Hemen Dutta
Ljubiša D. R. Kočinac
Hari M. Srivastava
Editors

Current Trends in Mathematical Analysis and Its Interdisciplinary Applications

Birkhäuser

Birkhäuser

Current Trends in Mathematical Analysis and Its Interdisciplinary Applications

Hemen Dutta • Ljubiša D. R. Kočinac •
Hari M. Srivastava
Editors

Current Trends in Mathematical Analysis and Its Interdisciplinary Applications

Birkhäuser

Editors
Hemen Dutta
Department of Mathematics
Gauhati University
Guwahati
Assam, India

Ljubiša D. R. Kočinac
University of Nis
Faculty of Sciences and Mathematics
Aleksandrovac, Serbia

Hari M. Srivastava
Dept. Mathematics and Statistics
University of Victoria
Victoria
BC, Canada

ISBN 978-3-030-15241-3 ISBN 978-3-030-15242-0 (eBook)
https://doi.org/10.1007/978-3-030-15242-0

Mathematics Subject Classification (2010): 33-XX, 34-XX, 35-XX, 45-XX, 46-XX, 47-XX, 49-XX, 58-XX, 76-XX

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

Preface

The book is for graduate and PhD students, researchers in mathematics and applied sciences, educators, and engineers. It contains research results on several important aspects of recent developments in interdisciplinary applications of mathematical analysis and also focuses on the uses and applications of mathematical analysis in many areas of scientific research. Each chapter aims at enriching the understanding of the research problems with sufficient material to understand the necessary theories, methods, and applications. Emphasis is given to present the basic developments concerning an idea in full detail and the most recent advances made in the area of study. The book shall also be useful for general readers having interest in recent developments in interdisciplinary applications of mathematical analysis. There are 23 chapters in the book, and they are organized as follows.

Chapter 1 is devoted to the study of stationary viscous incompressible fluid flow problems in a bounded domain with a subdifferential boundary condition of frictional type in the Orlicz spaces. It first investigates non-Newtonian fluid flow with a nonpolynomial growth of the viscous part of the Cauchy stress tensor together with a multivalued nonmonotone frictional boundary condition described by the Clarke subdifferential. Next, a Newtonian fluid flow with a multivalued nonmonotone boundary condition of a nonpolynomial growth between the normal velocity and normal stress is studied. In both cases, an abstract result on the existence and uniqueness of solution to a subdifferential operator inclusion and a hemivariational inequality in the reflexive Orlicz–Sobolev space is provided. The results obtained are applied to a hemivariational inequality that arises in the study of the flow phenomenon with frictional boundary conditions.

In Chap. 2, the classical identities of Jacobi theta functions are obtained from the multiplicities of the eigenvalues i^k, and corresponding eigenvectors of the *DFT* $\phi(n)$, expressed in terms of theta functions. An extended version of the classical Watson addition formula and Riemann's identity on theta functions is also derived. Watson addition formula and Riemann's identity are obtained as a particular case. An extension of some classical identities corresponding to the theta functions $\theta_{a,b}(x, \tau)$ with $a, b \in \frac{1}{3}Z$ is also derived.

Chapter 3 used combinatorial tools “color partitions,” “split-color partitions,” and “signed partitions” notion to define “signed color partitions” that are further used to derive one hundred Rogers–Ramanujan type identities. The chapter lists and provides combinatorial argument using signed colored partitions of $q$identities listed in ChuZhang and Slater’s compendium.

Chapter 4 discusses Stepanov-like almost automorphic function in the framework of impulsive systems. It establishes the existence and uniqueness of solution of a very general delayed model of impulsive neural network. The coefficients and forcing term are assumed to be Stepanov-like almost automorphic in nature and introduce the concept of piecewise continuous Stepanov-like almost automorphic function. First, some basic and important properties of these functions are established and then composition theorem is proved. Further, composition result and fixed-point theorem are used to investigate the existence, uniqueness, and stability of solution of the problem considered. A numerical example is given to illustrate the analytical findings.

Chapter 5 first discusses ideas to improve the speed of convergence of the secant method by means of iterative processes free of derivatives of the operator in their algorithms. For this, a previously constructed uniparametric family of secant-like methods is considered. The semilocal convergence of this uniparametric family of iterative processes is analyzed by using a technique that consists of a new system of recurrence relations.

Chapter 6 attempts to find answers to the questions like “Why is the manifold topology in a spacetime taken for granted?,” “Why do we prefer to use Riemann open balls as basic-open sets, while there also exists a Lorentz metric?,” “Which topology is the best candidate for a spacetime: a topology sufficient for the description of spacetime singularities or a topology which incorporates the causal structure? Or both?,” “Is it more preferable to consider a topology with as many physical properties as possible, whose description might be complicated and counterintuitive, or a topology which can be described via a countable basis but misses some important information?,” etc. The chapter aims to serve as a critical review of similar questions and contains a survey with remarks, corrections, and open questions.

Chapter 7 studies a generalized BBM equation from the point of view of the theory of symmetry reductions in partial differential equations. It first obtained the Lie symmetries and then used the transformation groups to reduce the equations into ordinary differential equations. Physical interpretation of these reductions and some exact solutions are also provided. It also derives all low-order conservation laws for the BBM equation by using the multiplier method.

Chapter 8 studies some Boussinesq equations with damping terms from the point of view of the Lie theory. It derives the classical Lie symmetries admitted by the equation as well as the reduced ordinary differential equations. The chapter also presents some exact solutions. Further, some nontrivial conservation laws for these equations are constructed by using the multiplier method.

Chapter 9 discusses on the weak solvability of some variable exponent problems via the critical point theory, which also includes the case of anisotropic exponents.

The author considers only a few powerful theorems as main tools that can be applied to all selected problems.

Chapter 10 is concerned with a coupled system of nonlinear viscoelastic wave equations that models the interaction of two viscoelastic fields. A new general decay result is established that improves most of the existing results in the literature related to the system of viscoelastic wave equations. The result of the chapter allows wider classes of relaxation functions.

Chapter 11 establishes local existence and uniqueness as well as blow-up criteria for solutions of the Navier–Stokes equations in Sobolev–Gevrey spaces. Precisely, if the maximal time of existence of solutions for these equations is finite, the chapter demonstrates the explosion, near this instant, of some limits superior and integrals involving specific usual Lebesgue spaces, and as a consequence, lower bounds related to Sobolev–Gevrey spaces are proved.

Chapter 12 deals with a survey and critical analysis focused on a variety of chemotaxis models in biology, namely the chemotaxis-(Navier)–Stokes system and its subsequent modifications, which, in several cases, have been developed to obtain models that prevent the nonphysical blow-up of solutions. First it focuses on the background of the models which is related to chemotaxis-(Navier)–Stokes system. Then, the chapter is devoted to the qualitative analysis of the (quasilinear) Keller–Segel model, the (quasilinear) chemotaxis-haptotaxis model, the (quasilinear) chemotaxis system with consumption of chemoattractant, and the (quasilinear) Keller–Segel–Navier–Stokes system.

Chapter 13 deals with the optimal control of a class of elliptic quasivariational inequalities. It started with an existence and uniqueness result for such inequalities. Then an optimal control problem is stated, the assumptions on the data are listed, and the existence of optimal pairs is proved. It further proceeds with a perturbed control problem for which a convergence result is established under general conditions. A particular case for which these conditions are satisfied is also presented. The use of the abstract results is illustrated in the study of a mathematical model which describes the equilibrium of an elastic body in frictional contact with an obstacle. The process is static and the contact is modeled with normal compliance and unilateral constraint, associated with the Coulomb's law of dry friction. The existence, uniqueness, and convergence results are proved together with the corresponding mechanical interpretation. These results are also illustrated in the study of a one-dimensional example.

In Chap. 14, master generalized sampling series expansion is presented for entire functions coming from a class, members of which satisfy an extended exponentially boundedness condition. First, estimates are given for the remainder of Maclaurin series of the functions and consequently derivative sampling results are obtained and discussed. The results thus obtained are employed in evaluating the related remainder term of functions which occur in sampling series expansion of stochastic processes and random fields of which spectral kernel satisfies the relaxed exponential boundedness. The derived truncation error upper bounds enable us to obtain mean-square master generalized derivative sampling series expansion formulae either for harmonizable Piranashvili-type stochastic processes or for random fields. Finally, the sampling series convergence rate being exponential, almost sure P sampling series expansion formulae are inferred.

Chapter 15 describes polygonal hybrid finite element formulation with fundamental solution kernels for two-dimensional elasticity in isotropic and homogeneous solids. The n-sided polygonal discretization is implemented by the Voronoi diagram in a given domain. Then, the element formulation is established by introducing two independent displacements respectively defined within the element domain and over the element boundary. The element interior fields approximated by the fundamental solutions of problem can naturally satisfy the governing equations, and the element frame fields approximated by one-dimensional shape functions are used to guarantee the conformity of elements. Finally, the present method is verified by three examples involving the usage of general and special n-sided polygonal hybrid finite elements.

Chapter 16 studies the existence of solutions for suitable Schrödinger equations in the whole space by means of variational methods. It considers a fractional version of the Schrödinger equation in the presence of a potential, which is studied in two different cases. The first one is when the potential is given a priori and the second one when the potential is unknown. These equations describe two physical models. In both cases, existence of multiple solutions is proved depending on some topological properties involving the set of minima of the potential.

Chapter 17 considers nonlinear elliptic equations driven by a nonhomogeneous differential operator plus an indefinite potential. The boundary condition is either Dirichlet or Robin. First it presents the corresponding regularity theory. Then the nonlinear maximum principle is developed and some nonlinear strong comparison principles are presented. Subsequently it is shown how these results together with variational methods, truncation and perturbation techniques, and Morse theory can be used to analyze different classes of elliptic equations, and special attention is given to $(p, 2)$-equations.

Chapter 18 investigates a new definition of convergence of a double sequence and a double series, which seems to be most suitable in the non-Archimedean context, and studies some of its properties. Then, a very brief survey of the results pertaining to the Nörlund, weighted mean, and $\left(M, \lambda_{m,n}\right)$ methods of summability for double sequences is presented. Further, a Tauberian theorem for the Nörlund method for double sequences is given.

Chapter 19 aims to develop effective approximate solution methods for the linear and nonlinear singular integral equations in Banach spaces. This chapter is devoted to investigating approximate solutions of linear and nonlinear singular integral equations in Banach spaces using technical methods such as collocation method, quadrature method, Newton–Kantorovich method, monotonic operator method, and fixed-point theory depending on the type of the equations. Sufficient conditions for the convergence of these methods are provided and some relevant properties are investigated.

In Chap. 20, a new generalized difference double sequence based on integer orders is defined. An application of the proposed operator, certain new related difference double sequence spaces have been presented and their corresponding topological properties have been discussed. The dual spaces related to the new difference double sequence spaces have been determined. The idea is also used to study the derivatives of single variable functions and also the partial derivatives of double variable functions.

Chapter 21 discusses m-singularity notion for double singular integral operators and presents several relevant results concerning pointwise convergence of nonlinear double m-singular integral operators. First, the reasons giving birth to m-singularity notion are explained and related theoretical background is mentioned. The well-definiteness of the operators on their domain is shown, and an auxiliary result and pointwise convergence theorem are proved. Then, the main theorem and Fatou-type convergence theorem are proved. Further, corresponding rates of convergences are evaluated.

Chapter 22 considers and surveys multifarious extensions of the p-adic integrals. q-analogues with diverse extensions of p-adic integrals are also considered such as the weighted p-adic q-integral on Zp. The two types of the weighted q-Boole polynomials and numbers are introduced and investigated in detail. Some generalized and classical q-polynomials and numbers are further obtained from the aforesaid extensions of p-adic integrals. The importance of these extensions is also analyzed.

Chapter 23 first discusses the concept of infiniteness and the development of summability methods. Then, ordinary and statistical versions of Cesàro and deferred Cesàro summability methods are demonstrated and the deferred Cesàro mean is applied to prove a Korovkin-type approximation theorem for the set of functions 1, e^{-x}, and e^{-2^x} defined on a Banach space $C[0, \infty)$. Further, a result for the rate of statistical deferred Cesàro summability mean with the help of the modulus of continuity is established, and some examples in support of the results are presented.

The editors are grateful to the contributors for their timely contribution and patience while the chapters were being processed and reviewed. We are greatly indebted to reviewers for their generous help and time given to review the chapters. Finally, the editors must thank the Birkhauser editor and staff for their support in bringing out this book.

