathrm{D}^{-1}\,F(\xi) = \frac{1}{2\pi}\int_\Omega \frac{1}{\sqrt{2(1-\xi\cdot\eta)}}F(\eta)d\omega(\eta), \quad \xi \in \Omega, \tag{96}$$

acting on a function F of class $\mathrm{C}^{(0)}(\Omega)$. Thus, equations (93)–(95) together with Theorem 3.5 and Remark 3.7 form integral representations of the Hardy–Hodge scalars.

4. Complete function systems

In the Euclidean setting, spherical harmonics form a complete function system in $\mathrm{L}^2(\Omega_R)$, and their harmonic extensions into the ball $\mathcal{B}_R = \{x \in \mathbb{R}^3 : |x| < R\}$ and its exterior $\mathcal{B}_R^c = \{x \in \mathbb{R}^3 : |x| > R\}$ (so-called inner and outer harmonics, respectively) form suitable function systems to approximate functions that are harmonic with respect to the Laplace operator. The limit- and jump-relations of layer potentials enable the extension of the completeness results to more general manifolds than the sphere. With the considerations of the previous sections at hand, we are now able to formulate analogous completeness results for function systems on general curves $\partial\Gamma$. We obtain completeness for certain function systems in $\mathrm{L}^2(\partial\Gamma)$ whose harmonic extensions into $\Gamma \subset \Omega$ are particularly well suited for the approximation of functions that are harmonic with respect to the Beltrami operator.

First, we need the notion of a fundamental system: Suppose that $\{\xi_k\}_{k\in\mathbb{N}} \subset \Gamma$ is a set of points satisfying

$$\mathrm{dist}(\{\xi_k\}_{k\in\mathbb{N}}, \partial\Gamma) > 0. \tag{97}$$

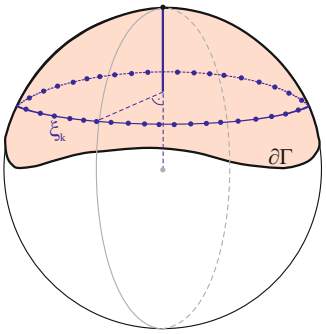


FIGURE 3. Example for a fundamental system $\{\xi_k\}_{k\in\mathbb{N}}$ (with respect to Γ).

If, for any harmonic function F in Γ, the condition $F(\xi_k) = 0$, $k \in \mathbb{N}$, implies that $F(\xi) = 0$ for all $\xi \in \Gamma$, then we call $\{\xi_k\}_{k\in\mathbb{N}}$ a *fundamental system* (*with respect to* Γ). Assuming that $\Sigma \subset \Gamma$ is a regular region with $\operatorname{dist}(\Sigma, \partial\Gamma) > 0$, an example for such a fundamental system is given by a dense point set $\{\xi_k\}_{k\in\mathbb{N}} \subset \partial\Sigma$. A particularly simple choice for Σ is a spherical cap within Γ (cf. Figure 3).

We begin with the completeness of function systems based on the fundamental solution for the Beltrami operator.

Theorem 4.1. *Let $\{\xi_k\}_{k\in\mathbb{N}}$ be a fundamental system with respect to Γ. Then the following statements hold true:*

(a) *The function system $\{G_k\}_{k\in\mathbb{N}_0}$ given by*

$$G_k(\xi) = \frac{1}{4\pi} \ln(1 - \xi_k \cdot \xi),\ k \in \mathbb{N}, \qquad G_0(\xi) = \frac{1}{4\pi},$$

is complete, and hence closed in $\mathrm{L}^2(\partial\Gamma)$.

(b) *The function system $\{\tilde{G}_k\}_{k\in\mathbb{N}_0}$, given by*

$$\tilde{G}_k(\xi) = \frac{1}{4\pi} \frac{\partial}{\partial\nu(\xi)} \ln(1 - \xi_k \cdot \xi),\ k \in \mathbb{N}, \qquad \tilde{G}_0(\xi) = \frac{1}{4\pi},$$

is complete, and hence closed in $\mathrm{L}^2(\partial\Gamma)$.

Remark 4.2. Let us assume that $\{\xi_k\}_{k\in\mathbb{N}}$ is a fundamental system with respect to Γ^c. Then the functions $\tilde{G}_k$ from Theorem 4.1 are harmonic in Γ and, thus, particularly suitably for the approximation of harmonic functions in Γ. The functions G_k from Theorem 4.1 need to be modified since they only satisfy $\Delta^* G_k(\xi) = -\frac{1}{4\pi}$, for $\xi \in \Gamma$ and $k \in \mathbb{N}$. Any auxiliary function G of class $\mathrm{C}^{(2)}(\overline{\Gamma})$ that satisfies $\Delta^* G(\xi) = \frac{1}{4\pi}$, $\xi \in \Gamma$, can be added to G_k without changing the completeness property. In other words, e.g.,

$$G_k^{(\mathrm{mod})}(\xi) = G_k(\xi) - \frac{1}{4\pi} \ln(1 - \xi \cdot \bar{\xi}),\ k \in \mathbb{N}, \qquad G_0^{(\mathrm{mod})}(\xi) = G_0(\xi),$$

with a fixed $\bar{\xi} \in \Gamma^c$, forms a complete function system in $\mathrm{L}^2(\partial\Gamma)$ that additionally satisfies $\Delta^* G_k^{(\mathrm{mod})}(\xi) = 0$, $\xi \in \Gamma$.

Next, we want to transfer the results from Theorem 4.1 to inner harmonics for spherical caps. In order to achieve this, we first need to clarify what we mean by inner harmonics for spherical caps. The sine and cosine functions obviously take the role of spherical harmonics on a circle in $\mathbb{R}^2$. Their harmonic continuations into the disc $\mathcal{D}_R = \{x \in \mathbb{R}^2 : |x| < R\}$ with radius $R > 0$ and into its exterior $\mathcal{D}_R^c = \{x \in \mathbb{R}^2 : |x| > R\}$ (the so-called inner and outer harmonics, respectively) are given by

$$H_{n,1}^{(\mathrm{int})}(R;x) = \frac{1}{R\sqrt{\pi}} \left(\frac{r}{R}\right)^n \cos(n\varphi),\ n \in \mathbb{N}_0, \quad x \in \mathcal{D}_R, \tag{98}$$

$$H_{n,2}^{(\mathrm{int})}(R;x) = \frac{1}{R\sqrt{\pi}} \left(\frac{r}{R}\right)^n \sin(n\varphi),\ n \in \mathbb{N}, \quad x \in \mathcal{D}_R, \tag{99}$$

$$H_{n,1}^{(\mathrm{ext})}(R;x) = \frac{1}{R\sqrt{\pi}} \left(\frac{R}{r}\right)^n \cos(n\varphi), n \in \mathbb{N}_0, \quad x \in \mathcal{D}_R^c, \tag{100}$$

$$H_{n,2}^{(\mathrm{ext})}(R;x) = \frac{1}{R\sqrt{\pi}} \left(\frac{R}{r}\right)^n \sin(n\varphi), n \in \mathbb{N}, \quad x \in \mathcal{D}_R^c, \tag{101}$$

where $x = (r\cos(\varphi), r\sin(\varphi))^T$, $r \geq 0$, $\varphi \in [0, 2\pi)$. Inner harmonics on a spherical cap $\Gamma_\rho(\zeta)$ with radius $\rho \in (0,2)$ and center $\zeta \in \Omega$ can then be obtained by a simple stereographic projection. More precisely,

$$H_{n,k}^{\rho,\zeta}(\xi) = H_{n,k}^{(\mathrm{int})}\left(\rho^{\frac{1}{4}}(2-\rho)^{\frac{1}{4}}; p_{\mathrm{stereo}}(\zeta;\xi)\right), \quad \xi \in \Gamma_\rho(\zeta), \tag{102}$$

denotes an *inner harmonic* (*of degree n and order k*) *on* $\Gamma_\rho(\zeta)$. The applied stereographic projection $p_{\mathrm{stereo}}(\zeta;\cdot) : \Omega \setminus \{-\zeta\} \to \mathbb{R}^2$ is defined via

$$p_{\mathrm{stereo}}(\zeta;\xi) = \left(\frac{2\xi \cdot (\mathbf{t}\varepsilon^1)}{1+\xi\cdot\zeta}, \frac{2\xi\cdot(\mathbf{t}\varepsilon^2)}{1+\xi\cdot\zeta}\right), \tag{103}$$

where $\varepsilon^1 = (1,0,0)^T$, $\varepsilon^2 = (0,1,0)^T$, $\varepsilon^3 = (0,0,1)^T$ denotes the canonical basis in $\mathbb{R}^3$ and $\mathbf{t} \in \mathbb{R}^{3\times 3}$ a rotation matrix with $\mathbf{t}\varepsilon^3 = \zeta$. From the harmonicity of $H_{n,k}^{(\mathrm{int})}(R;\cdot)$ in $\mathcal{D}_R$ it follows that $H_{n,k}^{\rho,\zeta}$ is harmonic in $\Gamma_\rho(\zeta)$. Note that, as always, harmonicity in the Euclidean space $\mathbb{R}^2$ is meant with respect to the Laplace operator while it is meant with respect to the Beltrami operator when we are intrinsic on the sphere Ω. Opposed to the Euclidean case, outer harmonics for spherical caps do not play a distinct role. Actually, for a spherical cap $\Gamma_\rho(\zeta)$, the corresponding outer harmonics coincide with the inner harmonics for the spherical cap $(\Gamma_\rho(\zeta))^c = \Gamma_{2-\rho}(-\zeta)$, which is why we do not consider them separately. The relation

$$\begin{aligned}\ln(1-\xi\cdot\eta) = &-\ln(2) + \ln(1+\xi\cdot\zeta) + \ln(1-\eta\cdot\zeta)\\ &- \sqrt{\rho(2-\rho)}\pi \sum_{n=1}^{\infty}\sum_{k=1}^{2} \frac{2}{n} H_{n,k}^{\rho,\zeta}(\xi) H_{n,k}^{2-\rho,-\zeta}(\eta),\end{aligned} \tag{104}$$

for $\xi \in \Omega\backslash\{-\zeta\}$, $\eta \in \Omega\backslash\{\zeta\}$, and $|p_{\text{stereo}}(\zeta;\xi)| < |p_{\text{stereo}}(\zeta;\eta)|$, eventually allows to transfer the completeness results from Theorem 4.1 to inner harmonics on spherical caps (for details, the reader is referred to [15]).

Theorem 4.3. *Let $\Gamma_\rho(\zeta)$ be a spherical cap with $\overline{\Gamma} \subset \Gamma_\rho(\zeta)$. Then the following statements hold true:*

(a) *The inner harmonics $\{H_{0,1}^{\rho,\zeta}\} \cup \{H_{n,k}^{\rho,\zeta}\}_{n\in\mathbb{N},k=1,2}$ form a complete, and hence closed function system in $\mathrm{L}^2(\partial\Gamma)$.*

(b) *The normal derivatives of the inner harmonics, i.e.,*

$$\{H_{0,1}^{\rho,\zeta}\} \cup \{\tfrac{\partial}{\partial\nu} H_{n,k}^{\rho,\zeta}\}_{n\in\mathbb{N},k=1,2}$$

form a complete and hence closed function system in $\mathrm{L}^2(\partial\Gamma)$.

We conclude this section by stating the use of the function systems from above for the approximation of solutions to the spherical boundary value problems (DP) and (NP) from Section 2.

Theorem 4.4. *Let $\{\Phi_k\}_{k\in\mathbb{N}_0}$ denote one of the function systems introduced in Theorem* 4.1(b)*, Remark* 4.2*, or Theorem* 4.3*, and $U \in \mathrm{C}^{(2)}(\Gamma) \cap \mathrm{C}^{(0)}(\overline{\Gamma})$ be a solution of one of the boundary value problems* (DP) *or* (NP)*. Then, for every $\varepsilon > 0$, there exist $M \in \mathbb{N}_0$ and coefficients $a_k \in \mathbb{R}$, $k = 0, 1, \ldots, M$, such that*

$$\left\| U - \sum_{k=0}^{M} a_k \Phi_k \right\|_{\mathrm{L}^2(\Gamma)} < \varepsilon. \tag{105}$$

The choice of M and the coefficients a_k, $k = 1, \ldots, M$, can be based solely on an approximation of U or $\frac{\partial}{\partial\nu}U$ on the boundary $\partial\Gamma$.

Remark 4.5. All the density and approximation results that were obtained in this section in an L^2-context also hold true in a $\mathrm{C}^{(0)}$-context with respect to the uniform topology and can be shown by the tools supplied throughout this chapter (see, e.g., [15, 17]).

5. Applications in geoscience

In this section, we present some applications of the previous tools to the approximation of different quantities of interest in physical geodesy. More precisely, we use techniques from Section 3.1 to reconstruct the disturbing potential from given vertical deflections over South America and the mean dynamic ocean topography (MDT) from given geostrophic ocean flow patterns over the Pacific Ocean, respectively. We will be rather brief about the geophysical derivations of the underlying spherical differential equations and refer the reader to classical literature such as [25, 28, 42, 44]. The particular formulations of our setting can also be found, e.g., in [10, 11, 13, 18]. Opposed to the latter, our reconstructions in Sections 5.1 and 5.2 are based on the approach in Section 3.1 via Neumann Green functions and does not require boundary information for the spherical caps under consideration.

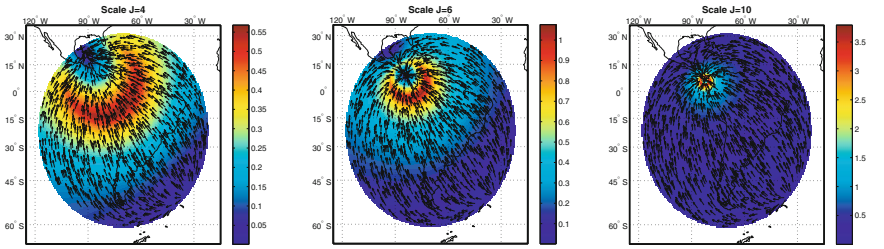


FIGURE 4. The kernel $\nabla^* G_N^J(\Delta^*;\xi,\cdot)$ for scales $J = 4, 6, 10$ and a fixed evaluation point ξ located at 7°N, 74°W (colors indicate the absolute value and arrows the orientation).

In Section 5.3, based on the results from Section 4, we address a model problem motivated by point vortex motion on the sphere.

5.1. Vertical deflections

The Earth's gravity potential $W = U + T$ is typically split into a normal gravity potential U corresponding to a reference ellipsoid $\mathcal{E}$ (i.e., $U(x) = \text{const.}$ for $x \in \mathcal{E}$) and a smaller remaining disturbing potential T. The vertical deflection $\Theta(x)$ measures the angular distance between the normal vector $\nu_{\mathcal{G}}(x)$ at a point x on the geoid $\mathcal{G}$ (i.e., $W(x) = \text{const.}$ for $x \in \mathcal{G}$) and the corresponding ellipsoidal normal vector $\nu_{\mathcal{E}}(x)$ with respect to $\mathcal{E}$. Assuming that $\nu_{\mathcal{G}} - \nu_{\mathcal{E}}$ and $\nu_{\mathcal{E}}$ are nearly orthogonal and that the deviation of the reference ellipsoid from a sphere is negligible, one can derive the following relation between the disturbing potential and the deflections of the vertical:

$$\nabla^* T(R\xi) = -\frac{GM}{R}\Theta(R\xi), \quad \xi \in \Omega, \tag{106}$$

where R is the Earth's mean radius, G the gravitational constant, and M the Earth's mass. For more details, the reader is referred to, e.g., [18, 25, 28]. We are particularly interested in solving (106) for the disturbing potential T in a subregion $\Gamma \subset \Omega$ (or, in other words, in a subregion Γ_R of the spherical Earth's surface Ω_R) from knowledge of the vertical deflections Θ only in that subregion. Theorem 3.3 yields the representation

$$T(R\xi) = \frac{1}{\|\Gamma\|}\int_\Gamma T(R\eta)d\omega(\eta) + \frac{GM}{R}\int_\Gamma \big(\nabla_\eta^* G_N(\Delta^*;\xi,\eta)\big)\cdot\Theta(R\eta)d\omega(\eta), \quad \xi \in \Gamma, \tag{107}$$

of which the first summand on the right-hand side simply represents the constant mean disturbing potential T^Γ_{mean} in Γ_R. We focus on the special case that $\Gamma = \Gamma_\rho(\zeta)$ is a spherical cap with center $\zeta \in \Omega$ and radius $\rho \in (0, 2)$, so that Theorem 2.7 supplies us with an explicit representation of the Neumann Green function $G_N(\Delta^*;\cdot,\cdot)$.

Concerning the numerical evaluation of (107), we first need to discretize the integral since Θ is typically only available in a discrete set of measurement points.

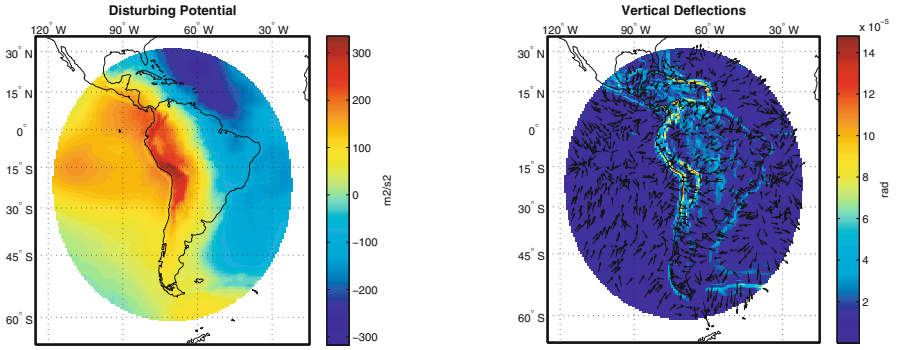


FIGURE 5. The 'true' disturbing potential T (left) and the corresponding vertical deflections Θ (right; colors indicate the absolute value and arrows the orientation).

For the tests in this section, we assume Θ to be given on a Gauss–Legendre grid in the spherical cap $\Gamma_\rho(\zeta)$, so that we can use the quadrature rule from [27]. Second, the numerical integration can become instable due to the singularity of the Neumann Green function $G_N(\Delta^*;\xi,\eta)$ at $\xi=\eta$ (originating in its contribution $\frac{1}{4\pi}\ln(1-\xi\cdot\eta)$). This can be circumvented by a regularization around this singularity via a truncated Taylor expansion. More precisely, for scaling parameters $J = 0,1,2,\ldots$, we define the regularized Neumann Green function

$$G_N^J(\Delta^*;\xi,\eta) = \begin{cases} \frac{1}{4\pi}\ln(1-\xi\cdot\eta) + \frac{1}{4\pi}\ln(\check{r}(1-\check{\xi}\cdot\eta)) \\ \quad + \frac{1-\rho}{2\pi\rho}\ln(1+\zeta\cdot\eta), & 1-\xi\cdot\eta \geq 2^{-J}, \\ \frac{2^J}{4\pi}(1-\xi\cdot\eta) - \frac{J}{4\pi}\ln(2) - \frac{1}{4\pi} \\ \quad + \frac{1}{4\pi}\ln(\check{r}(1-\check{\xi}\cdot\eta)) + \frac{1-\rho}{2\pi\rho}\ln(1+\zeta\cdot\eta), & 1-\xi\cdot\eta < 2^{-J}. \end{cases} \tag{108}$$

The regularization $G_N^J(\Delta^*;\cdot,\cdot)$ of the Neumann Green function $G_N(\Delta^*;\cdot,\cdot)$ closely relates to the regularization of the fundamental solution $G(\Delta^*;\cdot)$ briefly mentioned after Theorem 1.9. A stable approximation of T at scale J is then given by

$$T_J(R\xi) = T^\Gamma_{\text{mean}} + \frac{GM}{R}\int_{\Gamma_\rho(\zeta)} \left(\nabla^*_\eta G_N^J(\Delta^*;\xi,\eta)\right)\cdot\Theta(R\eta)d\omega(\eta), \quad \xi\in\Gamma_\rho(\zeta), \tag{109}$$

and satisfies $\lim_{J\to\infty}\sup_{\xi\in\tilde{\Gamma}}|T_J(R\xi)-T(R\xi)| = 0$ for every subset $\tilde{\Gamma}\subset\Gamma_\rho(\zeta)$ with $\operatorname{dist}(\tilde{\Gamma},\partial\Gamma_\rho(\zeta)) > 0$. Thus, higher scales J yield a more precise approximation of T and the difference $T_{J+1}-T_J$ between two consecutive scales reveals features of more and more local origin. The kernel $\nabla^* G_N^J(\Delta^*;\xi,\cdot)$ is illustrated in Figure 4.

In order to illustrate the reconstruction of the disturbing potential by the approximations T_J, we first compute a 'true' disturbing potential T from EGM2008

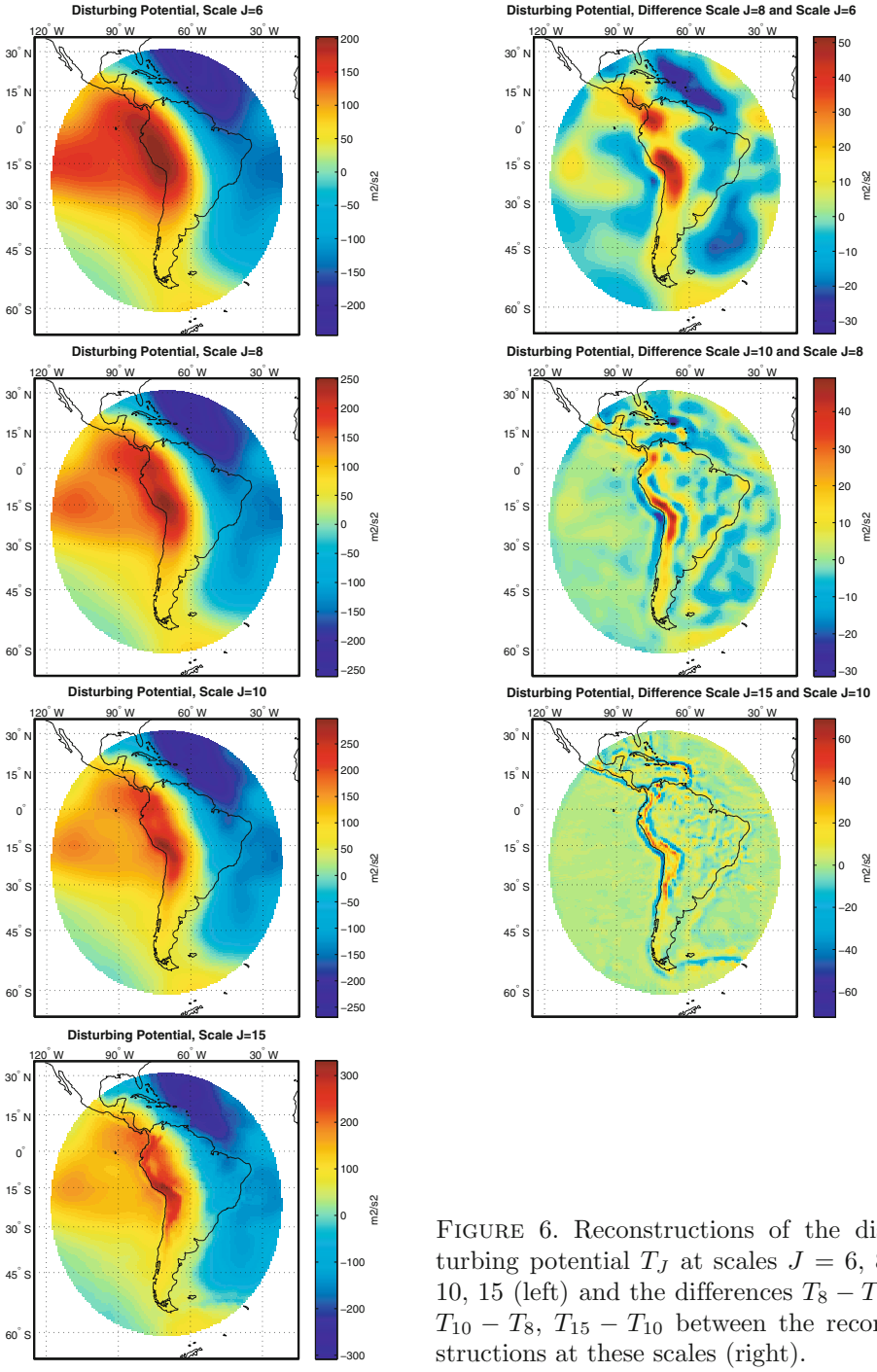


FIGURE 6. Reconstructions of the disturbing potential T_J at scales $J = 6$, 8, 10, 15 (left) and the differences $T_8 - T_6$, $T_{10} - T_8$, $T_{15} - T_{10}$ between the reconstructions at these scales (right).

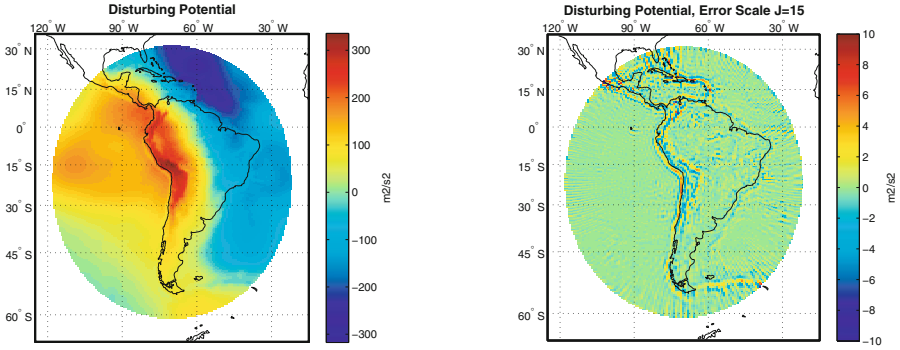


FIGURE 7. The 'true' disturbing potential T (left) and the reconstruction error $T - T_{15}$ (right).

(cf. [41]*) as a reference, using spherical harmonic degrees $n = 3, \ldots, 250$. From this T, we obtain our input vertical deflections Θ via (106) on a Gauss–Legendre grid of 63,252 points in a spherical cap over South America (cf. Figure 5). The approximations T_J for different scales J are shown in Figure 6. One can clearly see the refinement of the local features in the differences of the reconstructions T_J. Furthermore, the error $T - T_{15}$ in Figure 7 indicates a good approximation of T and does not reveal any artefacts due to the local reconstruction without use of any boundary information.

5.2. Geostrophic ocean flow

In subregions $\Gamma_R \subset \Omega_R$ of the ocean with a sufficiently large horizontal extent, away from the top and bottom Ekman layers and coastal regions, the geostrophic balance holds true: the horizontal pressure gradients in the ocean balance the Coriolis force resulting from horizontal currents. The Coriolis force term in a point $x \in \Gamma_R$ is given as the tangential contribution of $-2R\rho\, w \wedge v(x)$, where $v(x)$ is the horizontal ocean flow velocity and $w = |w|\varepsilon^3$ the Earth's rotation vector. ρ denotes the density and is assumed to be constant. The pressure $P(x)$ in $x \in \Gamma_R$ can be regarded as being proportional to the mean dynamic topography (MDT) $H(x)$, which denotes the height of the sea surface relative to the Geoid $\mathcal{G}$ and can be determined from altimetry measurements. More precisely, $P(x) = \rho G H(x)$, where G denotes the gravitational constant. Using the geostrophic balance, we therefore obtain

$$-2R\rho(w \cdot \xi)\, \xi \wedge v(R\xi) = \rho G \nabla^* H(R\xi), \qquad \xi \in \Gamma, \tag{110}$$

or, equivalently,

$$\frac{2R}{G}|w|(\xi \cdot \varepsilon^3) v(R\xi) = \mathrm{L}^* H(R\xi), \qquad \xi \in \Gamma. \tag{111}$$

*data accessed via
http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008/egm08_wgs84.html

For more details on the geophysical background, the reader is referred, e.g., to [42, 44]. In order to compute the MDT H from knowledge of the ocean flow velocity v in Γ, we need to solve Equation (111). Theorem 3.3 yields the representation

$$H(R\xi) = \frac{1}{\|\Gamma\|}\int_\Gamma H(R\eta)d\omega(\eta) - \frac{2R}{G}|w|\int_\Gamma (\eta\cdot\varepsilon^3)\left(\mathrm{L}_\eta^* G_N(\Delta^*;\xi,\eta)\right)\cdot v(R\eta)d\omega(\eta), \quad \xi\in\Gamma,$$

of which the first summand on the right-hand side simply represents the constant mean MDT H^Γ_{mean} in Γ_R. Again, we focus on the special case that $\Gamma = \Gamma_\rho(\zeta)$ is a spherical cap with center $\zeta\in\Omega$ and radius $\rho\in(0,2)$, so that we can apply the considerations from the previous section, i.e., we obtain an approximation at scale J by

$$H_J(R\xi) = H^\Gamma_{\mathrm{mean}} - \frac{2R}{G}|w|\int_{\Gamma_\rho(\zeta)} (\eta\cdot\varepsilon^3)\left(\mathrm{L}_\eta^* G_N^J(\Delta^*;\xi,\eta)\right)\cdot v(R\eta)d\omega(\eta), \quad \xi\in\Gamma_\rho(\zeta), \qquad (112)$$

where $G_N^J(\Delta^*;\cdot,\cdot)$ is given as in (108).

In order to illustrate the reconstruction of the MDT by the approximations H_J, we first compute a 'true' MDT H from [35]† as a reference. From this H, we can obtain our input ocean flow velocity v via (111) on a Gauss–Legendre grid of 63,252 points in a spherical cap over the Western Pacific Ocean (cf. Figure 8). The approximations H_J for different scales J are shown in Figure 9. The error $H - H_{15}$ in Figure 10 indicates a good approximation of H with larger errors only around the Hawaiian islands (where the geostrophic balance does not hold in the first place).

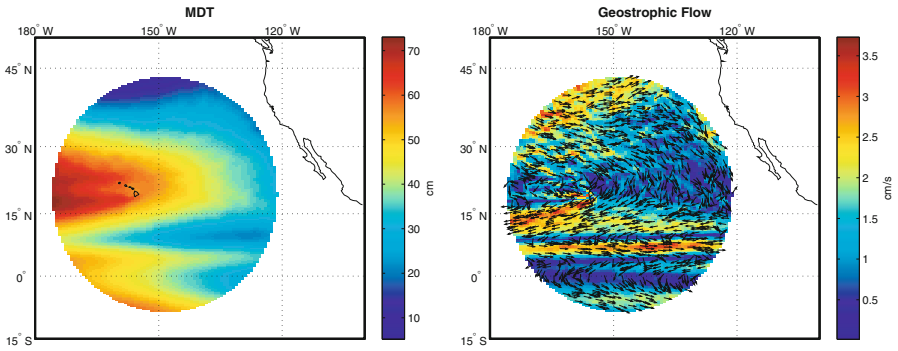


FIGURE 8. The 'true' MDT H (left) and the corresponding scaled geostrophic ocean flow velocity $\xi\mapsto(\xi\cdot\varepsilon^3)v(\xi)$ (right; colors indicate the absolute values and arrows the orientation).

†data accessed via http://apdrc.soest.hawaii.edu/projects/DOT

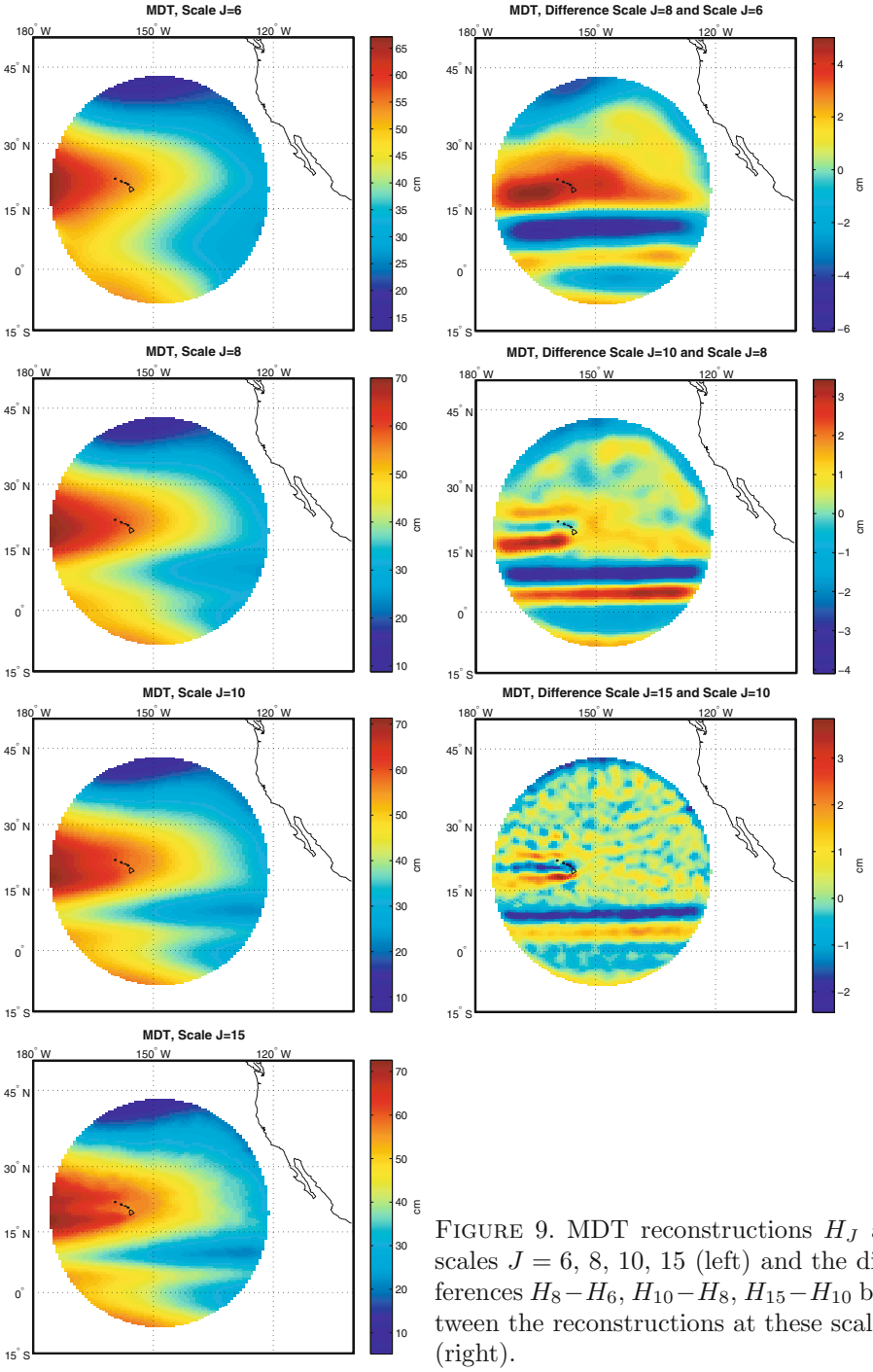


FIGURE 9. MDT reconstructions H_J at scales $J = 6$, 8, 10, 15 (left) and the differences $H_8 - H_6$, $H_{10} - H_8$, $H_{15} - H_{10}$ between the reconstructions at these scales (right).

