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