Guwahati, Assam, India — Hemen Dutta
Aleksandrovac, Serbia — Ljubiša D. R. Kočinac
Victoria, BC, Canada — Hari M. Srivastava
January, 2019

Contents

Contributors

Syed Abbas School of Basic Sciences, Indian Institute of Technology Mandi, Mandi, H.P., India

I. K. Argyros Department of Mathematics Sciences, Cameron University, Lawton, OK, USA

P. Baliarsingh Department of Mathematics, School of Applied Sciences, KIIT University, Bhubaneswar, India

Ezequiel Barbosa Departamento de Matemática, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

Maria-Magdalena Boureanu Department of Applied Mathematics, University of Craiova, Craiova, Romania

M. S. Bruzón Departamento de Matemáticas, Universidad de Cádiz, Puerto Real, Cádiz, Spain

R. de la Rosa Departamento de Matemáticas, Universidad de Cádiz, Puerto Real, Cádiz, Spain

Ugur Duran Department of the Basic Concepts of Engineering, Faculty of Engineering and Natural Sciences, İskenderun Technical University, Hatay, Turkey

Hemen Dutta Department of Mathematics, Gauhati University, Guwahati, Assam, India

Giovany Malcher Figueiredo Departamento de Matemática, Universidade de Brasilia, Brasilia, DF, Brazil

M. L. Gandarias Departamento de Matemáticas, Universidad de Cádiz, Puerto Real, Cádiz, Spain

T. M. Garrido Departamento de Matemáticas, Universidad de Cádiz, Puerto Real, Cádiz Spain

V. Gupta School of Mathematics, Thapar University, Patiala, Punjab, India

Jamilu Hashim Hassan Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

M. A. Hernández-Verón Department of Mathematics and Computation, University of La Rioja, Logroño, Spain

B. B. Jena Department of Mathematics, Veer Surendra Sai University of Technology, Burla, Odisha, India

Lakshman Mahto Department of Mathematics, Indian Institute of Information Technology Dharwad, Hubli, Karnataka, India

R. A. Malekar Department of Mathematics, National Defence Academy, Khadakwasla, Pune, India

Wilberclay G. Melo Departamento de Matemática, Universidade Federal de Sergipe, São Cristóvão, SE, Brazil

Salim A. Messaoudi Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

Stanisław Migórski Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland

Edwin Gonzalo Murcia Departamento de Matemáticas Pontificia, Universidad Javeriana, Bogotá, Colombia

Nizami Mustafa Kafkas University, Faculty of Science and Letters, Department of Mathematics, Kars, Turkey

P. N. Natarajan Department of Mathematics, Ramakrishna Mission Vivekananda College, Chennai, India

L. Nayak Department of Mathematics, School of Applied Sciences, KIIT University, Bhubaneswar, India

Veysel Nezir Kafkas University, Faculty of Science and Letters, Department of Mathematics, Kars, Turkey

Dariusz Pączka Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

S. K. Paikray Department of Mathematics, Veer Surendra Sai University of Technology, Burla, Odisha, India

Kyriakos Papadopoulos Department of Mathematics, Kuwait University, Safat, Kuwait

Nikolaos S. Papageorgiou National Technical University, Department of Mathematics, Athens, Greece

Zurab A. Piranashvili Vladimir Chavchanidze Institute of Cybernetics, Georgian Technical University, Tbilisi, Georgia

Tibor K. Pogány Faculty of Maritime Studies, University of Rijeka, Rijeka, Croatia

Institute of Applied Mathematics, Óbuda University, Budapest, Hungary

Qing-Hua Qin Research School of Engineering, Australian National University, Canberra, ACT, Australia

M. Rana School of Mathematics, Thapar University, Patiala, Punjab, India

Natã Firmino Rocha Departamento de Matemática, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

M. Rosa Departamento de Matemáticas, Universidad de Cádiz, Puerto Real, Cádiz, Spain

M. J. Rubio Department of Mathematics and Computation, University of La Rioja, Logroño, Spain

Fabio Scardigli Dipartimento di Matematica, Politecnico di Milano, Milano, Italy

Institute-Lorentz for Theoretical Physics, Leiden University, Leiden, The Netherlands

Gaetano Siciliano Instituto de Matemática e Estatística, Universidade de Brasilia, São Paulo, SP, Brazil

Mircea Sofonea Laboratoire de Mathématiques et Physique, University of Perpignan Via Domitia, Perpignan, France

Gumrah Uysal Department of Computer Technologies, Division of Technology of Information Security, Karabuk University, Karabuk, Turkey

Calogero Vetro University of Palermo, Department of Mathematics and Computer Science, Palermo, Italy

Francesca Vetro Nonlinear Analysis Research Group, Ton Duc Thang University, Ho Chi Minh, Vietnam

Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh, Vietnam

Hui Wang College of Civil Engineering & Architecture, Henan University of Technology, Zhengzhou, China

Jiashan Zheng School of Mathematics and Statistics Science, Ludong University, Yantai, People's Republic of China

Chapter 1
Frictional Contact Problems for Steady Flow of Incompressible Fluids in Orlicz Spaces

Stanisław Migórski and Dariusz Pączka

1.1 Introduction

The chapter is devoted to the study of steady-state flow problems of isotropic, isothermal, inhomogeneous, viscous, and incompressible fluids in a bounded domain with subdifferential boundary conditions in Orlicz spaces. Two general cases are investigated. First, we study the non-Newtonian fluid flow with a non-polynomial growth of the extra (viscous) part of the Cauchy stress tensor together with multivalued nonmonotone slip boundary conditions of frictional type described by the Clarke generalized gradient. Second, we analyze the Newtonian fluid flow with a multivalued nonmonotone leak boundary condition of frictional type which is governed by the Clarke generalized gradient with a non-polynomial growth between the normal velocity and normal stress. In both cases, we provide abstract results on existence and uniqueness of solution to subdifferential operator inclusions with the Clarke generalized gradient and the Navier–Stokes type operator which are associated with hemivariational inequalities in the reflexive Orlicz–Sobolev spaces. Moreover, our study, in both aforementioned cases, is supplemented by similar results for the Stokes flows where the convective term is negligible. Finally, the results are applied to examine hemivariational inequalities arising in the study of

S. Migórski
College of Applied Mathematics, Chengdu University of Information Technology, Chengdu, Sichuan Province, P.R. China

Chair of Optimization and Control, Jagiellonian University in Kraków, Kraków, Poland
e-mail: stanislaw.migorski@ii.uj.edu.pl

D. Pączka (✉)
Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland
e-mail: d.paczka@mini.pw.edu.pl

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_1

the flow phenomenon with frictional boundary conditions. The chapter is concluded with a continuous dependence result and its application to an optimal control problem for flows of Newtonian fluids under leak boundary condition of frictional type.

The steady-state generalized Navier–Stokes equation in a bounded regular domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is of the form

$$-\operatorname{div} \mathbf{S} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + \nabla\pi = \boldsymbol{f}, \quad \operatorname{div} \boldsymbol{u} = 0.$$

The general growth conditions are assumed for the stress deviator $\mathbf{S}$ in terms of the symmetric part of the velocity gradient $\mathbf{D}(\boldsymbol{u}) = \frac{1}{2}(\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^\top)$ formulated via an Orlicz function Φ. The special case $\Phi(t) = t^p$ with $1 < p < \infty$ leads to the power law model

$$\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) = (1 + \|\mathbf{D}(\boldsymbol{u})\|)^{p-2}\mathbf{D}(\boldsymbol{u}) \quad \text{in } \Omega$$

which is quite common to model non-Newtonian fluids. However, we use in this chapter the framework of Orlicz–Sobolev spaces which provides more flexibility. The model is supplemented by the nonstandard boundary condition

$$u_\nu = 0, \quad -\mathbf{S}_\tau \in \partial j_\tau(\boldsymbol{u}_\tau)$$

which is called the slip boundary condition of frictional type, or by the condition

$$\boldsymbol{u}_\tau = 0, \quad -\sigma_\nu \in \partial j_\nu(u_\nu)$$

called the leak boundary condition of frictional type. Here u_ν and $\boldsymbol{u}_\tau$ denote the normal and tangential part of the velocity, and σ_ν and $\mathbf{S}_\tau$ are the normal and tangential components of the stress tensor and the extra stress tensor, respectively. The notation ∂j_ν and ∂j_τ stands for the generalized gradient of locally Lipschitz functions j_ν and j_τ, respectively. We provide results on the solvability and unique solvability of the hemivariational inequalities which are weak formulations of the flow problems in the two aforementioned kinds of frictional boundary conditions.

The Orlicz and Orlicz–Sobolev spaces are suitable function spaces to describe fluid flow problems modeled by systems of nonlinear partial differential equations with nonlinearities of non-polynomial growth. There are examples of these nonlinearities in physics, for example models of fluids of Prandtl–Eyring [24], Powell–Eyring [79], and Sutterby [3]. In order to describe flows of anisotropic fluids with the rheology more general than power-law-type it is necessary to use the Musielak–Orlicz space, see [39, 40]. In the framework of Orlicz and Orlicz–Sobolev spaces, many problems in mechanics of solids and fluids have been considered, for instance [5, 9, 14, 19, 31, 32, 73]. Examples of N-functions Φ which generate reflexive Orlicz and Orlicz–Sobolev spaces are the following $\Phi(t) = t^p$, $\Phi(t) = t^p \log(1 + t^p)$, $\Phi(t) = t^p \log^q(1 + t)$ and $\Phi(t) = t^p \log^{q_1}(1 + t) \log^{q_2}(\log(1 + t))$ with $p, q, q_1, q_2 \in (1, \infty)$. Nonstandard examples of N-functions Φ which do not generate reflexive Orlicz and Orlicz–Sobolev spaces and occur in mechanics of solids and fluids are the following $\Phi_1(t) = t^\alpha \ln(1+t)$ for $t \geq 0$ and $1 \leq \alpha < 2$, $\Phi_2(t) = \int_0^t s^{1-\alpha}(\text{arcsinh}(s))^\alpha\, ds$ for $t \geq 0$ and $0 < \alpha \leq 1$, $\Phi_3(t) = t\ln(1+\ln(1+t))$ for $t \geq 0$ (see [31, 32] for details).

Our approach is based on a powerful technique on the surjectivity of pseudomonotone maps [8], the compactness of a trace operator in the Orlicz–Sobolev space [22], coercivity and pseudomonotonicity of the Navier–Stokes type operator [23], and a result on an integral representation of the Clarke subdifferential of locally Lipschitz integral functionals defined on the Orlicz space [71, 72]. In the treatment of this topic, we successfully use some techniques from the theory of hemivariational inequalities in Sobolev spaces of Panagiotopoulos [76, 78], Naniewicz and Panagiotopoulos [69], and Migórski et al. [67].

The results of this chapter are based on our research papers [65, 66]. In Sect. 1.5 we consider the constitutive relation which has a non-polynomial growth with respect to the stress tensor and it is not of explicit form in the context of the frictional contact law described by Clarke subgradient. The frictional contact boundary condition is also established for functions of a non-polynomial growth. Problem 1.5.1 has been studied in [23] in the 2D setting for a particular geometry of the domain in the context of the lubrication theory and with a polynomial growth for the stress deviator. Theorem 1.5.6 strengthens the conclusion of [23] in the 2D setting. In Sect. 1.6, the leak boundary conditions described by the Clarke subdifferential are considered for functions of a non-polynomial growth, and, in a consequence, the velocity has a non-polynomial growth. Problem 1.6.1 has been studied in [60] for Newtonian fluids in the Sobolev space $W^{1,2}(\Omega, \mathbb{R}^d)$ setting. Theorem 1.6.5 enhances the conclusion of [60] to Newtonian fluids with a non-polynomial growth in the reflexive Orlicz–Sobolev space $W^1L_\Phi(\Omega, \mathbb{R}^d)$. Furthermore, using the direct method of the calculus of variations, we deliver a result on existence of a solution to an optimal control problem for the hemivariational inequality. To this end, we prove a result on a dependence of the solution set of the hemivariational inequality on the density of external forces. Note that optimal control problems for hemivariational inequalities have been studied in several contributions, see [25, 41, 56–58, 63, 77] and the references therein. Besides, to the best of our knowledge, there are no results on existence and uniqueness of solution to hemivariational inequalities in the Orlicz–Sobolev space for contact problems arising in mechanics, including Newtonian or non-Newtonian fluid flow problems. Finally, note that the uniqueness of solutions is proved in Sects. 1.5 and 1.6 without the relaxed monotonicity condition (see Definition 1.2.1) for the superpotential, as previously required in [67] and other papers.

The frictional contact boundary conditions for steady/unsteady Newtonian or non-Newtonian fluid flows in Sobolev spaces have been studied, for instance, in [11–13, 23, 25, 45, 60–62]. Mathematical analysis of non-Newtonian fluids without friction can be found in [4, 5, 28, 29, 32, 81] in the stationary case and in [2, 9, 39, 40, 48, 50–53, 82] for the evolutionary case.

1.2 Preliminaries

In general, lowercase letters (Greek and Latin) are used for scalar quantities, upright boldface lowercase letters are used for vectors (for example, $\mathbf{n}$), and italic boldface lowercase letters are used for functions ranging in the multidimensional Euclidean space (for example, $\boldsymbol{u}$, $\boldsymbol{\xi}$). We denote by $\mathbf{a}\cdot\mathbf{b} = \sum_{i=1}^{n} a_i b_i$ the usual scalar product in $\mathbb{R}^n$ and by $|\mathbf{a}| = (\mathbf{a}\cdot\mathbf{a})^{1/2}$ the Euclidean norm. Matrices (tensors) and matrix-valued functions are represented by upright boldface uppercase letters such as $\mathbf{A}$ and $\mathbf{S}$. We set $\mathbf{A} : \mathbf{B} := \sum_{j,k=1}^{n} a_{jk} b_{jk}$, if $\mathbf{A} = (a_{jk})_{j,k=1}^{n}$ and $\mathbf{B} = (b_{jk})_{j,k=1}^{n}$. The symbol $\mathbb{S}^d$ stands for the space of symmetric matrices of order d. Let X, Y be a pair of vector spaces. The inner product on X will be denoted by $\langle\cdot,\cdot\rangle_X$, the canonical bilinear form on $X \times Y$ is usually denoted by $\langle\cdot,\cdot\rangle_{X\times Y}$ (or simply $\langle\cdot,\cdot\rangle$). By X^* we denote a topological dual space of a topological vector space X. The notation $X \hookrightarrow Y$ (resp. $X \hookrightarrow\hookrightarrow Y$) means that X and Y are normed spaces with X continuously (resp. compactly) embedded in Y. Arrows $\to$ and $\rightharpoonup$ are used to denote the strong and weak convergence, respectively, in the given topology. By X_ω (resp. X^*_ω, $X^*_{\omega^*}$) we denote the space X (resp. X^*) furnished with the weak (resp. weak, weak star) topology. We will denote by $\|A\|_{X\to Y}$ the norm of a continuous linear operator A between normed linear spaces X and Y. For a subset U of normed space X, we write $\|U\|_X = \sup\{\|u\|_X \mid u \in U\}$. The symbol $\overline{B}_X(x,r) = \{y \in X \mid \|y - x\|_X \le r\}$ stands for the closed ball of a real Banach space X centered at $x \in X$ and a radius $r > 0$, whereas $B_X(u,r)$ denotes the corresponding open ball.