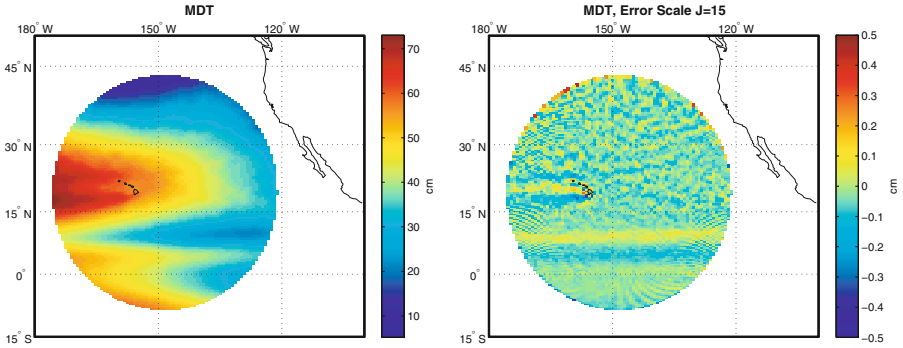


FIGURE 10. The true MDT H (left) and the reconstruction error $H - H_{15}$ (right).

5.3. Point vortex motion

Vorticity describes the rotational motion of a fluid. In the ocean, for horizontal flows v which extend over regions $\Gamma_R \subset \Omega_R$ at spatial scales of several tens or hundreds of kilometers, the following relation for the vorticity ω holds true:

$$\omega(R\xi) = \mathrm{L}^* \cdot v(R\xi), \quad \xi \in \Gamma. \tag{113}$$

The incompressible horizontal flow v itself can be represented by a stream function Ψ via $v = \mathrm{L}^*\Psi$, so that we obtain

$$\omega(R\xi) = \Delta^*\Psi(R\xi), \quad \xi \in \Gamma. \tag{114}$$

The geostrophic flow from Section 5.2 is an example for such a current. For more geophysical background on vorticity, the reader is again referred to [42, 44].

A single point vortex at location $R\eta$ on the sphere is associated with a vorticity $\omega(R\xi) = \bar{\omega}\left(\delta(1-\xi\cdot\eta) - \frac{1}{4\pi R}\right)$ of strength $\bar{\omega} \in \mathbb{R}$ (by δ we denote the Dirac distribution) and a corresponding stream function $\Psi(R\xi) = \frac{\bar{\omega}}{R}G(\Delta^*; \xi\cdot\eta)$, $\xi \in \Omega \setminus \{\eta\}$. If we consider a point vortex at location $R\eta$ in a subdomain $\Gamma_R \subset \Omega_R$ that produces no flow across the boundary $\partial\Gamma_R$ (e.g., a coastline), the vorticity would be $\omega(R\xi) = \bar{\omega}\left(\delta(1-\xi\cdot\eta)\right)$ and the corresponding stream function $\Psi(R\xi) = \frac{\bar{\omega}}{R}G_D(\Delta^*; \eta, \xi)$, $\xi \in \Gamma \setminus \{\eta\}$. In [20, 34], this motivated solving the model problem

$$\Delta^*\tilde{\Psi}(R\xi) = 0, \quad \xi \in \Gamma, \tag{115}$$

$$\tilde{\Psi}^-(R\xi) = \sum_{i=1}^{N} \frac{\bar{\omega}_i}{R} G(\Delta^*; \xi\cdot\eta_i) - \frac{\bar{\omega}_i}{4\pi R}\ln(1-\xi\cdot\bar{\xi}), \quad \xi \in \partial\Gamma, \tag{116}$$

for a fixed $\bar{\xi} \in \Gamma^c$ and point vortices of strengths $\bar{\omega}_i$ located at $R\eta_i \in \Gamma_R$, $i = 1, \ldots, N$. The actual stream function is then given by $\Psi(R\xi) = \sum_{i=1}^{N} \frac{\bar{\omega}_i}{R} G(\Delta^*; \xi\cdot\eta_i) - \frac{\bar{\omega}_i}{4\pi R}\ln(1-\xi\cdot\bar{\xi}) - \tilde{\Psi}(R\xi) = \sum_{i=1}^{N} \frac{\bar{\omega}_i}{R} G_D(\Delta^*; \eta_i, \xi)$, $\xi \in \Gamma \setminus \{\eta_1, \ldots, \eta_N\}$. More details on point vortex motion on the entire sphere (and more general closed

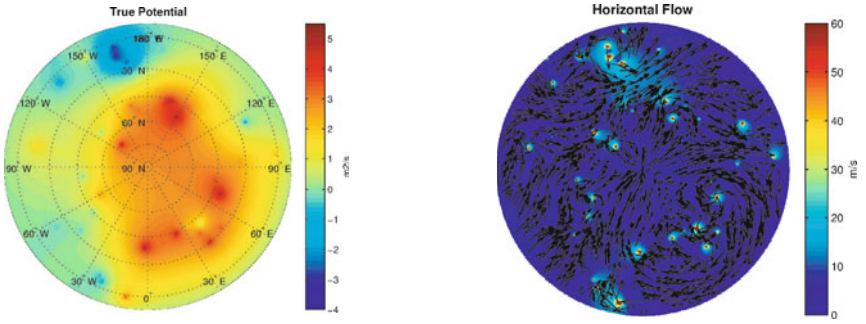


FIGURE 11. The 'true' potential Ψ (left) and the corresponding horizontal flow velocity v (right; colors indicate the absolute values and arrows the orientation).

manifolds) can be found, e.g., in [6, 7, 31, 33], and details on point vortex motion on subdomains of the sphere with impenetrable boundaries, e.g., in [24, 32].

In this section, we focus on the model problem (115), (116). Opposed to [20, 34], where boundary integral methods have been used, we want to solve it by the method of fundamental solutions based on the results of Section 4. More precisely, we choose Γ to be a spherical cap in the Northern hemisphere: $\Gamma = \Gamma_\rho(\zeta)$ with center $\zeta = (0,0,1)^T$ and radius $\rho = 0.9$. For simplicity, we set $R = 1$. The centers $\eta_i \in \Gamma_\rho(\zeta)$, $i = 1,\dots,N$, and the corresponding strengths $\bar{\omega}_i$ of the point vortices are chosen randomly. The point $\bar{\xi} \in (\Gamma_\rho(\zeta))^c$ from (116) is set to $\bar{\xi} = (0,0,-1)^T$. Furthermore, we assume the boundary data (116) to be given in equidistantly distributed points $\xi_i \in \partial\Gamma_\rho(\zeta)$, $i = 1,\dots,M$. Eventually, we interpolate the data by the functions $G_k^{(\mathrm{mod})}$, $k = 0,\dots,M-1$, from Theorem 4.1 and Remark 4.2, i.e.,

$$G_k^{(\mathrm{mod})}(\xi) = \frac{1}{4\pi}\ln(1-\xi\cdot\bar{\xi}_k) - \frac{1}{4\pi}\ln(1-\xi\cdot\bar{\xi}), \quad k = 1,\dots,M-1, \tag{117}$$

where the center points $\bar{\xi}_k$, $k = 1,\dots,M-1$, are chosen to be equidistantly distributed on $\partial\Gamma_{\bar{\rho}}(\zeta)$, for a radius $\bar{\rho} > \rho$. The resulting approximation $\Psi_{M,N,\bar{\rho}}$ of Ψ in $\Gamma_\rho(\zeta)$ is given by

$$\Psi_{M,N,\bar{\rho}}(\xi) = \sum_{i=1}^{N} \bar{\omega}_i G(\Delta^*; \xi\cdot\eta_i) - \frac{1}{4\pi}\ln(1-\xi\cdot\bar{\xi}) - \tilde{\Psi}_{M,N,\bar{\rho}}(\xi), \quad \xi \in \Gamma_\rho(\zeta), \tag{118}$$

$$\tilde{\Psi}_{M,N,\bar{\rho}}(\xi) = \sum_{k=0}^{M-1} a_k G_k^{(\mathrm{mod})}(\xi), \quad \xi \in \Gamma_\rho(\zeta), \tag{119}$$

where the coefficients a_k, $k = 0,\dots,M-1$, are obtained from the approximate solution of (115), (116) via interpolation of the boundary data. The resulting $\Psi_{M,N,\bar{\rho}}$ and the corresponding reconstruction errors are plotted in Figure 12 for

different settings of $M, \bar{\rho}$ (we fix the number of point vortices to $N = 40$). The actual potential Ψ and the underlying horizontal flow v are shown in Figure 11. We restrict our test example to a spherical cap $\Gamma = \Gamma_\rho(\zeta)$ because we then know an explicit representation of Ψ via the Dirichlet Green function $G_D(\Delta^*; \cdot, \cdot)$ from Section 4 and can compute the reconstruction errors. However, the approach can be easily adapted to more complex geometries of Γ.

The results in Figure 12 show a good performance for the test example of this easy to implement technique. The influence of the parameter $\bar{\rho}$ turns out to be fairly harmless for $M = 1000$ source points. A significant deterioration of the reconstruction error does not occur before $\bar{\rho} = 0.968$ (cf. Figure 12). However, in general, the method of fundamental solutions can be rather sensitive to the choice of the involved parameters, in particular of the source points $\bar{\xi}_k$ and the collocation points ξ_k. Furthermore, it can be advantageous to use a regularized least squares method instead of a simple interpolation. An overview on the method of fundamental solutions in general and its recent developments can be found, e.g., in [4, 9]. Latter, however, treat only the Euclidean setting. The current section is meant as a basic illustration of the method of fundamental solution for boundary value problems intrinsic on the sphere based on the techniques described in this chapter.

References

[1] O. Amm. Elementary currents for ionospheric fields. *J. Geomag. Geoelectr.*, 49:947–955, 1997.

[2] G. Backus, R. Parker, and C. Constable. *Foundations of Geomagnetism*. Cambridge University Press, 1996.

[3] L. Baratchart, D.P. Hardin, E.A. Lima, E.B. Saff, and B.P. Weiss. Characterizing kernels of operators related to thin plate magnetizations via generalizations of Hodge decompositions. *Inverse Problems*, 29:015004, 2013.

[4] C.S. Chen, A. Karageorghis, and Y.S. Smyrlis. *The Method of Fundamental Solutions – A Meshless Method*. Dynamic Publishers, Inc., 2008.

[5] R. Comblen, S. Legrand, E. Deleersnijdera, and V. Legata. A finite element method for solving the shallow water equations on the sphere. *Ocean Modelling*, 28:12–23, 2009.

[6] D.G. Dritschel. Contour dynamics/surgery on the sphere. *J. Comp. Phys.*, 78:477–483, 1988.

[7] D.G. Dritschel and S. Boatto. The motion of point vortices on closed surfaces. *Proc. R. Soc. A*, 471:20140890, 2015.

[8] R.L. Duduchava, D. Mitrea, and M. Mitrea. Differential operators and boundary value problems on hypersurfaces. *Math. Nachr.*, 279:996–1023, 2006.

[9] G. Fairweather and A. Karageorghis. The method of fundamental solutions for elliptic boundary value problems. *Adv. Comp. Math.*, 9:69–95, 1998.

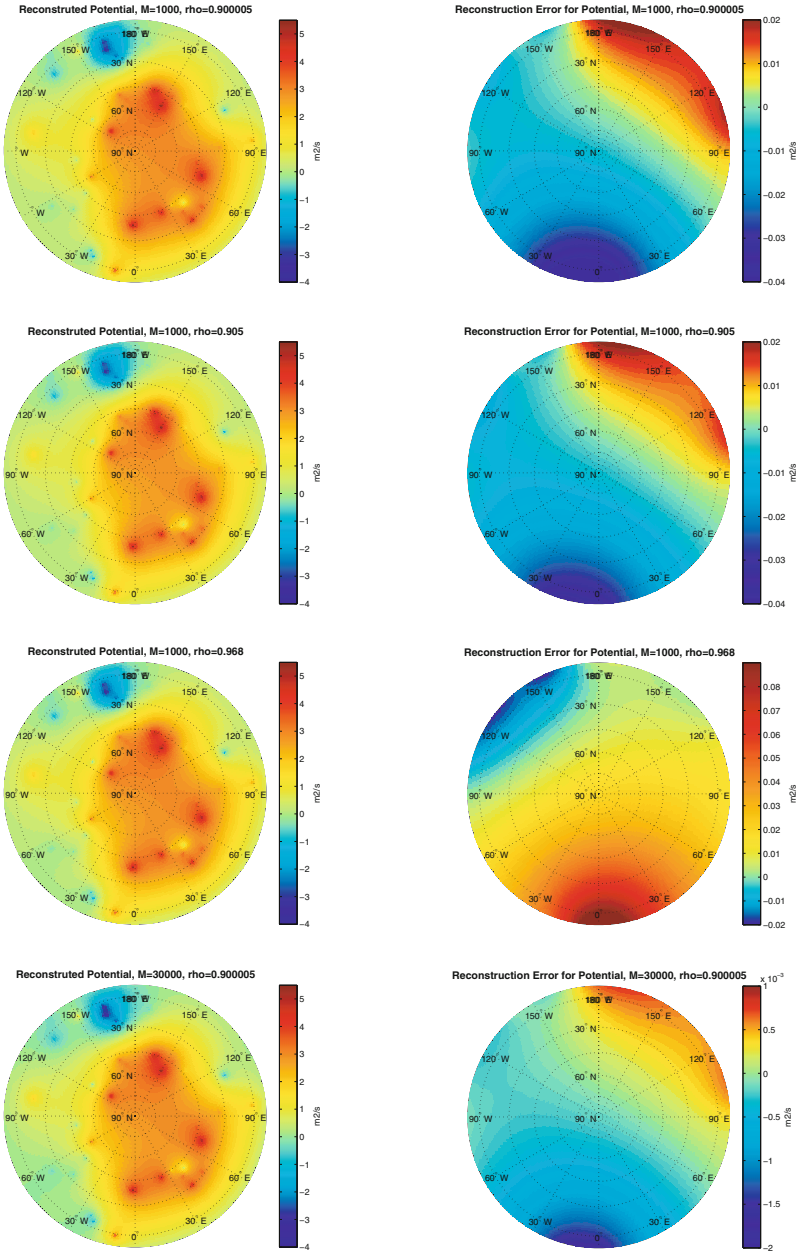


FIGURE 12. The reconstructed potential $\Psi_{M,N,\bar{\rho}}$ (left) and the corresponding reconstruction errors $\Psi - \Psi_{M,N,\bar{\rho}}$ (right) for $M = 1000, 30000$ and $\bar{\rho} = 0.900005, 0.905, 0.968$.

[10] T. Fehlinger, W. Freeden, C. Mayer, and M. Schreiner. On the local multiscale determination of the earths disturbing potential from discrete deflections of the vertical. *Comp. Geosc.*, 12:473–490, 2009.

[11] T. Fehlinger, W.S. Freeden, W. Freeden, C. Mayer, D. Michel, and M. Schreiner. Local modelling of sea surface topography from (geostrophic) ocean flow. *ZAMM*, 87:775–791, 2007.

[12] M.J. Fengler and W. Freeden. A nonlinear galerkin scheme involving vector and tensor spherical harmonics for solving the incompressible Navier–Stokes equation on the sphere. *SIAM J. Sci. Comp.*, 27:967–994, 2005.

[13] W. Freeden. Geomathematics: Its role, its aim, and its potential. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*. Springer, 2nd edition, 2015.

[14] W. Freeden, T. Fehlinger, M. Klug, D. Mathar, and K. Wolf. Classical globally reflected gravity field determination in modern locally oriented multiscale framework. *J. Geod.*, 83:1171–1191, 2009.

[15] W. Freeden and C. Gerhards. *Geomathematically Oriented Potential Theory*. Pure and Applied Mathematics. Chapman & Hall/CRC, 2012.

[16] W. Freeden, D. Michel, and V. Michel. Local multiscale approximations of geostrophic flow: Theoretical background and aspects of scientific computing. *Marine Geodesy*, 28:313–329, 2005.

[17] W. Freeden and V. Michel. *Multiscale Potential Theory* (*with Applications to Geoscience*). Birkhäuser, 2004.

[18] W. Freeden and M. Schreiner. *Spherical Functions of Mathematical Geosciences*. Springer, 2009.

[19] M. Ganesh, Q.T. LeGia, and I.H. Sloan. A pseudospectral quadrature method for Navier–Stokes equations on rotating spheres. *Math. Comp.*, 80:1397–1430, 2011.

[20] S. Gemmrich, N. Nigam, and O. Steinbach. Boundary integral equations for the Laplace–Beltrami operator. In H. Munthe-Kaas and B. Owren, editors, *Mathematics and Computation, a Contemporary View. Proceedings of the Abel Symposium* 2006. Springer, 2008.

[21] C. Gerhards. Spherical decompositions in a global and local framework: Theory and an application to geomagnetic modeling. *Int. J. Geomath.*, 1:205–256, 2011.

[22] C. Gerhards. Locally supported wavelets for the separation of spherical vector fields with respect to their sources. *Int. J. Wavel. Multires. Inf. Process.*, 10:1250034, 2012.

[23] C. Gerhards. Multiscale modeling of the geomagnetic field and ionospheric currents. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*. Springer, 2nd edition, 2015.

[24] E. Gutkin and K.P. Newton. The method of images and Green's function for spherical domains. *J. Phys. A: Math. Gen.*, 37:11989–12003, 2004.

[25] W.A. Heiskanen and H. Moritz. *Physical Geodesy*. W.H. Freeman and Company, 1967.

[26] L.L. Helms. *Introduction to Potential Theory*. Wiley-Interscience, 1969.

[27] K. Hesse and R.S. Womersley. Numerical integration with polynomial exactness over a spherical cap. *Adv. Comp. Math.*, 36:451–483, 2012.

[28] B. Hofmann-Wellenhof and H. Moritz. *Physical Geodesy*. Springer, 2nd edition, 2005.

[29] A.A. Il'in. The Navier–Stokes and Euler equations on two-dimensional closed manifolds. *Math. USSR Sb.*, 69:559–579, 1991.

[30] O.D. Kellogg. *Foundations of Potential Theory*. Springer, reprint edition, 1967.

[31] R. Kidambi and K.P. Newton. Motion of three point vortices on a sphere. *Physica D*, 116:143–175, 1998.

[32] R. Kidambi and K.P. Newton. Point vortex motion on a sphere with solid boundaries. *Phys. Fluids*, 12:581–588, 2000.

[33] Y. Kimura and H. Okamoto. Vortex motion on a sphere. *J. Phys. Soc. Jpn.*, 56:4203–4206, 1987.

[34] M.C.A. Kropinski and N. Nigam. Fast integral equation methods for the Laplace–Beltrami equation on the sphere. *Adv. Comp. Math.*, 40:577–596, 2014.

[35] N. Maximenko, P. Niiler, M.-H. Rio, L.C. Melnichenko, D. Chambers, V. Zlotnicki, and B. Galperin. Mean dynamic topography of the ocean derived from satellite and drifting buoy data using three different techniques. *J. Atm. Ocean. Tech.*, 26:1910–1919, 2009.

[36] C. Mayer and T. Maier. Separating inner and outer Earth's magnetic field from CHAMP satellite measurements by means of vector scaling functions and wavelets. *Geophys. J. Int.*, 167:1188–1203, 2006.

[37] M. Mitrea and M. Taylor. Boundary layer methods for Lipschitz domains in Riemannian manifolds. *J. Func. Anal.*, 163:181–251, 1999.

[38] M. Mitrea and M. Taylor. Potential theory on Lipschitz domains in Riemannian manifolds: Sobolev–Besov space results and the Poisson problem. *J. Func. Anal.*, 176:1–79, 2000.

[39] N. Olsen. Ionospheric F-region currents at middle and low latitudes estimated from MAGSAT data. *J. Geophys. Res.*, 102:4564–4576, 1997.

[40] N. Olsen, K.-H. Glassmeier, and X. Jia. Separation of the magnetic field into external and internal parts. *Space Sci. Rev.*, 152:135–157, 2010.

[41] N.K. Pavlis, S.A. Holmes, S.C. Kenyon, and J.K. Factor. The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *J. Geophys. Res.*, 117:B04406, 2012.

[42] J. Pedlosky. *Geophysical Fluid Dynamics*. Springer, 1979.

[43] L.M. Polvani and D.G. Dritschel. Wave and vortex dynamics on the surface of a sphere. *J. Fluid Mech.*, 255:35–64, 1993.

[44] R.H. Stewart. Introduction to physical oceanography. Online.

[45] J. Wermer. *Potential Theory*. Springer, 1974.

Christian Gerhards
Computational Science Center
University of Vienna
A-1090 Vienna, Autria
e-mail: christian.gerhards@univie.ac.at

Handbook of Mathematical Geodesy
Geosystems Mathematics, 855–882

Joint Inversion of Multiple Observations

Christian Gerhards, Sergiy Pereverzyev Jr., and Pavlo Tkachenko

Abstract. Joint inversion becomes increasingly important with the availability of various types of measurements related to the same quantity. Questions arising in this context are how to combine the different data sets in the first place and, secondly, how to choose the multiple parameters that naturally occur in such a combination. This chapter discusses some recently proposed techniques addressing these issues. Additionally, we distinguish the two cases when all underlying problems are ill posed (e.g., satellite data only) and when some of them are not ill posed (e.g., satellite data is complemented by data at the Earth surface). Theoretical discussions of the topics above are presented as well as numerical experiments with different settings of simulated data.

Keywords. Aggregation methods, combination of satellite and ground models.

1. Introduction

In various applications, especially in geoscience, one is provided with several data sets of observations of the same quantity of interest. These data sets may contain observations based on different physical principles (e.g., satellite-to-satellite tracking (SST) data as in the case of the GRACE satellite mission [29] and satellite gravity gradiometry (SGG) data as in the case of the GOCE satellite mission [27]) or observations of the same quantity at different locations (e.g., global magnetic field satellite measurements as by the CHAMP and Swarm missions [11, 23] and local magnetic field measurements at or near the Earth's surface). Depending on the kind of data, either the physical measurement principles have to be taken into account for the combination or properties originating from the location of the observations or possibly both.

The ill-posed problem of downward continuation of potential field data is an ever-present topic with gravitational and magnetic satellite data. It has been studied intensively, e.g., in [4, 18, 20, 21, 44, 46, 53, 56, 67, 68]. However, those studies treat downward continuation only for a single set of measurements. With multiple satellite observation models, such as SST and SGG, each providing approximations of the same quantity of interest, one is left with the choice of which to trust. A more advanced question is what to do with a less trustable approximation

or, in other words, whether an approximation that involves all available observations may actually serve as an effective way to reduce uncertainties in independent models. This question was and is discussed quite intensively in the geophysical literature, where the term 'joint inversion' was introduced by the authors of [69] for methods which provide a solution of various types of observation equations that are inverted simultaneously. A short overview about the application of joint inversion in geophysics may be found in [30]. To distinguish inversion methods based on different data combinations, researchers have introduced different names, e.g., aggregation [51], which was also used in the context of statistical regression analysis [37, 38]. Regardless of their names, what the above-mentioned approaches have in common is that they induce stability by simultaneously utilizing different types of indirect observations of the same phenomenon, which essentially limits the size of the class of possible solutions [2].

When additional data at or near the Earth's surface comes into play, not only ill-posed downward continuation influences the quality of the models but also the restricted local availability of data. A vast amount of research has been addressed to localization on the sphere. This involves, e.g., the development of spherical splines (e.g., [17, 61]), spherical cap harmonics (e.g., [31, 65]), Slepian functions (e.g., [58, 59, 62, 63]), as well as spherical multiscale methods (e.g., [7, 10, 14, 19, 22, 24, 34, 35, 39, 50, 52, 60]). However, methods that simultaneously address ill-posedness and localization are rather rare (e.g., [25, 59]). So, the task remains to find adequate ways of combining these different types of data and to appropriately choose the involved parameters.

We split this chapter into three sections that we have labeled according to what we believe their area of application might be rather than what mathematical methods have been used. Each of these brief sections is written in a way that should make it understandable without reading the other two.

Section 2 focuses on the combination of different models obtained from different ill-posed problems (e.g., the computation of the gravity potential at the Earth's surface from SST or SGG satellite measurements). It does not aim at obtaining the model directly from the supplied data but rather at aggregating different models into a more stable, trustworthy one. Joint inversion naturally leads to multi-parameter regularization (see, e.g., [40] for the regularization of geopotential determination from different types of satellite observations). At this point it is important to note that one should distinguish between multi-parameter schemes, where the regularization parameters penalize the norms of the approximant in different spaces and schemes where the parameters weigh the data misfits in different observation spaces. In the former schemes an observation space is fixed, and by changing the regularization parameters we try to find a suitable norm for the solution space, while in the latter schemes the situation is opposite: by changing the parameters we try to construct a common observation space as a weighted direct sum of given spaces. The choice of the regularization parameters for the former schemes has been extensively discussed in the literature. A few selected references are [8, 9, 13, 36, 45]. As to the latter schemes (schemes with a fixed solution space),

to the best of our knowledge, we can indicate only the paper [40], where a heuristic parameter choice rule is discussed, and the papers [15, 41], where the parameter choice is considered as a learning problem under the assumption that for similar inverse problems a suitable parameter choice has been known. It is clear that such approaches can be used only for particular classes of problems. In this section, we give a fairly general method that can be regarded as a parameter choice based on linear aggregation.

Section 3 provides a method on how to improve a global potential field model based on satellite data with additional discrete local data at the Earth's surface. The choice of the involved parameters is based on a method similar to that of Section 2.

Section 4 has a similar setup as Section 3. But while in Section 3 the goal is to obtain a global model that is refined by local data, the goal in Section 4 is to obtain a local model at the Earth's surface that is refined by global satellite data (and which eventually reveals a better reconstruction of the coarse features than a model purely based on local data). The presented method for the combination of the two data sets is based on the construction of convolution kernels that pay tribute to localization at the Earth's surface as well as to the ill-posedness of downward continuation (cf. [25]). Concerning the choice of the involved regularization parameters, the linear aggregation from Section 2 could generally be applied but the obtained error estimates do not hold true anymore when using the well-posed local ground data as a reference. The parameter choice method that we provide here is tailored for problems where direct measurements of the modeled quantity are available (cf. [26]).

2. Global combination of satellite models

In this section, we assume to have Hilbert spaces $\mathcal{X}, \mathcal{Y}_i$ and noisy observations $y_i^{\varepsilon_i} \in \mathcal{Y}_i$, $i = 1, \dots, m$, that are connected to the quantity of interest $x \in \mathcal{X}$ via

$$y_i^{\varepsilon_i} = \mathcal{A}_i x + e_i, \ i = 1, \dots, m. \tag{1}$$

The functions e_i represent the additive noise contained in the data and $\varepsilon_i > 0$ the noise level:

$$\|\mathcal{A}_i x - y_i^{\varepsilon_i}\|_{\mathcal{Y}_i} \leqslant \varepsilon_i, \quad i = 1, \dots, m, \tag{2}$$

The operators $\mathcal{A}_i : \mathcal{X} \to \mathcal{Y}_i$ are assumed to be compact, linear, and injective. The exact noise-free solution x of (1) will be denoted by $x^\dagger$. Classical choices for the situation above are $\mathcal{X} = L^2(\Omega_r)$ and $\mathcal{Y}_i = L^2(\Omega_{\rho_i})$, where $\Omega_r = \{\xi \in \mathbb{R}^3 : |x| = r\}$ denotes a spherical Earth surface of radius r and Ω_{ρ_i} spherical satellite orbits with radii $\rho_i > r$. Upward continuation operators can, e.g., have the form

$$\mathcal{A}_i x = \int_{\Omega_r} K_i(\cdot, \eta) x(\eta) dS(\eta), \quad i = 1, 2, \tag{3}$$

with kernels

$$K_1(\xi,\eta) = -\frac{1}{4\pi r}\frac{\partial}{\partial \rho_1}\left(\frac{\rho_1^2 - r^2}{(\rho_1^2 + r^2 - 2\xi\cdot\eta)^{3/2}}\right), \quad \xi \in \Omega_{\rho_1}, \eta \in \Omega_r,$$

for satellite-to-satellite tracking (SST) or

$$K_2(\xi,\eta) = \frac{1}{4\pi r}\frac{\partial^2}{\partial \rho_2^2}\left(\frac{\rho_2^2 - r^2}{(\rho_2^2 + r^2 - 2\xi\cdot\eta)^{3/2}}\right), \quad \xi \in \Omega_{\rho_2}, \eta \in \Omega_r,$$

for satellite gravity gradiometry (SGG). This setting will be used in the numerical examples in Section 2.2 later on.

The joint inversion of the multiple observation models (1), (2) can be formulated as the minimization problem

$$\min_{x\in\mathcal{X}} \sum_{i=1}^{m} \lambda_i \|\mathcal{A}_i x - y_i^{\varepsilon_i}\|_{\mathcal{Y}_i}^2 + \|x\|_{\mathcal{X}}^2. \tag{4}$$

The regularization parameters $\lambda_i > 0$ are introduced to adjust the contributions of the data misfit from the different observations. It is convenient to rewrite the objective functional of (4) in a compact form by introducing the direct sum $\mathcal{Y}_{\boldsymbol{\lambda}} = \bigoplus_{i=1}^m \mathcal{Y}_{i,\lambda_i}$ of the observation spaces. The inner product on $\mathcal{Y}_{\boldsymbol{\lambda}}$ is defined as $\langle y, \bar{y}\rangle_{\mathcal{Y}_{\boldsymbol{\lambda}}} = \sum_{i=1}^m \lambda_i \langle y_i, \bar{y}_i\rangle_{\mathcal{Y}_i}$, for $y = (y_i)_{i=1,\dots,m}, \bar{y} = (\bar{y}_i)_{i=1,\dots,m} \in \mathcal{Y}_{\boldsymbol{\lambda}}$, and $\boldsymbol{\lambda} = (\lambda_i)_{i=1,\dots,m}$. The norm $\|\cdot\|_{\mathcal{Y}_{\boldsymbol{\lambda}}}$ is defined according to this inner product. With these notations at hand, the minimization (4) can be rewritten as

$$\min_{x\in\mathcal{X}} \|\mathcal{A}_{\boldsymbol{\lambda}} x - y^{\boldsymbol{\varepsilon}}\|_{\mathcal{Y}_{\boldsymbol{\lambda}}}^2 + \|x\|_{\mathcal{X}}^2, \tag{5}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1,\dots,m}$ and $y^{\boldsymbol{\varepsilon}} = (y_i^{\varepsilon_i})_{i=1,\dots,m} \in \mathcal{Y}_{\boldsymbol{\lambda}}$.

The operator $\mathcal{A}_{\boldsymbol{\lambda}} = (\mathcal{A}_i)_{i=1,\dots,m} : \mathcal{X} \to \mathcal{Y}_{\boldsymbol{\lambda}}$ is again compact, linear, and injective. Representation (5) allows the classical Tikhonov–Phillips form of the minimizer $x_{\boldsymbol{\lambda}}^{\boldsymbol{\varepsilon}}$ of (4) and (5):

$$x_{\boldsymbol{\lambda}}^{\boldsymbol{\varepsilon}} = (I + \mathcal{A}_{\boldsymbol{\lambda}}^*\mathcal{A}_{\boldsymbol{\lambda}})^{-1}\mathcal{A}_{\boldsymbol{\lambda}}^* y^{\boldsymbol{\varepsilon}}, \tag{6}$$

where $I : \mathcal{X} \to \mathcal{X}$ is the identity operator and $\mathcal{A}_{\boldsymbol{\lambda}}^* : \mathcal{Y}_{\boldsymbol{\lambda}} \to \mathcal{X}$ is the adjoint of $\mathcal{A}_{\boldsymbol{\lambda}}$.

A similar representation as in (5) was obtained in [40] by Bayesian reasoning. It is also suggested in [40] to relate the values of the regularization parameters λ_i with the observation noise levels (variances) ε_i as follows:

$$\lambda_i = \lambda_1 \frac{\varepsilon_1^2}{\varepsilon_i^2}, \quad i = 1,\dots,m. \tag{7}$$

Note that this relation reduces the multi-parameter regularization (4) to a single-parameter regularization since only λ_1 needs to be chosen. The heuristic rule (7) can be motivated from a bound for the noise propagation error. Specifically, we

have that

$$\begin{aligned}\|x^0_{\boldsymbol{\lambda}} - x^{\boldsymbol{\varepsilon}}_{\boldsymbol{\lambda}}\|_{\mathcal{X}} &= \|(I + \mathcal{A}^*_{\boldsymbol{\lambda}}\mathcal{A}_{\boldsymbol{\lambda}})^{-1}\mathcal{A}^*_{\boldsymbol{\lambda}}(y - y^{\boldsymbol{\varepsilon}})\|_{\mathcal{X}} \\ &\leqslant \|(I + \mathcal{A}^*_{\boldsymbol{\lambda}}\mathcal{A}_{\boldsymbol{\lambda}})^{-1}\mathcal{A}^*_{\boldsymbol{\lambda}}\|_{\mathcal{Y}_{\boldsymbol{\lambda}}\to\mathcal{X}}\|y - y^{\boldsymbol{\varepsilon}}\|_{\mathcal{Y}_{\boldsymbol{\lambda}}} \\ &\leqslant \frac{1}{2}\|\mathcal{A}_{\boldsymbol{\lambda}}x^\dagger - y^{\boldsymbol{\varepsilon}}\|_{\mathcal{Y}_{\boldsymbol{\lambda}}},\end{aligned}$$

where $y^0 := (y^0_i)_{i=1,\dots,m} = (\mathcal{A}_i x^\dagger)_{i=1,\dots,m} \in \mathcal{Y}_{\boldsymbol{\lambda}}$ denotes noise-free input data. It follows from Assumption (2) that

$$\|x^0_{\boldsymbol{\lambda}} - x^{\boldsymbol{\varepsilon}}_{\boldsymbol{\lambda}}\|_{\mathcal{X}} \leqslant \frac{1}{2}\left(\sum_{i=1}^m \lambda_i \varepsilon_i^2\right)^{\frac{1}{2}}. \tag{8}$$

The heuristics behind the rule (7) is now clear: The choice (7) equates all the terms on the right-hand side of (8) and balances the data misfits against each other. The final balance may be achieved by making a choice of the remaining parameter $\lambda = \lambda_1$. The latter can be chosen by known single-parameter choice rules such as the quasi-optimality criterion [66] (let us label this strategy by M1 for later reference). Within this strategy, we choose the value λ from

$$\Lambda^q_N = \{\tilde{\lambda}_k = \bar{\lambda}q^k : k = 0, 1, \dots, N\},$$

for some fixed $q > 1$ and $\bar{\lambda} > 0$ such that

$$\left|x^{\boldsymbol{\varepsilon}}_{\lambda} - x^{\boldsymbol{\varepsilon}}_{\lambda q^{-1}}\right| = \min\left\{\left|x^{\boldsymbol{\varepsilon}}_{\tilde{\lambda}_k} - x^{\boldsymbol{\varepsilon}}_{\tilde{\lambda}_{k-1}}\right| : \tilde{\lambda}_k, \tilde{\lambda}_{k-1} \in \Lambda^q_N\right\}. \tag{9}$$

A multi-parameter version of the quasi-optimality (QO) criterion can be used as an alternative to the single-parameter reduction (7). We denote this strategy by M2 for later reference. For the sake of clarity we describe it here only for the case of two parameters: for each fixed regularization parameter, say $\lambda_1 \in \Lambda^q_N$, we chose the value $\lambda_2 = \lambda_2(\lambda_1) \in \Lambda^q_N$ according to the one-parameter QO criterion (9) with respect to $\lambda = \lambda_2$. Then we repeat the strategy (9) for the set of pairs $(\lambda_1, \lambda_2(\lambda_1))$ with respect to $\lambda = \lambda_1$.

2.1. The linear aggregation method

If a priori information like noise level of the measurements is given, then many studies are available on the choice of the regularization parameters. If this is not the case, one has to fall back on heuristic rules, e.g., quasi-optimality as mentioned above. In this section, we propose to 'aggregate' various approximations of $x^\dagger$ based on such heuristic rules in order to obtain an improved approximation. More details on the study of this scheme can be found in [12].

Let us assume for now that, from somewhere, we have obtained M different approximations $x^{\boldsymbol{\varepsilon}}_j$, $j = 1, \dots, M$. The goal is to find an optimal linear combination ('aggregation')

$$x^{\boldsymbol{\varepsilon}}_{\mathrm{ag},\beta^{\mathrm{opt}}} = \sum_{j=1}^M \beta_j^{\mathrm{opt}} x^{\boldsymbol{\varepsilon}}_j, \tag{10}$$

in the sense that $\beta^{\mathrm{opt}} = (\beta_j^{\mathrm{opt}})_{j=1,\dots,M}$ minimizes

$$\min_{\beta} \left\| x^\dagger - \sum_{j=1}^{M} \beta_j x_j^{\boldsymbol{\varepsilon}} \right\|_{\mathcal{X}} . \tag{11}$$

The solution β^{opt} of (11) is determined by solving a system of linear equations. To state this system of linear equations, we need the Gram matrix

$$\boldsymbol{G} = \left(\langle x_i^{\boldsymbol{\varepsilon}}, x_j^{\boldsymbol{\varepsilon}} \rangle_{\mathcal{X}}\right)_{i,j=1,\dots,M} \quad \text{and} \quad \boldsymbol{\kappa} = \left(\langle x_j^{\boldsymbol{\varepsilon}}, x^\dagger \rangle_{\mathcal{X}}\right)_{j=1,\dots,M}.$$

Then

$$\boldsymbol{G}\beta^{\mathrm{opt}} = \boldsymbol{\kappa}. \tag{12}$$

Remark 2.1. If the approximations $x_j^{\boldsymbol{\varepsilon}}$, $j = 1, \dots, M$, are linearly independent, then the Gram matrix $\boldsymbol{G}$ is positive definite and invertible such that

$$\|\boldsymbol{G}^{-1}\|_{\mathbb{R}^M \to \mathbb{R}^M} \leqslant c_1,$$

for some constant $c_1 > 0$.

At the technical level, we are actually more concerned with the degree of the linear correlation of $x_j^{\boldsymbol{\varepsilon}}$, $j = 1, \dots, M$, and the condition number of the Gram matrix $\boldsymbol{G}$. In principle, one may control this by excluding those members of the family $X_M = \{x_j^{\boldsymbol{\varepsilon}} : j = 1, \dots, M\}$ that are close to be linearly dependent of others. It is clear that their exclusion does not significantly change the value of (11). Gram–Schmidt orthogonalization combined with a thresholding technique can achieve this. The details can be found in [12].

The problem with the approach (10)–(12) is that the right-hand side $\boldsymbol{\kappa}$ of (12) is not known since $x^\dagger$ is not accessible. Summarizing our results so far, we propose the following method to obtain an improved approximation of $x^\dagger$.

Method 2.2. *From somewhere, we are given M different approximations $x_j^{\boldsymbol{\varepsilon}}$, $j = 1, \dots, M$ of $x^\dagger$ (based on the m observation equations (1), (2)). Beyond that, we assume to have a particularly trustworthy approximation $\tilde{x} \in \mathcal{X}$ that we use to define $\tilde{\boldsymbol{\kappa}} = \left(\langle x_j^{\boldsymbol{\varepsilon}}, \tilde{x} \rangle_{\mathcal{X}}\right)_{j=1,\dots,M}$. The Gram matrix $\boldsymbol{G}$ is given by $\boldsymbol{G} = \left(\langle x_i^{\boldsymbol{\varepsilon}}, x_j^{\boldsymbol{\varepsilon}} \rangle_{\mathcal{X}}\right)_{i,j=1,\dots,M}$. The parameter $\beta^* = (\beta_j^*)_{j=1,\dots,M}$ is chosen as the solution to*

$$\boldsymbol{G}\beta^* = \tilde{\boldsymbol{\kappa}} \tag{13}$$

and the corresponding aggregated approximation of $x^\dagger$ is given by

$$x_{\mathrm{ag},\beta^*}^{\boldsymbol{\varepsilon}} = \sum_{j=1}^{M} \beta_j^* x_j^{\boldsymbol{\varepsilon}},$$

It remains to study the influence of the approximation $\tilde{\boldsymbol{\kappa}}$, i.e., how well does $x_{\mathrm{ag},\beta^*}^{\boldsymbol{\varepsilon}}$ perform in comparison to the optimal aggregation $x_{\mathrm{ag},\beta^{\mathrm{opt}}}^{\boldsymbol{\varepsilon}}$ (based on the exact $\boldsymbol{\kappa}$)? To answer this, we regard the linear functional strategy as introduced in [1, 3, 42]. The essence of this strategy is that one is not interested in completely knowing $x^\dagger$ but only in knowing some quantity derived from it, such as the value

of a bounded linear functional $\langle \bar{x}, \cdot \rangle_{\mathcal{X}}$ applied to the solution $x^\dagger$. This quantity can be estimated more accurately than $x^\dagger$ itself. In order to allow a theoretically rigorous study, we will choose a particular $\tilde{\boldsymbol{\kappa}}$:

First, we pick the most trustworthy observation equation among those m ones available from (1), (2). The choice might, e.g., be the problem with the operator $\mathcal{A}_j$ that is least ill posed or the one with the data $y_j^{\varepsilon_j} \in \mathcal{Y}_j$ that has the lowest noise level ε_j. We abbreviate the chosen observation equation by

$$y^\varepsilon = \mathcal{A}x + e, \tag{14}$$

$$\|\mathcal{A}x - y^\varepsilon\|_{\mathcal{Y}} \leqslant \varepsilon. \tag{15}$$

Furthermore, x_α^ε denotes the Tikhonov–Phillips regularized solution to this problem, i.e.,

$$x_\alpha^\varepsilon = x_\alpha^\varepsilon(y^\varepsilon) = (\alpha I + \mathcal{A}^*\mathcal{A})^{-1}\mathcal{A}^* y^\varepsilon, \tag{16}$$

for some parameter $\alpha > 0$. From [32, 48, 49], we know that if $\varphi : [0, \|A\|^2_{\mathcal{X}\to\mathcal{Y}}] \to [0, \infty)$ is an index function (i.e., φ is continuous, strictly increasing, and satisfies $\varphi(0) = 0$), if $\frac{t}{\varphi(t)}$ is non-decreasing, and if $A_{\varphi,R}$ denotes the source condition set

$$A_{\varphi,R} = \{\bar{x} \in \mathcal{X} : \bar{x} = \varphi(\mathcal{A}^*\mathcal{A})v, \|v\|_{\mathcal{X}} \leqslant R\}, \quad R > 0 \text{ fixed}, \tag{17}$$

then the convergence rate

$$\sup_{x^\dagger \in \mathcal{A}_{\varphi,R}} \sup_{\substack{y^\varepsilon \in \mathcal{Y}; \\ \|\mathcal{A}x^\dagger - y^\varepsilon\|_{\mathcal{Y}} \leqslant \varepsilon}} \|x^\dagger - x_\alpha^\varepsilon(y^\varepsilon)\|_{\mathcal{X}} = O(\varphi(\theta^{-1}(\varepsilon))), \tag{18}$$

holds true for the choice $\alpha = \theta^{-1}(\varepsilon)$, with $\theta(t) = \varphi(t)\sqrt{t}$. Note that $O(\varphi(\theta^{-1}(\varepsilon)))$ is the best guaranteed order of accuracy for the reconstruction of $x^\dagger \in A_{\varphi,R}$ from the observation (14) and (15). Now, for each of the M available approximations $x_j^{\boldsymbol{\varepsilon}}$, $j = 1, \ldots, M$, we can approximate the component κ_j of the vector $\boldsymbol{\kappa}$ by

$$\tilde{\kappa}_j = \langle x_j^{\boldsymbol{\varepsilon}}, x_{\alpha_j}^\varepsilon \rangle_{\mathcal{X}}, \tag{19}$$

where $\alpha_j = \alpha = \theta^{-1}(\varepsilon)$. The approximation $\tilde{\boldsymbol{\kappa}}$ of $\boldsymbol{\kappa}$ is then defined via $\tilde{\boldsymbol{\kappa}} = (\tilde{\kappa}_j)_{j=1,\ldots,M}$. In [12] it has been shown that under the conditions above we have

$$\left| \langle x_j^{\boldsymbol{\varepsilon}}, x^\dagger \rangle_{\mathcal{X}} - \langle x_j^{\boldsymbol{\varepsilon}}, x_{\alpha_j}^\varepsilon \rangle_{\mathcal{X}} \right| = o(\varphi(\theta^{-1}(\varepsilon))) \tag{20}$$

However, in practice, the function φ describing the smoothness of the unknown solution $x^\dagger$ is not known. As a result, one cannot implement the *a priori* parameter choice $\alpha_j = \alpha = \theta^{-1}(\varepsilon)$. In principle, this difficulty may be resolved by use of the so-called Lepskii-type balancing principle, introduced in [6, 28] in the context of the linear functional strategy. But, the *a posteriori* parameter choice strategy presented in those papers requires the knowledge of an index functions ψ_j describing the smoothness of $x_j^{\boldsymbol{\varepsilon}}$ in terms of the source condition $x_j^{\boldsymbol{\varepsilon}} \in \mathrm{Range}(\psi_j(\mathcal{A}^*\mathcal{A}))$. This requirement may be too restrictive in some applications. To overcome this difficulty, we consider below a modification of the balancing principle to achieve the error bounds (20) without requiring the knowledge of φ or ψ_j.

The balancing principle is well known in the literature (see, for example, [28, 43] Section 1.1.5, and [47]). Following the general theory, we formulate a version of the balancing principle suitable for our context: We define a parameter set

$$\Lambda_N = \{\varepsilon^2 = \tilde{\alpha}_1 < \tilde{\alpha}_2 < \cdots < \tilde{\alpha}_N = 1\}$$

and choose α_j, $j = 1, \ldots, M$, according to the balancing principle

$$\alpha_j = \max\Big\{\tilde{\alpha}_k \in \Lambda_N : \big|\langle x_j^{\boldsymbol{\varepsilon}}, x_{\tilde{\alpha}_k}\rangle_{\mathcal{X}} - \langle x_j^{\boldsymbol{\varepsilon}}, x_{\tilde{\alpha}_l}\rangle_{\mathcal{X}}\big| \leqslant 4\varepsilon \left\|(\tilde{\alpha}_l I + \mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}x_j^{\boldsymbol{\varepsilon}}\right\|_{\mathcal{Y}}, l = 1, \ldots, k-1\Big\}. \tag{21}$$

Obviously, the choice of α_j as above does not require the knowledge of the index function φ. These considerations culminate in the following theorem. A detailed derivation can found in [12].

Theorem 2.3. *Suppose that $x^{\boldsymbol{\varepsilon}}_{\mathrm{ag},\beta^{\mathrm{opt}}}$ is the optimal aggregator in the sense of* (10)–(12) *and $x^{\boldsymbol{\varepsilon}}_{\mathrm{ag},\beta^*}$ is its approximation according to Method* 2.2. *The approximation $\tilde{\boldsymbol{\kappa}}$ required in Method* 2.2 *shall be constructed from* (16), (19), (21), *using only $y^\varepsilon, \mathcal{A}$ from* (14), (15), *and $x_j^{\boldsymbol{\varepsilon}}$, $j = 1, \ldots, M$. If $x^\dagger \in A_{\varphi,R}$, where φ is an index function such that the function $\frac{t}{\varphi(t)}$ is increasing and takes zero value at the origin, then*

$$\|x^\dagger - x^{\boldsymbol{\varepsilon}}_{\mathrm{ag},\beta^*}\|_{\mathcal{X}} - \|x^\dagger - x^{\boldsymbol{\varepsilon}}_{\mathrm{ag},\beta^{\mathrm{opt}}}\|_{\mathcal{X}} = o(\varphi(\theta^{-1}(\varepsilon))).$$

Theorem 2.3 tells us that the coefficients β_j^* of the aggregator $x^{\boldsymbol{\varepsilon}}_{\mathrm{ag},\beta^*}$ can be effectively obtained from the input data in such a way that the error $\|x^\dagger - x^{\boldsymbol{\varepsilon}}_{\mathrm{ag},\beta^*}\|_{\mathcal{X}}$ differs from the optimal error $\|x^\dagger - x^{\boldsymbol{\varepsilon}}_{\mathrm{ag},\beta^{\mathrm{opt}}}\|_{\mathcal{X}}$ by a quantity of higher order than the best guaranteed accuracy of the reconstruction of $x^\dagger$ from the most trustable observation (14), (15).

Remark 2.4. On the one hand, it is clear that involving more linearly independent solutions in the aggregation helps in reducing the approximation error. On the other hand, from the proof of Theorem 2.3 in [12], it can be seen that the coefficient implicitly involved in $o(\varphi(\theta^{-1}(\varepsilon)))$ increases with M at least as fast as $\sqrt{M}$. This means that in order to be effective, the aggregator $x^{\boldsymbol{\varepsilon}}_{\mathrm{ag},\beta^*}$ should be built on the basis of a modest number of approximations $x_j^{\boldsymbol{\varepsilon}}$. In our numerical illustrations, we use $M = 2$.

Also note that in our analysis, the balancing principle (21) has been used mainly for theoretical reasons. In the numerical experiments below, the regularization parameters α_j are chosen by the quasi-optimality criterion (compare (9)) such that

$$\begin{aligned}&\Big|\langle x_j^{\boldsymbol{\varepsilon}}, x^\varepsilon_{\alpha_j}\rangle_{\mathcal{X}} - \langle x_j^{\boldsymbol{\varepsilon}}, x^\varepsilon_{\alpha_j q^{-1}}\rangle_{\mathcal{X}}\Big| \\ &\quad = \min\left\{\Big|\langle x_j^{\boldsymbol{\varepsilon}}, x^\varepsilon_{\tilde{\alpha}_k}\rangle_{\mathcal{X}} - \langle x_j^{\boldsymbol{\varepsilon}}, x^\varepsilon_{\tilde{\alpha}_{k-1}}\rangle_{\mathcal{X}}\Big| : \tilde{\alpha}_k, \tilde{\alpha}_{k-1} \in \Lambda_N^q\right\}.\end{aligned} \tag{22}$$

A practical advantage of the quasi-optimality criterion is that it does not require the knowledge of the noise level ε. At the same time, as it was shown in [5], under

some assumptions on the noise spectral properties and on $x^\dagger$, the quasi-optimality criterion allows optimal order error bounds.

The strategy presented in this subsection, i.e., Method 2.2 with $\tilde{\boldsymbol{\kappa}}$ chosen according to (16), (19), and (22), can be seen as the third method for the parameter choice. We label this strategy by M3 for later use.

2.2. Numerical examples

We present in this section numerical experiments to demonstrate the efficiency of the proposed aggregation methodand to compare it with other known methods in the literature. All data are simulated in a way that they mimic the inputs of the SST-problem and the SGG-problem described by the equations (1), (3).

It is well known (see, e.g., [18]) that the integral operators $\mathcal{A}_i$ defined by (3) with the kernels $K_i,\ i = 1, 2$, act between the Hilbert spaces $\mathcal{X} = L^2(\Omega_r)$ and $\mathcal{Y}_i = L^2(\Omega_{\rho_i})$ of square-summable functions on the spheres $\Omega_r, \Omega_{\rho_i},\ i = 1, 2$, and admit the singular value expansions

$$\mathcal{A}_i x(\xi) = \sum_{k=0}^{\infty} a_k^{(i)} \sum_{l=0}^{2k+1} \frac{1}{\rho_i} Y_{k,l}\left(\frac{\xi}{\rho_i}\right) \left\langle \frac{1}{r} Y_{k,l}\left(\frac{\cdot}{r}\right), x \right\rangle_{L^2(\Omega_r)}, \quad \xi \in \Omega_{\rho_i}, \tag{23}$$

where the $Y_{k,l}$ denotes the orthonormal spherical harmonic of degree k and order l on the unit sphere Ω_1, and

$$a_k^{(1)} = \left(\frac{r}{\rho_1}\right)^k \frac{k+1}{\rho_1},\ a_k^{(2)} = \left(\frac{r}{\rho_2}\right)^k \frac{(k+1)(k+2)}{\rho_2^2}, \quad k \in \mathbb{N}_0. \tag{24}$$

The solution $x^\dagger$ to (1), (3) (i.e., the restriction of x to Ω_r) models the gravitational potential measured at the sphere Ω_r, that is expected to belong to the spherical Sobolev space $H_s(\Omega_r)$ with $s = \frac{3}{2}$ (see, e.g., [64]), which means that its Fourier coefficients $\widehat{x}(k,l) = \langle \frac{1}{r} Y_{k,l}\left(\frac{\cdot}{r}\right), x\rangle_{L^2(\Omega_r)}$ asymptotically decay at least as $(k+1)^{-\frac{3}{2}}$. Therefore, to produce the data for our numerical experiments we simulate the Fourier coefficients $\widehat{x}(k,l)$ of the solution $x^\dagger$ via

$$\widehat{x}(k,l) = (k+1)^{-\frac{3}{2}} g_{k,l}, \quad k \in \mathbb{N}_0, l = 1, \ldots, 2k+1,$$

where $g_{k,l}$ are uniformly distributed random values in $[-1, 1]$. In view of (23), the Fourier coefficients $\widehat{y_i^{\varepsilon_i}}(k,l) = \langle \frac{1}{\rho_i} Y_{k,l}\left(\frac{\cdot}{\rho_i}\right), y_i^{\varepsilon_i}\rangle_{L_2(\Omega_{\rho_i})}$ of the noisy data $y_i^{\varepsilon_i}$ are simulated as

$$\widehat{y_i^{\varepsilon_i}}(k,l) = a_k^{(i)} \widehat{x}(k,l) + e_{k,l}^{(i)}, \quad k \in \mathbb{N}_0, l = 1, \ldots, 2k+1,\ i = 1, 2,$$

where $e_{k,l}^{(i)}$ is Gaussian white noise which roughly correspond to (2) with noise level ratio $\frac{\varepsilon_1}{\varepsilon_2} = 3$. All random simulations are performed 500 times such that we have data for 1000 problems of the form (1), (3). Moreover, we take $r = 6371$km for the radius of the Earth, and $\rho_1 = 6621$km, $\rho_2 = 6771$km. All spherical Fourier coefficients are simulated up to the degree $N = 300$, which is in agreement with the dimension of the existing models, such as Earth Gravity Model 96 (EGM96). Thus, the set of simulated problems consists of 500 pairs of the SGG- and SST-

type problems (1), (3). In our experiments, each pair is inverted jointly by means of Tikhonov–Phillips regularization (4), (5) performed in a direct weighted sum of the observation spaces $\mathcal{Y}_i = L^2(\Omega_{\rho_i})$, $i = 1, 2$, and we use three methods for choosing the regularization parameters (weights) λ_1, λ_2:

In the first method (i.e., M1), we relate them according to (7). Recall that the data are simulated such that $\frac{\varepsilon_1}{\varepsilon_2} = 3$. Therefore, we have $\lambda_2 = 9\lambda_1$. Then the parameter λ_1 is chosen according to the standard quasi-optimality criterion from the geometric sequence $\Lambda_{30} = \{10^{\frac{40+j}{8}} : j = 0, \dots, 30\}$. As a result, for each of 500 pairs of the simulated problems we apply M1 and obtain a regularized approximation to the solution $x^\dagger$ that will play the role of the approximant $x_1^{\boldsymbol{\varepsilon}}$.

In the second method (i.e., M2), the parameters λ_1, λ_2 are selected from Λ_{30} according to the multi-parameter version of the quasi-optimality criterion. In this way, for each of 500 pairs of the simulated problems we apply M2 and obtain the second approximant $x_2^{\boldsymbol{\varepsilon}}$.

The third method (i.e., M3) consists in aggregating the approximants $x_1^{\boldsymbol{\varepsilon}}$, $x_2^{\boldsymbol{\varepsilon}}$ according to the methodology described at the end of the previous subsection. In our experiments the role of the most trustable observation equation (14) is played by the equations of the SGG-type (23), $i = 2$, and we label the aggregation based on them as M3(2). We choose these equations because the data for them are simulated with smaller noise intensity. Then the required regularization parameters α_1, α_2 are selected according to the quasi-optimality criterion (22) from the geometric sequence Λ_{30} in such a way that $\frac{1}{\alpha_1}, \frac{1}{\alpha_2} \in \Lambda_{30}$. Note that in general, no specific relation is required between the sets of possible values of the regularization parameters λ_j and α_j. In this test, we use the same set Λ_{30} for the sake of simplicity.

We have to admit that the decision, which model to select as the most trustable one, may contribute to the performance of the aggregation method M3. In our discussion, the 'most trustable model' might be either the least ill-posed observation equation or the equation with the smallest noise level. If one has a model with both of these features, then one can choose it. However, it may happen that the above features are not attributed to the same observation equation. For example, in our numerical illustrations for (3), (23), (24), the SST-type equation (3), (23), $i = 1$, is contaminated by more intensive noise, but it is less ill-posed than the SGG-type equation (3), (23), $i = 2$, which has been chosen by us as the most trustable model. This can be seen from (24) if one compares the rates of the decrease of the singular values $a_k^{(1)}$ and $a_k^{(2)}$ as $k \to \infty$: for the considered values $r = 6371$ km, $\rho_1 = 6621$ km, $\rho_2 = 6771$ km both $a_k^{(i)}$, $i = 1, 2$, decrease exponentially fast, but $a_k^{(1)}$ decreases slower than $a_k^{(2)}$.

To illustrate what happens when an alternative model is chosen as the most trustable one, we implement the aggregation method M3 on the base of the SST-type equation (23), $i = 1$, and label it as M3(1). All other implementation details are exactly as described for M3(2).

The performance of all four methods is compared in terms of the relative errors $\|x^\dagger - x_j^{\boldsymbol{\varepsilon}}\|_{\mathcal{X}} / \|x^\dagger\|_{\mathcal{X}}$, $j = 1, 2$, and $\|x^\dagger - x_{\mathrm{ag},\beta^*}^{\boldsymbol{\varepsilon}}\|_{\mathcal{X}} / \|x^\dagger\|_{\mathcal{X}}$. The results are

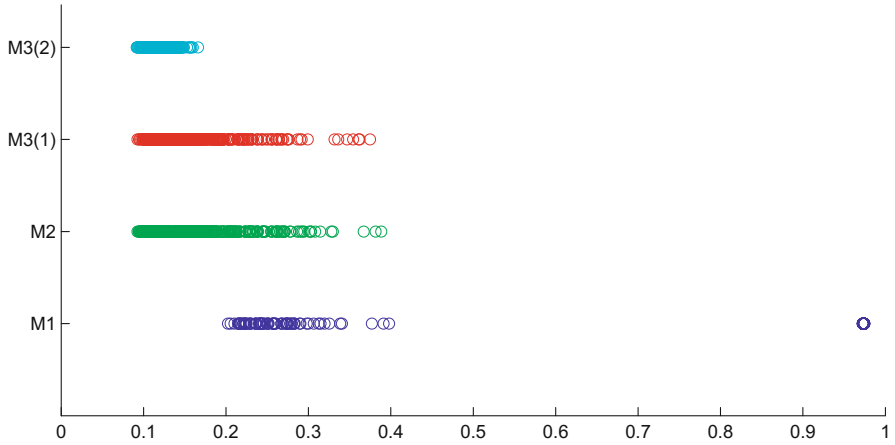


FIGURE 2.1. Examples of a joint regularization of two observation models. Relative errors of the regularization by a reduction to a single regularization parameter (M1), the regularization with a multi-parameter quasi-optimality criterion (M2), and the regularization by aggregation (M3(1), M3(2)).

displayed in Figure 2.1, where the projection of each circle onto the horizontal axis exhibits a value of the corresponding relative error of one of the methods M1, M2, M3(1), and M3(2), in the joint inversion of one of 500 pairs of the simulated problems. From this figure we can conclude that the aggregation by the linear functional strategy can essentially improve the accuracy of the joint inversion compared to M1 and M2. This conclusion is in agreement with our Theorem 2.3.

At the same time, Figure 2.1 also presents an evidence of the reliability of the proposed approach. Indeed, in the considered case, even with the use of an alternative (i.e., suboptimal) trusted reference model, the aggregation, this time M3(1), performs at least at the level of the best among the approximants M1 and M2.

3. Global combination of satellite and ground models

We now consider the situation when the global satellite data is complemented by regional/local measurements at or near the Earth's surface. The satellite data are well suited for the reconstruction of large-scale structures, but it fails for spatially localized features (due to the involved downward continuation). The opposite is true for locally/regionally available ground data. They are well suited to capture local phenomena but fails for global trends. Therefore, in order to obtain high-resolution gravitational models, such as EGM2008 [55], or geomagnetic models, such as NGDC-720 [54], it becomes necessary to combine both types of data. It is the goal of this section to introduce a two-parameter approximation reflecting such a combination.

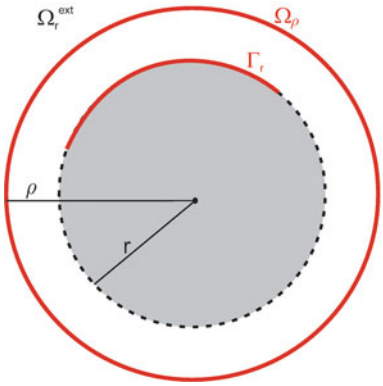


FIGURE 3.1. Schematic Description of the given data situation. The red areas indicate regions where data is available.

In the exterior $\Omega_r^{\text{ext}} = \{x \in \mathbb{R}^3 : |x| > r\}$ of the Earth, the gravity field and the crustal magnetic field can be described by a harmonic potential x. From satellite measurements we obtain data y_1 on a spherical orbit $\Omega_\rho = \{x \in \mathbb{R}^3 : |x| = \rho\}$ and from ground or near-ground measurements we obtain data y_2 in a subregion Γ_r of the spherical Earth surface Ω_r of radius $r < \rho$ (cf. Figure 3.1).

Of interest to us is the harmonic potential x, in particular, its restriction to the Earth surface Ω_r. It satisfies

$$\Delta x = 0, \qquad \text{in } \Omega_r^{\text{ext}}, \tag{25}$$

$$x = y_1, \qquad \text{on } \Omega_\rho, \tag{26}$$

$$x = y_2, \qquad \text{on } \Gamma_r. \tag{27}$$

The knowledge of y_1 on Ω_ρ already supplies information that can be used for approximating x. However, only a coarse structure of x can be reconstructed because of the presence of noise and ill-posedness of the involved downward continuation. Additional discrete measurements $\mathbf{y}_2$ of y_2 in Γ_r may improve the situation. By y_1, y_2, we typically denote functions while $\mathbf{y}_1 \in \mathbb{R}^{L_1}, \mathbf{y}_2 \in \mathbb{R}^{L_2}$ denote discrete evaluations of these functions at locations $\xi_i \in \Omega_\rho$, $i = 1, \ldots, L_1$, and $\eta_i \in \Gamma_r$, $i = 1, \ldots, L_2$, respectively.

3.1. The inversion procedure

The goal is to approximate x globally on Ω_r (or, since x is harmonic, in all of Ω_r^{ext}) from the situation (25)–(27). In general, we assume the following setting to be given:

Setting 3.1.

(a) *We have discrete measurements* $\mathbf{y}_2^{\varepsilon_2} \in \mathbb{R}^{L_2}$ *of* $y_2^{\varepsilon_2}$ *on* Γ_r *and we know* $y_1^{\varepsilon_1} \in L^2(\Omega_\rho)$. *Both quantities contain additive noise, i.e.,* $y_1^{\varepsilon_1} = y_1 + e_1$ *and* $y_2^{\varepsilon_2} = y_2 + e_2$ *with* $\|e_1\|_{L^2(\Omega_\rho)} \leqslant \varepsilon_1$ *and* $\|e_2\|_{L^2(\Omega_r)} \leqslant \varepsilon_2$.

(b) *The discretization operator* $\mathcal{D} : L^2(\Gamma_r) \to \mathbb{R}_w^{L_2}$ *maps a function* $\bar{x} \in L^2(\Gamma_r)$ *to the corresponding measurements* $\bar{\mathbf{x}} = \mathcal{D}\bar{x} = \big(\bar{x}(\eta_1), \ldots, \bar{x}(\eta_{L_2})\big) \in \mathbb{R}_w^{L_2}$. *By* $\mathbb{R}_w^{L_2}$ *we denote the space* $\mathbb{R}^{L_2}$ *equipped with the inner product*

$$\langle \bar{\mathbf{x}}, \bar{\mathbf{y}} \rangle_w = \sum_{i=1}^{L_2} w_i \bar{\mathbf{x}}_i \bar{\mathbf{y}}_i, \quad \bar{\mathbf{x}}, \bar{\mathbf{y}} \in \mathbb{R}^{L_2}, \tag{28}$$

and the corresponding norm $\|\cdot\|_w$.

(c) *Here,* $\mathcal{D}$ *and* w_i *are such that*

$$\langle \bar{x}, \bar{y} \rangle_{L^2(\Gamma_r)} = \langle \mathcal{D}\bar{x}, \mathcal{D}\bar{y} \rangle_w, \quad \text{for all } \bar{x}, \bar{y} \in \mathbb{P}_N, \tag{29}$$

where $\mathbb{P}_N$ *is the space of spherical polynomials up to degree* N.

(d) $\mathcal{H}_K \subset L^2(\Omega_r)$ *is a reproducing kernel Hilbert space (RKHS) with reproducing kernel of the form*

$$K(\xi, \eta) = \sum_{k=0}^{N} d_N(k) \sum_{l=1}^{2k+1} \frac{1}{r^2} Y_{k,l}\left(\frac{\xi}{r}\right) Y_{k,l}\left(\frac{\eta}{r}\right), \ \xi, \eta \in \Omega_r, \tag{30}$$

where $d_N = (d_N(1), d_N(2), \ldots, d_N(N))$ *is a monotone sequence. An inner product on* $\mathcal{H}_K$ *is defined by*

$$\langle \bar{x}, \bar{y} \rangle_{\mathcal{H}_K} = \sum_{k=0}^{N} \frac{1}{r^2 d_N(k)} \sum_{l=1}^{2k+1} \left\langle Y_{k,l}\left(\frac{\cdot}{r}\right), \bar{x} \right\rangle_{L_2(\Omega_r)} \left\langle Y_{k,l}\left(\frac{\cdot}{r}\right), \bar{y} \right\rangle_{L_2(\Omega_r)}, \quad \bar{x}, \bar{y} \in \mathcal{H}_K.$$

To combine global data on Ω_ρ and local data on Γ_r for the approximation of x in Ω_r^{ext}, we consider the following procedure:

Method 3.2. *Let* x *satisfy* (25)–(27) *and let the conditions of Setting* 3.1 *be satisfied. Then, the approximation* $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$ *of* x *is defined as the minimizer of*

$$\min_{\bar{x} \in \mathcal{H}_K} \lambda_1 \|\mathcal{A}\bar{x} - y_1^{\varepsilon_1}\|_{L_2(\Omega_\rho)}^2 + \lambda_2 \|\mathcal{D}\bar{x} - \mathcal{D}y_2^{\varepsilon_2}\|_w^2 + \|\bar{x}\|_{\mathcal{H}_K}^2, \tag{31}$$

where $\lambda_1, \lambda_2 > 0$ *are the regularization parameters.* $\mathcal{A} : \mathcal{H}_K \to L^2(\Omega_r^{\text{ext}})$ *denotes the upward continuation operator, i.e.,* $\mathcal{A}\bar{x}$ *is harmonic in* Ω_r^{ext}.

Note that, when only global data are used, i.e., $\lambda_2 = 0$, then (31) is reduced to the Regularized Collocation (RC) method [53]. On the other hand, if $\lambda_1 = 0$ (this means that we use only local data in Γ_r) then (31) defines the solution after denoising and is, in some sense, extrapolation. The solution of the minimization problem (31) is given by the following theorem:

Theorem 3.3. *Let Setting* 3.1 *hold true. Then the minimizer* $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$ *of* (31) *on* Ω_r *has the form*

$$x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2} = \sum_{k=0}^{N} \sum_{l=1}^{2k+1} \widehat{x}_{\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}(k,l) \frac{1}{r} Y_{k,l}\left(\frac{\cdot}{r}\right),$$

where

$$\begin{aligned}
\widehat{x}^{\varepsilon_1,\varepsilon_2}_{\lambda_1,\lambda_2}(k,l) &= \left\langle x^{\varepsilon_1,\varepsilon_2}_{N,\lambda_1,\lambda_2}, \frac{1}{r}Y_{k,l}\left(\frac{\cdot}{r}\right)\right\rangle_{L_2(\Omega_r)} \\
&= \frac{\lambda_1 a_k \widehat{y_1^{\varepsilon_1}}(k,l) + \lambda_2 d_N(k)\widehat{y_2^{\varepsilon_2}}(k,l)}{\lambda_1 a_k^2 + \lambda_2 d_N(k)\sum_{n=0}^{N}\sum_{m=1}^{2n+1} c_{k,l,n,m} + 1}, \\
\widehat{y_1^{\varepsilon_1}}(k,l) &= \left\langle y_1^{\varepsilon_1}, \frac{1}{\rho}Y_{k,j}\left(\frac{\cdot}{\rho}\right)\right\rangle_{L_2(\Omega_\rho)}, \\
\widehat{y_2^{\varepsilon_2}}(k,l) &= \sum_{i=1}^{L_2} w_i \frac{1}{r}Y_{k,l}\left(\frac{\eta_i}{r}\right)\mathbf{y}^{\varepsilon_2}_{2,i}, \\
c_{k,l,n,m} &= \left\langle \frac{1}{r}Y_{n,m}\left(\frac{\cdot}{r}\right), \frac{1}{r}Y_{k,j}\left(\frac{\cdot}{r}\right)\right\rangle_{L_2(\Gamma_r)}.
\end{aligned} \tag{32}$$

As mentioned earlier, $\mathbf{y}^{\varepsilon_2}_{2,i}$ *denotes the evaluation of* $y_2^{\varepsilon_2}$ *at the location* $\eta_i \in \Gamma_r$*, i.e.,* $\mathbf{y}^{\varepsilon_2}_{2,i} = y_2^{\varepsilon_2}(\eta_i) = \left(\mathcal{D}y_2^{\varepsilon_2}\right)_i$*, and* $a_k = \left(\frac{r}{\rho}\right)^k$ *represent the symbols of the upward continuation operator* $\mathcal{A}$*.*

3.2. Numerical example

In a test example, we use the following reference potential

$$x(\xi) = \sum_{k=0}^{30}\sum_{l=1}^{2k+1}\left(\frac{r}{|\xi|}\right)^k \widehat{x}(k,l)\frac{1}{|\xi|}Y_{k,l}\left(\frac{\xi}{|\xi|}\right), \quad \xi \in \Omega_r^{\mathrm{ext}}, \tag{33}$$

as our true solution to (25)–(27). The Fourier coefficients of x are given by $\widehat{x}(k,l) = (k+\frac{1}{2})^{-\frac{3}{2}} g_{k,l}$, where the $g_{k,l}$ are uniformly distributed random values from $[-1,1]$. This means that x mimics the Sobolev smoothness $s = \frac{3}{2}$ of the Earth's gravitational potential. The radius ρ is chosen such that $\frac{\rho}{r} = 1.48$ (if $r = 6371$km is the mean Earth radius, then $\rho = 9429$km).