1.2.1 Operators of Monotone Type

We recall now some definitions from set-valued (see, e.g., [17, 18]). Given a Suslin locally convex space S (e.g., $S = E$ or $S = E^*_{\omega^*}$, where $E^*_{w^*}$ stands for the dual space of a separable Banach space E with the weak star topology $w^* = \sigma(E^*, E)$), we denote by $\mathcal{B}(S)$ the σ-algebra of Borel subsets of S. A *set-valued map*, or a *multifunction* F from a set $\mathcal{O}$ to S, is a map that associates with any $\omega \in \mathcal{O}$ a nonempty subset $\mathsf{F}(\omega)$ of S, and we write $\mathsf{F}\colon \mathcal{O} \multimap S$. Let $(\mathcal{O}, \mathfrak{A})$ be a measurable space. The multifunction F is called *measurable* if $\mathsf{F}^-(C) := \{\omega \in \mathcal{O} \mid \mathsf{F}(\omega) \cap C \ne \emptyset\} \in \mathfrak{A}$ for $C \in \mathcal{B}(S)$. By a *measurable selection* of F we mean a (single-valued) function f such that $f(\omega) \in \mathsf{F}(\omega)$ for almost all $\omega \in \mathcal{O}$. We will denote by $\operatorname{Sel}\mathsf{F}$ the set of all measurable selections of F. Given a function $u\colon \mathcal{O} \to E$ and $\mathsf{F}\colon \mathcal{O} \times E \multimap E^*_{\omega^*}$, define the *multivalued superposition operator* $\mathcal{N}_{\mathsf{F}}(u) := \operatorname{Sel}\mathsf{F}(\cdot, u(\cdot))$. Let X and Y be metric spaces. A multifunction $\mathsf{F}\colon X \multimap Y$ is called *closed* if its graph $\operatorname{Gr}(\mathsf{F}) = \{(x,y) \in X \times Y \mid y \in \mathsf{F}(x)\}$ is a closed subset of $X \times Y$; *upper semicontinuous* (or *u.s.c.*) at $x \in X$ if, for any open set $V \subset Y$ with $\mathsf{F}(x) \subseteq V$, one may find an open neighborhood $U \subseteq X$ of x such that $\mathsf{F}(x) \subseteq V$ for all $x \in U$. Now, let Z be a vector metric space. A multifunction

$\mathsf{F}\colon X \multimap Z$ is called *sequentially strongly-weakly closed*, if $\mathrm{Gr}(\mathsf{F})$ is sequentially closed in $X \times Z_\omega$, where Z_ω is endowed with the weak topology $\omega = \sigma(Z, Z^*)$.

Definition 1.2.1 Let V be a reflexive Banach space. A multivalued operator $\mathsf{A}\colon V \multimap V^*$ is called:

(a) *bounded*, if A maps bounded subsets of V into bounded subsets of V^*;
(b) *relaxed monotone*, if there exists a constant $m \geq 0$ satisfying the inequality $\langle u^* - v^*, u - v\rangle_{V^*\times V} \geq -m\,\|u - v\|_V^2$ for all $(u, v), (u^*, v^*) \in \mathrm{Gr}(\mathsf{A})$.
(c) *pseudomonotone*, if the following conditions hold:

   1) A has values which are closed and convex sets;
   2) A is u.s.c. from each finite-dimensional subspace of V to V_w^*;
   3) for any sequence $(v_n, v_n^*) \subset \mathrm{Gr}(\mathsf{A})$ satisfying the conditions $v_n \rightharpoonup v$ in $\sigma(V, V^*)$ and $\limsup_{n\to\infty}\langle v_n^*, v_n - v\rangle_{V^*\times V} \leq 0$, it follows that for every $y \in V$ there exists $(y, v^*(y)) \in \mathrm{Gr}(\mathsf{A})$ such that $\langle v^*(y), v - y\rangle_{V^*\times V} \leq \liminf_{n\to\infty}\langle v_n^*, v_n - y\rangle_{V^*\times V}$;

(d) *generalized pseudomonotone*, if for any sequence $(v_n, v_n^*) \subset \mathrm{Gr}(\mathsf{A})$ such that $v_n \rightharpoonup v$ in $\sigma(V, V^*)$, $v_n^* \rightharpoonup v^*$ in $\sigma(V^*, V)$ and $\limsup_{n\to\infty}\langle v_n^*, v_n - v\rangle_{V^*\times V} \leq 0$, we have $(v, v^*) \in \mathrm{Gr}(\mathsf{A})$ and $\langle v_n^*, v_n\rangle_{V^*\times V} \to \langle v^*, v\rangle_{V^*\times V}$;
(e) *coercive*, if there exists a function $\ell\colon (0, \infty) \to \mathbb{R}$ satisfying the conditions $\lim_{t\to\infty}\ell(t) = \infty$ and $\langle w, u\rangle_{V^*\times V} \geq \ell(\|u\|_V)\,\|u\|_V$ for all $(u, w) \in \mathrm{Gr}(\mathsf{A})$.

Definition 1.2.2 Let V be a reflexive Banach space. A single-valued operator $\mathcal{A}\colon V \to V^*$ is called:

(a) *bounded*, if $\mathcal{A}$ maps bounded subsets of V into bounded subsets of V^*;
(b) *strongly monotone*, if $\langle \mathcal{A}v_1 - \mathcal{A}v_2, v_1 - v_2\rangle_{V^*\times V} \geq m\,\|v_1 - v_2\|_V^2$ for all $v_1, v_2 \in V$ with $m > 0$;
(c) *pseudomonotone*, if $\mathcal{A}$ is bounded and for any sequence $(v_n) \subset V$ satisfying the conditions $v_n \rightharpoonup v$ in $\sigma(V, V^*)$ and $\limsup_{n\to\infty}\langle \mathcal{A}v_n, v_n - v\rangle_{V^*\times V} \leq 0$, it holds $\langle \mathcal{A}v, v - y\rangle_{V^*\times V} \leq \liminf_{n\to\infty}\langle \mathcal{A}v_n, v_n - y\rangle_{V^*\times V}$ for all $y \in V$;
(d) (α, ρ)-*coercive*, if there exist constants $\alpha > 0$ and $\rho \in (1, \infty)$ such that $\langle \mathcal{A}v, v\rangle_{V^*\times V} \geq \alpha\,\|v\|_V^\rho$ for all $v \in V$;
(e) *hemicontinuous*, if $t \mapsto \langle \mathcal{A}(u + tv), w\rangle_{V^*\times V}$ is continuous on $[0, 1]$ for all $u, v, w \in V$;
(f) *weakly sequentially continuous*, if $v_n \rightharpoonup v$ in $\sigma(V, V^*)$ implies $\mathcal{A}v_n \rightharpoonup \mathcal{A}v$ in $\sigma(V^*, V)$.

Remark 1.2.3 Note that (c) of Definition 1.2.2 is equivalent to the following one: a single-valued operator $\mathcal{A}\colon V \to V^*$ is called *pseudomonotone*, if $\mathcal{A}$ is bounded and for any sequence $(v_n) \subset V$ satisfying the conditions $v_n \rightharpoonup v$ in $\sigma(V, V^*)$ and $\limsup_{n\to\infty}\langle \mathcal{A}v_n, v_n - v\rangle_{V^*\times V} \leq 0$, we have $\mathcal{A}u_n \to \mathcal{A}u$ in $\sigma(V^*, V)$ and $\lim_{n\to\infty}\langle \mathcal{A}u_n, u_n - u\rangle = 0$.

Definition 1.2.4 Let V be a separable and reflexive Banach space. An operator $\mathcal{N}\colon V \to V^*$ is called a *Navier–Stokes type operator* if $\mathcal{N}\boldsymbol{v} = \mathcal{A}\boldsymbol{v} + \mathcal{B}[\boldsymbol{v}]$, where

(1) $\mathcal{A}\colon V \to V^*$ is pseudomonotone and (α, ρ)-coercive;
(2) $\mathcal{B}[\boldsymbol{v}] = \mathcal{B}(\boldsymbol{v}, \boldsymbol{v})$, where $\mathcal{B}\colon V \times V \to V^*$ is a bilinear continuous operator with

 (2a) $\langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{v}), \boldsymbol{v}\rangle_{V^* \times V} = 0$ for all $\boldsymbol{u}, \boldsymbol{v} \in V$;
 (2b) $\mathcal{B}[\cdot]\colon V \to V^*$ is weakly sequentially continuous.

Lemma 1.2.5 ([23, Lemma 10]) *The Navier–Stokes type operator $\mathcal{N}\colon V \to V^*$ is pseudomonotone and (α, ρ)-coercive.*

Proof The coercivity of $\mathcal{N}$ is a consequence of conditions (1) and (2a) of Definition 1.2.4, namely for every $\boldsymbol{v} \in V$, we have

$$\langle \mathcal{N}\boldsymbol{v}, \boldsymbol{v}\rangle = \langle \mathcal{A}\boldsymbol{v}, \boldsymbol{v}\rangle + \langle \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}), \boldsymbol{v}\rangle \geq \alpha \, \|\boldsymbol{v}\|_V^\rho .$$

Now, we show that $\mathcal{N}$ is pseudomonotone. First, the boundedness of $\mathcal{N}$ follows from the facts that $\mathcal{A}$ is bounded and $\mathcal{B}$ is bilinear and continuous. Second, let $\boldsymbol{v}_n \rightharpoonup \boldsymbol{v}$ in $\sigma(V, V^*)$ and $\limsup_{n\to\infty} \langle \mathcal{A}\boldsymbol{v}_n, \boldsymbol{v}_n - \boldsymbol{v}\rangle_{V^* \times V} \leq 0$, and $\boldsymbol{v} \in V$. By (2a) and (2b) of Definition 1.2.4, we have

$$\begin{aligned}\langle \mathcal{B}[\boldsymbol{u}_n], \boldsymbol{u}_n - \boldsymbol{v}\rangle - \langle \mathcal{B}[\boldsymbol{u}], \boldsymbol{u} - \boldsymbol{v}\rangle &= \langle \mathcal{B}[\boldsymbol{u}_n], \boldsymbol{u}_n\rangle - \langle \mathcal{B}[\boldsymbol{u}_n], \boldsymbol{v}\rangle - \langle \mathcal{B}[\boldsymbol{u}], \boldsymbol{u}\rangle + \langle \mathcal{B}[\boldsymbol{u}], \boldsymbol{v}\rangle \\ &= \langle \mathcal{B}[\boldsymbol{u}], \boldsymbol{v}\rangle - \langle \mathcal{B}[\boldsymbol{u}_n], \boldsymbol{v}\rangle\end{aligned}$$

which implies $\lim_{n\to\infty} \langle \mathcal{B}[\boldsymbol{u}_n], \boldsymbol{u}_n - \boldsymbol{v}\rangle = \langle \mathcal{B}[\boldsymbol{u}], \boldsymbol{u} - \boldsymbol{v}\rangle$ for all $\boldsymbol{v} \in V$. Hence, in particular, we have $\lim_{n\to\infty} \langle \mathcal{B}[\boldsymbol{u}_n], \boldsymbol{u}_n - \boldsymbol{u}\rangle = 0$. Thus,

$$\begin{aligned}\limsup_{n\to\infty} \langle \mathcal{A}\boldsymbol{u}_n, \boldsymbol{u}_n - \boldsymbol{u}\rangle &= \limsup_{n\to\infty} \langle \mathcal{A}\boldsymbol{u}_n, \boldsymbol{u}_n - \boldsymbol{u}\rangle + \lim \langle \mathcal{B}[\boldsymbol{u}_n], \boldsymbol{u}_n - \boldsymbol{u}\rangle \\ &= \limsup_{n\to\infty} \langle \mathcal{N}\boldsymbol{u}_n, \boldsymbol{u}_n - \boldsymbol{u}\rangle \leq 0.\end{aligned}$$

From the pseudomonotonicity of $\mathcal{A}$, we obtain

$$\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{u} - \boldsymbol{v}\rangle \leq \liminf_{n\to\infty} \langle \mathcal{A}\boldsymbol{u}_n, \boldsymbol{u}_n - \boldsymbol{v}\rangle$$

for all $\boldsymbol{v} \in V$ which yields $\langle \mathcal{N}\boldsymbol{u}, \boldsymbol{u} - \boldsymbol{v}\rangle \leq \liminf_{n\to\infty} \langle \mathcal{N}\boldsymbol{u}_n, \boldsymbol{u}_n - \boldsymbol{v}\rangle$ for all $\boldsymbol{v} \in V$. □

1.2.2 Orlicz and Orlicz–Sobolev Spaces

We recall definitions of Orlicz and Orlicz–Sobolev spaces and some of their properties (see [22, 46, 55, 68, 80]). A function $\Phi\colon [0, \infty) \to [0, \infty)$ is called an *N-function* if it is convex and such that $\Phi(t) > 0$ for $t > 0$ and $\Phi(t)/t \to 0$

as $t \to 0$, $\Phi(t)/t \to \infty$ as $t \to \infty$. The *complementary function* Φ^* to an N-function Φ is defined by $\Phi^*(v) = \sup\{uv - \Phi(u) \mid u \geq 0\}$ for all $v \geq 0$. It is also an N-function. Furthermore, $(\Phi^*)^* = \Phi$ and the *Young inequality* satisfies $st \leq \Phi(t) + \Phi^*(s)$ for $s, t \in [0, \infty)$. An N-function Φ is said to satisfy the Δ_2-*condition near infinity* denoted by $\Phi \in \Delta_2^\infty$, if there exists $k > 0$ such that $\Phi(2t) \leq k\,\Phi(t)$ for all $t \geq t_0 > 0$. For two N-functions Φ_1 and Φ_2, we say that Φ_2 *dominates* Φ_1 *near infinity* and write $\Phi_1 \prec \Phi_2$, if there exists $k > 0$ such that $\Phi_1(t) \leq \Phi_2(kt)$ for all $t \geq t_0 > 0$. The N-functions Φ_1 and Φ_2 are called *equivalent near infinity* and write $\Phi_1 \sim \Phi_2$ if each dominates the other near infinity. If $\lim_{t\to\infty} \Phi_1(t)/\Phi_2(kt) = 0$ for all $k > 0$, we say that Φ_2 *grows essentially faster near infinity* than Φ_1 and write $\Phi_1 \ll \Phi_2$. If $\Phi_1 \ll \Phi_2$, then $\Phi_1 \prec \Phi_2$.