The function y_1 on Ω_ρ is given as the restriction of x to Ω_ρ. Its noisy counterpart $y_1^{\varepsilon_1}$ is generated by adding Gaussian white noise of intensity $\varepsilon_1 = 0.05$ at the knots $\{\xi_i\}_{i=1,\dots,L_1}$ of a Gauss–Legendre cubature grid on the sphere Ω_ρ. For the subregion Γ_r we choose the spherical cap $\Gamma_r = \Gamma_r(\bar{\xi},\bar{r}) = \{\eta \in \Omega_r : 1 - \frac{\bar{\xi}}{r}\cdot\frac{\eta}{r} < \bar{r}\}$ with spherical radius $\bar{r} \in (0,2)$ and the North Pole $\bar{\xi} = (0,0,r)$ as center. The function y_2 denotes the restriction of x to Γ_r. The noisy discrete values $\mathbf{y}_2^{\varepsilon_2} \in \mathbb{R}^{L_2}$ are simulated by adding a Gaussian white noise of intensity $\varepsilon_2 = 0.1$ to the values of y_2 at the knots $\{\eta_i\}_{i=1,\dots,L_2}$ of a Gauss–Legendre cubature grid on the cap Γ_r. For Gauss–Legendre grids, positive cubature weights are known that satisfy the polynomial exactness condition required in (29).

We apply Method 3.2 to obtain approximations $x^{\varepsilon_1,\varepsilon_1}_{N,\lambda_1,\lambda_2}$ of x in the setting above. $N = 30$ is fixed and λ_1, λ_2 are chosen from the set $\Lambda_{200} = \{10^6 \times 0.95^j : j = 0,\dots,200\}$. To assess the performance of the considered schemes, we measure

the relative error

$$\mathrm{err}_{\mathrm{rel}} = \frac{\left\| x - x^{\varepsilon_1,\varepsilon_1}_{N,\lambda_1,\lambda_2} \right\|_{L^2(\Omega_r)}}{\|x\|_{L^2(\Omega_r)}} = \frac{\left(\sum_{k=0}^{30} \sum_{l=1}^{2k+1} \left(\widehat{x}(k,l) - \widehat{x}^{\varepsilon_1,\varepsilon_2}_{\lambda_1,\lambda_2}(k,l) \right)^2 \right)^{1/2}}{\left(\sum_{k=0}^{30} \sum_{l=1}^{2k+1} \left(\widehat{x}(k,l) \right)^2 \right)^{1/2}},$$

where $\widehat{x}^{\varepsilon_1,\varepsilon_2}_{\lambda_1,\lambda_2}(k,l)$ are given by (32). We test method for different radii $\bar{r}$ of the spherical cap $\Gamma_r = \Gamma_r(\bar{\xi}, \bar{r})$. Spherical radius $\bar{r} = 2$ means that we regard the entire sphere $\Gamma_r = \Omega_r$.

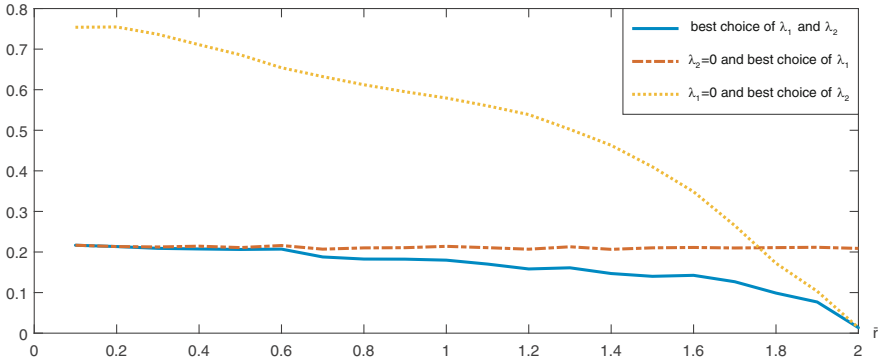


FIGURE 3.2. Optimal relative errors $\mathrm{err}_{\mathrm{rel}}$ for the method (32) with RC ($\lambda_2 = 0$), denoising + extrapolation ($\lambda_1 = 0$), and no additional conditions on λ_1, λ_2.

The results are displayed in Figure 3.2, where the vertical axis represents the global relative errors $\mathrm{err}_{\mathrm{rel}}$. The relative errors are plotted for each of the three methods: regularized collocation (RC) method corresponding to the case when $\lambda_2 = 0$, denoising + extrapolation (the case when $\lambda_1 = 0$), and the two-parameter scheme (32). For all methods, the 'optimal' regularization parameters are chosen, i.e., the choice of the parameter is performed such that $\mathrm{err}_{\mathrm{rel}}$ is minimal (with respect to the possible constraints $\lambda_1 = 0$ or $\lambda_2 = 0$).

Of course, the relative errors displayed in Figure 3.2 require the knowledge of the 'true' solution x, which cannot be obtained in practice. However, Figure 3.2 shows that even for rather small subregions Γ_r (i.e., small spherical radii $\bar{r}$) of the sphere, where the additional noisy data $\mathbf{y}_2^{\varepsilon_2}$ are available, the use of these additional data allows to improve the reconstruction compared to the standard Tikhonov method (which only uses information $y_1^{\varepsilon_1}$ at satellite altitude).

Next, we consider the more realistic case when the regularization parameters are chosen by an *a posteriori* procedure, which does not require the knowledge of the 'true' x. We can choose the regularization parameters by the analogue of the well-known quasi-optimality criterion for the two-parameter case (see strategy M2 in Section 2.1).

However, our tests show that within this strategy we may only reach the accuracy corresponding to local data denoising + extrapolation ($\lambda_1 = 0$). In this situation the use of the quasi-optimality criterion makes sense only for a large amount of regional data (almost full coverage of the sphere Ω_r).

Potentially, one may perform global data smoothing with direct inversion. This strategy has been presented in [57] for pointwise computations. However, it does not use local data and its accuracy cannot be improved regardless the amount of local information.

In this context, the idea is to use M procedures and aggregate them by a linear functional strategy as presented in Section 2 as Method 2.2, where $\mathcal{X} = L^2(\Omega_r)$ and $\tilde{x} = x_{\tilde{j}}^{\varepsilon} \in \left\{x_j^{\varepsilon}, j = 1, 2, \ldots, M\right\}$ is a trustworthy approximant that we use to define $\tilde{\boldsymbol{\kappa}}$ in (13).

We now apply this version of Method 2.2 to our previous test example. In particular, we will aggregate $M = 2$ solutions x_1^{ε}, x_2^{ε}: the first one is given by the two-parameter quasi-optimality (QO) criterion (strategy M2 in Section 2) applied to the already obtained approximations $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$ of x, the second one is the solution after global data smoothing with direct inversion. This choice leads to a linear system of two equations with two unknowns β_1 and β_2. For the approximation $\tilde{\boldsymbol{\kappa}}$ of the right-hand side of (13) we use the second involved approximation (smoothing+inversion). In Figure 3.3, we compare the relative error of the aggregated

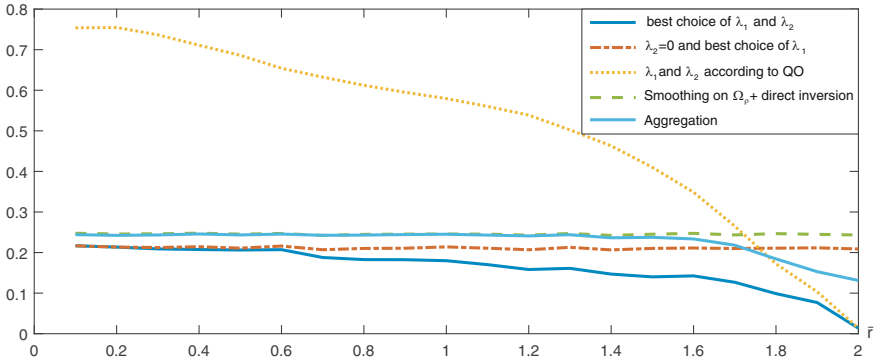


FIGURE 3.3. Relative errors $\mathrm{err}_{\mathrm{rel}}$ for RC ($\lambda_2 = 0$) with the best choice of regularization parameters (based on the knowledge of x), for global data smoothing + direct inversion, for QO, and for the aggregation of the latter two.

approximation $x_{\mathrm{ag},\beta^*}^{\varepsilon}$ with the approximation $x_{N,\lambda_1^*,\lambda_2^*}^{\varepsilon_1,\varepsilon_2}$ (where λ_1^*, λ_2^* are chosen according to QO) and with the actual best choice of $x_{N,\lambda_1^{\mathrm{opt}},\lambda_2^{\mathrm{opt}}}^{\varepsilon_1,\varepsilon_2}$ (based on the knowledge of the 'true' x). One can conclude that in most situations the aggregation of the two considered regularization methods provides a better approximation than each of the involved strategies separately. QO is only better than the aggregated result if Γ_r is close to the entire sphere Ω_r (i.e., if the spherical radius $\bar{r}$

is close to 2). Moreover, it is instructive to observe that the aggregated solution "ignores" the two-parameter QO for small spherical caps and uses it only for larger spherical caps.

4. Local combination of satellite and ground models

We assume to be in the same situation as in Section 3 and have data $\mathbf{y}_1$ and $\mathbf{y}_2$ given on Ω_ρ and Γ_r, respectively (cf. Figure 3.1). Yet, opposed to Section 3, we are not interested in a global model on Ω_r but in a local model of on Γ_r. One might say that Section 3 uses local data in Γ_r to refine a global model while now we use global data on Ω_ρ to improve the coarse features of a local model.

A possibility to incorporate the local data set on Γ_r and the global data set on Ω_ρ is to use spherical basis functions that are built upon satisfying certain localization constraints as well as ameliorating the ill-posed downward continuation (see, e.g., [59]). However, in this section, we are aiming at designing convolution kernels $K_N, \tilde{K}_N$ that reflect the trade-off between localization and downward continuation (see, e.g., [25]). An approximation of x is then given by the convolution of the data against these kernels. Opposed to the spherical basis function approach, the convolution approach allows to treat the input data $\mathbf{y}_1$ and $\mathbf{y}_2$ consecutively and indicates the influence of each one to the overall approximation of x. The approximation does not hold globally but only in Γ_r. Therefore, we denote the restriction of x to Γ_r by $x^\dagger$.

Just like in Section 3, we illustrate our approach for the Dirichlet problem (25)–(27). The approximation of $x^\dagger$ (which now denotes the restriction of x to Γ_r) from noisy data is denoted by $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$, with unknown parameters N, λ_1, λ_2, and will be explained in Section 4.1. The approach is not limited to the problem (25)–(27) but can also be applied to other problems where the singular values for the underlying problems are known. Vectorial problems that are of relevance in geomagnetic applications can be handled as well. For details on the theoretical aspects of this section, the reader is referred to [25].

4.1. The inversion procedure

We first regard only the problem (25), (26). Then x can be approximated on Ω_r by

$$x_N = \mathcal{T}_N[y_1] = \int_{\Omega_\rho} K_N(\cdot,\eta) y_1(\eta) dS(\eta), \tag{34}$$

with a kernel K_N of the form

$$K_N(\xi,\eta) = \sum_{k=0}^{N} \sum_{l=1}^{2k+1} d_N(k) \frac{1}{r} Y_{k,l}\left(\frac{\xi}{r}\right) \frac{1}{\rho} Y_{k,l}\left(\frac{\eta}{\rho}\right), \quad \xi \in \Omega_r, \eta \in \Omega_\rho, \tag{35}$$

and symbols $d_N(k)$ satisfying

$$\lim_{N\to\infty} d_N(k) = \left(\frac{\rho}{r}\right)^k, \tag{36}$$

i.e., they converge towards the singular values of downward continuation. On the other hand, regarding only the problem (25), (27), with y_2 given on all of Ω_r, x can be approximated on Ω_r by

$$x_N = \tilde{\mathcal{T}}_N[y_2] = \int_{\Omega_r} \tilde{K}_N(\cdot,\eta) y_2(\eta) dS(\eta), \tag{37}$$

with a kernel $\tilde{K}_N$ of the form

$$\tilde{K}_N(\xi,\eta) = \sum_{k=0}^{N} \sum_{l=1}^{2k+1} \tilde{d}_N(k) \frac{1}{r} Y_{k,l}\left(\frac{\xi}{r}\right) \frac{1}{r} Y_{k,l}\left(\frac{\eta}{r}\right), \quad \xi \in \Omega_r, \eta \in \Omega_r, \tag{38}$$

and symbols $\tilde{d}_N(k)$ satisfying

$$\lim_{N\to\infty} \tilde{d}_N(k) = 1. \tag{39}$$

However, according to (25)–(27), ground data y_2 is only given in the subregion $\Gamma_r \subset \Omega_r$. Thus, the symbols $\tilde{d}_N(k)$ do not only have to be chosen to satisfy (39) but also such that $\tilde{K}_N(\xi,\cdot)$ is fairly well localized in Γ_r for every $\xi \in \Gamma_r$. To compensate this localization constraint, we have to include an approximation of the form (34), (35) that uses satellite data on Ω_ρ. Therefore, we use the following approximation $x_{N,\lambda_1,\lambda_2}$ of $x^\dagger$:

Method 4.1. *Let $x^\dagger$ denote the restriction of x to Γ_r, where x satisfies* (25)–(27). *The approximation $x_{N,\lambda_1,\lambda_2}$ of $x^\dagger$ is defined by*

$$x_{N,\lambda_1,\lambda_2} = \mathcal{T}_N[y_1] + \tilde{\mathcal{T}}_N[y_2]. \tag{40}$$

$\mathcal{T}_N$ and the corresponding kernel K_N are given as in (34), (35). *The operator $\tilde{\mathcal{T}}_N$ and the kernel $\tilde{K}_N$ are slightly modified in the sense*

$$\tilde{\mathcal{T}}_N[y_2] = \int_{\Gamma_r} \tilde{K}_N(\cdot,\eta) y_2(\eta) dS(\eta),$$

$$\tilde{K}_N(\xi,\eta) = \sum_{k=0}^{\lfloor \kappa N \rfloor} \sum_{l=1}^{2k+1} \left(\tilde{d}_N(k) - d_N(k) \left(\frac{r}{\rho}\right)^k \right) \frac{1}{r} Y_{k,l}\left(\frac{\xi}{r}\right) \frac{1}{r} Y_{k,l}\left(\frac{\eta}{r}\right), \quad \xi \in \Omega_r, \eta \in \Omega_r,$$

for some fixed $\kappa > 1$. The symbols $d_N(k)$, $\tilde{d}_N(k)$ of K_N and $\tilde{K}_N$, respectively, are chosen to minimize the functional

$$\mathcal{F}(d_N, \tilde{d}_N) = \lambda_1 \sum_{k=0}^{\lfloor \kappa N \rfloor} \left| 1 - \tilde{d}_N(k) \right|^2 + \lambda_1 \sum_{k=0}^{N} \left| 1 - d_N(n) \left(\frac{r}{\rho}\right)^k \right|^2 + \lambda_2 \sum_{k=0}^{N} \left| d_N(k) \right|^2 + \left\| \tilde{K}_N \right\|^2_{L^2([-1,1-\varrho)}. \tag{41}$$

The parameters $\lambda_1, \lambda_2 > 0$ are not known a priori. The radius $\varrho > 0$ is fixed in advance and chosen to reflect the 'spherical radius' of the set Γ_r, and

$$\left\|\tilde{K}_N\right\|^2_{L^2([-1,1-\varrho])} = \int_{-1}^{1-\varrho} |\tilde{K}_N(t)|^2 dt,$$

where it has to be noted that the kernel $\tilde{K}_N(\xi,\eta)$ actually only depends on the scalar product $t = \frac{\xi}{r} \cdot \frac{\eta}{r}$, i.e.,

$$\tilde{K}_N(\xi,\eta) = \tilde{K}_N(t) = \sum_{k=0}^{\lfloor \kappa N \rfloor} \frac{2k+1}{4\pi r^2} \left(\tilde{d}_N(k) - d_N(k) \left(\frac{r}{\rho}\right)^k \right) P_k(t),$$

where P_k is the Legendre polynomial of degree k.

Remark 4.2. The functional $\mathcal{F}$ in (41) reflects the properties that we would like to imply on the kernels K_N and $\tilde{K}_N$. The first term on the right-hand side of (41) represents the overall approximation error (under the assumption that undisturbed global data is available on Ω_ρ as well as on Ω_r), the second term reflects the approximation error under the assumption that only undisturbed data on Ω_ρ is available. The third and fourth term act as penalty terms. While the third term is meant to regularize the ill-posed downward continuation, the fourth term penalizes the kernel $\tilde{K}_N$ if it is not localized in Γ_r (more precisely, it penalizes the contributions of $\tilde{K}_N(\cdot,\eta)$ in the exterior spherical cap $\Omega_r \setminus \Gamma_r(\eta,\varrho)$, where $\Gamma_r(\eta,\varrho) = \{\xi \in \Omega_r : 1 - \frac{\xi}{r} \cdot \frac{\eta}{r} < \varrho\}$ is a spherical cap with center $\eta \in \Gamma_r$ and fixed radius $\varrho > 0$ such that $\Gamma_r(\eta,\varrho) \subset \Gamma_r$). The parameters λ_1, λ_2 allow to weigh the approximation property against the regularization and the localization penalty.

Eventually, the obtained symbols $d_N(k)$ of K_N reflect the regularization of downward continuation while the $\tilde{d}_N(k)$ offer some control over the localization of $\tilde{K}_N$ in Γ_r. An illustration of the optimized symbols is shown in Figure 4.1. One can see that large parameters λ_2 (relative to λ_1) typically cause a strong damping of $d_N(k)$ (i.e., satellite data y_1 is damped and ground data y_2 has more influence) while small parameters λ_2 (again, relative to λ_1) are in favour of $d_N(k)$ (i.e., satellite data y_1 has more influence). The question whether a stronger influence of satellite data or a stronger influence of ground data is adequate depends on the noise levels ε_1, ε_2. However, the noise levels are typically not known, so that other means of determining λ_1, λ_2 become necessary.

The minimizer of (41) can easily be computed by solving the system of linear equations

$$\mathbf{M}\mathbf{d} = \boldsymbol{\lambda},$$

where $\mathbf{d} = \left(d_N(0), d_N(1)\frac{r}{\rho}, \ldots, d_N(N)\left(\frac{r}{\rho}\right)^N, \tilde{d}_N(0), \ldots, \tilde{d}_N(\lfloor \kappa N \rfloor)\right) \in \mathbb{R}^{N+\lfloor \kappa N \rfloor + 2}$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_1) \in \mathbb{R}^{N+\lfloor \kappa N \rfloor + 2}$, and

$$\mathbf{M} = \left(\begin{array}{c|c} \mathbf{D}_1 + \mathbf{P}_1 & -\mathbf{P}_2 \\ \hline -\mathbf{P}_3 & \mathbf{D}_2 + \mathbf{P}_4 \end{array}\right).$$

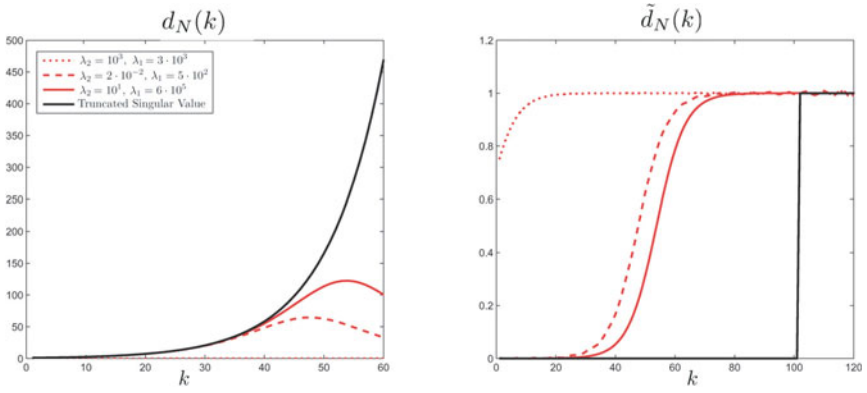


FIGURE 4.1. The symbols $d_N(k)$ (left) and $\tilde{d}_N(k)$ (right) for different choices of λ_1, λ_2.

The diagonal matrices $\mathbf{D}_1$, $\mathbf{D}_2$ are given by

$$\mathbf{D}_1 = \operatorname{diag}\left(\lambda_2 \left(\frac{\varrho}{r}\right)^{2k} + \lambda_1\right)_{k=0,\dots,N}, \qquad \mathbf{D}_2 = \operatorname{diag}\left(\lambda_1\right)_{k=0,\dots,\lfloor \kappa N \rfloor},$$

whereas $\mathbf{P}_1, \dots, \mathbf{P}_4$ are submatrices of the Gram matrix $\left(\mathbf{P}^{\varrho}_{k,l}\right)_{k,l=0,\dots,\lfloor \kappa N \rfloor}$. More precisely,

$$\mathbf{P}_1 = \left(\mathbf{P}^{\varrho}_{k,l}\right)_{k,l=0,\dots,N}, \qquad \mathbf{P}_2 = \left(\mathbf{P}^{\varrho}_{k,l}\right)_{\substack{k=0,\dots,N;\\ l=0,\dots,\lfloor \kappa N \rfloor}},$$

$$\mathbf{P}_3 = \left(\mathbf{P}^{\varrho}_{k,l}\right)_{\substack{k=0,\dots,\lfloor \kappa N \rfloor;,\\ l=0,\dots,N}}, \qquad \mathbf{P}_4 = \left(\mathbf{P}^{\varrho}_{k,l}\right)_{k,l=0,\dots,\lfloor \kappa N \rfloor}.$$

with

$$\mathbf{P}^{\varrho}_{k,l} = \frac{(2k+1)(2l+1)}{2} \int_{-1}^{1-\varrho} P_k(t) P_l(t) dt.$$

Under appropriate conditions on λ_1, λ_2, it can be shown that $x_{N,\lambda_1,\lambda_2}$ converges to $x^\dagger$ as $N \to \infty$. This convergence also holds true for $x^{\varepsilon_1,\varepsilon_2}_{N,\lambda_1,\lambda_2}$ (the approximation based on noisy input data $y_1^{\varepsilon_1}$, $y_2^{\varepsilon_2}$) as $\varepsilon_1, \varepsilon_2 \to 0$ if an appropriate connection of ε_1, ε_2, and N is assumed. Details can be found in [25].

If only discrete data $\mathbf{y}_1^{\varepsilon_1}$, $\mathbf{y}_2^{\varepsilon_2}$ is available, then the evaluation of the integrals in $x^{\varepsilon_1,\varepsilon_2}_{N,\lambda_1,\lambda_2}$ has to be done numerically. In the examples later on we use the cubature rules [16, 33].

4.2. A parameter choice method for direct observations

Let us now assume that we have obtained a set of various approximations $\{x^{\varepsilon_1,\varepsilon_2}_{N,\lambda_1,\lambda_2}\}_{(N,\lambda_1,\lambda_2)\in\Lambda}$ with $\Lambda \subset \mathbb{N} \times \mathbb{R}_+ \times \mathbb{R}_+$ denoting the finite set of parameters. The standard question in such a situation is which $x^{\varepsilon_1,\varepsilon_2}_{N,\lambda_1,\lambda_2}$ to choose as a 'good' approximation of $x^\dagger$. Typically, we do not know the noise levels ε_1, ε_2 and obviously not the true $x^\dagger$. But we have the direct discrete observations $\mathbf{y}_2^{\varepsilon_2}$ of $x^\dagger$ in

Γ_r as a reference. Since those are the non-ill-posed contributions to our problem, a natural choice would be to choose $N^*, \lambda_1^*, \lambda_2^*$ as the minimizers of

$$\min_{(N,\lambda_1,\lambda_2)\in\Lambda} \|\mathbf{x}_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2} - \mathbf{y}_2^{\varepsilon_2}\|_w, \tag{42}$$

where $\mathbf{x}_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2} = \big(x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}(\eta_i)\big)_{i=1,\dots,L_2} \in \mathbb{R}^{L_2}$ represents the vector of the evaluations of $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$ at the locations $\eta_i \in \Gamma_r$, $i = 1,\dots,L_2$. However, although it is easy to find a minimizer of (42), it is not clear how to derive an adequate estimate of the error $\|x^\dagger - x_{N^*,\lambda_1^*,\lambda_2^*}^{\varepsilon_1,\varepsilon_2}\|_{L^2(\Gamma_r)}$. Therefore, instead of (42), we use a modified procedure to choose $N^*, \lambda_1^*, \lambda_2^*$ that allows such an error estimate (cf. [26]).

We begin by formulating a more general setting than in the previous paragraph:

Setting 4.3.

(a) *Let $\Gamma_r \subset \Omega_r$ be a subdomain where discrete measurements $\mathbf{y} \in \mathbb{R}^L$ of the underlying quantity x are available. Let $\mathcal{D} : L^2(\Gamma_r) \to \mathbb{R}_w^L$ be the discretization operator of Setting* 3.1 *and assume that it satisfies condition* (29).

(b) *The measurements y in Γ_r may be blurred by additive noise (i.e., $y^\varepsilon = y + e$ for some $e \in L^2(\Gamma_r)$) and we assume that*

$$\big\|\mathcal{D}x^\dagger - \mathcal{D}y^\varepsilon\big\|_w \leqslant \varepsilon.$$

(c) *We assume that from somewhere, a set $\{x_j^\varepsilon\}_{j=1,2,\dots,M}$ of approximations of $x^\dagger$ is available and that all these approximations belong to the space $\mathbb{P}_{\lfloor \kappa N\rfloor}$ of spherical polynomials up to degree $\lfloor \kappa N\rfloor$.*

Remark 4.4. If Γ_r is, for instance, a spherical cap, one can find a system of nodes $\{\eta_i\}_{i=1,\dots,L} \subset \Gamma_r$ and positive weights $w = (w_1,\dots,w_L)$ such that

$$\int_{\Gamma_r} \bar{x}(\zeta)dS(\zeta) = \sum_{i=1}^{L} w_i\, \bar{x}(\eta_i), \quad \text{for all } \bar{x} \in \mathbb{P}_{2N}.$$

Then it is clear that the discretization operator

$$\mathcal{D}\bar{x} = \bar{\mathbf{x}} = \big(\bar{x}(\eta_i)\big)_{i=1,\dots,L} \in \mathbb{R}^L, \quad \bar{x} \in L^2(\Gamma_r),$$

meets condition. If x satisfies the problem (25)–(27) and the approximations x_j^ε of $x^\dagger$ are chosen to be of the form $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$ as in Method 4.1, then we see that the parameter choice problem for the inversion procedure described in Section 4.1 fits into Setting 4.3.

The optimal parameter j_{opt} is the one that minimizes $\min_{j=1,2,\dots,M} \|x^\dagger - x_j^\varepsilon\|_{L^2(\Gamma_r)}$. Of course, $x_{j_{\text{opt}}}^\varepsilon$ cannot be found without knowledge of $x^\dagger$. But we observe

$$\begin{aligned}\Big\|x_j^\varepsilon - x_{j_{\text{opt}}}^\varepsilon\Big\|_{L^2(\Gamma_r)} &= \sup_{a\in L^2(\Gamma_r),\|a\|_{L^2(\Gamma_r)}=1} \Big\langle x_j^\varepsilon - x_{j_{\text{opt}}}^\varepsilon, a\Big\rangle_{L^2(\Gamma_r)} \\ &= \max_{a\in A} \Big\langle x_j^\varepsilon - x_{j_{\text{opt}}}^\varepsilon, a\Big\rangle_{L^2(\Gamma_r)}, \end{aligned} \tag{43}$$

where the finite set A is defined as follows

$$A = \left\{ \frac{x_j^\varepsilon - x_i^\varepsilon}{\left\| x_j^\varepsilon - x_i^\varepsilon \right\|_{L^2(\Gamma_r)}} : i, j = 1, 2, \ldots, M \right\} \subset \mathbb{P}_{\lfloor \kappa N \rfloor}. \tag{44}$$

Then, for any $j = 1, 2, \ldots, M$ and $a \in A$, the right-hand side of (43) can be rewritten

$$\begin{aligned} \left\langle x_j^\varepsilon - x_{j_{\mathrm{opt}}}^\varepsilon, a \right\rangle_{L^2(\Gamma_r)} &= \left\langle x_j^\varepsilon, a \right\rangle_{L^2(\Gamma_r)} - \left\langle x_{j_{\mathrm{opt}}}^\varepsilon, a \right\rangle_{L^2(\Gamma_r)} \\ &= \left\langle x_j^\varepsilon, a \right\rangle_{L^2(\Gamma_r)} - \left\langle \mathcal{D} x_{j_{\mathrm{opt}}}^\varepsilon, \mathcal{D} a \right\rangle_w \\ &\approx \left\langle x_j^\varepsilon, a \right\rangle_{L^2(\Gamma_r)} - \left\langle \mathcal{D} y^\varepsilon, \mathcal{D} a \right\rangle_w . \end{aligned}$$

The computations above motivate the following choice of j^* as a substitute to the choice $j = j_{\mathrm{opt}}$:

Method 4.5. *Let Setting* 4.3 *hold true and the set A be given as in* (44)*. Then we choose the parameter j^* to be the minimizer of*

$$\min_{j=1,\ldots,M} \max_{a \in A} \left| \left\langle x_j^\varepsilon, a \right\rangle_{L^2(\Gamma_r)} - \left\langle \mathcal{D} y^\varepsilon, \mathcal{D} a \right\rangle_w \right|$$

Choosing j^* according to Method 4.5 eventually allows the estimate:

Theorem 4.6. *Let Setting* 4.3 *hold true. Then the minimizer j^* according to Method* 4.5 *satisfies*

$$\left\| x^\dagger - x_{j^*}^\varepsilon \right\|_{L^2(\Gamma_r)} \leqslant \left\| x^\dagger - x_{j_{\mathrm{opt}}}^\varepsilon \right\|_{L^2(\Gamma_r)} + 2 \left\| \mathcal{D} x^\dagger - \mathcal{D} x_{j_{\mathrm{opt}}}^\varepsilon \right\|_w + 2\varepsilon.$$

In other words, we can reasonably bound the error of $x_{j^*}^\varepsilon$ against the error of the optimal $x_{j_{\mathrm{opt}}}^\varepsilon$ among all available approximations x_j^ε. Knowledge of the noise level ε is not required for the choice of the parameter j^*. More details can be found in [26].

4.3. Numerical example

In this section, we illustrate the numerical performance of the previously described joint inversion and parameter choice method for the problem (25)–(27). The approximations $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$ are obtained as described in Method 4.1, the choice of λ_1^*, λ_2^* is performed according to Method 4.5 and Remark 4.4. The precise procedure for the numerical test is as follows:

(a) From the EGM2008 gravity potential model (cf. [55]), we generate a reference potential x up to spherical harmonic degree 130 on a sphere Ω_ρ, $\rho = 7,071$km, and on a spherical cap $\Gamma_r = \Gamma_r(\bar{\xi}, \bar{r})$, $r = 6,371$km, with center $\bar{\xi} = (0, 0, r)$ and spherical radius $\bar{r} = 0.3$ (corresponding to a spherical radius of approximately $5,000$km around the North Pole at the Earth's surface). The truncation degree in Method 4.1 will be fixed to $\lfloor \kappa N \rfloor = 150$.

(b) From EGM2008, we generate different sets of corresponding noisy measurements $\mathbf{y}_1^{\varepsilon_1}$, $\mathbf{y}_2^{\varepsilon_2}$, where the noise levels ε_2 of the ground data and ε_1 of the satellite data are varied among 0.001, 0.1. The data on Ω_ρ are computed on an equiangular grid according to [16] while the data on the spherical cap Γ_r are computed on a Gauss–Legendre grid according to [33]. In a first set of examples we choose the cubature rules [16, 33] to be polynomially exact up to degree 300, which guarantees condition (29) and implies $\mathbf{y}_1^{\varepsilon_1} \in \mathbb{R}^{L_1}$, $\mathbf{y}_2^{\varepsilon_2} \in \mathbb{R}^{L_2}$ with $L_1 \approx L_2 \approx 90,000$. In two further examples, we choose the polynomial exactness of the cubature rules to be only of degree 130 and degree 80, respectively (i.e., $\mathbf{y}_1^{\varepsilon_1} \in \mathbb{R}^{L_1}$, $\mathbf{y}_2^{\varepsilon_2} \in \mathbb{R}^{L_2}$ with $L_1 \approx L_2 \approx 18,000$ and $L_1 \approx L_2 \approx 6,500$, respectively). In these cases, the condition (29) is violated.

(c) For the different sets of input data from part (b), we compute approximations $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$ of $x^\dagger$ on Γ_r via the procedure of Method 4.1. The parameter λ_1 is varied in the interval $[10^1, 10^8]$ and λ_2 is varied in the interval $[10^{-2}, 10^3]$. This way, we obtain $M = 100$ different approximations $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$ for each fixed noise level $\varepsilon_1, \varepsilon_2$ and for each set of input data $\mathbf{y}_1^{\varepsilon_1}$, $\mathbf{y}_2^{\varepsilon_2}$.

(d) Among the M approximations $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$, we use the proposed procedure from Method 4.5 to choose a 'good' approximation $x_{N,\lambda_1^*,\lambda_2^*}^{\varepsilon_1,\varepsilon_2}$. Afterwards, we compare the relative error of our parameter choice, i.e.,

$$\mathrm{err}_* = \|x_{N,\lambda_1^*,\lambda_2^*}^{\varepsilon_1,\varepsilon_2} - x^\dagger\|_{L^2(\Gamma_r)} / \|x^\dagger\|_{L^2(\Gamma_r)},$$

with the actual best relative error, i.e.,

$$\mathrm{err}_{\mathrm{opt}} = \|x_{N,\lambda_1^{\mathrm{opt}},\lambda_2^{\mathrm{opt}}}^{\varepsilon_1,\varepsilon_2} - x^\dagger\|_{L^2(\Gamma_r)} / \|x^\dagger\|_{L^2(\Gamma_r)},$$

which can be computed because we know $x^\dagger$ from part (a).

The results of the tests are shown in Figures 4.2 and 4.3. Each figure shows the relative errors err_* and $\mathrm{err}_{\mathrm{opt}}$ for every test run. Additionally, we plotted $\mathrm{err}_{\max} = \max_{\lambda_1,\lambda_2} \|x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2} - x^\dagger\|_{L^2(\Gamma_r)} / \|x^\dagger\|_{L^2(\Gamma_r)}$ and the average errors $\mathrm{err}_{\mathrm{av}} = \frac{1}{M} \sum_{\lambda_1,\lambda_2} \|x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2} - x^\dagger\|_{L^2(\Gamma_r)} / \|x^\dagger\|_{L^2(\Gamma_r)}$ in order to illustrate the performance.

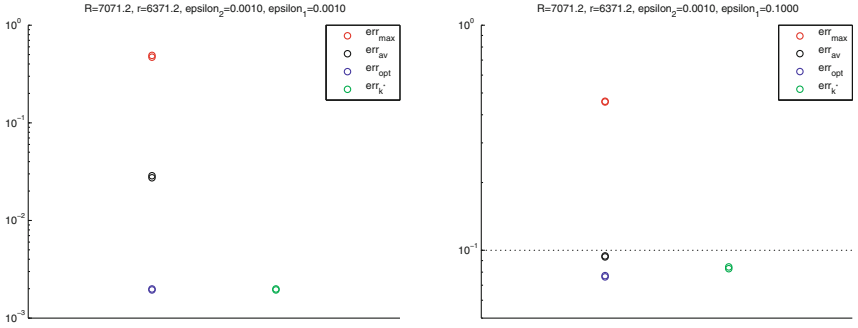


FIGURE 4.2. Relative errors for $\varepsilon_1 = \varepsilon_2 = 0.001$ (left) and $\varepsilon_1 = 0.1$, $\varepsilon_2 = 0.001$ (right; the dotted black line marks the noise level $\varepsilon_1 = 0.1$) and cubature rules with polynomial exactness of degree 300.

In Figure 4.2, it can be seen that the algorithm works particularly well for the setting $\varepsilon_1 = \varepsilon_2$ and that the oracle error err_* is nearly identical with the minimum error $\mathrm{err}_{\mathrm{opt}}$. The situation is different when $\varepsilon_2 \gg \varepsilon_1$. The minimum error $\mathrm{err}_{\mathrm{opt}}$ is smaller than the noise level ε_2. Thus, since our parameter choice strategy is based on comparing $x_{N,\lambda_1,\lambda_2}^{\varepsilon_1,\varepsilon_2}$ to $\mathbf{y}_2^{\varepsilon_2}$, we cannot expect that err_* is as good as $\mathrm{err}_{\mathrm{opt}}$. Yet, astonishingly enough, it seems that err_* is still slightly smaller than ε_2 for our test setting.

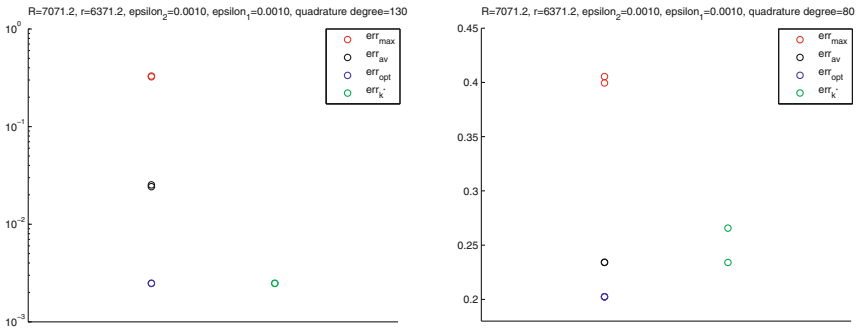


FIGURE 4.3. Relative errors for $\varepsilon_1 = \varepsilon_2 = 0.001$ and cubature rules with polynomial exactness of degree 130 (left) and of degree 80 (right).

Figure 4.3 show that the parameter choice algorithm is fairly stable with respect to violation of condition (29). For cubature rules with polynomial exactness only of degree 130, we see that err_* is still nearly identical to $\mathrm{err}_{\mathrm{opt}}$. For exactness of degree 80, however, we see that the parameter choice rules fails. The good stability of the rule in our example with respect to violation of (29) is due to the fact that the strongest contributions of the EGM2008 model are at low spherical harmonic degrees. For tests with synthetic data that has approximately equal strength at all spherical harmonic degrees, we refer the reader to [26]. In such situations the parameter choice rule is less stable but 'small' violations of condition (29) still yield good results.

Acknowledgment

Pavlo Tkachenko gratefully acknowledges the support of the Austrian Science Fund (FWF): projects P25424, I1669, and of the consortium AMMODIT funded within EU H2020-MSCA-RICE. Sergiy Pereverzyev Jr. gratefully acknowledges the support of the Austrian Science Fund (FWF): project P29514-N32.

References

[1] R.S. Anderssen. The linear functional strategy for improperly posed problems. In J.R. Cannon and U. Hornung, editors, *Proceedings of the Conference held at the Mathematical Research Institute at Oberwolfach, Black Forest, May* 18–24, 1986, pages 11–30, 1986.

[2] R.S. Anderssen. Inverse problems: A pragmatist's approach to the recovery of information from indirect measurements. *ANZIAM*, 45:C588–C622, 2005.

[3] R.S. Anderssen and H.W. Engl. The role of linear functionals in improving convergence rates for parameter identification via Tikhonov regularization. In M. Yamaguti, K. Hayakawa, Y. Iso, M. Mori, T. Nishida, K. Tomoeda, and M. Yamamoto, editors, *Inverse Problems in Engineering Sciences, ICM*-90 *Satellite Conf. Proc.*, pages 1–10, 1991.

[4] F. Bauer, M. Gutting, and M.A. Lukas. Evaluation of parameter choice methods for regularization of ill-posed problems in geomathematics. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*. Springer, 2nd edition, 2015.

[5] F. Bauer and S. Kindermann. The quasi-optimality criterion for classical inverse problems. *Inverse Problems*, 24:035002, 2009.

[6] F. Bauer, P. Mathe, and S. Pereverzev. Local solutions to inverse problems in geodesy. *J. Geod.*, 81:39–51, 2007.

[7] M. Bayer, W. Freeden, and T. Maier. A vector wavelet approach in iono- and magnetospheric geomagnetic satellite data. *J. Atm. Solar-Terr. Phys.*, 63:581–597, 2001.

[8] M. Brenner, Y. Jiang, and Y. Xu. Multiparameter regularization for Volterra kernel identification via multiscale collocation methods. *Adv. Comp. Math.*, 31:421–455, 2008.

[9] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, and S. Seatzu. Multi-parameter regularization techniques for ill-conditioned linear systems. *Numerische Mathematik*, 94:203–228, 2003.

[10] A. Chambodut, I. Panet, M. Mandea, M. Diamet, M. Holschneider, and O. Jamet. Wavelet frames: an alternative to spherical harmonic representation of potential fields. *Geophys. J. Int.*, 163:875–899, 2005.

[11] CHAMP. http://www.gfz-potsdam.de/champ/.

[12] J. Chen, S. Pereverzyev Jr., and Y. Xu. Aggregation of regularized solutions from multiple observation models. *Inverse Problems*, 31:075005, 2015.

[13] Z. Chen, Y. Lu, Y. Xu, and H. Yang. Multi-parameter Tikhonov regularization for linear ill-posed operator equations. *J. Comp. Math*, 26:37–55, 2008.

[14] S. Dahlke, W. Dahmen, E. Schmitt, and I. Weinreich. Multiresolution analysis and wavelets on s^2 and s^3. *Num. Func. Anal. Appl.*, 16:19–41, 1995.

[15] J.C. De Los Reyes and C.-B. Schönlieb. Image denoising: Learning the noise model via nonsmooth PDE-constrained optimization. *Inverse Probl. Imaging*, 7:1183–1214, 2013.

[16] J.R. Driscoll and M.H. Healy Jr. Computing fourier transforms and convolutions on the 2-sphere. *Adv. Appl. Math.*, 15:202–250, 1994.

[17] W. Freeden. On approximation by harmonic splines. *Manuscr. Geod.*, 6:193–244, 1981.

[18] W. Freeden. *Multiscale Modelling of Spaceborne Geodata*. Teubner, 1999.

[19] W. Freeden and C. Gerhards. Poloidal and toroidal field modeling in terms of locally supported vector wavelets. *Math. Geosc.*, 42:818–838, 2010.

[20] W. Freeden and S. Pereverzev. Spherical Tikhonov regularization wavelets in satellite gravity gradiometry with random noise. *J. Geod.*, 74:730–736, 2001.

[21] W. Freeden and F. Schneider. Regularization wavelets and multiresolution. *Inverse Problems*, 14:225–243, 1998.

[22] W. Freeden and U. Windheuser. Combined spherical harmonics and wavelet expansion – a future concept in earth's gravitational potential determination. *Appl. Comp. Harm. Anal.*, 4:1–37, 1997.

[23] E. Friis-Christensen, H. Lühr, and G. Hulot. Swarm: A constellation to study the Earth's magnetic field. *Earth Planets Space*, 58:351–358, 2006.

[24] C. Gerhards. Locally supported wavelets for the separation of spherical vector fields with respect to their sources. *Int. J. Wavel. Multires. Inf. Process.*, 10:1250034, 2012.

[25] C. Gerhards. A combination of downward continuation and local approximation for harmonic potentials. *Inverse Problems*, 30:085004, 2014.

[26] C. Gerhards, S. Pereverzyev Jr., and P. Tkachenko. A parameter choice strategy for the inversion of multiple observations. *Adv. Comput. Math.*, 43:101–112, 2017.

[27] GOCE. http://www.esa.int/our_activities/observing_the_earth/goce.

[28] A. Goldenshluger and S. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Rel. Fields*, 118:169–186, 2000.

[29] GRACE. http://www.jpl.nasa.gov/missions/gravity-recovery-and-climate-experiment-grace/.

[30] A. Gyulai, M.K. Baracza, and E.E. Tolnai. The application of joint inversion in geophysical exploration. *Int. J. Geosc.*, 4:283–289, 2013.

[31] G.V. Haines. Spherical cap harmonic analysis. *J. Geophys. Res.*, 90:2583–2591, 1985.

[32] M. Hegland. An optimal order regularization method which does not use additional smoothness assumptions. *SIAM J. Num. Anal.*, 29:1446–1461, 1992.

[33] K. Hesse and R.S. Womersley. Numerical integration with polynomial exactness over a spherical cap. *Adv. Comp. Math.*, 36:451–483, 2012.

[34] M. Holschneider. Continuous wavelet transforms on the sphere. *J. Math. Phys.*, 37:4156–4165, 1996.

[35] M. Holschneider, A. Chambodut, and M. Mandea. From global to regional analysis of the magnetic field on the sphere using wavelet frames. *Phys. Earth Planet. Inter.*, 135:107124, 2003.

[36] K. Ito, B. Jin, and Takeuchi. Multi-parameter Tikhonov regularization. Technical report, arXiv 1102.1173, 2011.

[37] L.K. Jones. A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training. *Ann. Statist.*, 20:608–613, 1992.

[38] A. Juditsky and A. Nemirovski. Functional aggregation for nonparametric regression. *Ann. Statist.*, 28:681–712, 2000.

[39] R. Klees and T. Wittwer. Local gravity field modeling with multipole wavelets. In P. Tregoning and C. Rizos, editors, *Dynamic Planet, International Association of Geodesy Symposia*, volume 130. Springer, 2007.

[40] K.R. Koch and J. Kusche. Regularization of geopotential determination from satellite data by variance components. *J. Geod.*, 76:259.268, 2002.

[41] K. Kunisch and T. Pock. A bilevel optimization approach for parameter learning in variational models. *SIAM J. Imag. Sci.*, 6:938–983, 2013.

[42] A.K. Louis. Approximate inverse for linear and nonlinear problems. *Inverse Problems*, 12:175–190, 1996.

[43] S. Lu and S. Pereverzev. *Regularization Theory for ill-posed problems – Selected Topics*. de Gruyter, 2013.

[44] S. Lu and S. Pereverzyev. Multiparameter regularization in downward continuation of satellite data. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*. Springer, 2nd edition, 2015.

[45] S. Lu, S. Pereverzyev Jr., and S. Sampath. Multiparameter regularization for construction of extrapolating estimators in statistical learning theory. In X. Shen and A.I. Zayed, editors, *Multiscale Signal Analysis and Modeling*. Springer, 2013.

[46] T. Maier and C. Mayer. Multiscale downward continuation of champ fgm-data for crustal field modelling. In C. Rizos, H. Lühr, and P. Schwintzer, editors, *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*. Springer, 2003.

[47] P. Mathe. The Lepskii principle revisited. *Inverse Problems*, 22:L11–L15, 2006.

[48] P. Mathe and B. Hofmann. How general are general source conditions? *Inverse Problems*, 24:015009, 2008.

[49] P. Mathe and S. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19:789, 2003.

[50] C. Mayer and T. Maier. Separating inner and outer Earth's magnetic field from CHAMP satellite measurements by means of vector scaling functions and wavelets. *Geophys. J. Int.*, 167:11881203, 2006.

[51] J.R.W. Merrick. Aggregation of forecasts from multiple simulation models. In *Proceedings of the* 2013 *Winter Simulation Conference*, pages 533–542. IEEE, 2013.

[52] V. Michel. Scale continuous, scale discretized and scale discrete harmonic wavelets for the outer and the inner space of a sphere and their application to an inverse problem in geomathematics. *Appl. Comp. Harm. Anal.*, 12:77–99, 2002.

[53] V. Naumova, S. Pereverzev, and P. Tkachenko. Regularized collocation for spherical harmonics gravitational field modeling. *Int. J. Geomath.*, 5:81–98, 2014.

[54] NGDC-720. http://geomag.org/models/ngdc720.html.

[55] N.K. Pavlis, S.A. Holmes, S.C. Kenyon, and Factor J.K. The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *J. Geophys. Res.*, 117:B04406, 2012.

[56] S. Pereverzyev and E. Schock. Error estimates for band-limited spherical regularization wavelets in an inverse problem of satellite geodesy. *Inverse Problems*, 15:881–890, 1999.

[57] S. Pereverzyev and P. Tkachenko. Pointwise computation in an ill-posed spherical pseudo-differential equation. *Comp. Meth. Appl. Math.*, 15:213–219, 2015.

[58] A. Plattner and F.J. Simons. Spatiospectral concentration of vector fields on a sphere. *Appl. Comp. Harm. Anal.*, 36:1–22, 2014.

[59] A. Plattner and F.J. Simons. Potential-field estimation from satellite data using scalar and vector Slepian functions. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*. Springer, 2nd edition, 2015.

[60] P. Schröder and W. Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. In S.G. Mair and R. Cook, editors, *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, 1995.

[61] L. Shure, R.L. Parker, and G.E. Backus. Harmonic splines for geomagnetic modeling. *Phys. Earth Planet. Inter.*, 28:215–229, 1982.

[62] F.J. Simons, F.A. Dahlen, and M.A. Wieczorek. Spatiospectral localization on a sphere. *SIAM Review*, 48:505–536, 2006.

[63] F.J. Simons and A. Plattner. Scalar and vector Slepian functions, spherical signal estimation and spectral analysis. In W. Freeden, M.Z. Nashed, and T. Sonar, editors, *Handbook of Geomathematics*. Springer, 2nd edition, 2015.

[64] S.L. Svensson. Pseudodifferential operators – a new approach to the boundary problems of physical geodesy. *Manuscr. Geod.*, 8:1–40, 1983.

[65] E. Thébault, J.J. Schott, and M. Mandea. Revised spherical cap harmonic analysis (R-SCHA): Validation and properties. *J. Geophys. Res.*, 111:B01102, 2006.

[66] A.N. Tikhonov and V.B. Glasko. Use of the regularization method in non-linear problems. *USSR Comput. Math. Math. Phys.*, 5:93–107, 1965.

[67] H. Trompat, F. Boschetti, and P. Hornby. Improved downward continuation of potential field data. *Explor. Geophys.*, 34:249–256, 2003.

[68] I.N. Tziavos, V.D. Andritsanos, R. Forsberg, and A.V. Olesen. Numerical investigation of downward continuation methods for airborne gravity data. In C. Jekeli, L. Bastos, and J. Fernandes, editors, *Gravity, Geoid and Space Missions – GGSM 2004 IAG International Symposium Porto*. Springer, 2005.

[69] K. Vozoff and D.L.B. Jupp. Joint inversion of geophysical data. *Geophys. J. Int.*, 42:977–991, 1975.

Christian Gerhards
Computational Science Center
University of Vienna
A-1090 Vienna, Austria
e-mail: christian.gerhards@univie.ac.at

Sergiy Pereverzyev Jr.
Department of Mathematics
University of Innsbruck
A-6020 Innsbruck, Austria
e-mail: sergiy.pereverzyev@uibk.ac.at

Pavlo Tkachenko
Institute for Design and Control
of Mechatronical Systems
Johannes Kepler University
A-4040 Linz, Austria
e-mail: pavlo.tkachenko@jku.at

Handbook of Mathematical Geodesy
Geosystems Mathematics, 883–919

On the Non-uniqueness of Gravitational and Magnetic Field Data Inversion (Survey Article)

Sarah Leweke, Volker Michel, and Roger Telschow

Abstract. The gravitational and the magnetic field of the Earth represent some of the most important observables of the geosystem. The inversion of these fields reveals hidden structures and dynamics at the surface or in the interior of the Earth (or other celestial bodies). However, the inversions of both fields suffer from a severe non-uniqueness of the solutions. In this paper, we present a generalized approach which includes the inversion of gravitational and magnetic field data. Amongst others, uniqueness constraints are proposed and compared. This includes the surface density ansatz (also known as the thin layer assumption). We characterize the null space of the considered class of inverse problems via an appropriate orthonormal basis system. Further, we expand the reconstructable part of the solution by means of orthonormal bases and reproducing kernels. One result is that information on the radial dependence of the solution is lost in the observables. As an illustration of the non-uniqueness, we show examples of anomalies which cannot be disclosed from the inversion of gravitational data. This paper is intended to be a theoretical reference work on the inversion of gravitational but also magnetic field data of the Earth.

1. Introduction

Numerous tasks in mathematical geodesy involve the regularization of ill-posed inverse problems. The reason is obvious: neither the interior of the Earth nor the Earth's surface in its entirety are accessible for exploration. However, the demand for more accurate and more localized models has dramatically increased for the last decades. As a consequence, numerous large data sets of various observables have been generated, also by means of satellite missions. These data sets often provide us with the possibility to derive models for non-observable, but urgently needed geodetic fields. Examples are the quantification of mass transports due to climate change or other phenomena (GRACE (Gravity Recovery And Climate Experiment) data are well appropriate for this purpose, see, e.g., [11, 26, 29, 52, 53])

and the modeling of those layers of the Earth which contribute to the magnetic field (this can be done with SWARM data, see, e.g., [37, 43, 50]).

This survey article presents a generalized approach which comprises, in particular, the inversion of gravitational or magnetic field data. In the former case, the unknown is the mass density distribution of the Earth's body or its surface. In the latter case, the unknown is considered to be the electric current distribution inside. In this sense, this paper is an extension of the survey article [33] on inverse gravimetry. One benefit of the generalized approach is that it makes it easier to transfer theoretical knowledge and numerical methods from one problem to the other within the considered class of problems. For example, it was shown in [23] and [33] that such a transfer yields novel achievements. Furthermore, our generalized approach also enables us to set the surface mass density approach (also known as the thin layer assumption) into the same concept with the inversion for volumetric density distribution – two approaches which have often been used parallelly and independently (see, e.g., [33, 52]). Since this paper addresses primarily a geodetic audience, we focus on the relevant facts and their interpretation. For the detailed mathematical theory including the proofs, we recommend to use the paper [34] as a supplement.

Note that the considered inversion of magnetic field data is motivated by the inversion of MEG (magnetoencephalography) data, as it occurs in medical imaging (see also [23] and the references therein). Thus, it does not represent a typical inverse problem in geomagnetics, where, for instance, material parameters like the magnetization or the susceptibility are the unknowns and not the current (see, e.g., [46]). However, the inversion of the magnetic field for currents in the interior might be interesting for investigating the outer core. Nevertheless, there is still an obvious limitation of our generalized approach with respect to the practical applicability in geomagnetics. On the other hand, reversing the point of view, the generalized approach shows a perspective how methods from medical imaging (which exist in a vast variety) could be transferred to geodetic and geophysical inverse problems.

The content and the outline of the paper are as follows: in Section 2, we summarize some basic fundamentals, like the definition of the function spaces and the orthogonal polynomials which we need.

In Section 3, we formulate the generalized class of inverse problems which represents the central theme of this paper. Then, we discuss two particular cases: the inversion of the gravitational field (this is known as the inverse gravimetric problem) and the inversion of the Bio–Savart operator of a magnetic field for getting the current distribution inside (we call this the inverse magnetic problem). With this in mind, every theoretical result that we present here for the generalized problem is valid for these two particular applications, and the derived formulae can be directly used for the precise problem by inserting the associated parameters. In Subsection 3.2, we derive a spectral relation between the given field and the unknown field. This relation directly shows the problem of the non-uniqueness which is linked to the insufficiently identifiable radial parametrization of the solution.

In Section 4, we introduce a class of orthonormal basis systems on a 3-dimensional ball. One particular instance of this class yields the well-known system of harmonic and anharmonic functions which have been used for the inverse gravimetric problem. We include some plots of the basis functions and show that the basis is appropriate for separating the solution into its projections on the null space of the solution (i.e., the indeterminable part of the solution) and on the orthogonal complement (the components of the solution which are uniquely constrained by the given data). We also show graphical illustrations of phantoms which occur, that is, examples of anomalies inside the Earth which cannot be distinguished if only gravitational data are available.

In Section 5, we discuss several modeling assumptions which can be used to obtain a unique solution: a minimum norm constraint, a harmonicity constraint, a layer density constraint and the surface density (i.e., thin layer) constraint, which is common for the identification of water mass transports.

2. Preliminaries

In this work, the set of positive integers is denoted by $\mathbb{N}$, where $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Moreover, $\mathbb{R}$ represents the set of real numbers. The Euclidean standard $\mathbb{R}^3$-scalar product (dot product) is denoted by $\cdot$ and the cross product by $\times$. The norm associated to the Euclidean dot product is represented by $|x| := \sqrt{x \cdot x}$, $x \in \mathbb{R}^3$. Furthermore, the sphere with radius R is denoted by $\Omega_R := \{x \in \mathbb{R}^3 \mid |x| = R\}$ and the corresponding (closed) ball is denoted by $\mathcal{B} := \{x \in \mathbb{R}^3 \mid |x| \leq R\}$. For $R = 1$, we often use the abbreviation $\Omega := \Omega_1$. By $\mathcal{S} := \Omega_\beta$, with $\beta > R$, we denote a particular sphere in the exterior of $\mathcal{B}$. This could, for example, represent a satellite altitude or the location of airborne data.

A function $F\colon \mathcal{G} \to \mathbb{R}$ possessing k continuous derivatives on the open set $\mathcal{G} \subset \mathbb{R}^n$ is of class $\mathrm{C}^{(k)}(\mathcal{G})$, for $0 \leq k \leq \infty$. Furthermore, for a measurable set $\mathcal{G} \subset \mathbb{R}^n$, $\mathrm{L}^2(\mathcal{G})$ stands for the space of all square-integrable functions (more precisely, some equivalence classes of such functions). $\mathrm{L}^2(\mathcal{G})$ is a Hilbert space with the inner product

$$\langle F, G\rangle_{\mathrm{L}^2(\mathcal{G})} := \int_{\mathcal{G}} F(x)G(x)\,\mathrm{d}\mu(x), \qquad F,\ G \in \mathrm{L}^2(\mathcal{G}),$$

and the norm

$$\|F\|_{\mathrm{L}^2(\mathcal{G})} = \left(\int_{\mathcal{G}} F(x)^2\,\mathrm{d}\mu(x)\right)^{1/2}, \qquad F \in \mathrm{L}^2(\mathcal{G}),$$

where μ is an appropriate measure, like a surface measure ω if $\mathcal{G}$ is a surface. For a mathematically accurate definition of the space, see, for example, [42].

With $P_m^{(\alpha,\beta)}$, we denote the Jacobi polynomials, where $\alpha, \beta > -1$. They are uniquely determined by the conditions that

1. each $P_m^{(\alpha,\beta)}$ is a polynomial of degree m,
2. for all $m, n \in \mathbb{N}_0$ with $m \neq n$,

$$\left\langle P_m^{(\alpha,\beta)}, P_n^{(\alpha,\beta)} \right\rangle_{\alpha,\beta} := \int_{-1}^{1} (1-x)^\alpha (1+x)^\beta P_m^{(\alpha,\beta)}(x) P_n^{(\alpha,\beta)}(x)\,\mathrm{d}x = 0, \quad (1)$$

3. and for each $m \in \mathbb{N}_0$, we set $P_m^{(\alpha,\beta)}(1) = \binom{m+\alpha}{m}$.

For $\alpha = \beta = 0$, the Jacobi polynomials coincide with the Legendre polynomials. For further properties and the $\mathrm{L}^2[0,R]$-norm of Legendre, or (more generally) Jacobi polynomials, see [24, 36, 49].

3. Generalization of gravitational and magnetic field inversion

3.1. A class of inverse problems and examples

Within this paper, we consider a class of inverse problems which are given by a Fredholm integral operator of the first kind $T\colon \mathrm{L}^2(\mathcal{B}) \to \mathrm{L}^2(\mathcal{S})$

$$T\colon D \mapsto \int_{\mathcal{B}} D(x) k(x,\cdot)\,\mathrm{d}x = V \quad (2)$$

with an integral kernel $k\colon \mathcal{B} \times \mathcal{S} \to \mathbb{R}$ of the form

$$k(x,y) := \sum_{i=0}^{\infty} c_i \frac{|x|^{l_i}}{|y|^{i+1}} P_i\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right), \quad (3)$$

which is defined for all $(x,y) \in \mathrm{dom}(k)$, where the domain of the kernel k is given by

$$\mathrm{dom}(k) := \{(x,y) \in \mathcal{B} \times \mathcal{S} \mid x \neq 0 \text{ if there exists } i \in \mathbb{N}_0 \text{ with } l_i < 0\}.$$

In this setting, the right-hand side V in Equation (2) is given and the function D is unknown. It is the aim to reconstruct D in $\mathcal{B}$ from knowledge of V on $\mathcal{S}$. In order to have a well-defined integral kernel, which means that the series representation in (3) converges, k has to fulfil certain assumptions:

Assumption 3.1.

1. *The sequence* $(c_i)_{i\in\mathbb{N}_0}$ *is a real and bounded sequence* (*i.e., there exists* $c \in \mathbb{R}^+$ *such that* $\sup_{i\in\mathbb{N}_0} |c_i| \leq c$).
2. *The sequence of real exponents* $(l_i)_{i\in\mathbb{N}_0}$ *satisfies* $\inf_{i\in\mathbb{N}_0} l_i \geq -1$.
3. *The sequence* $(l_i)_{i\in\mathbb{N}_0}$ *fulfils the condition* $\sup_{i\in\mathbb{N}_0} R^{l_i - i} < \infty$.

Note, that the third condition implies

$$R^{i-l_i} = \frac{1}{R^{l_i-i}} \geq \frac{1}{\sup_{i\in\mathbb{N}} R^{l_i-i}} > 0.$$

This kind of integral equation arises in many areas, for example, in geosciences and medical imaging. Two examples for this inverse problem are given below.

For both, Example 3.2 (inverse gravimetric problem) and Example 3.3 (inverse magnetic problem), the conditions of Assumption 3.1 are fulfilled. In the first particular case, that is, $l_i = i$ and $c_i = \gamma$ for all $i \in \mathbb{N}_0$, the integral kernel is well known. In this case, we directly obtain $k(x,y) = \frac{\gamma}{|x-y|}$ for $|x| < |y|$, due to the identity

$$\sum_{i=0}^{\infty} \frac{|x|^i}{|y|^{i+1}} P_i\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right) = \frac{1}{|x-y|} \qquad \text{for } |x| < |y|. \tag{4}$$

Example 3.2 (The Inverse Gravimetric Problem). For the inverse gravimetric problem, the kernel and the integral operator are given by

$$T^{\mathrm{G}} \colon D \mapsto \int_{\mathcal{B}} D(x) k^{\mathrm{G}}(x,\cdot)\,\mathrm{d}x,$$
$$k^{\mathrm{G}}(x,y) \coloneqq \frac{\gamma}{|x-y|} = \gamma \sum_{i=0}^{\infty} \frac{|x|^i}{|y|^{i+1}} P_i\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right),$$

where $x \in \mathcal{B}$, $y \in \mathcal{S}$, P_i denotes the Legendre polynomial of degree i and γ is the gravitational constant. $T^{\mathrm{G}}D$ is known as the gravitational potential or the Newton potential. The associated inverse problem $T^{\mathrm{G}}D = V$ represents the reconstruction of a (volumetric) mass density function from the gravitational potential, which is important, for example, for the detection of particular anomalies or mass transports. For the latter, time series of potential models have been provided, for instance, by the GRACE mission, see [10]. Note that the determination of a surface density can be regarded as a particular modeling in this context.

This problem first occurs in the works of Stokes [47] and has been widely discussed since then (see also the survey article [33]).

Example 3.3 (The Inverse Magnetic Problem). To compute the magnetic field B caused by electric sources inside a body, the quasi-static approximation of Maxwell's equation is often used, see [39].

$$E = -\nabla U \quad \text{on } \mathcal{B}, \qquad \nabla \cdot B = 0 \quad \text{on } \mathcal{B},$$
$$\nabla \times B = 0 \quad \text{on } \mathcal{S}, \qquad \nabla \times B = \mu_0 J^{\mathrm{T}} \quad \text{on } \mathcal{B},$$

where E is the electric field, U is the electric potential, $J^{\mathrm{T}} = J^{\mathrm{P}} + \sigma E$ is the total current with the primary current J^{P} and the Ohmic current σE, σ is the conductivity, and μ_0 is the permeability. It is common to use the Biot–Savart operator instead of Maxwell's equations to describe the relation between the current and the magnetic field

$$B(x) = \frac{\mu_0}{4\pi} \int_{\mathcal{B}} J^{\mathrm{T}}(y) \times \frac{x-y}{|x-y|}\,\mathrm{d}y. \tag{5}$$

In this case, we want to recover a particular component of the electric current inside $\mathcal{B}$ (which could be the Earth (in particular the outer core)). Note that

this geophysical problem is closely related to a problem in medical imaging, where neuronal currents are determined from magnetoencephalography (MEG) data, see, for example, [19]. In some applications, only the reconstruction of the primary current instead of the total current or the induced current is of interest. After splitting the current in this sense and assuming a ball-shaped conductor consisting of spherical shells Ω_j with constant conductivities σ_j, one obtains the Geselowitz' formula (see [25])

$$B(x) = \frac{\mu_0}{4\pi} \int_{\mathcal{B}} J^{\mathrm{P}}(y) \times \frac{x-y}{|x-y|^3} \,\mathrm{d}y \\ - \frac{\mu_0}{4\pi} \sum_j (\sigma_{j-1} - \sigma_j) \int_{\Omega_{j-1}} V(y) n(y) \times \frac{x-y}{|x-y|^3} \,\mathrm{d}\omega(y),$$

where n is the normal vector on the surface Ω_j. With the identity in (4) and after further calculations, see [23], one gets a relation for the magnetic potential ($B = \nabla V$)

$$V(y) = \frac{1}{4\pi} \int_{\mathcal{B}} \nabla_x \cdot (J^{\mathrm{P}}(x) \times x) \sum_{i=0}^{\infty} \frac{|x|^i}{|y|^{i+1}\,(i+1)} P_i\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right) \mathrm{d}x.$$

More precisely, the vectorial current J^{P} inside $\mathcal{B}$ can be decomposed via two scalar-valued (up to an additional constant unique) functions F and G and a scalar-valued unique function J^r (see, e.g., [23]) as follows:

$$J^{\mathrm{P}}(r\xi) = \frac{1}{r} \nabla^*_\xi G(r\xi) - \frac{1}{r} \mathrm{L}^*_\xi F(r\xi) + J^r(r\xi)\xi.$$

Here, $\mathcal{B} \setminus \{0\} \ni x = r\xi$ with $\xi \in \Omega$ and $r = |x|$, ∇^*_ξ is the surface gradient, and $\mathrm{L}^*_\xi := \xi \times \nabla^*_\xi$ is the surface curl operator on the unit sphere. Due to [45] and the above decomposition, the relation between the current and the magnetic potential V in a spherical model can be described by

$$V(y) = \frac{1}{4\pi} \int_{\mathcal{B}} \Delta^*_{\frac{x}{|x|}} F(x) \sum_{i=0}^{\infty} \frac{|x|^{i-1}}{|y|^{i+1}\,(i+1)} P_i\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right) \mathrm{d}x,$$

where $\Delta^*_{\frac{x}{|x|}}$ denotes the Beltrami operator.

Hence, only the function F and, therefore, only one tangential component of the current can be reconstructed. We use now the abbreviation $D(x) := \Delta^*_{\frac{x}{|x|}} F(x)$ such that for the inverse magnetic problem (as we call the problem here), the kernel and the integral operator are given by

$$T^{\mathrm{M}} \colon D \mapsto \int_{\mathcal{B}} D(x) k^{\mathrm{M}}(x, \cdot) \,\mathrm{d}x, \tag{6}$$

$$k^{\mathrm{M}}(x,y) := \frac{1}{4\pi} \sum_{i=0}^{\infty} \frac{|x|^{i-1}}{|y|^{i+1}\,(i+1)} P_i\left(\frac{x}{|x|} \cdot \frac{y}{|y|}\right), \tag{7}$$

where $x \in \mathcal{B} \setminus \{0\}$, $y \in \mathcal{S}$.

This particular modeling of the inversion of magnetic data has been used for data from MEG, as we indicated above. For two reasons, we consider a discussion to be useful: The magnetic field of a ball-shaped domain with a current in the interior is also relevant in geodesy, and there is a close link to the inverse gravimetric problem as our generalized approach suggests.

We can find further properties of the integral kernel in (3). An estimate shows that the kernel function $k(\cdot, y)$, for each fixed $y \in \mathcal{S}$, is a function in $\mathrm{L}^2(\mathcal{B})$. Indeed (with $x = r\xi$, $r \in [0, R]$, $\xi \in \Omega$) we get, using Assumption 3.1 and the fact that $|P_i(t)| \leq 1$ for all $i \in \mathbb{N}_0$ and all $t \in [-1, 1]$, the estimate

$$\begin{aligned}
\int_{\mathcal{B}} (k(x,y))^2 \,\mathrm{d}x &= \int_{\mathcal{B}} \left(\sum_{i=0}^{\infty} c_i \frac{|x|^{l_i}}{|y|^{i+1}} P_i\left(\frac{x}{|x|} \cdot \frac{y}{|y|} \right) \right)^2 \mathrm{d}x \\
&\leq c^2 \int_{\mathcal{B}} \left(\sum_{i=0}^{\infty} \frac{|x|^{l_i}}{|y|^{i+1}} \right)^2 \mathrm{d}x = 4\pi c^2 \int_0^R r^2 \left(\sum_{i=0}^{\infty} \frac{r^{l_i}}{|y|^{i+1}} \right)^2 \mathrm{d}r \\
&= 4\pi c^2 \int_0^R \left(\sum_{i=0}^{\infty} \frac{r^{l_i+1}}{|y|^{i+1}} \right)^2 \mathrm{d}r \leq 4\pi R c^2 \left(\sup_{n \in \mathbb{N}_0} R^{l_n - n} \right)^2 \left(\sum_{i=0}^{\infty} \frac{R^{i+1}}{|y|^{i+1}} \right)^2 < \infty.
\end{aligned}$$

The last series is convergent and, hence, finite, since it is a geometric series. With similar calculations one can prove that the interchanging between the series and the integration over $\mathcal{B}$ was allowed.