Let $(\mathcal{O}, \mathfrak{A}, \mu)$ be a positive finite complete measure space and Φ be an N-function. The *Orlicz space* $L_\Phi(\mathcal{O}, \mathbb{R})$ is the space of (equivalence classes of) measurable functions $u\colon \mathcal{O} \to \mathbb{R}$ which satisfy $\int_{\mathcal{O}} \Phi(\lambda\,|u(\omega)|)\,\mathrm{d}\mu(\omega) < \infty$ for some $\lambda > 0$. It is a Banach space with the *Luxemburg norm*

$$\|u\|_{L_\Phi(\mathcal{O})} := \inf\left\{\lambda > 0 \,\middle|\, \int_{\mathcal{O}} \Phi(|u(\omega)|/\lambda)\,\mathrm{d}\mu(\omega) \leq 1\right\}.$$

The space $L_\Phi(\mathcal{O}, \mathbb{R})$ is reflexive if and only if $\Phi, \Phi^* \in \Delta_2^\infty$, and $L_\Phi(\mathcal{O}, \mathbb{R})$ is separable if and only if $\Phi \in \Delta_2^\infty$ and μ is nonatomic. Furthermore, $(L_\Phi(\mathcal{O}, \mathbb{R}))^* = L_{\Phi^*}(\mathcal{O}, \mathbb{R})$ if and only if $\Phi \in \Delta_2^\infty$. The embedding $L_{\Phi_2}(\mathcal{O}, \mathbb{R}) \hookrightarrow L_{\Phi_1}(\mathcal{O}, \mathbb{R})$ holds if and only if $\Phi_1 \prec \Phi_2$ near infinity. It is well known that if $\Phi(t) = t^p$ and $p \in (1, \infty)$ then $L_\Phi(\mathcal{O}, \mathbb{R}) = L^p(\mathcal{O}, \mathbb{R})$. The *Hölder inequality* in Orlicz spaces has the form $\|uv\|_{L^1(\mathcal{O},\mathbb{R})} \leq 2\,\|u\|_{L_\Phi(\mathcal{O},\mathbb{R})}\,\|v\|_{L_{\Phi^*}(\mathcal{O},\mathbb{R})}$ for $u \in L_\Phi(\mathcal{O}, \mathbb{R})$ and $v \in L_{\Phi^*}(\mathcal{O}, \mathbb{R})$.

Given a separable Banach space E and $L_\Phi(\mathcal{O}, \mathbb{R})$, the *Orlicz–Bochner space* $L_\Phi(\mathcal{O}, E)$ is defined as the normed space of (equivalence classes of) strongly measurable functions $u\colon \mathcal{O} \to E$ such that the function $\omega \in \mathcal{O} \mapsto \|u(\omega)\|_E$ belongs to $L_\Phi(\mathcal{O}, \mathbb{R})$ with the norm $\|u\|_{L_\Phi(\mathcal{O},E)} := \|\,\|u(\cdot)\|_E\,\|_{L_\Phi(\mathcal{O})}$. Recall that $u\colon \mathcal{O} \to E$ is said to be a *strongly measurable function* if there exists a sequence (u_n) of simple functions such that $\lim_{n\to\infty} \|u_n(\omega) - u(\omega)\|_E = 0$ for almost all $\omega \in \mathcal{O}$.

Let $(\Omega, \mathfrak{A}, \mathrm{d}\mathbf{x})$ be a measure space with an open bounded set $\Omega \subset \mathbb{R}^n$, $\mathfrak{A} = \mathcal{B}(\Omega)$ (with $\mathcal{B}(\Omega)$ being the Borel σ-algebra on Ω) and the n-dimensional Lebesgue measure $\mathrm{d}\mathbf{x}$ on $\mathcal{B}(\Omega)$. The *Orlicz–Sobolev* space $W^1L_\Phi(\Omega, \mathbb{R}^d)$ is the space of all $\boldsymbol{u} \in L_\Phi(\Omega, \mathbb{R}^d)$ such that $\nabla\boldsymbol{u} \in L_\Phi(\Omega, \mathbb{R}^{n\times d})$, where $\nabla\boldsymbol{u}$ is a matrix-valued function whose all components are distributional partial derivatives of $\boldsymbol{u}$. It is a Banach space endowed with the norm

$$\|\boldsymbol{u}\|_{W^1L_\Phi(\Omega,\mathbb{R}^d)} = \|\boldsymbol{u}\|_{L_\Phi(\Omega,\mathbb{R}^d)} + \|\nabla\boldsymbol{u}\|_{L_\Phi(\Omega,\mathbb{R}^{n\times d})}\,.$$

The symbol $C_c^\infty(\Omega, \mathbb{R}^d)$ means the space of all C^∞-functions $u\colon \Omega \to \mathbb{R}^d$ with a compact support in Ω. If Ω has finite measure and $\Phi \in \Delta_2^\infty$, then

$\mathring{W}^1 L_\Phi(\Omega, \mathbb{R}^d)$ is defined as the norm-closure of $C_c^\infty(\Omega, \mathbb{R}^d)$ in $W^1 L_\Phi(\Omega, \mathbb{R}^d)$ and $W^{-1} L_\Phi(\Omega, \mathbb{R}^d) := (\mathring{W}^1 L_\Phi(\Omega, \mathbb{R}^d))^*$. Furthermore, if $\Phi, \Phi^* \in \Delta_2^\infty$, then the spaces $W^1 L_\Phi(\Omega, \mathbb{R}^d)$ and $W^{-1} L_\Phi(\Omega, \mathbb{R}^d)$ are reflexive and separable. It is well known that if $\Phi(t) = t^p$ and $p \in (1, \infty)$ then $W^1 L_\Phi(\Omega, \mathbb{R}^d) = W^{1,p}(\Omega, \mathbb{R}^d)$.

Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain, that is, a bounded connected open set in $\mathbb{R}^d$ with a Lipschitz boundary $\partial\Omega = \Gamma$. We recall (see [22]) that if an N-function Φ satisfies the following conditions

$$\int_0^1 \frac{\Phi^{-1}(t)}{t^{1+1/d}}\, \mathrm{d}t < \infty, \qquad \int_1^\infty \frac{\Phi^{-1}(t)}{t^{1+1/d}}\, \mathrm{d}t = \infty,$$

then the *Sobolev conjugate N-function* Φ_* of Φ is defined by

$$\Phi_*^{-1}(t) = \int_0^t \frac{\Phi^{-1}(\tau)}{\tau^{1+1/d}}\, \mathrm{d}\tau, \qquad t \geq 0.$$

Since $\Phi \ll \Phi_*$ (cf. [36, Lemma 4.14]), it follows from [15, Theorem 3] that

$$W^1 L_\Phi(\Omega, \mathbb{R}^m) \hookrightarrow\hookrightarrow L_\Phi(\Omega, \mathbb{R}^m). \tag{1.1}$$

By [22, Theorem 2.3], it follows that as long as $\Phi \in \Delta_2^\infty$, then the space $C^\infty(\overline{\Omega}, \mathbb{R}^m)$ is dense in $W^1 L_\Phi(\Omega, \mathbb{R}^m)$ with respect to the norm convergence. Hence, by Fougères [27] (see also [37]), there exists a unique linear continuous operator

$$\gamma\colon W^1 L_\Phi(\Omega, \mathbb{R}^m) \to L_\Phi(\Gamma, \mathbb{R}^m)$$

such that $\gamma \boldsymbol{u} = \boldsymbol{u}|_\Gamma$ for all $\boldsymbol{u} \in C^\infty(\overline{\Omega}, \mathbb{R}^m)$ and the kernel of γ is $\mathring{W}^1 L_\Phi(\Omega, \mathbb{R}^m)$. The function $\gamma \boldsymbol{u}$ is called the *trace* of the function $\boldsymbol{u}$ on $\partial\Omega = \Gamma$ and the operator γ is called the *trace operator*. By [22, Theorem 3.8 and Corollary 3.3] we have a compact embedding for the trace operator, i.e., if $\Phi \ll (\Phi_*)^{1-1/d}$ with $d > 1$ then

$$W^1 L_\Phi(\Omega, \mathbb{R}^m) \hookrightarrow\hookrightarrow L_\Phi(\Gamma, \mathbb{R}^m). \tag{1.2}$$

Theorem 1.2.6 (Korn's Inequality, [16, Corollary 3.6], [6, Theorem 1.1]) *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain and Φ be an N-function. There exists $c_K > 0$ such that*

$$\|\nabla \boldsymbol{u}\|_{L_\Phi(\Omega, \mathbb{R}^{d\times d})} \leq c_K \, \|\mathbf{D}(\boldsymbol{u})\|_{L_\Phi(\Omega, \mathbb{S}^d)}, \qquad \forall \boldsymbol{u} \in \mathring{W}^1 L_\Phi(\Omega, \mathbb{R}^d)$$

if and only if $\Phi, \Phi^* \in \Delta_2^\infty$, where $\mathbf{D}(\boldsymbol{u}) = \frac{1}{2}\left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^\top\right)$ denotes the symmetric part of the gradient $\nabla \boldsymbol{u}$.

Lemma 1.2.7 ([66, Lemma A.5]) *Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain. For each N-function Ψ such that $\Psi, \Psi^* \in \Delta_2^\infty$ the following inequalities hold:*

$$\int_\Omega \Psi(|u(\mathbf{x})|)\,\mathrm{d}\mathbf{x} \geq \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{a_\Psi^\infty-\alpha} \quad \text{with } \|u\|_{L_\Psi(\Omega,\mathbb{R})} > 1, \tag{1.3}$$

$$\int_\Omega \Psi(|u(\mathbf{x})|)\,\mathrm{d}\mathbf{x} \leq \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{a_\Psi^\infty-\alpha} \quad \text{with } \|u\|_{L_\Psi(\Omega,\mathbb{R})} \leq 1, \tag{1.4}$$

$$\int_\Omega \Psi(|u(\mathbf{x})|)\,\mathrm{d}\mathbf{x} \leq \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{b_\Psi^\infty+\mu} \quad \text{with } \|u\|_{L_\Psi(\Omega,\mathbb{R})} > 1, \tag{1.5}$$

$$\int_\Omega \Psi(|u(\mathbf{x})|)\,\mathrm{d}\mathbf{x} \geq \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{b_\Psi^\infty+\mu} \quad \text{with } \|u\|_{L_\Psi(\Omega,\mathbb{R})} \leq 1, \tag{1.6}$$

where $a_\Psi^\infty := \liminf_{t\to\infty} \frac{t\Psi'(t)}{\Psi(t)}$ and $b_\Psi^\infty := \limsup_{t\to\infty} \frac{t\Psi'(t)}{\Psi(t)}$ satisfy $1 < a_\Psi^\infty \leq b_\Psi^\infty < \infty$; $\mu > 0$ and $\alpha \in (0, a_\Psi^\infty)$.

Proof It is known (see [80, Corollary 4, p. 26]) that $\Psi, \Psi^* \in \Delta_2^\infty$ if and only if

$$1 < a_\Psi^\infty \leq b_\Psi^\infty < \infty, \tag{1.7}$$

where the numbers a_Ψ^∞ and b_Ψ^∞ are called the Simonenko indices of the N-function Ψ (see [54, p. 20]).

Firstly, we prove that (1.3) and (1.4) hold. By the definition of a_Ψ^∞, there exist $\alpha \in (0, a_\Psi^\infty)$ and $t_1 \geq t_0 > 0$ such that

$$\frac{t\Psi'(t)}{\Psi(t)} \geq a_\Psi^\infty - \alpha, \quad \forall t \geq t_1. \tag{1.8}$$

Hence, for $\sigma \in (1, \infty)$, we obtain

$$\log \frac{\Psi(\sigma t)}{\Psi(t)} = \int_t^{\sigma t} \frac{\Psi'(s)}{\Psi(s)}\mathrm{d}s \geq \int_t^{\sigma t} \frac{a_\Psi^\infty - \alpha}{s}\mathrm{d}s = (a_\Psi^\infty - \alpha)\log\frac{\sigma t}{t}.$$

Therefore,

$$\Psi(\sigma t) \geq \sigma^{a_\Psi^\infty-\alpha}\Psi(t), \quad \forall t \geq t_1. \tag{1.9}$$

From (1.9) with $\|u\|_{L_\Psi(\Omega,\mathbb{R})} > 1$ and [80, Proposition 6, p.77], we have

$$\begin{aligned}\int_\Omega \Psi\left(|u(\mathbf{x})|\right)\,\mathrm{d}\mathbf{x} &= \int_\Omega \Psi\left(\|u\|_{L_\Psi(\Omega,\mathbb{R})}\frac{|u(\mathbf{x})|}{\|u\|_{L_\Psi(\Omega,\mathbb{R})}}\right)\mathrm{d}\mathbf{x} \\ &\geq \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{a_\Psi^\infty-\alpha}\int_\Omega \Psi\left(\frac{|u(\mathbf{x})|}{\|u\|_{L_\Psi(\Omega,\mathbb{R})}}\right)\mathrm{d}\mathbf{x} = \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{a_\Psi^\infty-\alpha}.\end{aligned} \tag{1.10}$$

On the other hand, by (1.8) for $\sigma \in (0, 1]$, it follows

$$\log \frac{\Psi(t/\sigma)}{\Psi(t)} = \int_t^{t/\sigma} \frac{\Psi'(s)}{\Psi(s)} \mathrm{d}s \geq \int_t^{t/\sigma} \frac{a_\Psi^\infty - \alpha}{s} \mathrm{d}s = (a_\Psi^\infty - \alpha) \log \frac{t/\sigma}{t}.$$

Therefore,

$$\Psi(t) \leq \sigma^{a_\Psi^\infty - \alpha} \Psi(t/\sigma), \quad \forall t \geq t_1. \tag{1.11}$$

By (1.11) for $\|u\|_{L_\Psi(\Omega,\mathbb{R})} \leq 1$ and [80, Proposition 6, p.77], we obtain

$$\int_\Omega \Psi\left(|u(\mathbf{x})|\right) \, \mathrm{d}\mathbf{x} \leq \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{a_\Psi^\infty - \alpha} \int_\Omega \Psi\left(\frac{|u(\mathbf{x})|}{\|u\|_{L_\Psi(\Omega,\mathbb{R})}}\right) \mathrm{d}\mathbf{x} = \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{a_\Psi^\infty - \alpha}. \tag{1.12}$$

By (1.10) and (1.12) the proof is complete.