Besides the well-definition of the integral kernel, we need the existence of the integral in (2) to obtain a well-defined problem. We will later see that this is achieved if some technical conditions are fulfilled. On the other hand, for the well-posedness of the problem (in the sense of Hadamard), three questions are important.

- Does, for every right-hand side V in (2), a solution D exist?
- Is there not more than one solution D for a given V?
- Is the problem stable, that is, does D depend continuously on the data V?

The question about the non-uniqueness of the solution for the above mentioned problems has been discussed comprehensively in literature. One of the first works is the paper due to Stokes [47] for the inverse gravimetric problem. Further publications are, for example, [4, 6, 8, 48]. For a survey article on this topic, see [33]. For the inverse magnetic problem (with a focus on medical imaging), see [13–15, 19–22, 45].

In the following sections, we want to derive a possibility to characterize the null space, or in other words we want to describe the part of the solution which is non-reconstructable. We also want to formulate additional conditions to guarantee the uniqueness of the solution. For this, we need more knowledge of the forward problem.

3.2. Derivation of a spectral relation

In this subsection, it is our aim to derive an equation which connects the spherical harmonics coefficients of the given function V and the unknown function D. With this spectral relation, we are able to give answers to the questions concerning the ill-posedness of the problem. For this purpose, we analyze the forward problem. The following considerations are motivated by a similar result for the particular case of the inverse gravimetric problem, see [33]. We assume that we can choose basis functions for D which are separable into a radial and an angular part such that D is expandable in an $\mathrm{L}^2(\mathcal{B})$-convergent spherical harmonics series

$$D(x) = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} D_{n,j}(|x|) Y_{n,j}\left(\frac{x}{|x|}\right). \tag{8}$$

Here, $Y_{n,j}$ denotes the spherical harmonics of degree n and order j, which are an orthonormal basis for $\mathrm{L}^2(\Omega)$. Furthermore, $D_{n,j}(r)$, $r \in [0, R]$, represents the spherical harmonics coefficients for the case that D is restricted to the sphere around the origin with radius r.

By virtue of the weak convergence in Hilbert spaces, we know that

$$\begin{aligned}\int_{\mathcal{B}} D(x)F(x)\,\mathrm{d}x &= \int_{\mathcal{B}} \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} D_{n,j}(|x|) Y_{n,j}\left(\frac{x}{|x|}\right) F(x)\,\mathrm{d}x \\ &= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \int_{\mathcal{B}} D_{n,j}(|x|) Y_{n,j}\left(\frac{x}{|x|}\right) F(x)\,\mathrm{d}x\end{aligned}$$

for all functions $F \in \mathrm{L}^2(\mathcal{B})$. In particular, this holds true for the integral kernel $k(\cdot, y) \in \mathrm{L}^2(\mathcal{B})$ for all $y \in \mathcal{S}$. Inserting the expansion (8) in (2) and using the abbreviation $y = |y|\,\eta$, $x = r\xi$ with $\eta, \xi \in \Omega$, we get

$$\begin{aligned}V(y) &= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \sum_{i=0}^{\infty} \int_0^R r^2 D_{n,j}(r) \frac{c_i r^{l_i}}{|y|^{i+1}}\,\mathrm{d}r \int_{\Omega} P_i\left(\xi \cdot \eta\right) Y_{n,j}\left(\xi\right)\,\mathrm{d}\omega(\xi) \\ &= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \sum_{i=0}^{\infty} \frac{c_i}{|y|^{i+1}} \int_0^R r^{l_i+2} D_{n,j}(r)\,\mathrm{d}r\, \frac{4\pi}{2n+1}\,\delta_{i,n} Y_{n,j}\left(\eta\right) \\ &= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \left(\int_0^R r^{l_n+2} D_{n,j}(r)\,\mathrm{d}r \right) \frac{4\pi c_n}{(2n+1)\,|y|^{n+1}} Y_{n,j}\left(\eta\right). \end{aligned} \tag{9}$$

In the first step the reproducing property of the reproducing kernel for the spherical harmonics of degree n, given by

$$\Omega^2 \ni (\xi, \eta) \mapsto \frac{2n+1}{4\pi} P_n(\xi \cdot \eta),$$

is used. More precisely,

$$\frac{2n+1}{4\pi}\int_\Omega P_i(\xi\cdot\eta)Y_{n,j}(\xi)\,\mathrm{d}\omega(\xi)=Y_{n,j}(\eta)\delta_{i,n} \tag{10}$$

for all $\eta\in\Omega$. We also remark that the existence of the integral in (2) only depends on the existence of the integral of the radial part and the convergence of the series in (9). Regarding the latter, we obtain a pointwise convergence of (9) for $y\in\mathcal{S}$, since the following estimate of the summands in (9) (note that $\max_{\xi\in\Omega}|Y_{n,j}(\xi)|\leq\sqrt{(2n+1)/(4\pi)}$ for all $n\in\mathbb{N}_0$) holds true:

$$\begin{aligned}
&\left|\int_0^R r^{l_n+2}D_{n,j}(r)\,\mathrm{d}r\;\frac{4\pi c_n}{2n+1}\;|y|^{-n-1}Y_{n,j}\left(\frac{y}{|y|}\right)\right|\\
&\quad\leq\left(\frac{R^{2l_n+3}}{2l_n+3}\int_0^R r^2\left(D_{n,j}(r)\right)^2\,\mathrm{d}r\right)^{1/2}\frac{4\pi c}{2n+1}\;|y|^{-n-1}\sqrt{\frac{2n+1}{4\pi}}\\
&\quad\leq c\left(\frac{R^{2l_n+3}}{R^{2n+2}(2l_n+3)}\int_0^R r^2\left(D_{n,j}(r)\right)^2\,\mathrm{d}r\frac{4\pi}{2n+1}\right)^{1/2}\left(\frac{R}{|y|}\right)^{n+1}.
\end{aligned}$$

The right-hand side is bounded for all $n\in\mathbb{N}_0$, due to the conditions on $(l_n)_{n\in\mathbb{N}}$ (see Assumption 3.1, items 2 and 3) and the convergence of the Parseval identity of $D\in\mathrm{L}^2(\mathcal{B})$. Hence, the series (9) is dominated by a geometric series for all $y\in\mathcal{S}$ (i.e., $|y|>R$).

We are also able to extend the function V onto Ω_R. In addition, for $V|_{\Omega_R}$, we obtain the $\mathrm{L}^2(\Omega_R)$-convergence of the series representation in Equation (9). This convergence is a direct consequence of the Cauchy–Schwarz inequality and the Parseval identity (note that $\{\frac{1}{R}Y_{n,j}(\frac{\cdot}{R})\}_{n\in\mathbb{N}_0,j=1,\dots,2n+1}$ is an orthonormal basis of $\mathrm{L}^2(\Omega_R)$), since

$$\begin{aligned}
\|V|_{\Omega_R}\|^2_{\mathrm{L}^2(\Omega_R)}&=\sum_{n=0}^\infty\sum_{j=1}^{2n+1}\left(\int_0^R r^{l_n+2}D_{n,j}(r)\,\mathrm{d}r\right)^2\left(\frac{4\pi c_n}{(2n+1)R^n}\right)^2\\
&\leq\sum_{n=0}^\infty\sum_{j=1}^{2n+1}\left(\int_0^R\frac{r^{2l_n+2}}{R^{2n}}\,\mathrm{d}r\right)\left(\int_0^R r^2(D_{n,j}(r))^2\,\mathrm{d}r\right)\left(\frac{4\pi c}{2n+1}\right)^2\\
&\leq 16\pi^2c^2\sum_{n=0}^\infty\sum_{j=1}^{2n+1}\frac{R^{2l_n+3-2n}}{2l_n+3}\left(\int_0^R r^2(D_{n,j}(r))^2\,\mathrm{d}r\right)\\
&\leq 16\pi^2c^2R^3\sup_{n\in\mathbb{N}_0}\frac{R^{2l_n-2n}}{2l_n+3}\,\|D\|^2_{\mathrm{L}^2(\mathcal{B})}<\infty.
\end{aligned}$$

Hence, Equation (9) is valid pointwise on $\mathcal{S}$ and in the sense of $\mathrm{L}^2(\Omega_R)$ on Ω_R.

In order to find a direct relation between the Fourier coefficients of the given function V and the unknown function D, we consider the Fourier coefficients of V restricted to the sphere Ω_R. This relation can be seen directly from (9).

Theorem 3.4. *Consider the orthonormal basis system on Ω_R given by the set of functions $\{\frac{1}{R}Y_{n,j}(\frac{\cdot}{R})\}_{n\in\mathbb{N}_0,j=1,\dots,2n+1}$. Then, the Fourier coefficients of V defined by $V_{n,j} \coloneqq \left\langle V|_{\Omega_R}, \frac{1}{R}Y_{n,j}\left(\frac{\cdot}{R}\right)\right\rangle_{\mathrm{L}^2(\Omega_R)}$ satisfy the identity*

$$V_{n,j} = \left(\int_0^R r^{l_n+2} D_{n,j}(r)\,\mathrm{d}r\right)\frac{4\pi c_n}{(2n+1)R^n}.$$

for all $n \in \mathbb{N}_0$, $j = 1,\dots,2n+1$. This yields the equation

$$\frac{(2n+1)R^n}{4\pi c_n}V_{n,j} = \int_0^R r^{l_n+2} D_{n,j}(r)\,\mathrm{d}r, \qquad \text{if } c_n \neq 0, \tag{11}$$

otherwise $V_{n,j} = 0$ with $j = 1,\dots,2n+1$, respectively.

The relation from Theorem 3.4 allows an infinite number of choices for $D_{n,j}$ and, hence, the solution D cannot be uniquely determined by the function $V|_{\Omega_R}$. For the inverse gravimetric problem, the last relation is well known, see, for example, [35, 38, 41], and for the inverse magnetic problem for $R = 1$, see for instance [21]. Analogously, we obtain with (remember that $\mathcal{S} = \Omega_\beta$, $\beta > R$) for all $n \in \mathbb{N}_0$, $j = 1,\dots,2n+1$

$$\begin{aligned} V^{\mathcal{S}}_{n,j} &\coloneqq \left\langle V|_{\mathcal{S}}, \frac{1}{\beta}Y_{n,j}\left(\frac{\cdot}{\beta}\right)\right\rangle_{\mathrm{L}^2(\mathcal{S})} \\ &= \left(\int_0^R r^{l_n+2} D_{n,j}(r)\,\mathrm{d}r\right)\frac{4\pi c_n}{(2n+1)\beta^n} = V_{n,j}\left(\frac{R}{\beta}\right)^n \end{aligned} \tag{12}$$

the spherical harmonics coefficients of V with respect to an orthonormal basis system on $\mathcal{S}$. Hence, we have a direct relation between the singular values of the Fredholm integral operator T and the spherical harmonic coefficients $V_{n,j}$. The additional factor $(\frac{R}{\beta})^n$ symbolizes the upward continuation from $\mathcal{S}$ to Ω_R. The upward continuation does not effect the null space of the operator T at all. Due to this property and the aim to keep the formulae simple, we analyze Equation (11) further and keep in mind that we can consequently deduce properties of T via Equation (12).

Note that (11) shows, in particular, the degree of freedom with respect to the radial part of D, since $V_{n,j}$ is some weighted radial mean of $D_{n,j}(r)$. On the other hand, one can expect a one-to-one relation for the angular dependence of V and D.

4. Investigation of the homogeneous problem

In order to obtain a unique solution, an appropriate modeling is required, that is, the solution space has to be restricted by certain constraints. Before this can be done (in Section 5), we have to study the null space $\ker T$, that is, the space of all D with $TD = 0$. Note that, due to the linearity, all solutions of $TD = V$ are given by $\tilde{D} + D_0$, with arbitrary $D_0 \in \ker T$, for a particular solution $\tilde{D}$ of $TD = V$.

4.1. Some orthonormal basis functions on the ball

It is our aim to characterize the null space, that is, the so-called kernel of the Fredholm integral operator of the first kind, in order to describe the non-reconstructable parts of the solution. For the separation of $\mathrm{L}^2(\mathcal{B})$ into the null space and the orthogonal complement we need an appropriate basis for $\mathrm{L}^2(\mathcal{B})$.

For the ball, there are several known basis systems available. For the construction of these systems see, for example, [1, 7, 17, 30, 32, 51]. We analogously use the idea to combine an orthonormal basis system on the unit sphere with one on the interval $[0, R]$, to construct a basis system on the ball. The $\mathrm{L}^2(\mathcal{B})$-orthonormal system used here is a generalization of the system which was introduced in [17] and [7].

For $x \in \mathcal{B} \setminus \{0\}$, it is given by

$$G_{m,n,j}(x) \coloneqq \gamma_{m,n} P_m^{(0,l_n+1/2)}\left(2\frac{|x|^2}{R^2} - 1\right) \frac{|x|^{l_n}}{R^{l_n}} Y_{n,j}\left(\frac{x}{|x|}\right), \tag{13}$$

with $m, n \in \mathbb{N}_0$, $j = 1, \ldots, 2n+1$, where $\{P_m^{(\alpha,\beta)}\}_{m\in\mathbb{N}_0}$ are the Jacobi polynomials and $\gamma_{m,n}$ are normalization constants with

$$\gamma_{m,n} \coloneqq \sqrt{\frac{4m + 2l_n + 3}{R^3}}\,. \tag{14}$$

Since $\alpha = 0$ in Equation (13) and $P_m^{(0,l_n+1/2)}(1) = 1$ for all $m, n \in \mathbb{N}_0$, we get $G_{m,n,j}|_{\Omega_R} = \gamma_{m,n} Y_{n,j}(\frac{\cdot}{R})$.

The functions in (13) were called $G^{\mathrm{I}}_{m,n,j}$ in [31] and [32] in the case of $l_n = n$ (remember that this setting corresponds to the inverse gravimetric problem).

A continuous expansion of our functions $G_{m,n,j}$ on the domain $\mathcal{B}$ is possible, if all exponents l_n, $n \in \mathbb{N}$, are positive. Otherwise we obtain a singularity at the origin of the functions $G_{m,n,j}$ for negative values of l_n and a discontinuity at the very same place in the case $l_n = 0$ for $n > 0$. For the theory stated in this paper, this is not a problem, since $G_{m,n,j}$ remains square-integrable for $-1 \le l_n$ (as we required).

As we claimed above, the functions $G_{m,n,j}$ for $m, n \in \mathbb{N}_0$, $j = 1, \ldots, 2n+1$ given in (13) build an orthonormal basis for $\mathrm{L}^2(\mathcal{B})$. This property can easily be verified by calculating the inner products and using a formula for a weighted L^2-norm of Jacobi polynomials (see, e.g., [36]). With the $\mathrm{L}^2(\Omega)$-orthogonality of the spherical harmonics and the substitution $r = R\sqrt{(1+z)/2}$, we obtain

$$\begin{aligned}
&\langle G_{m,n,j}, G_{\mu,\nu,\iota}\rangle_{\mathrm{L}^2(\mathcal{B})} \\
&\quad = \gamma_{m,n}\gamma_{\mu,\nu}\delta_{\nu,n}\delta_{\iota,j} \int_0^R \frac{r^{2l_n+2}}{R^{2l_n}} P_m^{(0,l_n+1/2)}\left(2\frac{r^2}{R^2} - 1\right) P_\mu^{(0,l_n+1/2)}\left(2\frac{r^2}{R^2} - 1\right) \mathrm{d}r \\
&\quad = \gamma_{m,n}\gamma_{\mu,n}\delta_{\nu,n}\delta_{\iota,j} \frac{R^3}{2^{l_n+5/2}} \int_{-1}^1 (1+z)^{l_n+1/2} P_m^{(0,l_n+1/2)}(z) P_\mu^{(0,l_n+1/2)}(z)\,\mathrm{d}z
\end{aligned}$$

$$= \gamma_{m,n}\gamma_{\mu,n}\delta_{\nu,n}\delta_{\iota,j}\frac{R^3}{2^{l_n+5/2}}\frac{2^{l_n+3/2}}{2m+l_n+3/2}\,\delta_{\mu,m}$$
$$= \gamma_{m,n}^2\delta_{\nu,n}\delta_{\iota,j}\frac{R^3}{4m+2l_n+3}\,\delta_{\mu,m} = \delta_{\mu,m}\delta_{\nu,n}\delta_{\iota,j}.$$

Thus, the set $\{G_{m,n,j}\}_{m,n\in\mathbb{N}_0,j=1,\dots,2n+1}$ is $\mathrm{L}^2(\mathcal{B})$-orthonormal. Moreover, the spherical harmonics are complete in $\mathrm{L}^2(\Omega)$ and the Jacobi polynomials are complete with respect to the inner product in (1) such that the system

$$\{G_{m,n,j}\}_{m,n\in\mathbb{N}_0,j=1,\dots,2n+1}$$

is complete in $\mathrm{L}^2(\mathcal{B})$ and constitutes an orthonormal basis.

Some of the functions $G^{\mathrm{I}}_{m,n,j}$ (i.e., in the case of $l_n = n$) are shown in Figures 1 and 2. For $m = 0$, the functions $G^{\mathrm{I}}_{0,n,j}$ are inner harmonics, hence they are harmonic, and attain their maximum and minimum on the boundary. A selection of the functions corresponding to the inverse magnetic problem, where $l_n = n-1$, is shown in Figures 3 and 4. The singularity (for $n = 0$, i.e., $l_0 = -1$) at the origin is visible in Figures 3 (A) and (C) and Figure 4 (B).

4.2. Splitting the basis into the null space and its complement

With the orthonormal basis introduced in Subsection 4.1, we are now able to expand the functions $D_{n,j}$ in (8) for all $n \in \mathbb{N}_0$ and $j = 1,\dots,2n+1$ and we obtain

$$D_{n,j}(r) = \frac{r^{l_n}}{R^{l_n}}\sum_{m=0}^{\infty} d_{m,n,j}\gamma_{m,n}P_m^{(0,l_n+1/2)}\left(2\frac{r^2}{R^2}-1\right), \tag{15}$$

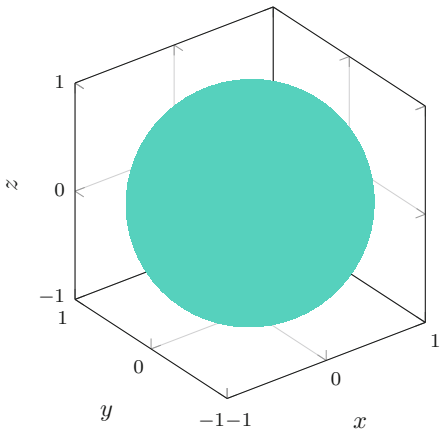
where $d_{m,n,j} \coloneqq \langle D, G_{m,n,j}\rangle_{\mathrm{L}^2(\mathcal{B})}$ and $\gamma_{m,n}$ is given in (14).

For further investigations of the forward problem, we use the representation of (the known function) V in (9), where we have already calculated the integral over the angular part. For the remaining integral over the radial part, we use the precise representation of $D_{n,j}$ in (15) and the orthogonality of the Jacobi polynomials.With the substitution $r = R\sqrt{(1+z)/2}$, $\mathrm{d}r = \frac{R}{4}\left(\frac{2}{1+z}\right)^{1/2}\mathrm{d}z$, we get

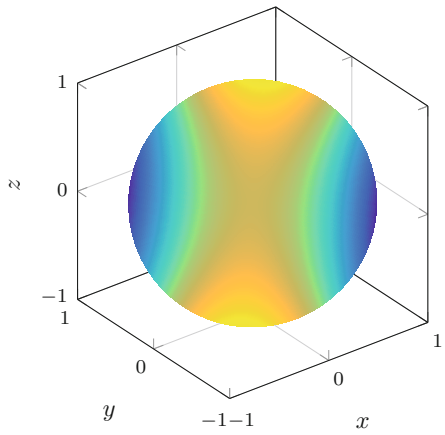
$$\int_0^R r^{l_n+2}D_{n,j}(r)\,\mathrm{d}r = \int_0^R \frac{r^{2l_n+2}}{R^{l_n}}\sum_{m=0}^{\infty} d_{m,n,j}\gamma_{m,n}P_m^{(0,l_n+1/2)}\left(2\frac{r^2}{R^2}-1\right)\mathrm{d}r$$
$$= \frac{R^{3+l_n}}{2^{l_n+5/2}}\sum_{m=0}^{\infty} d_{m,n,j}\gamma_{m,n}\int_{-1}^{1}(1+z)^{l_n+1/2}\,P_m^{(0,l_n+1/2)}(z)\,\mathrm{d}z = \frac{R^{3+l_n}}{2l_n+3}d_{0,n,j}\gamma_{0,n}.$$

Inserting the latter result in (9), we eventually obtain (remember the definition of $\gamma_{m,n}$ in (14))

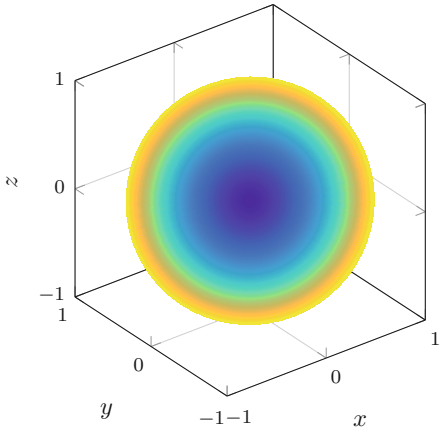
$$V(y) = \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\left(\int_0^R r^{l_n+2}D_{n,j}(r)\,\mathrm{d}r\right)\frac{4\pi c_n}{(2n+1)\,|y|^{n+1}}Y_{n,j}\left(\frac{y}{|y|}\right)$$



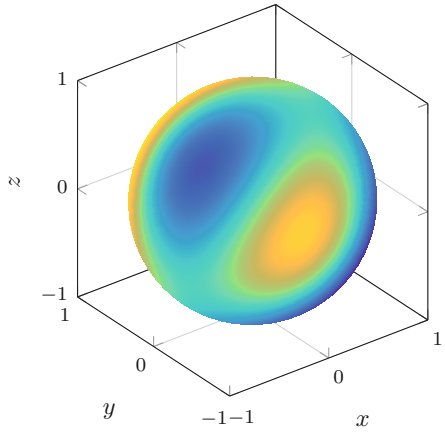
(A) $G^{\mathrm{I}}_{m,n,j}$ for $m = 0$, $n = 0$, $j = 1$



(B) $G^{\mathrm{I}}_{m,n,j}$ for $m = 0$, $n = 2$, $j = 5$



(C) $G^{\mathrm{I}}_{m,n,j}$ for $m = 1$, $n = 0$, $j = 1$

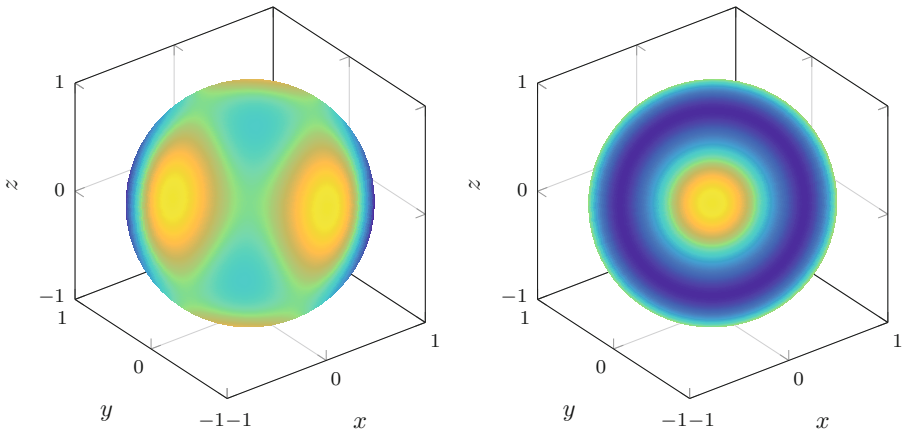


(D) $G^{\mathrm{I}}_{m,n,j}$ for $m = 1$, $n = 1$, $j = 3$

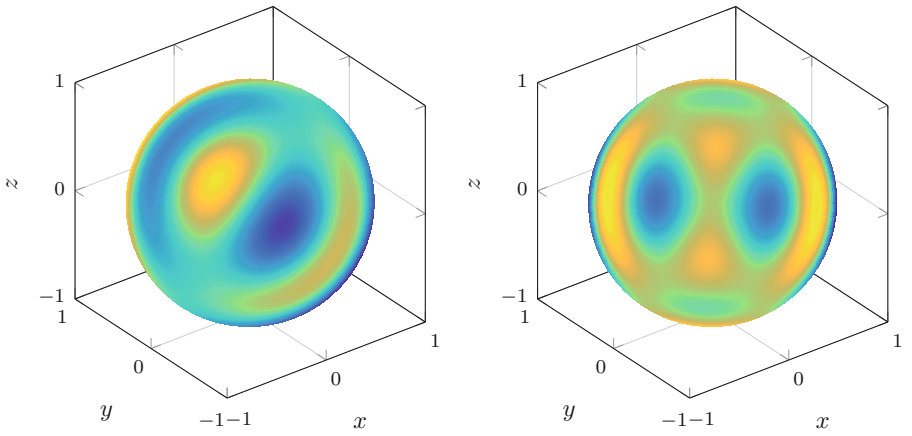
FIGURE 1. The functions $G_{m,n,j}$ in the case $l_n = n$ (also called $G^{\mathrm{I}}_{m,n,j}$) for different parameters m, n, j are plotted at the plane through the origin with normal vector $(1, 1, -1)^{\mathrm{T}}$. For the particular parameters, see the respective caption. The maximum is always yellow and the minimum is blue (see also [32, 34]).

$$= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \frac{R^{3+l_n} 4\pi c_n}{(2l_n+3)(2n+1)|y|^{n+1}} d_{0,n,j} \gamma_{0,n} Y_{n,j}\left(\frac{y}{|y|}\right)$$

$$= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \frac{4\pi c_n R^{l_n}}{(2n+1)\,|y|^{n+1}} d_{0,n,j} \gamma_{0,n}^{-1} Y_{n,j}\left(\frac{y}{|y|}\right). \tag{16}$$



(A) $G^{\mathrm{I}}_{m,n,j}$ for $m = 1$, $n = 2$, $j = 5$ (B) $G^{\mathrm{I}}_{m,n,j}$ for $m = 2$, $n = 0$, $j = 1$

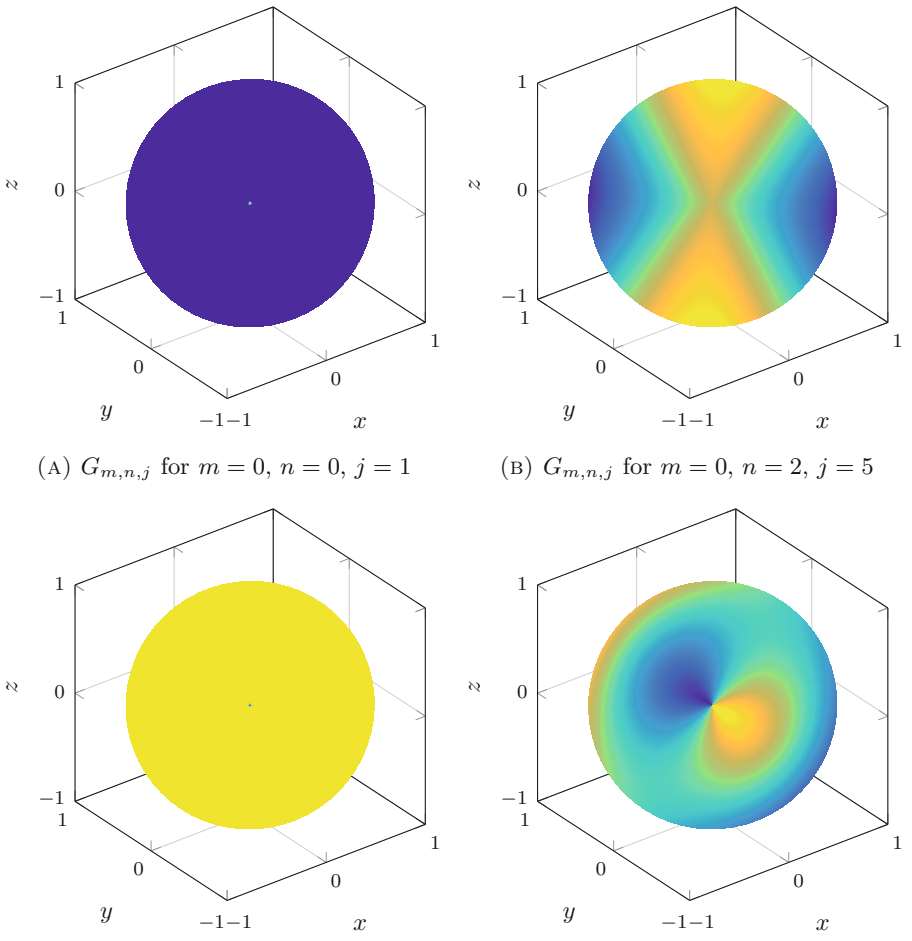


(C) $G^{\mathrm{I}}_{m,n,j}$ for $m = 2$, $n = 1$, $j = 3$ (D) $G^{\mathrm{I}}_{m,n,j}$ for $m = 2$, $n = 2$, $j = 5$

FIGURE 2. The functions $G_{m,n,j}$ in the case $l_n = n$ (also called $G^{\mathrm{I}}_{m,n,j}$) for different parameters m, n, j are plotted at the plane through the origin with normal vector $(1, 1, -1)^{\mathrm{T}}$. For the particular parameters, see the respective caption. The maximum is always yellow and the minimum is blue (see also [32, 34]).

Hence, $G_{m,n,j}$ is in the null space of the operator T with the kernel from (3), if and only if $m > 0$ or $c_n = 0$. Examples of functions in the null space are given in Figures 2 and 4 (for different inverse problems). The function plotted in Figure 5 is not in the null space.

Since $\mathrm{L}^2(\mathcal{B})$ is the direct sum of the null space $\ker T$ and its orthogonal complement, the obtained result allows a precise characterization of the null space



(A) $G_{m,n,j}$ for $m = 0$, $n = 0$, $j = 1$

(B) $G_{m,n,j}$ for $m = 0$, $n = 2$, $j = 5$

(C) $G_{m,n,j}$ for $m = 1$, $n = 0$, $j = 1$

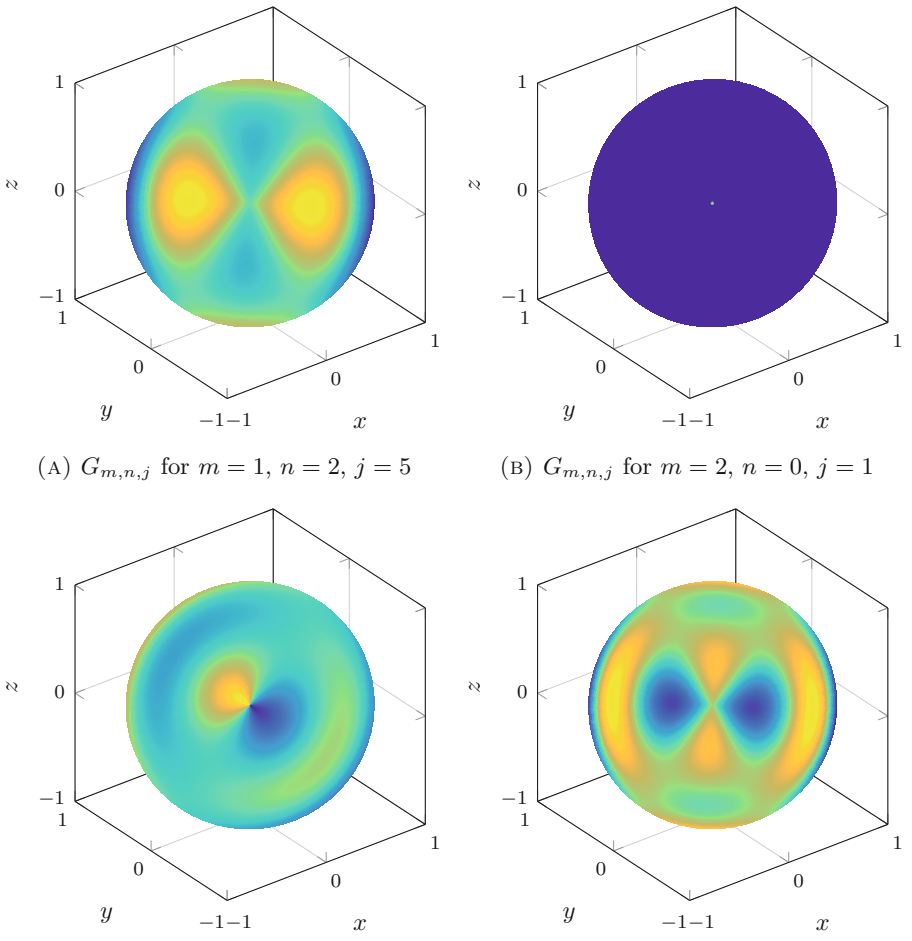
(D) $G_{m,n,j}$ for $m = 1$, $n = 1$, $j = 3$

FIGURE 3. The functions $G_{m,n,j}$ in the case $l_n = n - 1$ for different parameters m, n, j are plotted at the plane through the origin with normal vector $(1, 1, -1)^{\mathrm{T}}$. For the particular parameters, see the respective caption. The maximum is always yellow and the minimum is blue (see also [34]).

of the corresponding Fredholm integral operator as

$$\ker T = \overline{\operatorname{span}\left\{G_{m,n,j} \;\middle|\; m \geq 1, n \in \mathbb{N}_0, j = 1, \ldots, 2n+1 \text{ or } c_n = 0\right\}}^{\|\cdot\|_{\mathrm{L}^2(\mathcal{B})}}. \quad (17)$$

For the inverse gravimetric problem ($l_n = n$), we can deduce the well-known fact that the null space can be described as the set of all anharmonic functions, which are the elements of the orthogonal complement of the set of all harmonic



(A) $G_{m,n,j}$ for $m = 1$, $n = 2$, $j = 5$

(B) $G_{m,n,j}$ for $m = 2$, $n = 0$, $j = 1$

(C) $G_{m,n,j}$ for $m = 2$, $n = 1$, $j = 3$

(D) $G_{m,n,j}$ for $m = 2$, $n = 2$, $j = 5$

FIGURE 4. The functions $G_{m,n,j}$ in the case $l_n = n - 1$ for different parameters m, n, j are plotted at the plane through the origin with normal vector $(1, 1, -1)^{\mathrm{T}}$. For the particular parameters, see the respective caption. The maximum is always yellow and the minimum is blue (see also [34]).

functions. That is,

$$\ker T^{\mathrm{G}} = \overline{\operatorname{span}\left\{ G^{\mathrm{I}}_{m,n,j} \;\middle|\; m \geq 1 \right\}}^{\|\cdot\|_{\mathrm{L}^2(\mathcal{B})}} = \left\{ F \in \mathrm{C}^{(2)}(\mathcal{B}) \;\middle|\; \Delta F = 0 \right\}^{\perp_{\mathrm{L}^2(\mathcal{B})}},$$

where $\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}$ represents the Laplace operator. In this case, the functions $G^{\mathrm{I}}_{0,n,j}$, $n \in \mathbb{N}_0$, $j = 1, \ldots, 2n+1$ are the inner harmonics and, therefore,

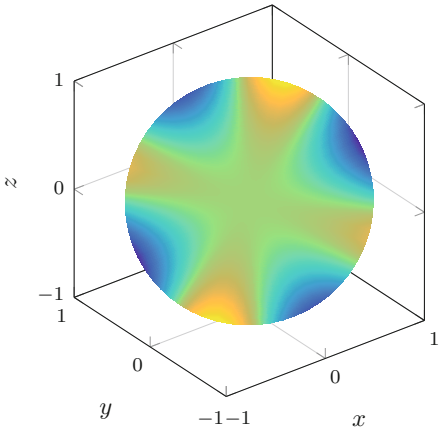


FIGURE 5. The function $G_{0,4,8}$, which is not in the null space of the Fredholm integral operator T for $l_n = n - 1$.

form a basis for the set of all harmonic functions on the ball:

$$G^{\mathrm{I}}_{0,n,j}(x) = \sqrt{\frac{2n+3}{R}}\frac{|x|^n}{R^{n+1}}Y_{n,j}\left(\frac{x}{|x|}\right), \qquad x \in \mathcal{B}.$$

For some particular cases of the considered Fredholm integral operators, we are also able to find a characterization of the null space via an elliptic partial differential equation.

For this purpose, we consider the particular integral kernel

$$k(x,y) := \sum_{i=0}^{\infty} c_i \frac{|x|^{i+\kappa}}{|y|^{i+1}} P_i\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right), \qquad (x,y) \in \mathrm{dom}(k),$$

for a fixed $\kappa \in [-1,\infty)$ and $c_i \neq 0$ for all $i \in \mathbb{N}_0$. Note that in the case of the inverse gravimetric problem $\kappa = 0$ and in the case of the inverse magnetic problem $\kappa = -1$. We have already proven that the orthogonal complement of the null space of the corresponding operator T is given by the set

$$(\ker T)^{\perp_{\mathrm{L}^2(\mathcal{B})}} = \overline{\mathrm{span}\left\{G_{m,n,j} \;\middle|\; m = 0, n \in \mathbb{N}_0, j = 1,\dots,2n+1 \text{ and } c_n \neq 0\right\}}^{\|\cdot\|_{\mathrm{L}^2(\mathcal{B})}}.$$

Now, we define an elliptic partial differential operator $\tilde{\Delta}$ by

$$\tilde{\Delta}F(r\xi) := \Delta\left(r^{-\kappa}F(r\xi)\right) = \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta^*_\xi\right)\left(r^{-\kappa}F(r\xi)\right).$$

Using the product rule for the derivative, we get

$$\begin{aligned}\tilde{\Delta}F(r\xi) &= \left(-\kappa(-\kappa-1)r^{-\kappa-2} - 2\kappa r^{-\kappa-1}\frac{\partial}{\partial r} + r^{-\kappa}\frac{\partial^2}{\partial r^2} - 2\kappa r^{-\kappa-2}\right.\\ &\qquad \left. +2r^{-\kappa-1}\frac{\partial}{\partial r} + r^{-\kappa-2}\Delta_\xi^*\right) F(r\xi)\\ &= \left(r^{-\kappa}\frac{\partial^2}{\partial r^2} + 2(1-\kappa)r^{-\kappa-1}\frac{\partial}{\partial r} + \kappa(\kappa-1)r^{-\kappa-2} + r^{-\kappa-2}\Delta_\xi^*\right) F(r\xi).\end{aligned}$$

In the particular case of the inverse gravimetric problem (i.e., $\kappa = 0$) this reduces to

$$\tilde{\Delta}F(r\xi) = \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta_\xi^*\right) F(r\xi) = \Delta F(r\xi),$$

and the differential operator corresponding to the inverse magnetic problem (i.e., $\kappa = -1$) is given by

$$\tilde{\Delta}F(r\xi) = \left(r\frac{\partial^2}{\partial r^2} + 4\frac{\partial}{\partial r} + \frac{2}{r} + \frac{1}{r}\Delta_\xi^*\right) F(r\xi).$$

In order to get a new characterization of the null space, we apply the differential operator to the basis functions $G_{0,n,j}$ for $n \in \mathbb{N}_0$, $j = 1, \ldots, 2n+1$ and obtain

$$\begin{aligned}\tilde{\Delta}G_{0,n,j}(r\xi) &= \tilde{\Delta}\left(\gamma_{0,n}\left(\frac{r}{R}\right)^{n+\kappa} Y_{n,j}(\xi)\right) = \gamma_{0,n}\Delta\left(r^{-\kappa}\left(\frac{r}{R}\right)^{n+\kappa} Y_{n,j}(\xi)\right)\\ &= \frac{\gamma_{0,n}}{R^{n+\kappa}}\,\Delta\,(r^n Y_{n,j}(\xi)) = 0,\end{aligned}$$

since the mapping $r\xi \mapsto r^n Y_{n,j}(\xi)$ is a harmonic function for all $n \in \mathbb{N}_0$, $j = 1, \ldots, 2n+1$. In analogy, $\tilde{\Delta}G_{m,n,j} \not\equiv 0$ for $m \geq 1$, $n \in \mathbb{N}_0$, $j = 1, \ldots, 2n+1$ follows by similar considerations. This means that $\tilde{\Delta}F$ is equal to zero if and only if $r\xi \mapsto r^{-\kappa}F(r\xi)$ is a harmonic function, that is, is contained in

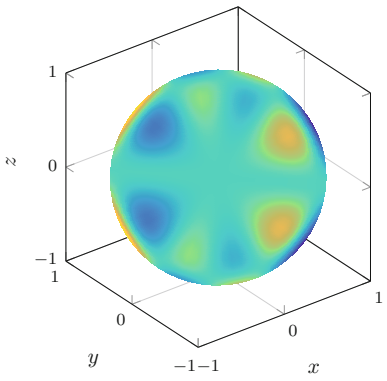
$$\operatorname{span}\left\{G^{\mathrm{I}}_{0,n,j}\right\}_{n\in\mathbb{N}_0, j=1\ldots,2n+1}.$$

Since this is equivalent to expanding $F(r\xi)$ in terms of $r^\kappa G^{\mathrm{I}}_{0,n,j}(r\xi)$ and $l_n = n+\kappa$ here, our definition in (13) leads us to the following result.

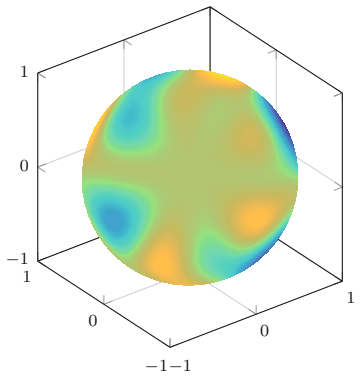
Theorem 4.1. *If we assume that there exists a fixed parameter $\kappa \geq -1$ such that $l_n = n + \kappa$ for all $n \in \mathbb{N}_0$ and that $c_n \neq 0$ for all $n \in \mathbb{N}_0$, then*

$$\begin{aligned}\ker T &= \overline{\operatorname{span}\{G_{m,n,j} \mid m > 0, n \in \mathbb{N}_0, j = 1, \ldots, 2n+1\}}^{\|\cdot\|_{\mathrm{L}^2(\mathcal{B})}}\\ &= \left\{F\colon \mathcal{B} \to \mathbb{R} \;\middle|\; (r\xi \mapsto r^{-\kappa}F(r\xi)) \in \mathrm{C}^{(2)}(\mathcal{B}) \text{ and } \tilde{\Delta}F = 0\right\}^{\perp_{\mathrm{L}^2(\mathcal{B})}}. \qquad (18)\end{aligned}$$

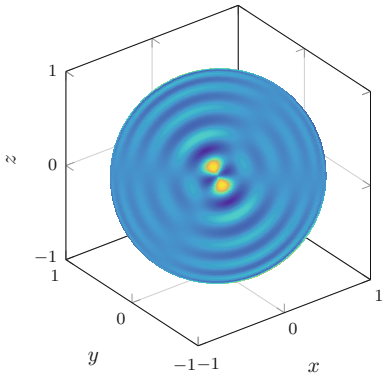
After having given two mathematical characterizations of the null space in Equation (18) for a particular case (i.e., $l_n = n + \kappa$) and one characterization in Equation (17) for the general case, we want to demonstrate what kind of functions D generate the same forward solution V.



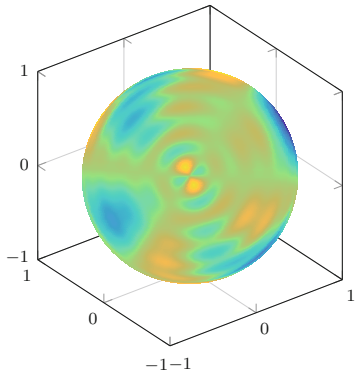
(A) $G_{1,5,10}$



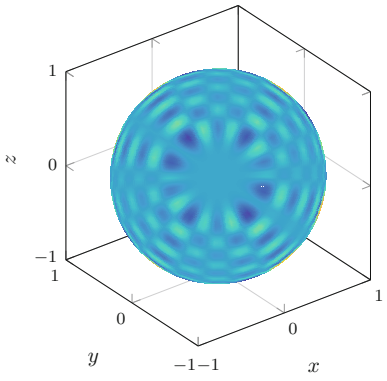
(B) Sum of the function in Fig. 5 and (A)



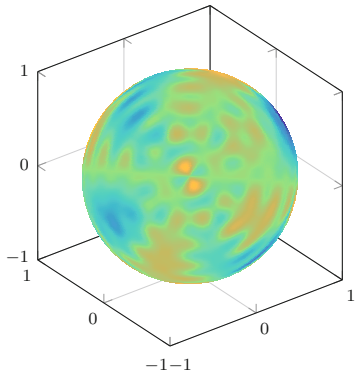
(C) $G_{10,2,4}$



(D) Sum of the function in (B) and (C)

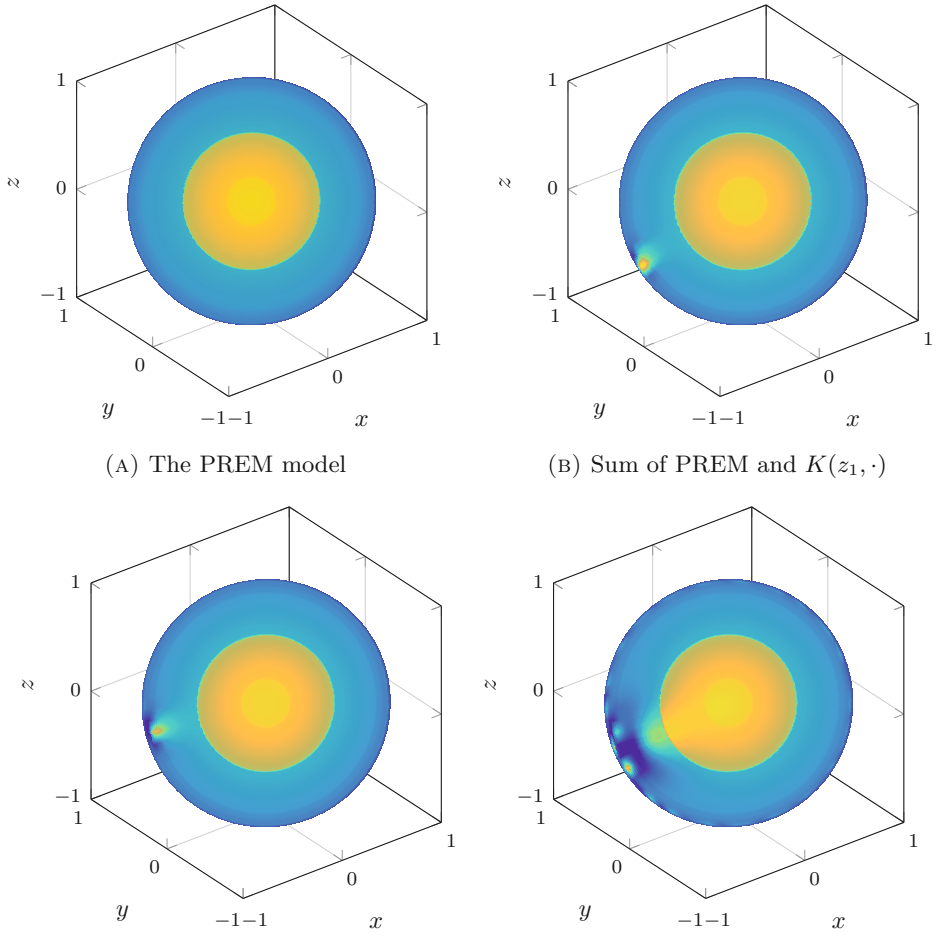


(E) $G_{7,10,4}$



(F) Sum of the function in (D) and (E)

FIGURE 6. Several functions from the null space of T, that is, they generate the solution $V = 0$ (left column), and the sum of these functions with $G_{0,4,8} \notin \ker T$ (right column) which generate the same right-hand side $V = TG_{0,4,8}$, that is, the same data for the inverse problem.



(A) The PREM model

(B) Sum of PREM and $K(z_1, \cdot)$

(C) Sum of PREM and $K(z_2, \cdot)$

(D) Sum of PREM and $K(z_1, \cdot)+K(z_2, \cdot)+K(z_3, \cdot)$

FIGURE 7. The density of the PREM model added to several functions from the null space of T^{G}. They all generate the same gravitational potential. Here, $z_i \in \mathcal{B}$, $i = 1, 2, 3$ are fixed.

For this purpose, we consider the function $G_{0,4,8}$ plotted in Figure 5, which is not in the null space of the operator, that means this function generates the result $TG_{0,4,8} = V \neq 0$. Then, we add several functions from the null space (see Figures 6 (A), (C), and (E)) to $G_{0,4,8}$. The results are shown in Figures 6 (B), (D), and (F). Keep in mind that all functions in the left column of Figure 6 generate the zero potential and all functions in the right column of Figure 6 generate the same forward solution $V = TG_{0,4,8}$. Similarly, we proceed in Figure 7, where linear

combinations of functions $K(z_i, \cdot)$, $z_i \in \mathcal{B} \setminus \{0\}$, with

$$K(z_i, x) := \sum_{n=0}^{100} \sum_{j=1}^{2n+1} (0.95)^{1+n} G_{1,n,j}(x) G_{1,n,j}(z_i), \qquad x \in \mathcal{B} \setminus \{0\},$$

are added to the density D of the PREM model, see [18]. Again, $K(z_i, \cdot)$ can be extended onto $\mathcal{B}$, if $l_n \geq 0$ for all $n \in \mathbb{N}_0$. Note that $K(z_i, \cdot) \in \ker T$ for all $z_i \in \mathcal{B} \setminus \{0\}$ such that, again, there is no difference between the potentials generated by PREM (see Figure 7 (A)) and the potentials generated by the perturbed mass densities in Figures 7 (B), (C), and (D).

Hence, the solution of the inverse problem from Equation (2) is not unique, since we can always add functions from the null space to it without changing the function V. In particular, Figure 7 shows that certain kinds of mass anomalies (in the interior of the Earth) remain completely concealed if gravitational data are used solely.

Now we can sum up our results and give an answer to the three questions about the well-posedness of the problem posed in Section 3.

Theorem 4.2. *Let the operator* $T \colon \mathrm{L}^2(\mathcal{B}) \to \mathrm{L}^2(\mathcal{S})$ *be given by*

$$T \colon D \mapsto \int_{\mathcal{B}} D(x) k(x, \cdot) \, \mathrm{d}x. \tag{19}$$

with an integral kernel $k \colon \mathcal{B} \times \mathcal{S} \to \mathbb{R}$ *of the form*

$$k(x, y) := \sum_{i=0}^{\infty} c_i \frac{|x|^{l_i}}{|y|^{i+1}} P_i \left(\frac{x}{|x|} \cdot \frac{y}{|y|} \right), \qquad x \in \mathcal{B} \setminus \{0\}, y \in \mathcal{S},$$

satisfying Assumption 3.1. *Moreover, let the following three conditions be fulfilled (by the function* V*):*

- *The restriction* $V|_{\Omega_R}$ *of* V *is an* $\mathrm{L}^2(\Omega_R)$*-function.*
- *The spherical harmonics coefficients* $V_{n,j}$ *of* V *fulfil a summability condition*

$$\sum_{\substack{n=0 \\ c_n \neq 0}}^{\infty} n^2 (2l_n + 3) R^{2n-2l_n} c_n^{-2} \sum_{j=1}^{2n+1} V_{n,j}^2 < \infty.$$

- *The function* V *is harmonic in the exterior of* $\mathcal{B}$*, that is,* $\Delta V(y) = 0$ *for all* $y \in \mathbb{R}^3 \setminus \mathcal{B}$*, and regular at infinity, that is,* $|V(y)| = \mathcal{O}(|y|^{-1})$ *and* $|\nabla V(y)| = \mathcal{O}(|y|^{-2})$ *for* $|y| \to \infty$*.*

Then both inverse problems, which are, the recovery of $D \in \mathrm{L}^2(\mathcal{B})$ *from either given values of* $V|_{\Omega_R}$ *or the upward continued potential* $V|_{\mathcal{S}}$ *are ill posed, since their solutions are not unique. However, in both cases, the solution exists under these conditions but is not stable.*

The second condition in Theorem 4.2 is also known as the Picard condition.In several cases, for example, the inverse gravimetric problem (i.e., $l_n = n$, $c_n = \gamma$ for all $n \in \mathbb{N}_0$), the Picard condition implies $V|_{\Omega_R} \in \mathrm{L}^2(\Omega_R)$. For the inverse

gravimetric problem the Picard condition is satisfied, since the (empirical) Kaula rule of thumb holds:

$$\sum_{j=1}^{2n+1} \langle V|_{\Omega_R}, Y_{n,j}\rangle^2_{\mathrm{L}^2(\Omega_R)} = \mathcal{O}(\vartheta^{n+1} n^{-3}), \qquad n \to \infty,$$

for a constant $\vartheta \in]0,1[$, see, for example, [28] or [44]. Note that the Picard condition is necessary for the existence of the solution. Since this condition is not necessarily satisfied by every $V|_{\Omega_R} \in \mathrm{L}^2(\Omega_R)$, also this criterion by Hadamard may be violated.

We want to discuss the instability of the solution in detail using the following example.

Example 4.3. Let a family of functions be defined by

$$V_n(y) \coloneqq \frac{1}{n} \frac{\beta^n}{|y|^{n+1}} Y_{n,1}\left(\frac{y}{|y|}\right), \qquad y \in \mathcal{S}, \text{ for all } n \in \mathbb{N}_0.$$

Since $\{\frac{1}{\beta} Y_{n,1}(\frac{\cdot}{\beta})\}_{n\in\mathbb{N}_0}$ is an $\mathrm{L}^2(\mathcal{S})$-orthonormal system, we get

$$\|V_n\|_{\mathrm{L}^2(\mathcal{S})} = \frac{1}{n} \to 0 \quad \text{as } n \to \infty.$$

Hence, the norms build a null sequence. Using Equation (16), we see that

$$D_n(x) \coloneqq \frac{\sqrt{2n+3}\,(2n+1)}{4\pi R^{3/2} n} \left(\frac{\beta}{R}\right)^n G_{0,n,1}(x)$$

yields $TD_n = V_n$ in the case of $l_n = n$, $c_n = 1$. In addition, we obtain that the sequence of norms diverges, since $\beta > R$ and

$$\begin{aligned} \|D_n\|_{\mathrm{L}^2(\mathcal{B})} &= \frac{\sqrt{2n+3}(2n+1)}{4\pi R^{3/2} n} \left(\frac{\beta}{R}\right)^n \|G_{0,n,1}\|_{\mathrm{L}^2(\mathcal{B})} \\ &= \frac{\sqrt{2n+3}(2n+1)}{4\pi R^{3/2} n} \left(\frac{\beta}{R}\right)^n \to \infty \quad \text{as } n \to \infty. \end{aligned}$$

Thus, small changes in the potential V yield large changes in the solution D and, hence, the problem is not stable. Note that this instability is already given for the case of terrestrial data, which means that it is not (only) caused by the instability of the downward continuation.

4.3. Expansion of the solution in reproducing kernel based functions

In certain cases, it can be of interest to expand the unknown function D in terms of appropriate reproducing kernels instead of orthonormal basis functions. Reproducing kernels are localized in contrast to the global orthonormal basis functions from the previous subsection (see also the paper by Freeden, Michel and Simons in this handbook). In addition, the problems due to the discontinuity at the origin can be avoided by using this approach. For a more general introduction into reproducing kernels and reproducing kernel Hilbert spaces, see, for a general setting, [2, 3, 5, 16], for reproducing kernel Hilbert spaces on the ball.

Let $\mathcal{H} := \mathcal{H}((A_{m,n}), \mathcal{B}) \subset \mathrm{L}^2(\mathcal{B})$, with the real sequence $(A_{m,n})_{m,n\in\mathbb{N}_0}$, be defined as

$$\mathcal{H}((A_{m,n}), \mathcal{B}) := \overline{\left\{ F \in \mathrm{L}^2(\mathcal{B}) \;\middle|\; \sum_{m,n=0}^{\infty} \sum_{j=1}^{2n+1} A_{m,n}^2 \langle F, G_{m,n,j}\rangle_{\mathrm{L}^2(\mathcal{B})}^2 < \infty \right\}}^{\|\cdot\|_{\mathcal{H}}},$$

with

$$\|F\|_{\mathcal{H}}^2 := \sum_{m,n=0}^{\infty} \sum_{j=1}^{2n+1} A_{m,n}^2 \langle F, G_{m,n,j}\rangle_{\mathrm{L}^2(\mathcal{B})}^2, \qquad F \in \mathcal{H}.$$

The inner product in $\mathcal{H}$ is then given by

$$\langle F, G\rangle_{\mathcal{H}} = \sum_{m,n=0}^{\infty} \sum_{j=1}^{2n+1} A_{m,n}^2 \langle F, G_{m,n,j}\rangle_{\mathrm{L}^2(\mathcal{B})} \langle G, G_{m,n,j}\rangle_{\mathrm{L}^2(\mathcal{B})} \tag{20}$$

for all F, $G \in \mathcal{H}$.

If the sequence $(A_{m,n})_{m,n\in\mathbb{N}_0}$ fulfils a certain summability condition, see, for more details, [34], then $\mathcal{H}$ is a reproducing kernel Hilbert space. Due to the property of the sequence $(A_{m,n})_{m,n\in\mathbb{N}_0}$, the evaluation functional in $\mathcal{H}$ is continuous. The reproducing kernel of $\mathcal{H}$ is given by $K\colon (\mathcal{B}\setminus\{0\}) \times (\mathcal{B}\setminus\{0\}) \to \mathbb{R}$ with

$$K(z, x) := \sum_{\substack{m,n=0;\\ A_{m,n}\neq 0}}^{\infty} \sum_{j=1}^{2n+1} A_{m,n}^{-2} G_{m,n,j}(x) G_{m,n,j}(z), \qquad z,\ x \in \mathcal{B}\setminus\{0\}. \tag{21}$$

Again, in certain cases of $G_{m,n,j}$, the definition of K on $\mathcal{B}\times\mathcal{B}$ is valid.

The kernel K has the reproducing property, that is,

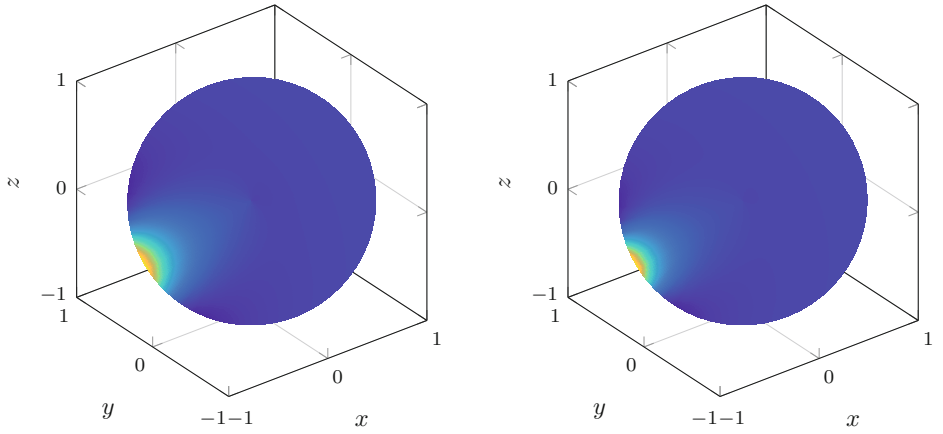
$$\langle F, K(z,\cdot)\rangle_{\mathcal{H}} = F(z) \qquad \text{for all } F \in \mathcal{H} \text{ and all } z \in \mathcal{B}\setminus\{0\}.$$

In our setting, the first input argument z denotes the (fixed) centre of the kernel, that is, the position in the ball where the kernel is located. Some examples of reproducing kernels with the same centre and different sequences $(A_{m,n})_{m,n\in\mathbb{N}_0}$ are plotted in Figure 8. As one can see, the discontinuity at the origin is, at least visibly, smoothed away.

Let the set $\{y_1, \ldots, y_\ell\} \subset \mathcal{S}$, $\ell \in \mathbb{N}$, contain our measuring positions. We define linear functionals by $\mathcal{F}^\nu F := (TF)(y_\nu)$ for $\nu = 1, \ldots, \ell$. In other words, the functionals $\mathcal{F}^\nu$ are the evaluations of our operator T applied to an (unknown) function F at the measuring positions y_ν, $\nu = 1, \ldots, \ell$. The data collected at the sensor positions are given by $v_\nu = V(y_\nu)$. The functionals $\mathcal{F}^\nu$ are linear, since they are the composition of the linear operator T and the linear evaluation functional.

If the function F is an element of the Sobolev space $\mathcal{H}((A_{m,n}), \mathcal{B})$ with a sequence $(A_{m,n})_{m,n\in\mathbb{N}_0}$ fulfilling the summability condition

$$\sum_{n=0}^{\infty} A_{0,n}^{-2} \frac{1}{(2n+1)(2l_n+3)} < \infty,$$



(A) $h = 0.75$ (B) $h = 0.8$

(C) $h = 0.9$ (D) $h = 0.92$

FIGURE 8. Reproducing kernel $K(z_1, \cdot)$ for several $(A_{m,n})_{m,n\in\mathbb{N}_0}$ with $A_{m,n}^{-2} = (Cn+1)h^{2(m+n)}\delta_{m,0}$ at a fixed centre $z_1 \in \mathcal{B}\backslash\{0\}$, a sufficiently large constant C, and the functions $G_{m,n,j}$ in the case $l_n = n-1$.

then the functionals $\mathcal{F}^\nu$ are also continuous (with $y_\nu = r_\nu\xi_\nu$) for $\nu = 1, \dots, \ell$, since

$$
\begin{aligned}
|\mathcal{F}^\nu F|^2 &= |(TF)(y_\nu)|^2 = \left|\left(T\left(\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\langle F, G_{m,n,j}\rangle_{\mathrm{L}^2(\mathcal{B})}G_{m,n,j}\right)\right)(y_\nu)\right|^2 \\
&= (4\pi)^2\left|\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\langle F, G_{0,n,j}\rangle_{\mathrm{L}^2(\mathcal{B})}\frac{c_n R^{l_n}}{(2n+1)|r_\nu|^{n+1}}\gamma_{0,n}^{-1}Y_{n,j}(\xi_\nu)\frac{A_{0,n}}{A_{0,n}}\right|^2 \\
&\le (4\pi)^2\left(\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\langle F, G_{0,n,j}\rangle_{\mathrm{L}^2(\mathcal{B})}^2 A_{0,n}^2\right)\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\left(\frac{c_n R^{l_n}}{A_{0,n}(2n+1)|r_\nu|^{n+1}}\gamma_{0,n}^{-1}Y_{n,j}(\xi_\nu)\right)^2
\end{aligned}
$$

$$\leq (4\pi c)^2 \|F\|_{\mathcal{H}}^2 \left(\sup_{n\in\mathbb{N}_0} R^{l_n-n} \right)^2 \sum_{n=0}^{\infty} \frac{R}{A_{0,n}^2 (2n+1)^2 (2l_n+3)} \frac{2n+1}{4\pi}$$

$$\leq 4\pi c^2 R \|F\|_{\mathcal{H}}^2 \left(\sup_{n\in\mathbb{N}_0} R^{l_n-n} \right)^2 \sum_{n=0}^{\infty} \frac{1}{A_{0,n}^2 (2n+1)(2l_n+3)} < \infty,$$

due to (16), the Cauchy–Schwarz inequality, the definition of the inner product in $\mathcal{H}$ in (20), and Assumption 3.1.

We can apply these functionals to the kernel with respect to z and obtain the following result by using Equation (16) and the addition theorem for spherical harmonics. The interchanging of limits (in the series) and the integral, which is needed in this calculation, is allowed due to the previous estimates. Hence,

$$\begin{aligned}
\mathcal{F}_z^{\nu} K(z,x) &= \left[\int_{\mathcal{B}} K(z,x) \sum_{i=0}^{\infty} c_i \frac{|z|^{l_i}}{|y|^{i+1}} P_i \left(\frac{z}{|z|} \cdot \frac{y}{|y|} \right) \mathrm{d}z \right]_{y=y_\nu} \\
&= \sum_{\substack{m,n=0;\\ A_{m,n}\neq 0}}^{\infty} \sum_{j=1}^{2n+1} A_{m,n}^{-2} G_{m,n,j}(x) \mathcal{F}_z^{\nu} G_{m,n,j}(z) \\
&= \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} A_{0,n}^{-2} \gamma_{0,n} \frac{|x|^{l_n}}{R^{l_n}} Y_{n,j}\left(\frac{x}{|x|} \right) \gamma_{0,n}^{-1} \frac{4\pi c_n R^{l_n}}{(2n+1)|y_\nu|^{n+1}} Y_{n,j}\left(\frac{y_\nu}{|y_\nu|} \right) \\
&= \sum_{n=0}^{\infty} A_{0,n}^{-2} c_n \frac{|x|^{l_n}}{|y_\nu|^{n+1}} P_n \left(\frac{x}{|x|} \cdot \frac{y_\nu}{|y_\nu|} \right).
\end{aligned}$$

It is known that we can construct an expansion for the solution, see [3, 23] by

$$D(x) = \sum_{\nu=1}^{\ell} a_\nu \mathcal{F}_z^{\nu} K(z,x). \tag{22}$$

Our aim is to determine the corresponding coefficients a_ν, $\nu = 1, \ldots, \ell$. Applying the functional on both sides, we obtain for $\iota = 1, \ldots, \ell$

$$\begin{aligned}
\mathcal{F}_x^{\iota} D(x) = v_\iota &= \sum_{\nu=1}^{\ell} a_\nu \mathcal{F}_x^{\iota} \mathcal{F}_z^{\nu} K(z,x) \\
&= \sum_{\nu=1}^{\ell} a_\nu \sum_{n=0}^{\infty} A_{0,n}^{-2} \gamma_{0,n}^{-2} \frac{4\pi c_n^2}{2n+1} \frac{R^{2l_n}}{|y_\iota|^{n+1}|y_\nu|^{n+1}} P_n \left(\frac{y_\iota}{|y_\iota|} \cdot \frac{y_\nu}{|y_\nu|} \right).
\end{aligned}$$

This linear system is uniquely solvable, which means that the expansion in (22) is unique, if the linear and continuous functionals $\mathcal{F}^{\nu}$, $\nu = 1, \ldots, \ell$ are linearly independent, see [2]. Among all solutions $D \in \mathcal{H}$ with $\mathcal{F}^{\nu} D = v_\nu$ for $\nu = 1, \ldots, \ell$, the solution in (22) uniquely minimizes the norm $\|\cdot\|_{\mathcal{H}}$ induced by the inner product in (20). These are basically the ideas of a spline interpolation method (for further details, see [2] and [9]).

5. Constraints for the uniqueness of the solution

In the previous section, we have shown that we cannot expect a unique solution of the Fredholm integral equation of the first kind stated in (2). Hence, in practice, additional conditions are necessary to impose uniqueness. Some possible uniqueness constraints are now discussed. The most approaches are generalizations of the results in [33]. More precisely, we present the minimum norm condition, a generalization of the harmonicity constraint, and the layer density constraint. In addition, we discuss the surface density approach.

5.1. Minimum norm constraint

As we have seen, we are not able to obtain a uniquely determined solution without additional assumptions or information. A widespread approach to force uniqueness is the minimum norm condition (see, e.g., [40]). The following result is a generalization of the theorem concerning the minimum norm solution of the inverse gravimetric problem, see [33] and the references therein. Throughout this subsection, we assume that the conditions in Theorem 4.2 are fulfilled and, hence, a solution of the inverse problem exists.

Recall Equation (11), which is repeated below for convenience:

$$\frac{(2n+1)R^n}{4\pi c_n} V_{n,j} = \int_0^R r^{l_n+2} D_{n,j}(r)\,\mathrm{d}r, \qquad \text{if } c_n \neq 0,$$

$V_{n,j} = 0$ for all $j = 1, \ldots, 2n+1$ otherwise. $D_{n,j}$ is originated by the (in $\mathrm{L}^2(\mathcal{B})$ convergent) series

$$D(x) = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} D_{n,j}(|x|) Y_{n,j}\left(\frac{x}{|x|}\right).$$

The minimum norm conditionis fulfilled, if among all $D \in \mathrm{L}^2(\mathcal{B})$ with $V = \int_{\mathcal{B}} D(x) k(x, \cdot)\,\mathrm{d}x$, we choose the one with the minimum (squared) norm

$$\|D\|^2_{\mathrm{L}^2(\mathcal{B})} = \int_{\mathcal{B}} (D(x))^2\,\mathrm{d}x = \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \int_0^R r^2 (D_{n,j}(r))^2\,\mathrm{d}r.$$

If we minimize this expression, we obtain the following minimization problem for each $n \in \mathbb{N}_0$ and $j = 1, \ldots, 2n+1$:

$$\begin{aligned} &\text{minimize} && \int_0^R r^2 (D_{n,j}(r))^2\,\mathrm{d}r, \\ &\text{subject to} && \int_0^R r^{l_n+2} D_{n,j}(r)\,\mathrm{d}r = \frac{2n+1}{4\pi c_n} R^n V_{n,j}, \qquad \text{if } c_n \neq 0. \end{aligned}$$

Note that the side condition drops out in the case $c_n = 0$ such that the unconstrained minimizer $D_{n,j} \equiv 0$ occurs. With the substitution $F_{n,j}(r) := r D_{n,j}(r)$,

the problem above is equivalent to

$$\begin{aligned} &\text{minimize} && \int_0^R (F_{n,j}(r))^2 \,\mathrm{d}r, \\ &\text{subject to} && \int_0^R r^{l_n+1} F_{n,j}(r)\,\mathrm{d}r = \frac{2n+1}{4\pi c_n} R^n V_{n,j}, \qquad \text{if } c_n \neq 0. \end{aligned}$$

We now apply an orthogonal decomposition in $\mathrm{L}^2[0,R]$ to $F_{n,j}$ in the sense that $F_{n,j}(r) = \alpha_{n,j} r^{l_n+1} + H_{n,j}(r)$, where $\int_0^R r^{l_n+1} H_{n,j}(r)\,\mathrm{d}r = 0$. With this ansatz, our minimization problem reads

$$\begin{aligned} &\text{minimize} && \alpha_{n,j}^2 \int_0^R r^{2l_n+2}\,\mathrm{d}r + \|H_{n,j}\|^2_{\mathrm{L}^2[0,R]}, \\ &\text{subject to} && \alpha_{n,j} \int_0^R r^{2l_n+2}\,\mathrm{d}r = \frac{2n+1}{4\pi c_n} R^n V_{n,j}, \qquad \text{if } c_n \neq 0. \end{aligned}$$

Since the side condition is independent of $H_{n,j}$, we see that $H_{n,j} \equiv 0$ yields the unique minimum, for which we have

$$\alpha_{n,j} = (2l_n+3)\frac{2n+1}{4\pi c_n}\frac{R^n}{R^{2l_n+3}} V_{n,j}, \qquad \text{if } c_n \neq 0$$

and $\alpha_{n,j} = 0$ for all $j = 1, \ldots, 2n+1$, if $c_n = 0$. We summarize our results in the following theorem.