Secondly, we prove that (1.5) and (1.6) hold. By the definition of b_Φ^∞, there exist $\mu > 0$ and $t_1 \geq t_0 > 0$ such that

$$\frac{t\Psi'(t)}{\Psi(t)} \leq b_\Psi^\infty + \mu, \quad \forall t \geq t_1. \tag{1.13}$$

Hence, for $\sigma \in (1, \infty)$, we have

$$\log \frac{\Psi(\sigma t)}{\Psi(t)} = \int_t^{\sigma t} \frac{\Psi'(s)}{\Psi(s)} \mathrm{d}s \leq \int_t^{\sigma t} \frac{b_\Psi^\infty + \mu}{s} \mathrm{d}s = (b_\Psi^\infty + \mu) \log \frac{\sigma t}{t}.$$

Therefore,

$$\Psi(\sigma t) \leq \sigma^{b_\Psi^\infty + \mu} \Psi(t), \quad \forall t \geq t_1. \tag{1.14}$$

Using (1.14) with $\|u\|_{L_\Psi(\Omega,\mathbb{R})} > 1$ and [80, Proposition 6, p.77], we obtain

$$\begin{aligned} \int_\Omega \Psi\left(|u(\mathbf{x})|\right) \, \mathrm{d}\mathbf{x} &= \int_\Omega \Psi\left(\|u\|_{L_\Psi(\Omega,\mathbb{R})} \frac{|u(\mathbf{x})|}{\|u\|_{L_\Psi(\Omega,\mathbb{R})}}\right) \mathrm{d}\mathbf{x} \\ &\leq \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{b_\Psi^\infty + \mu} \int_\Omega \Psi\left(\frac{|u(\mathbf{x})|}{\|u\|_{L_\Psi(\Omega,\mathbb{R})}}\right) \mathrm{d}\mathbf{x} = \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{b_\Psi^\infty + \mu}. \end{aligned} \tag{1.15}$$

On the other hand, by (1.13) for $\sigma \in (0, 1]$, we deduce

$$\log \frac{\Psi(t/\sigma)}{\Psi(t)} = \int_t^{t/\sigma} \frac{\Psi'(s)}{\Psi(s)} \mathrm{d}s \leq \int_t^{t/\sigma} \frac{b_\Psi^\infty + \mu}{s} \mathrm{d}s = (b_\Psi^\infty + \mu) \log \frac{t/\sigma}{t}.$$

Therefore,

$$\Psi(t) \geq \sigma^{b_\Psi^\infty+\mu}\Psi(t/\sigma), \quad \forall t \geq t_1. \tag{1.16}$$

By (1.16) for $\|u\|_{L_\Psi(\Omega,\mathbb{R})} \leq 1$ and [80, Proposition 6, p.77], we get

$$\int_\Omega \Psi\left(|u(\mathbf{x})|\right)\, \mathrm{d}\mathbf{x} \geq \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{b_\Psi^\infty+\mu} \int_\Omega \Psi\left(\frac{|u(\mathbf{x})|}{\|u\|_{L_\Psi(\Omega,\mathbb{R})}}\right) \mathrm{d}\mathbf{x} = \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{b_\Psi^\infty+\mu}. \tag{1.17}$$

By (1.17) and (1.15) the proof is complete. □

Corollary 1.2.8 ([66, Corollary A.6]) *Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain. For each N-function Ψ such that $\Psi^* \in \Delta_2^\infty$ the following inequality holds:*

$$\int_\Omega \Psi(|u(\mathbf{x})|)\, \mathrm{d}\mathbf{x} \leq c_1(1 + \|u\|_{L_\Psi(\Omega,\mathbb{R})}^\rho),$$

for all $\rho > 1$ and $c_1 > 0$.

Proof By (1.14) with $\sigma = 1 + \|u\|_{L_\Psi(\Omega,\mathbb{R})}$ and [80, Proposition 6, p.77], we deduce

$$\begin{aligned}
\int_\Omega \Psi\left(|u(\mathbf{x})|\right)\, \mathrm{d}\mathbf{x} &= \int_\Omega \Psi\left((1 + \|u\|_{L_\Psi(\Omega,\mathbb{R})})\frac{|u(\mathbf{x})|}{1 + \|u\|_{L_\Psi(\Omega,\mathbb{R})}}\right) \mathrm{d}\mathbf{x} \\
&\leq (1 + \|u\|_{L_\Psi(\Omega,\mathbb{R})})^{b_\Psi^\infty+\mu} \int_\Omega \Psi\left(\frac{|u(\mathbf{x})|}{1 + \|u\|_{L_\Psi(\Omega,\mathbb{R})}}\right) \mathrm{d}\mathbf{x} \\
&\leq (1 + \|u\|_{L_\Psi(\Omega,\mathbb{R})})^{b_\Psi^\infty+\mu} \int_\Omega \Psi\left(\frac{|u(\mathbf{x})|}{\|u\|_{L_\Psi(\Omega,\mathbb{R})}}\right) \mathrm{d}\mathbf{x} \\
&\leq 2^{b_\Psi^\infty+\mu-1}(1 + \|u\|_{L_\Psi(\Omega,\mathbb{R})}^{b_\Psi^\infty+\mu}),
\end{aligned}$$

where $b_\Psi > 1$ and $\mu > 0$. □

1.2.3 Generalized Gradient

We recall now some definitions and results from nonsmooth analysis [17].

Definition 1.2.9 (Lipschitz Function) Let U be a subset of a Banach space E. A function $f : U \to \mathbb{R}$ is said to be *Lipschitz* on U, if there exists $L > 0$ such that

$$|f(y) - f(z)| \leq L\, \|y - z\|_E\,, \quad \forall y, z \in U.$$

The constant L is called the *Lipschitz constant*.

Definition 1.2.10 (Locally Lipschitz Function) Let U be a subset of a Banach space E. A function $f: U \to \mathbb{R}$ is said to be *locally Lipschitz* on U, if for all $x \in U$ there exist a neighborhood $N(x)$ and $L_x > 0$ such that

$$|f(y) - f(z)| \le L_x \|y - z\|_E\,, \quad \forall y, z \in N(x).$$

The constant L_x is called the *Lipschitz constant*.

Definition 1.2.11 (Generalized Directional Derivative) Let U be an open subset of a Banach space E. If $f: U \to \mathbb{R}$ is locally Lipschitz, then f has the *generalized directional derivative* at the point $x \in U$ in the direction $v \in U$, denoted $f^0(x; v)$, which is defined by

$$f^0(x; v) = \limsup_{y \to x, \lambda \to 0^+} \lambda^{-1}(f(y + \lambda v) - f(y)).$$

Definition 1.2.12 (Regular Function) A function f is said to be *regular* (in the sense of Clarke) at x if the directional derivative $f'(x; v)$ of f at x along v exists and $f'(x; v) = f^0(x, v)$ for every $v \in U$.

Definition 1.2.13 (Generalized Gradient) The *Clarke subdifferential* or the *generalized gradient* in the sense of Clarke of f at x is the set

$$\partial f(x) = \left\{ \zeta \in E^* \,\middle|\, \langle \zeta, v \rangle \le f^0(x; v),\ \forall v \in U \right\}.$$

Theorem 1.2.14 (Generalized Gradient) *The Clarke subdifferential $\partial f(x)$ is a nonempty convex compact set in the weak star topology $\omega^* = \sigma(E^*, E)$. The multifunction $\partial f: U \multimap E^*_{\omega^*}$ is upper semicontinuous. Furthermore, if f is continuous (Fréchet) differentiable, i.e., $f \in C^1$, then the Clarke subdifferential $\partial f(x)$ reduces to a singleton, namely $\partial f(x) = \{f'(x)\}$. Also, for every $v \in U$ we have $f^0(x; v) = \max\{\langle \zeta, v \rangle \mid \zeta \in \partial f(x)\}$.*

Next, we present a result on an integral representation of the Clarke subdifferential of locally Lipschitz integral functionals defined on the Orlicz–Bochner space. We adopt the following conditions from [71, Theorem 4.3], [72, Corollary 6.2, Lemma 6.1].

Hypotheses 1.2.15 Let $(\mathcal{O}, \mathfrak{A}, \mu)$ be a positive finite complete measure space and E be a separable Banach space. Assume that $\Phi, \Phi^*: [0, \infty) \to [0, \infty)$ are a pair of complementary N-functions and $g: \Omega \times E \to \mathbb{R}$ is a Carathéodory function satisfying:

- (Φ1) $\Phi, \Phi^* \in \Delta_2^\infty$;
- (Φ2) $g(x, \cdot)$ is locally Lipschitz for a.e. $x \in \Omega$;
- (Φ3) For some $\kappa > 0$ there exist $a_\kappa \in L^1(\Omega, [0, \infty))$ and positive constants b_κ and d_κ such that $\Phi^*(\|a^*\|_{E^*}/d_\kappa) \le a_\kappa(x) + b_\kappa \Phi(\|a\|_E/\kappa)$ holds for all $(a, a^*) \in \mathrm{Gr}(\partial g)$ and a.e. $x \in \Omega$.

Remark 1.2.16 By hypothesis ($\Phi 1$), the Orlicz spaces $L_\Phi(\Omega, E)$ are separable and reflexive. Hypothesis ($\Phi 3$) is natural for applications. In fact, it implies that the set-valued superposition operator $\mathcal{N}_{\partial g}: L_\Phi(\Omega, E) \multimap L_{\Phi^*}(\Omega, E^*_{\omega^*})$ is bounded on $B_{L_\Phi(\Omega,E)}(0, \kappa)$. On the other hand, it follows from [70] that if the measure μ is continuous, ∂g is Carathéodory multifunction and $\mathcal{N}_{\partial g}: L_\Phi(\Omega, E) \multimap L_{\Phi^*}(\Omega, E^*_{\omega^*})$ is bounded on $B_{L_\Phi(\Omega,E)}(0, \kappa)$, then hypothesis ($\Phi 3$) is satisfied. Finally, note that each one of the following conditions implies hypothesis ($\Phi 3$):

(1) For some $\kappa > 0$ there exist positive constants b_κ, d_κ and functions $a_\kappa \in L^1(\Omega, [0, \infty))$ and $h_\kappa: \Omega \times [0, \infty) \to [0, \infty)$ such that

 (a) $|g(x, u) - g(x, v)| \le h_\kappa(x, \|u\|_E + \|v\|_E)\, \|u - v\|_E$ for all $u, v \in E$ and for a.e. $x \in \Omega$;
 (b) $\Phi^*(h_\kappa(x, \alpha)/d_\kappa) \le a_\kappa(x) + b_\kappa \Phi(\alpha/\kappa)$ for a.e. $x \in \Omega$ and $\alpha \in [0, \infty)$.

(2) For some $\kappa > 0$ there exist $c_\kappa \in L_{\Phi^*}(\Omega, [0, \infty))$ and a positive constant b_κ such that $\|a^*\|_{E^*} \le c_\kappa(x) + b_\kappa \Phi'_+(\|a\|_E)$ holds for all $(a, a^*) \in \mathrm{Gr}(\partial g)$ and a.e. $x \in \Omega$, where Φ'_+ is the right derivative of the N-function Φ.

where Φ and g are such as in Hypotheses 1.2.15.

The following result is an integral representation of the Clarke subdifferential of locally Lipschitz integral functionals defined on the Orlicz space (see [71, Theorem 4.3], [72, Corollary 6.2] and [72, Lemma 6.1]).

Theorem 1.2.17 *Under Hypotheses 1.2.15, if the functional*

$$G(u) := \int_\Omega g(x, u(x))\, dx \quad \textit{for } u \in L_\Phi(\Omega, E),$$

is finite at least for one $\bar{u}$ *in* $\overline{B}_{L_\Phi(\Omega,E)}(0, \kappa/2)$*, then*

(1) *G is Lipschitz on* $\overline{B}_{L_\Phi(\Omega,E)}(0, \kappa/2)$*;*
(2) $G^0(u; v) \le \int_\Omega g^0(x, u(x); v(x))\, dx$ *for* $u \in B_{L_\Phi(\Omega,E)}(0, \kappa/2)$ *and* $v \in L_\Phi(\Omega, E)$*;*
(3) $\partial G(u) \subset \mathcal{N}_{\partial g}(u)$ *for all* $u \in B_{L_\Phi(\Omega,E)}(0, \kappa/2)$*, where the multivalued superposition operator* $\mathcal{N}_{\partial g}: L_\Phi(\Omega, E) \multimap L_{\Phi^*}(\Omega, E^*_{\omega^*})$ *is bounded, that is, if* $\zeta \in \partial J(u) \subset L_{\Phi^*}(\Omega, E^*_{\omega^*})$*, then*

$$\langle \zeta, v \rangle = \int_\Omega \xi(x) \cdot v(x)\, dx$$

 for all $v \in L_\Phi(\Omega, E)$ *and for some* $\xi \in \mathcal{N}_{\partial g}(y) = \mathrm{Sel}\, \partial g(\cdot, y(\cdot))$*;*
(4) *if additionally the function* $g(x, \cdot)$ *is regular (in Clarke's sense) at* $u(x)$ *for a.e.* $x \in \Omega$*, then the functional* G *is regular at* u *and* $\partial G(u) = \mathcal{N}_{\partial g}(u)$*.*

1.3 Subdifferential Operator Inclusions

In this section we prove existence and uniqueness of solution in a reflexive Orlicz–Sobolev space to a subdifferential operator inclusion with the Clarke subdifferential operator and the Navier–Stokes type operator.