Theorem 5.1. *Let the conditions on V from Theorem 4.2 be fulfilled. Then, among all $D \in \mathrm{L}^2(\mathcal{B})$ with $V = \int_{\mathcal{B}} D(x)k(x,\cdot)\,\mathrm{d}x$, the $\mathrm{L}^2(\mathcal{B})$-convergent series,*

$$\begin{aligned} D(x) &= \sum_{\substack{n=0 \\ c_n \neq 0}}^{\infty} \sum_{j=1}^{2n+1} (2l_n+3)\frac{2n+1}{4\pi c_n} R^{n-l_n-3} V_{n,j} \frac{|x|^{l_n}}{R^{l_n}} Y_{n,j}\left(\frac{x}{|x|}\right) \\ &= \sum_{\substack{n=0 \\ c_n \neq 0}}^{\infty} \sum_{j=1}^{2n+1} \sqrt{\frac{2l_n+3}{R^3}}\frac{2n+1}{4\pi c_n} R^{n-l_n} V_{n,j} G_{0,n,j}(x), \end{aligned} \tag{23}$$

is the unique minimizer of the functional

$$\mathcal{F}(D) := \int_{\mathcal{B}} (D(x))^2\,\mathrm{d}x.$$

In the particular case of $l_n = n$ and $c_n = \gamma$ for all $n \in \mathbb{N}_0$, it can be proven that the harmonic solution is equivalent to the minimum norm solution, see [33]. This particular solution of the inverse gravimetric problem is then given by

$$D(x) = \frac{1}{\gamma} \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \sqrt{\frac{2n+3}{R^3}}\frac{2n+1}{4\pi} V_{n,j} G^{\mathrm{I}}_{0,n,j}(x), \qquad x \in \mathcal{B}.$$

The convergence of the series in (23) can be proven using the orthonormality of the $G_{m,n,j}$ functions, since the Parseval identity yields

$$\|D\|^2_{\mathrm{L}^2(\mathcal{B})} = \sum_{\substack{n=0 \\ c_n \neq 0}}^{\infty} \sum_{j=1}^{2n+1} \frac{2l_n+3}{R^3} \left(\frac{2n+1}{4\pi c_n}\right)^2 R^{2n-2l_n} V_{n,j}^2.$$

Comparing this with Theorem 4.2, we achieve that the series in (23) converges if and only if V fulfils the Picard condition, that is,

$$\sum_{\substack{n=0 \\ c_n \neq 0}}^{\infty} \frac{n^2(2l_n+3)}{c_n^2 R^{2(l_n-n)}} \sum_{j=1}^{2n+1} V_{n,j}^2 < \infty. \tag{24}$$

5.2. A generalization of the harmonicity constraint

In [33], the quasi-harmonic solution, which had already been discussed in the literature, was seized on. In this case, functions of the kind $x \mapsto |x|^{n+p} Y_{n,j}(\frac{x}{|x|})$, $x \in \mathcal{B}$, for a fixed $p \in \mathbb{R}_0^+$ are used as basis functions. We consider here the generalized case of a basis $\{B_{n,j}\}_{n\in\mathbb{N}_0, j=1,\dots,2n+1}$ given by

$$B_{n,j}(x) \coloneqq \frac{|x|^{k_n}}{R^{k_n+1}} Y_{n,j}\left(\frac{x}{|x|}\right), \qquad n \in \mathbb{N}_0, j = 1, \dots, 2n+1$$

with a preliminarily chosen sequence $(k_n)_{n\in\mathbb{N}_0} \subset \mathbb{R}$ and the additional condition that $2k_n + 3 > 0$ for all $n \in \mathbb{N}_0$. This condition guarantees that these functions have a finite $\mathrm{L}^2(\mathcal{B})$-norm. The orthogonality is a direct consequence of the $\mathrm{L}^2(\Omega)$-orthogonality of the spherical harmonics $Y_{n,j}$, since

$$\begin{aligned}
\langle B_{n,j}, B_{\nu,\iota}\rangle_{\mathrm{L}^2(\mathcal{B})} &= \int_{\mathcal{B}} \frac{|x|^{k_n}}{R^{k_n+1}} Y_{n,j}\left(\frac{x}{|x|}\right) \frac{|x|^{k_\nu}}{R^{k_\nu+1}} Y_{\nu,\iota}\left(\frac{x}{|x|}\right) \mathrm{d}x \\
&= \int_0^R \frac{r^{2k_n+2}}{R^{2k_n+2}} \,\mathrm{d}r \;\delta_{n,\nu}\delta_{j,\iota} \\
&= \frac{R^{2k_n+3}}{(2k_n+3)R^{2k_n+2}} \;\delta_{n,\nu}\delta_{j,\iota} \\
&= \frac{R}{2k_n+3} \;\delta_{n,\nu}\delta_{j,\iota}.
\end{aligned}$$

In the case $k_n = n$, the subspace spanned by this basis is the set of all harmonic functions and in the case $k_n = n + p$ we get the quasi-harmonic setting.

In contrast to the previous subsection, we have to assume slightly different properties of V. However, note that Assumption 3.1 is still valid.

Assumption 5.2. *We suppose that*

- *the restriction $V|_{\Omega_R}$ of V is an $\mathrm{L}^2(\Omega_R)$-function,*

- *the summability condition*

$$\sum_{\substack{n=0\\ c_n\neq 0}}^{\infty} R^{2n-2l_n}\frac{n^2(l_n+k_n+3)^2}{c_n^2(2k_n+3)}\sum_{j=1}^{2n+1}V_{n,j}^2<\infty$$

is fulfilled,
- *V is harmonic in the outer space, that is, $\Delta V(y)=0$ for all $y\in\mathbb{R}^3\setminus\mathcal{B}$,*
- *V is regular at infinity.*

With the orthogonal basis $\{B_{n,j}\}_{n\in\mathbb{N}_0,j=1,\dots,2n+1}$, the density D can be represented by the expansion

$$D(x)=\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}d_{n,j}\sqrt{\frac{2k_n+3}{R}}B_{n,j}(x),\qquad x\in\mathcal{B}\setminus\{0\},\tag{25}$$

in the sense of $\mathrm{L}^2(\mathcal{B})$. In accordance with the notations above, we have

$$D_{n,j}(r)=d_{n,j}\sqrt{\frac{2k_n+3}{R}}\frac{r^{k_n}}{R^{k_n+1}},\qquad r\in[0,R].$$

Thus, the relation between the Fourier coefficients of V and $D_{n,j}$ in (11) becomes for all $j=1,\dots,2n+1$

$$\begin{aligned}\frac{(2n+1)R^n}{4\pi c_n}V_{n,j}&=\int_0^R d_{n,j}\sqrt{\frac{2k_n+3}{R}}\frac{r^{l_n+k_n+2}}{R^{k_n+1}}\,\mathrm{d}r\\&=d_{n,j}\sqrt{\frac{2k_n+3}{R}}\frac{R^{l_n+k_n+3}}{(l_n+k_n+3)R^{k_n+1}}\\&=d_{n,j}\sqrt{\frac{2k_n+3}{R}}\frac{R^{l_n+2}}{(l_n+k_n+3)},\qquad\text{if } c_n\neq 0,\end{aligned}\tag{26}$$

and $V_{n,j}=0$, if $c_n=0$. Solving (26) for $d_{n,j}$ and inserting the result in (25), we obtain

$$\begin{aligned}D(x)&=\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}d_{n,j}\sqrt{\frac{2k_n+3}{R}}\frac{|x|^{k_n}}{R^{k_n+1}}Y_{n,j}\left(\frac{x}{|x|}\right)\\&=\sum_{\substack{n=0\\ c_n\neq 0}}^{\infty}\frac{2n+1}{4\pi c_n}(l_n+k_n+3)R^{n-l_n-2}\frac{|x|^{k_n}}{R^{k_n+1}}\sum_{j=1}^{2n+1}V_{n,j}Y_{n,j}\left(\frac{x}{|x|}\right)+\tilde{D}\\&=\sum_{\substack{n=0\\ c_n\neq 0}}^{\infty}\frac{2n+1}{4\pi c_n}(l_n+k_n+3)R^{n-l_n-2}\sum_{j=1}^{2n+1}V_{n,j}B_{n,j}(x)+\tilde{D},\end{aligned}$$

where $\tilde{D}\in\overline{\operatorname{span}\{B_{n,j}\mid n\in\mathbb{N}_0\text{ with }c_n=0,\,j=1,\dots,2n+1\}}^{\|\cdot\|_{\mathrm{L}^2(\mathcal{B})}}$ is arbitrary. The convergence of the series is guaranteed by the summability conditions on V. Summarizing these results, we get the next theorem.

Theorem 5.3. *Let $c_n \neq 0$ for all $n \in \mathbb{N}_0$, and let Assumptions* 3.1 *and* 5.2 *be fulfilled. Then the unique solution $D \in U$, where the $\mathrm{L}^2(\mathcal{B})$-subspace U has the basis $\{B_{n,j}\}_{n\in\mathbb{N}_0,j=1,\dots,2n+1}$, of the inverse problem*

$$\int_{\mathcal{B}} D(x)k(x,y)\,\mathrm{d}x = V(y) \qquad \text{in } \overline{\mathbb{R}^3\setminus\mathcal{B}},$$

with $(x,y) \in \mathrm{dom}(k)$ is given by

$$D(x) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi c_n}(l_n+k_n+3)R^{n-l_n}\frac{|x|^{k_n}}{R^{k_n+3}}\sum_{j=1}^{2n+1} V_{n,j}Y_{n,j}\left(\frac{x}{|x|}\right),$$

in the sense of $\mathrm{L}^2(\mathcal{B})$.

In [33], the biharmonic solution was also considered. In this case, the needed radial basis is given by the sum of two radial parts. An approach for a generalization of this ansatz is given by the sum of $K \in \mathbb{N}$ different radial parts, that is, $\{(\sum_{i=1}^{K} |\cdot|^{k_{i,n}})Y_{n,j}(\frac{\cdot}{|\cdot|})\}_{n\in\mathbb{N}_0,j=1,\dots,2n+1}$. However, without any additional information, a unique solution cannot be obtained in this case (see also the result for the biharmonic solution in [33]).

5.3. Layer density constraint

As we have seen above, the non-uniqueness is primarily a matter of the radial parametrization of the solution D. For this reason and in view of the fact that, for example, lithospheric heterogeneities are particularly interesting with respect to their lateral structure, we consider here the (thin) spherical shell

$$\Omega_{[\tau,\tau+\varepsilon]} \coloneqq \left\{x \in \mathbb{R}^3 : 0 < \tau \leq |x| \leq \tau+\varepsilon \leq R\right\},$$

for $\tau > 0$ and $\varepsilon > 0$. We are interested in finding a solution D which consists of purely laterally inhomogeneous anomalies in $\Omega_{[\tau,\tau+\varepsilon]}$. This kind of uniqueness constraint was, for example, used in [23] for the inverse magnetic problem.

For the layer density constraint, we assume that the density $D \in \mathrm{L}^2(\mathcal{B})$ has (again) the form

$$D(x) = \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} D_{n,j}(|x|)Y_{n,j}\left(\frac{x}{|x|}\right), \qquad x \in \mathcal{B}, \tag{27}$$

where now

$$D_{n,j}(r) \coloneqq \kappa d_{n,j}\chi_{[\tau,\tau+\varepsilon]}(r), \qquad r \in [0,R], \tag{28}$$

for all $n \in \mathbb{N}_0$, $j = 1,\dots,2n+1$, and χ is the characteristic function (i.e., $\chi_A(x) = 0$ if $x \notin A$ and $\chi_A(x) = 1$ if $x \in A$). The normalization constant κ is chosen as

$$\kappa \coloneqq \sqrt{\frac{3}{(\tau+\varepsilon)^3-\tau^3}}.$$

Assumption 5.4. *For the function V, we now assume that*

- *the restriction $V|_{\Omega_R}$ of V is an $\mathrm{L}^2(\Omega_R)$-function,*

- *the summability condition*

$$\sum_{\substack{n=0\\ c_n\neq 0}}^{\infty} \frac{n^2 l_n^2 R^{2n}}{((\tau+\varepsilon)^{l_n+3}-\tau^{l_n+3})^2 c_n^2} \sum_{j=1}^{2n+1} V_{n,j}^2 < \infty$$

is fulfilled,
- *V is harmonic in the outer space, that is, $\Delta V(y)=0$ for all $y\in\mathbb{R}^3\setminus\mathcal{B}$,*
- *V is regular at infinity.*

Using (11) and the desired representation of D, we have

$$\begin{aligned}\frac{(2n+1)R^n}{4\pi c_n} V_{n,j} &= \int_0^R r^{l_n+2} D_{n,j}(r)\,\mathrm{d}r\\ &= \kappa \int_0^R r^{l_n+2} d_{n,j}\chi_{[\tau,\tau+\varepsilon]}(r)\,\mathrm{d}r\\ &= \kappa d_{n,j}\,\frac{(\tau+\varepsilon)^{l_n+3}-\tau^{l_n+3}}{l_n+3}.\end{aligned}$$

This yields, for all $j=1,\dots,2n+1$,

$$\kappa d_{n,j} = \frac{(2n+1)R^n}{4\pi c_n} V_{n,j}\,\frac{l_n+3}{(\tau+\varepsilon)^{l_n+3}-\tau^{l_n+3}}, \qquad \text{if } c_n\neq 0,$$

and $V_{n,j}=0$, if $c_n=0$. We insert this in Equations (27) and (28) and directly obtain, for all $x\in\mathcal{B}$,

$$\begin{aligned}D(x) &= \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} d_{n,j}\kappa\chi_{[\tau,\tau+\varepsilon]}(|x|)Y_{n,j}\left(\frac{x}{|x|}\right)\\ &= \sum_{\substack{n=0\\ c_n\neq 0}}^{\infty}\sum_{j=1}^{2n+1}\frac{(2n+1)(l_n+3)}{4\pi c_n}\frac{R^n}{(\tau+\varepsilon)^{l_n+3}-\tau^{l_n+3}} V_{n,j}\chi_{[\tau,\tau+\varepsilon]}(|x|)Y_{n,j}\left(\frac{x}{|x|}\right)+\tilde{D},\end{aligned}$$

where

$$\tilde{D}\in \overline{\operatorname{span}\left\{ D_{n,j}(|\cdot|)Y_{n,j}\left(\frac{\cdot}{|\cdot|}\right)\;\middle|\; n\in\mathbb{N}_0 \text{ with } c_n=0,\ j=1,\dots,2n+1\right\}}^{\|\cdot\|_{\mathrm{L}^2(\mathcal{B})}}$$

can be chosen arbitrarily.

Theorem 5.5. *Let $c_n\neq 0$ for all $n\in\mathbb{N}_0$ and let Assumptions* 3.1 *and* 5.4 *be fulfilled. Then the unique solution under the layer density constraint is given by*

$$D(x)=\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1}\frac{R^n}{4\pi c_n}\frac{(2n+1)(l_n+3)}{(\tau+\varepsilon)^{l_n+3}-\tau^{l_n+3}} V_{n,j}\,\chi_{[\tau,\tau+\varepsilon]}(|x|)Y_{n,j}\left(\frac{x}{|x|}\right) \tag{29}$$

in the sense of $\mathrm{L}^2(\mathcal{B})$.

Moreover, under the conditions in Assumption 5.4, the corresponding potential V possesses the following outer harmonics expansion

$$V(y) = \kappa \sum_{n=0}^{\infty} \frac{4\pi c_n}{(2n+1)(l_n+3)} \left((\tau+\varepsilon)^{l_n+3} - \tau^{l_n+3}\right) |y|^{-n-1} \sum_{j=1}^{2n+1} d_{n,j} Y_{n,j}\left(\frac{y}{|y|}\right).$$

This series fulfils the condition of Assumption 5.4, that is, $V|_{\Omega_R} \in \mathrm{L}^2(\Omega_R)$:

$$\begin{aligned}
\|V|_{\Omega_R}\|^2_{\mathrm{L}^2(\Omega_R)} &= \kappa^2 \sum_{n=0}^{\infty} \left(\frac{4\pi c_n}{(2n+1)(l_n+3)} \left((\tau+\varepsilon)^{l_n+3} - \tau^{l_n+3}\right) \right)^2 R^{-2n} \sum_{j=1}^{2n+1} d_{n,j}^2 \\
&\leq 16\pi^2 c^2 \kappa^2 \sum_{n=0}^{\infty} \frac{\left(R^{l_n+3} + R^{l_n+3}\right)^2}{(2n+1)^2 (l_n+3)^3 R^{2n}} \sum_{j=1}^{2n+1} d_{n,j}^2 \\
&\leq 64\pi^2 c^2 \kappa^2 \sup_{n\in\mathbb{N}_0} \left(\frac{R^{2l_n+6-2n}}{(2n+1)^2 (l_n+3)^2} \right) \sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} d_{n,j}^2 < \infty.
\end{aligned}$$

For this estimate, we used the boundedness of the sequence $(c_n)_{n\in\mathbb{N}_0}$ (given by Assumption 3.1, item 1), the boundedness of the supremum in the latter estimate (given by Assumption 3.1, items 2 and 3), and the square-integrability of D.

5.4. Surface density

In inverse gravimetry, in particular, it is reasonable to consider a surface density instead of a density on the entire ball $\mathcal{B}$. In a time-variable gravity field (with relatively short time scales) most of the changes occur on the (Earth's) surface or at least on layers very close to it. So, if one is interested in anomalies as deviations from a reference model, which could be an annual mean, for instance, these anomalies can be typically found on the surface of the underlying body.

So far, in our general setup, we have

$$V(y) = (TD)(y) = \int_{\mathcal{B}} D(x) k(x,y) \,\mathrm{d}x. \tag{30}$$

Since the operator T is linear and continuous, we can also read the equation above in distributional sense. For the mathematical theory of distributions and, in this context, the definition of test functions, the reader is referred to [27]. In other words, we can look at Equation (30) as an application of a regular distribution D applied to the kernel k, that is

$$V(y) = (TD)(y) = \langle D, k(\cdot, y)\rangle. \tag{31}$$

Actually, we have a regular distribution $\mathfrak{D}$ with

$$\mathfrak{D}\varphi \coloneqq \langle D, \varphi\rangle$$

for all test functions[1] φ, which is uniquely determined by the function D and vice versa (at least almost everywhere). Thus, the distribution can be, in fact,

[1]Actually, the function $k(\cdot, y)$ is not a test function, but the domain of $\mathfrak{D}$ can be extended such that $\mathfrak{D}k(\cdot, y)$ makes sense and equals $(TD)(y)$.

represented by the function D itself and the distinction is commonly omitted. Now, one can think of replacing the regular distribution and also allow singular distributions. For our purposes, a very useful singular distribution is $F\delta_{\Omega_R}$, which is a variation of the well known delta distribution and is given by

$$\langle F\delta_{\Omega_R}, \varphi\rangle := \int_{\Omega_R} F(x)\varphi(x)\,\mathrm{d}\omega(x),$$

for an arbitrary, over Ω_R square-integrable, function F and for every test function φ. In that case, we have (cf. Equation (31))

$$\tilde{V}(y) := \langle D\delta_{\Omega_R}, k(\cdot, y)\rangle = \int_{\Omega_R} D(x)k(x,y)\,\mathrm{d}\omega(x).$$

Conclusively, with our previous considerations, we get

$$\begin{aligned}\tilde{V}(y) &= \sum_{n=0}^{\infty} c_n \frac{R^{l_n}}{|y|^{n+1}} \int_{\Omega_R} D(x) P_n\left(\frac{x}{|x|}\cdot\frac{y}{|y|}\right)\mathrm{d}\omega(x)\\ &= \sum_{n=0}^{\infty} c_n \frac{R^{l_n}}{|y|^{n+1}} \int_{\Omega} D(R\xi) P_n\left(\xi\cdot\frac{y}{|y|}\right) R^2\,\mathrm{d}\omega(\xi).\end{aligned}$$

With the addition theorem for spherical harmonics and the ansatz (8), it follows that

$$\begin{aligned}\tilde{V}(y) &= \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} \frac{4\pi c_n}{2n+1}\frac{R^{l_n+2}}{|y|^{n+1}} Y_{n,j}\left(\frac{y}{|y|}\right)\int_{\Omega} D(R\xi)Y_{n,j}(\xi)\,\mathrm{d}\omega(\xi)\\ &= \sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} \frac{4\pi c_n}{2n+1} R^{l_n-n+2} D_{n,j}(R)\left(\frac{R}{|y|}\right)^n \frac{1}{|y|} Y_{n,j}\left(\frac{y}{|y|}\right).\end{aligned}$$

Consequently, we find the Fourier coefficients

$$\tilde{V}_{n,j} = \frac{4\pi c_n}{2n+1} R^{l_n-n+2} D_{n,j}(R)$$

which in other words means that, for $c_n \neq 0$,

$$\frac{(2n+1)R^n}{4\pi c_n}\tilde{V}_{n,j} = R^{l_n+2} D_{n,j}(R). \tag{32}$$

As we see, this problem is again uniquely solvable (if $c_n \neq 0$ for all $n \in \mathbb{N}_0$) and in the particular case of the inverse gravimetric problem, the coefficients read

$$\frac{2n+1}{4\pi\gamma}\tilde{V}_{n,j} = R^2 D_{n,j}(R). \tag{33}$$

Theorem 5.6. *Let $D_{n,j}$ be given according to* (32) *and $c_n \neq 0$ for all $n \in \mathbb{N}_0$. Further, let $\tilde{V}$ be a harmonic function in the exterior of Ω_R which is regular at infinity with $\tilde{V}|_{\Omega_R} \in \mathrm{L}^2(\Omega_R)$ and*

$$\sum_{n=0}^{\infty}\sum_{j=1}^{2n+1} \frac{n^2 R^{2n-2l_n}}{c_n^2}\tilde{V}_{n,j}^2 < \infty.$$

Then a distributional solution of the Fredholm integral equation of the first kind in (2) *is given by*

$$D\delta_{\Omega_R} = \left(\sum_{n=0}^{\infty} \sum_{j=1}^{2n+1} \frac{(2n+1)R^{n-l_n}}{4\pi R^2 c_n} \tilde{V}_{n,j} Y_{n,j}\left(\frac{\cdot}{R}\right) \right) \delta_{\Omega_R}.$$

In the inverse gravimetric problem, as the typical application of the surface density approach, we have the following setting. Let $\bar{\rho}\colon \mathcal{B} \to \mathbb{R}$ be a density given by an arbitrary reference model of the Earth, for example, the Preliminary Reference Earth Model (PREM), see [18]. The corresponding gravitational potential is given by

$$\bar{V} = \gamma \int_{\mathcal{B}} \frac{\bar{\rho}(x)}{|x - \cdot|} \, \mathrm{d}x$$

and describes a part of the potential that does not change in the associated time span. The entire measured potential is given by $V = \bar{V} + \tilde{V}$, where $\tilde{V}$ are the relevant occurring changes in the gravitational potential. That is, we are here looking for a surface density $\sigma\colon \Omega_R \to \mathbb{R}$ with

$$\tilde{V} = V - \bar{V} = \gamma R^2 \int_{\Omega_R} \frac{\sigma(x)}{|x - \cdot|} \, \mathrm{d}\omega(x),$$

which causes these changes of the potential. By virtue of Equation (33), we know that the Fourier coefficients of the surface density are given by

$$\sigma_{n,j} = \frac{(2n+1)}{4\pi\gamma R^2} \tilde{V}_{n,j} \tag{34}$$

for all $n \in \mathbb{N}_0$ and all $j = 1, \ldots, 2n+1$. Chao [12] also proved that this problem is uniquely solvable. The obtained formula (34) coincides with the formulae which are commonly used in geodesy for a surface density ansatz or thin layer assumption, respectively, as originally proposed in [52].

6. Conclusions

We observed similarities between the inverse gravimetric and the inverse magnetic problem by considering both as particular cases of a kind of a master inverse problem. With this approach, a larger class of data inversion problems can be analyzed and solved all at once. A particular focus of the paper was the complete analysis of the non-uniqueness of the solution of all inverse problems of the investigated type. This analysis was based on something like a fundamental equation for the Fourier coefficients of the given data and the solution. The construction of a particular and appropriate orthonormal system on the ball enabled us to further understand the relation of the solution and the data. With this basis system and an adequate expansion in the data space, we characterized the null space of the Fredholm integral operator of the first kind in detail and calculated the singular system. Such a knowledge is an essential prerequisite for a series of regularization methods for inverse problems.

Furthermore, using the derived singular value decomposition, we also proved that this kind of inverse problem is unstable, that is, the inverse operator is unbounded. It also turned out that all considered problems have in common that most of the radial information gets lost. The ill-posedness of the considered problems is severely aggravated by the fact that the null space of the operator is infinite-dimensional, and, hence, the solution of the inverse problem is not unique. For this reason, we discussed four different additional conditions in order to obtain a unique solution: the minimum norm condition, a generalization of the harmonicity constraint, the layer density condition, and the surface density approach. In the particular case of the inverse gravimetric problem, our results coincide with the corresponding well-known results and in the case of the inverse magnetic problem, we found new results.

References

[1] Abramo, L.R., Reimberg, P.H., Xavier, H.S. (2010) CMB in a box: causal structure and the Fourier–Bessel expansion. Phys. Rev. D 82:043510

[2] Amirbekyan, A. (2007) The Application of Reproducing Kernel Based Spline Approximation to Seismic Surface and Body Wave Tomography: Theoretical Aspects and Numerical Results. Ph.D.-thesis, University of Kaiserslautern, Geomathematics Group, https://kluedo.ub.uni-kl.de/frontdoor/index/index/docId/1872

[3] Amirbekyan, A., Michel, V. (2008) Splines on the three-dimensional ball and their application to seismic body wave tomography. Inverse Probl. 24:015022

[4] Anger, G. (1990) Inverse Problems in Differential Equations. Akademie-Verlag, Berlin

[5] Aronszajn, N. (1950) Theory of reproducing kernels. Trans. Am. Math. Soc. 68:337–404

[6] Ballani, L., Stromeyer, D. (1992) On the structure of uniqueness in linear inverse source problems, Vieweg, Braunschweig, pp. 85–98. Theory and Practice of Geophysical Data Inversion

[7] Ballani, L., Engels, J., Grafarend, E.W. (1993) Global base functions for the mass density in the interior of a massive body (Earth). Manuscr. Geodaet. 18:99–114

[8] Barzaghi, R., Sansò, F. (1986) Remarks on the inverse gravimetric problem. Boll. geod. e sci. affini XLV:203–216

[9] Berkel, P. (2009) Multiscale Methods for the Combined Inversion of Normal Mode and Gravity Variations. Ph.D.-thesis, University of Kaiserslautern, Geomathematics Group

[10] Center for Space Research (2002) GRACE. http://www.csr.utexas.edu/grace/overview.html, [Online; accessed 30-August-2015]

[11] Chao, B.F., Dehant, V., Gross, R.S., Ray, R.D., Salstein, D., Watkins, M., Wilson, C. (2000) Space geodesy monitors mass transports in global geophysical fluids. EOS 81:247–250

[12] Chao, B.F. (2005) On inversion for mass distribution from global (time-variable) gravity field, J. Geodyn., 39:223–230

[13] Dassios, G., Fokas, A.S. (2009) Electro-magneto-encephalography for a three-shell model: dipoles and beyond for the spherical geometry. Inverse Probl. 25:035001
[14] Dassios, G., Fokas, A.S. (2013) The definite non-uniqueness results for deterministic EEG and MEG data. Inverse Probl. 29:065012
[15] Dassios, G., Fokas, A.S., Kariotou, F. (2005) On the non-uniqueness of the inverse MEG problem. Inverse Probl. 21:L1–L5
[16] Davis, P.J. (1975) Interpolation and Approximation. Dover Publications, New York
[17] Dufour, H.M. (1977) Fonctions orthogonales dans la sphère. Résolution théorique du problème du potential terrestre. B. Geod. 51:227–237
[18] Dziewonski, A.M., Anderson, D.L. (1981) Preliminary reference Earth model. Physics of the Earth and Planetary Interiors 25:297–356
[19] Fokas, A.S. (2009) Electro-magneto-encephalography for a three-shell model: distributed current in arbitrary, spherical and ellipsoidal geometries. J. R. Soc. Interface 6:479–488
[20] Fokas, A.S., Kurylev, Y. (2012) Electro-magneto-encephalography for the three-shell model: minimal L^2-norm in spherical geometry. Inverse Probl. 28:035010
[21] Fokas, A.S., Gel-fand, I.M., Kurylev, Y. (1996) Inversion method for magnetoencephalography. Inverse Probl. 2:L9–L11
[22] Fokas, A.S., Kurylev, Y., Marinakis, V. (2004) The unique determination of neuronal currents in the brain via magnetoencephalography. Inverse Probl. 20:1067–1082
[23] Fokas, A.S., Hauk, O., Michel, V. (2012) Electro-magneto-encephalography for the three-shell model: numerical implementation via splines for distributed current in spherical geometry. Inverse Probl. 28:035009
[24] Freeden, W., Gutting, M. (2013) Special Functions of Mathematical (Geo-)Physics. Birkhäuser, Basel
[25] Geselowitz, D.B. (1970) On the magnetic field generated outside an inhomogeneous volume conductor by internal current sources. IEEE Trans. Magn. 6:346–347
[26] Han, S., Shum, C., Bevis, M., Ji, C., Kuo, C. (2006) Crustal dilatation observed by GRACE after the 2004 Sumatra-Andaman earthquake. Science 313:658–662
[27] Hörmander, L. (1983) The Analysis of Linear Partial Differential Operators I. Springer, Berlin
[28] Kaula, W.M. (1966) Theory of Satellite Geodesy. Blaisdell, Waltham
[29] Kusche, J., Schrama, E. (2005) Surface mass redistribution inversion from global GPS deformation and Gravity Recovery and Climate Experiment (GRACE) gravity data. J. Geophys. Res. 110:B09409
[30] Leistedt, B., McEwen, J.D. (2012) Exact wavelets on the ball. IEEE Trans. Signal Process 60:6257–6269
[31] Michel, V. (2005) Wavelets on the 3-dimensional ball. Proc. Appl. Math. Mech. 5:775–776
[32] Michel, V. (2013) Lectures on Constructive Approximation. Fourier, Spline, and Wavelet Methods on the Real Line, the Sphere, and the Ball. Birkhäuser, Boston
[33] Michel, V., Fokas, A.S. (2008) A unified approach to various techniques for the non-uniqueness of the inverse gravimetric problem and wavelet-based methods. Inverse Probl. 24:045019
[34] Michel, V., Orzlowski, S. (2016) On the null space of a class of Fredholm integral equations of the first kind. J Inverse Ill-Posed Probl 24:687–710

[35] Moritz, H. (1990) The Figure of the Earth. Theoretical Geodesy of the Earth's Interior. Wichmann Verlag, Karlsruhe
[36] Nikiforov, A.F., Uvarov, V.B. (1988) Special Functions of Mathematical Physics. A Unified Introduction with Applications. Birkhäuser, Basel
[37] Olsen, N., et al (2013) The Swarm Satellite Constellation Application and Research Facility (SCARF) and Swarm data products. Earth Planets Space 65:1189–1200
[38] Pizzetti, P. (1910) Intorno alle possibili distribuzioni della massa nell'interno della terra. Annali di Mat., Milano XVII:225–258
[39] Plonsey, R. (1969) Biomagnetic Phenomena. McGraw-Hill, New York
[40] Rieder, A. (2003) Keine Probleme mit Inversen Problemen. Vieweg, Wiesbaden
[41] Rubincam, D.P. (1979) Gravitational potential energy of the Earth: a spherical harmonics approach. J. Geophys. Res.-Sol. Ea 84:6219–6225
[42] Rudin, W. (1991) Functional Analysis, 2nd edn. McGraw-Hill, Inc.
[43] Sabaka, T., Olsen, N. (2006) Enhancing comprehensive inversions using the Swarm constellation. Earth Planets Space 58:371–395
[44] Sansò, F., Rummel, R. (1997) Geodetic boundary value problems in view of the one centimeter geoid, Lect. Notes Earth Sci., vol 65. Springer, Berlin, Heidelberg
[45] Sarvas, J. (1987) Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Phy. Med. Biol. 32:11–22
[46] Schnetzler, C. (1985) An estimation of continental crust magnetization and susceptibility from Magsat data for the conterminous United States. J. Geophys. Res. 90:2617–2620
[47] Stokes, G.G. (1867) On the internal distribution of matter which shall produce a given potential at the surface of a gravitating mass. Proc. Royal Soc. 15:482–486
[48] Stromeyer, D., Ballani L. (1984) Uniqueness of the inverse gravimetric problem for point mass models. Manuscr. Geodaet. 9:125–136
[49] Szegö, G. (1975) Orthogonal Polynomials. American Mathematical Society, Providence, Rhode Island
[50] Thébault, E., Purucker, M., Whaler, K.A., Langlais, B., Sabako, T.J. (2010) The magnetic field of the Earth's lithosphere. Space Sci. Rev. 155:95–127
[51] Tscherning, C.C. (1996) Isotropic reproducing kernels for the inner of a sphere or spherical shell and their use as density covariance functions. Math. Geol. 28:161–168
[52] Wahr, J., Molenaar, M., Bryan, F. (1998) Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE. J. Geophys. Res. 103B(12):30205–30229
[53] Wouters, B., Chambers, D., Schrama, E.J.O., (2008) GRACE observes small-scale mass loss in Greenland. Geophys. Res. Lett. 35:L20501

Sarah Leweke and Volker Michel
University of Siegen
Geomathematics Group
Walter-Flex-Str. 3
D-57068 Siegen, Germany

emails: leweke@mathematik.uni-siegen.de
michel@mathematik.uni-siegen.de

Roger Telschow
Computational Science Center
University of Vienna
Oskar Morgenstern-Platz 1, Room 07.131
A-1090 Vienna, Austria

e-mail: roger.telschow@univie.ac.at

Handbook of Mathematical Geodesy
Geosystems Mathematics, 921–932

Index