Problem 1.3.1 Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain, $V := W^1 L_\Phi(\Omega, \mathbb{R}^m)$ and $Y := L_\Phi(\Gamma, \mathbb{R}^m)$ be reflexive and separable. Assume that $V \hookrightarrow\hookrightarrow Y$ and $\boldsymbol{f} \in V^*$. Find $\boldsymbol{u} \in V$ such that

$$\mathcal{N}\boldsymbol{u} + \gamma^* \partial J(\gamma \boldsymbol{u}) \ni \boldsymbol{f},$$

where $\mathcal{N} \colon V \to V^*$ is the Navier–Stokes type operator, $\gamma \colon V \to Y$ is the trace operator, $\partial J \colon Y \multimap Y^*$ is the set-valued subdifferential operator, and $\gamma^* \colon Y^* \to V^*$ is the adjoint operator to γ.

We complete the statement of Problem 1.3.1 with the following definition.

Definition 1.3.2 An element $\boldsymbol{u} \in V$ is solution to Problem 1.3.1 if and only if there exists $\boldsymbol{\eta} \in V^*$ such that $\mathcal{N}\boldsymbol{u} + \boldsymbol{\eta} = \boldsymbol{f}$ and $\boldsymbol{\eta} \in \gamma^* \partial J(\gamma \boldsymbol{u})$.

Hypotheses 1.3.3 Let $Y := L_\Phi(\Gamma, \mathbb{R}^m)$ and $J \colon Y \to \mathbb{R}$ be a functional such that:

(I1) J is well-defined and Lipschitz on bounded subsets of Y;
(I2) $\|\partial J(\boldsymbol{y})\|_{Y^*} \le c_2 + c_3 \|\boldsymbol{y}\|_Y^{\rho-1}$ for all $\boldsymbol{y} \in Y$ with $1 < \rho < \infty$ and $c_2, c_3 > 0$.

The existence and uniqueness result in study of Problem 1.3.1 reads as follows.

Theorem 1.3.4 *Under Hypotheses 1.3.3, Problem 1.3.1 has a solution $\boldsymbol{u} \in V$ provided $\alpha > c_6$, and*

(i1) *the solution satisfies the estimate $\|\boldsymbol{u}\|_V \le C$ with $C := \left(\frac{\|\boldsymbol{f}\|_{V^*}+c_5}{\alpha - c_6}\right)^{1/(\rho-1)}$;*
(i2) *the solution is unique if, in addition, the operator $\mathcal{A}$ is strongly monotone with the positive constant $m_{\mathcal{A}}$ satisfying the smallness condition:*

$$m_{\mathcal{A}} > 2c_2 \|\gamma\|^2_{V \to Y} + 2c_3 \frac{\|\boldsymbol{f}\|_{V^*}+c_5}{\alpha - c_6} \|\gamma\|^{\rho+1}_{V \to Y} + c_4 \left(\frac{\|\boldsymbol{f}\|_{V^*}+c_5}{\alpha - c_6}\right)^{1/(\rho-1)};$$

where $\rho \in (1, \infty)$, α is the coercivity constant of the operator $\mathcal{A}$, c_4 is the continuity constant of the trilinear form b associated with the operator $\mathcal{B}$, $c_5 := c_2 \|\gamma\|_{V \to Y}$ and $c_6 := c_3 \|\gamma\|^{\rho}_{V \to Y}$ with $c_2, c_3 > 0$ from hypothesis (I2).

Proof We apply the surjectivity result for pseudomonotone maps [8, Theorem 3]. To this end, we define a multivalued operator $\mathsf{G} \colon V \multimap V^*$ by

$$\mathsf{G}(\boldsymbol{v}) = \mathcal{N}\boldsymbol{v} + \mathsf{F}(\boldsymbol{v}), \quad \forall \boldsymbol{v} \in V,$$

where $\mathsf{F}\colon V \multimap V^*$ is the multivalued operator given by

$$\mathsf{F}(\boldsymbol{v}) = \gamma^* \partial J(\gamma \boldsymbol{v}), \quad \forall \boldsymbol{v} \in V. \tag{1.18}$$

We prove that the multivalued operator $\mathsf{G}\colon V \multimap V^*$ is pseudomonotone and coercive. We divide this proof into Steps 1.3.1 to 1.3.2.

Step 1.3.1 We show that F *has convex and weakly compact values*. By hypothesis (I1) and [17, Proposition 2.1.2], we obtain that $\mathsf{F}(\boldsymbol{v})$ are nonempty and convex for all $\boldsymbol{v} \in V$. In order to show that the values of F are weakly compact, let $\boldsymbol{v} \in V$ and $(\boldsymbol{\eta}_n)$ be a sequence in $\mathsf{F}(\boldsymbol{v})$. Thus $\boldsymbol{\eta}_n = \gamma^* \boldsymbol{\zeta}_n$ with $\boldsymbol{\zeta}_n \in \partial J(\gamma \boldsymbol{v})$. Since $\partial J(\gamma \boldsymbol{v})$ is a weakly compact subset of $Y^*_{\omega^*}$, there exists a subsequence $\boldsymbol{\zeta}_{n_j}$ and $\boldsymbol{\zeta} \in \partial J(\gamma \boldsymbol{v})$ such that $\boldsymbol{\zeta}_{n_j} \rightharpoonup \boldsymbol{\zeta}$ in $\sigma(Y^*, Y)$. Since Y is a separable and reflexive Banach space, it follows that $\Phi, \Phi^* \in \Delta_2^\infty$ and $(L_{\Phi^*}(\Gamma, \mathbb{R}^m))^* = (L_{(\Phi^*)^*}(\Gamma, \mathbb{R}^m)) = L_\Phi(\Gamma, \mathbb{R}^m)$ with equivalent norms (see, e.g., [68, Theorem 8.17]). Therefore, $\boldsymbol{\zeta}_{n_j} \rightharpoonup \boldsymbol{\zeta}$ in $\sigma(Y^*, (Y^*)^*)$. Since V is a separable and reflexive Banach space, from the continuity of the operator $\gamma^*\colon Y^* \to V^*$, we infer that $\boldsymbol{\eta}_{n_j} = \gamma^* \boldsymbol{\zeta}_{n_j} \rightharpoonup \gamma^* \boldsymbol{\zeta} := \boldsymbol{\eta}$ in $\sigma(V^*, (V^*)^*)$. Thus $\boldsymbol{\eta} = \gamma^* \boldsymbol{\zeta}$ and $\boldsymbol{\zeta} \in \partial J(\gamma \boldsymbol{v})$. So $\boldsymbol{\eta} \in \mathsf{F}(\boldsymbol{v})$.

Step 1.3.2 We show that F *is sequentially strongly-weakly closed*. Indeed, let $(\boldsymbol{v}_n)$ and $(\boldsymbol{\eta}_n)$ be sequences such that $\boldsymbol{\eta}_n \in \mathsf{F}(\boldsymbol{v}_n)$ together with $\boldsymbol{v}_n \to \boldsymbol{v} \in V$ and $\boldsymbol{\eta}_n \rightharpoonup \boldsymbol{v}^*$ in $\sigma(V^*, V)$. We prove that $\boldsymbol{v}^* \in \mathsf{F}(\boldsymbol{v})$. As $\boldsymbol{\eta}_n \in \mathsf{F}(\boldsymbol{v}_n)$, we have

$$\boldsymbol{\eta}_n = \gamma^* \boldsymbol{\zeta}_n \text{ and } \boldsymbol{\zeta}_n \in \partial J(\gamma \boldsymbol{v}_n) \subset Y^*. \tag{1.19}$$

By $\boldsymbol{v}_n \to \boldsymbol{v} \in V$ and $\gamma\colon V \hookrightarrow\hookrightarrow Y$, we obtain

$$\gamma \boldsymbol{v}_n \to \gamma \boldsymbol{v} \text{ in } Y. \tag{1.20}$$

Hence, $(\gamma \boldsymbol{\eta}_n)$ lies in a bounded subset of Y. It follows from (1.19) and hypothesis (I2) that $(\boldsymbol{\zeta}_n)$ remains in a bounded subset of Y^*. The Banach–Alaoglu theorem shows that

$$\boldsymbol{\zeta}_{n_j} \rightharpoonup \boldsymbol{\zeta} \text{ in } \sigma(Y^*, Y). \tag{1.21}$$

Since the set-valued map ∂J is sequentially strongly-weakly closed (see [17, Propsition 2.1.5]), we infer that ∂J is closed in $Y \times Y^*_{w^*}$ topology. Hence, by (1.19) and (1.21) we obtain

$$\boldsymbol{\zeta}_{n_j} \rightharpoonup \boldsymbol{\zeta} \text{ in } \sigma(Y^*, Y) \text{ and } \boldsymbol{\zeta} \in \partial J(\gamma \boldsymbol{v}) \subset Y^*, \tag{1.22}$$

Letting $n_j \to \infty$ in (1.19), from $\boldsymbol{\eta}_{n_j} \rightharpoonup \boldsymbol{v}^*$ in $\sigma(V^*, V)$ along with (1.20) and (1.22) we obtain $\boldsymbol{v}^* = \gamma^* \boldsymbol{\zeta}$ and $\boldsymbol{\zeta} \in \partial J(\gamma \boldsymbol{v})$, which gives $\boldsymbol{v}^* \in \mathsf{F}(\boldsymbol{v})$.

Step 1.3.3 We show that $\|\mathsf{F}(\boldsymbol{v})\|_{V^*} \leq c_5 + c_6 \|\boldsymbol{v}\|_V^{\rho-1}$ *for all* $\boldsymbol{v} \in V$ *with* $\rho \in (1, \infty)$, $c_5 := c_2 \|\gamma\|_{V \to Y} > 0$ *and* $c_6 := c_3 \|\gamma\|^\rho_{V \to Y} > 0$, *where* c_2, c_3 *are constants in* *hypothesis* (I2). Let $\boldsymbol{\eta} \in \mathsf{F}(\boldsymbol{v})$ for $\boldsymbol{v} \in V$. Thus $\boldsymbol{\eta} = \gamma^*\boldsymbol{\zeta}$ and $\boldsymbol{\zeta} \in \partial J(\gamma \boldsymbol{v})$. Hence,

$$|\langle \boldsymbol{\eta}, \boldsymbol{v}\rangle_{V^*\times V}| = \left|\langle \gamma^*\boldsymbol{\zeta}, \boldsymbol{v}\rangle_{V^*\times V}\right| \le \left\|\gamma^*\right\|_{Y^*\to V^*} \|\boldsymbol{\zeta}\|_{Y^*} \|\boldsymbol{v}\|_V , \quad \forall \boldsymbol{v} \in V. \tag{1.23}$$

By (1.23), hypothesis (I2) and $\|\gamma^*\|_{Y^*\to V^*} = \|\gamma\|_{V\to Y}$, we obtain

$$\begin{aligned} \|\boldsymbol{\eta}\|_{V^*} &= \left\|\gamma^*\boldsymbol{\zeta}\right\|_{V^*} \le \left\|\gamma^*\right\|_{Y^*\to V^*} \left(c_2 + c_3 \|\gamma \boldsymbol{v}\|_Y^{\rho-1}\right) \\ &\le \left\|\gamma^*\right\|_{Y^*\to V^*} \left(c_2 + c_3 \|\gamma\|_{V\to Y}^{\rho-1} \|\boldsymbol{v}\|_V^{\rho-1}\right) \\ &= \|\gamma\|_{V\to Y} \left(c_2 + c_3 \|\gamma\|_{V\to Y}^{\rho-1} \|\boldsymbol{v}\|_V^{\rho-1}\right) \le c_5 + c_6 \|\boldsymbol{v}\|_V^{\rho-1} . \end{aligned} \tag{1.24}$$

Step 1.3.4 We show that F *is a pseudomonotone operator*. Because of Step 1.3.1 and the Mazur theorem (see, e.g., [67, Theorem 1.33]), the operator $\mathsf{F}\colon V \multimap V^*$ on a reflexive Banach space V has closed and convex values. By [8, Proposition 4], it is enough to prove that F is a generalized pseudomonotone operator. To this end, let $(\boldsymbol{v}_n)$ and $(\boldsymbol{\eta}_n)$ be sequences such that $\boldsymbol{\eta}_n \in \mathsf{F}(\boldsymbol{v}_n)$ together with $\boldsymbol{v}_n \rightharpoonup \boldsymbol{v}$ in $\sigma(V, V^*)$, $\boldsymbol{\eta}_n \rightharpoonup \boldsymbol{v}^*$ in $\sigma(V^*, V)$ and $\limsup_{n\to\infty} \langle \boldsymbol{\eta}_n, \boldsymbol{v}_n - \boldsymbol{v}\rangle_{V^*\times V} \le 0$. We prove that $\boldsymbol{v}^* \in \mathsf{F}(\boldsymbol{v})$ and $\langle \boldsymbol{\eta}_n, \boldsymbol{v}_n\rangle_{V^*\times V} \to \langle \boldsymbol{v}^*, \boldsymbol{v}\rangle_{V^*\times V}$. Since $\boldsymbol{\eta}_n \in \mathsf{F}(\boldsymbol{v}_n)$, we have

$$\boldsymbol{\eta}_n = \gamma^*\boldsymbol{\zeta}_n \text{ and } \boldsymbol{\zeta}_n \in \partial J(\gamma \boldsymbol{v}_n) \subset Y^*. \tag{1.25}$$

By $\boldsymbol{v}_n \rightharpoonup \boldsymbol{v}$ in $\sigma(V, V^*)$ and $\gamma\colon V \hookrightarrow\hookrightarrow Y$, we obtain

$$\gamma \boldsymbol{v}_n \to \gamma \boldsymbol{v} \text{ in } Y. \tag{1.26}$$

Hence, $(\gamma \boldsymbol{\eta}_n)$ is in a bounded subset of Y. It follows from (1.25) and hypothesis (I2) that $(\boldsymbol{\zeta}_n)$ lies in a bounded subset of Y^*. The Banach–Alaoglu theorem shows that

$$\boldsymbol{\zeta}_{n_j} \rightharpoonup \boldsymbol{\zeta} \text{ in } \sigma(Y^*, Y). \tag{1.27}$$

As before, since the map ∂J is sequentially strongly-weakly closed (see [17, Propsition 2.1.5]), we infer that ∂J is closed in $Y \times Y^*_{w^*}$ topology. By (1.25) to (1.27), we obtain

$$\boldsymbol{\zeta}_{n_j} \rightharpoonup \boldsymbol{\zeta} \text{ in } \sigma(Y^*, Y) \text{ and } \boldsymbol{\zeta} \in \partial J(\gamma \boldsymbol{v}) \subset Y^*. \tag{1.28}$$

Letting $n_j \to \infty$ in (1.25), from $\boldsymbol{\eta}_{n_j} \rightharpoonup \boldsymbol{v}^*$ in $\sigma(V^*, V)$ together with (1.26) and (1.28) we obtain $\boldsymbol{v}^* = \gamma^*\boldsymbol{\zeta}$ and $\boldsymbol{\zeta} \in \partial J(\gamma \boldsymbol{v})$, which gives $\boldsymbol{v}^* \in \mathsf{F}(\boldsymbol{v})$.

By (1.26) and (1.28), we infer that for each subsequence $(\boldsymbol{\eta}_{n_j}), (\boldsymbol{v}_{n_j})$ of $(\boldsymbol{\eta}_n), (\boldsymbol{v}_n)$, we have

$$\lim_{j\to\infty} \langle \boldsymbol{\eta}_{n_j}, \boldsymbol{v}_{n_j}\rangle_{V^*\times V} = \lim_{j\to\infty} \langle \gamma^*\boldsymbol{\zeta}_{n_j}, \boldsymbol{v}_{n_j}\rangle_{V^*\times V} = \lim_{j\to\infty} \langle \boldsymbol{\zeta}_{n_j}, \gamma\boldsymbol{v}_{n_j}\rangle_{Y^*\times Y}$$
$$= \langle \boldsymbol{\zeta}, \gamma\boldsymbol{v}\rangle_{Y^*\times Y} = \langle \gamma^*\boldsymbol{\zeta}, \boldsymbol{v}\rangle_{V^*\times V} = \langle \boldsymbol{v}^*, \boldsymbol{v}\rangle_{V^*\times V},$$

which completes the proof of the generalized pseudomonotonicity of F.

Step 1.3.5 We show that the multivalued operator $\mathsf{G}\colon V \multimap V^*$ *is pseudomonotone and coercive provided* $\alpha > c_6$. Since the class of multivalued pseudomonotone operators is closed under addition of mappings (see [18, Proposition 6.3.68]), it follows that G is pseudomonotone due to Lemma 1.2.5 and Step 1.3.4. Furthermore, Lemma 1.2.5 shows that for all $\boldsymbol{v} \in V$ and $\boldsymbol{\eta} \in \mathsf{F}(\boldsymbol{v})$, we have

$$\langle \mathsf{G}(\boldsymbol{v}), \boldsymbol{v}\rangle_{V^*\times V} = \langle \mathcal{N}(\boldsymbol{v}), \boldsymbol{v}\rangle_{V^*\times V} + \langle \boldsymbol{\eta}, \boldsymbol{v}\rangle_{V^*\times V} \geq \alpha \|\boldsymbol{v}\|_V^2 + \langle \boldsymbol{\eta}, \boldsymbol{v}\rangle_{V^*\times V}.$$

By Step 1.3.3, we have $|\langle \boldsymbol{\eta}, \boldsymbol{v}\rangle_{V^*\times V}| \leq (c_5 + c_6 \|\boldsymbol{v}\|_V^{\rho-1}) \|\boldsymbol{v}\|_V$ for all $\boldsymbol{v} \in V$ and $\boldsymbol{\eta} \in \mathsf{F}(\boldsymbol{v})$. Hence,

$$\langle \boldsymbol{v}^*, \boldsymbol{v}\rangle_{V^*\times V} \geq \alpha \|\boldsymbol{v}\|_V^{\rho} - \left(c_5 + c_6 \|\boldsymbol{v}\|_V^{\rho-1}\right) \|\boldsymbol{v}\|_V = \ell(\|\boldsymbol{v}\|_V) \|\boldsymbol{v}\|_V$$

for all $(\boldsymbol{v}, \boldsymbol{v}^*) \in \mathrm{Gr}(\mathsf{G})$, where $\ell(t) := \alpha t^{\rho-1} - c_6 t^{\rho-1} - c_5$ and $\lim_{t\to\infty} \ell(t) = \infty$ if $1 < \rho < \infty$ provided $\alpha - c_6 > 0$. Therefore ℓ is the coercivity function of the operator G, and G is coercive as claimed.

In conclusion, because of [8, Theorem 3], G is surjective, i.e. $\mathsf{G}(V) = V^*$. Hence, for every $\boldsymbol{f} \in V^*$, there exists $\boldsymbol{u} \in V$ such that $\mathcal{N}\boldsymbol{u} + \mathsf{F}(\boldsymbol{u}) \ni \boldsymbol{f}$. Furthermore, from the coercivity of the operator G, we have

$$\alpha \|\boldsymbol{u}\|_V^{\rho} - (c_6 \|\boldsymbol{u}\|_V^{\rho-1} + c_5) \|\boldsymbol{u}\|_V \leq \|\boldsymbol{f}\|_{V^*} \|\boldsymbol{u}\|_V .$$

Thus, $\alpha \|\boldsymbol{u}\|_V^{\rho-1} \leq c_6 \|\boldsymbol{u}\|_V^{\rho-1} + c_5 + \|\boldsymbol{f}\|_{V^*}$ which implies the estimate in (i1).

Step 1.3.6 We show the uniqueness of a solution of Problem 1.3.1. Let $\boldsymbol{u}_1, \boldsymbol{u}_2 \in V$ be solutions of Problem 1.3.1. By (i1), we have $\|\boldsymbol{u}_j\|_V \leq C$ for $j = 1, 2$ and $C > 0$ is the constant as in (i1). Thus, there exist $\boldsymbol{\eta}_j = \gamma^*\boldsymbol{\zeta}_j \in V^*$, $\boldsymbol{\zeta}_j \in \partial J(\gamma\boldsymbol{u}_j) \subset Y^*$ with $\|\gamma\boldsymbol{u}_j\|_Y \leq \|\gamma\|_{V\to Y} \|\boldsymbol{u}_j\|_V \leq C \|\gamma\|_{V\to Y}$ for $j = 1, 2$ such that

$$\mathcal{N}\boldsymbol{u}_j + \boldsymbol{\eta}_j = \boldsymbol{f} \quad \text{for } j = 1, 2.$$

Subtracting the above two equations, multiplying the result by $\boldsymbol{u}_1 - \boldsymbol{u}_2$, and using the strong monotonicity of the operator $\mathcal{A}$ (see (i2)), we have

$$m_{\mathcal{A}} \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V^2 + \langle \mathcal{B}[\boldsymbol{u}_1] - \mathcal{B}[\boldsymbol{u}_2], \boldsymbol{u}_1 - \boldsymbol{u}_2\rangle_{V^*\times V} + \langle \boldsymbol{\eta}_1 - \boldsymbol{\eta}_2, \boldsymbol{u}_1 - \boldsymbol{u}_2\rangle_{V^*\times V} \leq 0.$$

In virtue of Definition 1.2.4, we obtain the estimate

$$|\langle \mathcal{B}[\boldsymbol{u}_1] - \mathcal{B}[\boldsymbol{u}_2], \boldsymbol{u}_1 - \boldsymbol{u}_2 \rangle_{V^* \times V}| = |\langle \mathcal{B}(\boldsymbol{u}_1 - \boldsymbol{u}_2, \boldsymbol{u}_2), \boldsymbol{u}_1 - \boldsymbol{u}_2 \rangle_{V^* \times V}|$$
$$= |b(\boldsymbol{u}_1 - \boldsymbol{u}_2, \boldsymbol{u}_2, \boldsymbol{u}_1 - \boldsymbol{u}_2)| \le c_4 \|\boldsymbol{u}_2\|_V \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V^2 \le c_4 C \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V^2 ,$$

where $c_4 > 0$ is the continuity constant of the trilinear form b associated with the operator $\mathcal{B}$ and $C > 0$ is the constant such as in (i1). By hypothesis (I2), we have $\|\boldsymbol{\zeta}_j\|_{Y^*} \le c_2 + c_3 C^{\rho-1} \|\gamma\|_{V\to Y}^{\rho-1}$ for $\boldsymbol{\zeta}_j \in \partial J(\gamma \boldsymbol{u}_j)$ and $j = 1, 2$. Hence,

$$\left|\langle \boldsymbol{\eta}_1 - \boldsymbol{\eta}_2, \boldsymbol{u}_1 - \boldsymbol{u}_2 \rangle_{V^* \times V}\right| = \left|\langle \boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2, \gamma \boldsymbol{u}_1 - \gamma \boldsymbol{u}_2 \rangle_{Y^* \times Y}\right|$$
$$\le \|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|_{Y^*} \|\gamma \boldsymbol{u}_1 - \gamma \boldsymbol{u}_2\|_Y \le (\|\boldsymbol{\zeta}_1\|_{Y^*} + \|\boldsymbol{\zeta}_2\|_{Y^*}) \|\gamma \boldsymbol{u}_1 - \gamma \boldsymbol{u}_2\|_Y .$$

Thus, we infer that $\langle \boldsymbol{\eta}_1 - \boldsymbol{\eta}_2, \boldsymbol{u}_1 - \boldsymbol{u}_2 \rangle_{V^* \times V} \ge -2r \|\gamma\|_{V\to Y}^2 \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V^2$ with $r := c_2 + c_3 C^{\rho-1} \|\gamma\|_{V\to Y}^{\rho-1} > 0$. Therefore,

$$(m_{\mathcal{A}} - c_4 C - 2r \|\gamma\|_{V\to Y}^2) \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V^2 \le 0.$$

This implies $\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V = 0$ provided $m_{\mathcal{A}} > c_4 C + 2r \|\gamma\|_{V\to Y}^2$. Thus, the solution to Problem 1.3.1 is unique. □

1.4 Hemivariational Inequalities

In this section we use the results of Sect. 1.3 to provide existence and uniqueness results to hemivariational inequalities (1.29) and (1.32) below, which will be applied to fluid flow problems in Sects. 1.5 and 1.6.

1.4.1 Tangential Superpotential

Problem 1.4.1 Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain, $V := W^1 L_\Phi(\Omega, \mathbb{R}^m)$, $Y := L_\Phi(\Gamma, \mathbb{R}^m)$, and $\boldsymbol{f} \in V^*$. Find $\boldsymbol{u} \in V$ such that

$$\langle \mathcal{N}\boldsymbol{u}, \boldsymbol{v} \rangle_{V^* \times V} + \int_\Gamma j_\tau^0(\mathbf{x}, (\gamma \boldsymbol{u})_\tau; (\gamma \boldsymbol{v})_\tau)\, \mathrm{d}\Gamma \ge \langle \boldsymbol{f}, \boldsymbol{v} \rangle_{V^* \times V}, \quad \forall \boldsymbol{v} \in V, \tag{1.29}$$

where $\mathcal{N}\colon V \to V^*$, $\mathcal{N}\boldsymbol{v} = \mathcal{A}\boldsymbol{v} + \mathcal{B}[\boldsymbol{v}]$ is the Navier–Stokes type operator, whereas $(\gamma \boldsymbol{u})_\tau \in Y$ and $(\gamma \boldsymbol{v})_\tau \in Y$ are tangential components of traces $\gamma \boldsymbol{u} \in Y$ and $\gamma \boldsymbol{v} \in Y$ of functions $\boldsymbol{u}$ and $\boldsymbol{v}$ on boundary $\partial\Omega = \Gamma$, respectively, and j_τ^0 stands for the generalized directional derivative of $j_\tau(\mathbf{x}, \cdot)$ for $j_\tau : \Gamma \times \mathbb{R}^m \to \mathbb{R}$.

For the above problem we assume the following hypotheses.

Hypotheses 1.4.2 Let $\Phi, \Phi^* \colon [0,\infty) \to [0,\infty)$ be a pair of complementary N-functions and $j_\tau \colon \Gamma \times \mathbb{R}^m \to \mathbb{R}$ be a Carathéodory function such that:

(V1) $\Phi, \Phi^* \in \Delta_2^\infty$ and $\Phi \lll (\Phi_*)^{1-1/d}$ with $d \geq 2$, where Φ_* is the Sobolev conjugate N-function of Φ;
(V2) $j_\tau(\mathbf{x}, \cdot)$ is locally Lipschitz for a.e. $\mathbf{x} \in \Gamma$;
(V3) For some $\kappa > 0$ there exist $a_\kappa \in L^1(\Gamma, [0,\infty))$ and positive constants b_κ and d_κ such that $\Phi^*(\|\mathbf{a}^*\|_{\mathbb{R}^m}/d_\kappa) \leq a_\kappa(\mathbf{x}) + b_\kappa \Phi(\|\mathbf{a}\|_{\mathbb{R}^m}/\kappa)$ holds for all $(\mathbf{a}, \mathbf{a}^*) \in \mathrm{Gr}(\partial j_\tau)$ and a.e. $\mathbf{x} \in \Gamma$.

Remark 1.4.3 With hypothesis (V1), the spaces $V = W^1 L_\Phi(\Omega, \mathbb{R}^m)$ and $Y = L_\Phi(\Gamma, \mathbb{R}^m)$ are separable and reflexive satisfying $V \hookrightarrow\hookrightarrow Y$ (cf. (1.2)). Hypotheses (V2) and (V3) (cf. Hypotheses 1.2.15) allow to use the results of Sect. 1.3.

In order to establish existence of solutions to Problem 1.4.1, we associate with it, a subdifferential operator inclusion as in Sect. 1.3. It works owing to Theorem 1.2.17 on an integral representation of the Clarke subdifferential of locally Lipschitz integral functionals defined on the Orlicz space.

Lemma 1.4.4 *Under Hypotheses 1.4.2, if the functional $J \colon Y \to \mathbb{R}$, defined by*

$$J(\boldsymbol{y}) = \int_\Gamma j_\tau(\mathbf{x}, \boldsymbol{y}(\mathbf{x}))\, \mathrm{d}\Gamma \quad \textit{for } \boldsymbol{y} \in Y = L_\Phi(\Gamma, \mathbb{R}^m), \tag{1.30}$$

is finite at least for one $\boldsymbol{y} \in \overline{B}_Y(0, \kappa/2)$, then

(t1) *J is Lipschitz continuous on $\overline{B}_Y(0,\kappa/2)$;*
(t2) $J^0(\boldsymbol{y}; \boldsymbol{z}) \leq \int_\Gamma j_\tau^0(\mathbf{x}, \boldsymbol{y}(\mathbf{x}); \boldsymbol{z}(\mathbf{x}))\, \mathrm{d}\Gamma$ *for all $\boldsymbol{y} \in B_Y(0,\kappa/2)$ and $\boldsymbol{z} \in Y$;*
(t3) $\partial J(\boldsymbol{y}) \subset \mathcal{N}_{\partial j_\tau}(\boldsymbol{y})$ *for all $\boldsymbol{y} \in B_Y(0,\kappa/2)$, where the multivalued superposition operator $\mathcal{N}_{\partial j_\tau} \colon Y \multimap Y^*$ is bounded, that is, if $\boldsymbol{\zeta} \in \partial J(\boldsymbol{y}) \subset Y^*$ then*

$$\langle \boldsymbol{\zeta}, \boldsymbol{z} \rangle_{Y^* \times Y} = \int_\Gamma \boldsymbol{\xi}(\mathbf{x}) \cdot \boldsymbol{z}(\mathbf{x})\, \mathrm{d}\Gamma$$

for all $\boldsymbol{z} \in Y$ and for some $\boldsymbol{\xi} \in \mathcal{N}_{\partial j_\tau}(\boldsymbol{y}) = \mathrm{Sel}\,\partial j_\tau(\cdot, \boldsymbol{y}(\cdot))$.
(t4) $\|\partial J(\boldsymbol{y})\|_{Y^*} \leq c_8 + c_7 \|\boldsymbol{y}\|_Y^{\rho-1}$ *for all $\boldsymbol{y} \in Y$ with $\rho \in (1,\infty)$, where $c_7 := \frac{b_\kappa d_\kappa}{\kappa} > 0$ and $c_8 := 2d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + c_7 > 0$.*
(t5) *If, in addition, either j_τ or $-j_\tau$ is regular (in the sense of Clarke) at $\boldsymbol{y}(\mathbf{x})$ a.e., then J or $-J$ is regular (in the sense of Clarke) at $\boldsymbol{y}$, respectively.*
(t6) *If, in addition, either j_τ or $-j_\tau$ is regular (in the sense of Clarke) at $\boldsymbol{y}(\mathbf{x})$ a.e., then* (t2) *and* (t3) *hold with equalities.*

Proof We only show that (t4) is satisfied. First, we prove that a superposition multivalued operator $\mathcal{N}_{\partial j_\tau} \colon Y \multimap Y^*$, $\mathcal{N}_{\partial j_\tau}(\boldsymbol{y}) := \mathrm{Sel}\,\partial j_\tau(\cdot, \boldsymbol{y}(\cdot))$ is bounded, i.e., for every $\kappa > 0$ there exists $r(\kappa) > 0$ such that $\|\boldsymbol{y}\|_Y \leq \kappa$ implies $\|\boldsymbol{\xi}\|_{Y^*} \leq r(\kappa)$ for $\boldsymbol{\xi} \in \mathcal{N}_{\partial j_\tau}(\boldsymbol{y})$. Indeed, fix $\boldsymbol{y}$ and $\boldsymbol{\xi}$ such that $\|\boldsymbol{y}\|_Y \leq \kappa$ and $\boldsymbol{\xi}(\mathbf{x}) \in \partial j_\tau(\mathbf{x}, \boldsymbol{y}(\mathbf{x}))$

a.e. $\mathbf{x} \in \Gamma$. Note that $\int_\Gamma \Phi\left(\|\boldsymbol{y}(\mathbf{x})\|_{\mathbb{R}^m}/\kappa\right) d\Gamma \le 1$ due to [47, Lemma 3.8.4]. By hypothesis (V3), for each $\boldsymbol{\xi} \in \mathcal{N}_{\partial j_\tau}(\boldsymbol{y})$, we obtain

$$\int_\Gamma \Phi^*\left(\|\boldsymbol{\xi}(\mathbf{x})\|_{\mathbb{R}^m}/d_\kappa\right) d\Gamma \le \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + b_\kappa \int_\Gamma \Phi\left(\|\boldsymbol{y}(\mathbf{x})\|_{\mathbb{R}^m}/\kappa\right) d\Gamma$$
$$\le \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + b_\kappa + 1 < \infty.$$

Since Φ^* is convex, we have $\int_\Gamma \Phi^*\left(\|\boldsymbol{\xi}(\mathbf{x})\|_{\mathbb{R}^m}/[d_\kappa(\|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + b_\kappa + 1)]\right) d\Gamma < 1$. Therefore, $\|\boldsymbol{\xi}\|_{Y^*} \le r(\kappa) := d_\kappa(\|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + b_\kappa + 1)$.

Next, observe that for all $\boldsymbol{\zeta} \in \partial J(\boldsymbol{y})$ with $\boldsymbol{y} \in Y$, by (t2), Theorem 1.2.14, Hölder's inequality, and hypothesis (V3), we obtain

$$|\langle \boldsymbol{\zeta}, \boldsymbol{z}\rangle| \le \int_\Gamma \left| j^0(\mathbf{x}, \boldsymbol{y}(\mathbf{x}); \boldsymbol{z}(\mathbf{x}))\right| d\Gamma \le \int_\Gamma \sup\left\{\|\boldsymbol{\xi}(\mathbf{x})\|_{\mathbb{R}^m} \|\boldsymbol{z}(\mathbf{x})\|_{\mathbb{R}^m} \,\middle|\, \boldsymbol{\xi} \in \mathcal{N}_{\partial j_\tau}(\boldsymbol{y})\right\} d\Gamma$$
$$\le 2(\left\|d_\kappa(\Phi^*)^{-1}(a_\kappa)\right\|_{Y^*} + \left\|b_\kappa d_\kappa(\Phi^*)^{-1}\Phi(\|\boldsymbol{y}(\mathbf{x})\|_{\mathbb{R}^m}/\kappa))\right\|_{Y^*})\, \|\boldsymbol{z}\|_Y$$
$$\le 2(d_\kappa \left\|(\Phi^*)^{-1}(a_\kappa)\right\|_{Y^*} + \tfrac{b_\kappa d_\kappa}{\kappa} \left\|(\Phi^*)^{-1}\Phi(\|\boldsymbol{y}(\mathbf{x})\|_{\mathbb{R}^m}))\right\|_{Y^*})\, \|\boldsymbol{z}\|_Y,$$

where $(\Phi^*)^{-1}$ of Φ^* is defined by $(\Phi^*)^{-1}(s) = \sup\{t \mid \Phi^*(t) \le s\}$ for $s \in [0, \infty)$. Since $\Phi^*((\Phi^*)^{-1}(s)) \le s$, it follows that

$$\left\|(\Phi^*)^{-1}(a_\kappa)\right\|_{Y^*} \le \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, \quad \left\|(\Phi^*)^{-1}\Phi(\|\boldsymbol{y}(\mathbf{x})\|_{\mathbb{R}^m})\right\|_{Y^*} \le \|\boldsymbol{y}\|_Y.$$

By [47, Theorem 3.8.5, Formula 3.6.3.1, p.145], we have $\|\boldsymbol{y}\|_Y \le 1 + \int_\Gamma \Phi(|\boldsymbol{y}(\mathbf{x})|)\, d\Gamma$. Therefore,

$$\|\boldsymbol{\zeta}\|_{Y^*} \le c_8 + c_7 \int_\Gamma \Phi(|\boldsymbol{y}(\mathbf{x})|)\, d\Gamma, \quad \forall \boldsymbol{\zeta} \in \partial J(\boldsymbol{y})$$

with $c_7 := \frac{b_\kappa d_\kappa}{\kappa} > 0$ and $c_8 := 2d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + c_7 > 0$. Hence, by Lemma 1.2.7 (see (1.4)) we obtain

$$\|\boldsymbol{\zeta}\|_{Y^*} \le c_8 + c_7 \|\boldsymbol{y}\|_Y^{\rho-1}, \quad \forall \boldsymbol{\zeta} \in \partial J(\boldsymbol{y}),\ 1 < \rho < \infty \text{ and } c_7, c_8 > 0. \tag{1.31}$$

This is our claim (t4). □

The existence and uniqueness result in study of Problem 1.4.1 is the following.

Theorem 1.4.5 *Under Hypotheses 1.4.2, Problem 1.4.1 has a solution* $\boldsymbol{u} \in V$ *provided* $\alpha > c_6$, *and*

(v1) *the solution satisfies the estimate* $\|\boldsymbol{u}\|_V \le C$ *with* $C := \left(\frac{\|\boldsymbol{f}\|_{V^*}+c_5}{\alpha-c_6}\right)^{1/(\rho-1)}$;

(v2) *the solution is unique if, in addition, the operator* $\mathcal{A}$ *is strongly monotone with the positive constant* $m_{\mathcal{A}}$ *satisfying the smallness condition:*

$$m_{\mathcal{A}} > 2c_8 \|\gamma\|^2_{V\to Y} + 2c_7 \frac{\|\boldsymbol{f}\|_{V^*}+c_5}{\alpha-c_6} \|\gamma\|^{\rho+1}_{V\to Y} + c_4 \left(\frac{\|\boldsymbol{f}\|_{V^*}+c_5}{\alpha-c_6}\right)^{1/(\rho-1)};$$

where $\rho \in (1, \infty)$, α *is the coercivity constant of the operator* $\mathcal{A}$, c_4 *is the continuity constant of the trilinear form* b *associated with the operator* $\mathcal{B}$, $c_5 := c_8 \|\gamma\|_{V\to Y}$ *and* $c_6 := c_7 \|\gamma\|^{\rho}_{V\to Y}$ *with* $c_7 := \frac{2b_\kappa d_\kappa}{\kappa} > 0$ *and* $c_8 := 2d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + c_7 > 0$, *whereas* $b_\kappa, d_\kappa, a_\kappa$ *such as in hypothesis* (V3).

Proof We apply Theorem 1.3.4. By Lemma 1.4.4, we conclude that hypothesis (I1) and (I2) are satisfied. □

Lemma 1.4.6 *Every solution of Problem 1.3.1 is also a solution of Problem 1.4.1. Furthermore, if either* j *or* $-j$ *is regular (in the sense of Clarke), then the converse is also true.*

Proof Let $\boldsymbol{u} \in V$ be a solution of Problem 1.3.1. Hence, there exists $\boldsymbol{\eta} = \gamma^*\boldsymbol{\zeta} \in V^*$ and $\boldsymbol{\zeta} \in \partial J(\gamma\boldsymbol{u}) \subset Y^*$ such that $\mathcal{A}\boldsymbol{u} + \boldsymbol{\eta} = \boldsymbol{f}$. By [17, Proposition 2.1.5], we have $\langle \boldsymbol{\zeta}, \gamma\boldsymbol{v}\rangle_{Y^*\times Y} \le J^0(\gamma\boldsymbol{u}, \gamma\boldsymbol{v})$ for all $\gamma\boldsymbol{v} \in Y$. Thus

$$\begin{aligned}\langle \boldsymbol{f} - \mathcal{A}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} &= \langle \boldsymbol{\eta}, \boldsymbol{v}\rangle_{V^*\times V} = \langle \gamma^*\boldsymbol{\zeta}, \boldsymbol{v}\rangle_{V^*\times V} = \langle \boldsymbol{\zeta}, \gamma\boldsymbol{v}\rangle_{Y^*\times Y}\\ &\le J^0(\gamma\boldsymbol{u}; \gamma\boldsymbol{v}) \le \int_\Gamma j^0_\tau(\mathbf{x}, \gamma\boldsymbol{u}(\mathbf{x}); \gamma\boldsymbol{v}(\mathbf{x}))\, d\Gamma, \quad \forall \boldsymbol{v} \in V,\end{aligned}$$

where the last inequality holds due to (t2) (see Lemma 1.4.4). Therefore, $\boldsymbol{u}$ is also a solution of Problem 1.4.1. Furthermore, we show that if either j_τ or $-j_\tau$ is regular (in the sense of Clarke), then every solution of Problem 1.4.1 is a solution of Problem 1.3.1. From the hemivariational inequality (1.29) and (t6) (see Lemma 1.4.4), we obtain

$$\langle \boldsymbol{f} - \mathcal{A}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} \le \int_\Gamma j^0_\tau(\mathbf{x}, \gamma\boldsymbol{u}(\mathbf{x}); \gamma\boldsymbol{v}(\mathbf{x}))\, d\Gamma = J^0(\gamma\boldsymbol{u}; \gamma\boldsymbol{v}).$$

It follows from [17, Theorem 2.3.10, Corollary p. 47] that

$$J^0(\gamma\boldsymbol{u}; \gamma\boldsymbol{v}) = (J \circ \gamma)^0(\boldsymbol{u}; \boldsymbol{v}) \quad \text{and} \quad \partial(J \circ \gamma)(\boldsymbol{u}) = \gamma^*\partial J(\gamma\boldsymbol{u}).$$

Therefore, $\langle \boldsymbol{f} - \mathcal{A}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} \le (J \circ \gamma)^0(\boldsymbol{u}; \boldsymbol{v})$ and $\boldsymbol{f} - \mathcal{A}\boldsymbol{u} \in \partial(J \circ \gamma)(\boldsymbol{u}) = \gamma^*\partial J(\gamma\boldsymbol{u})$, which implies our assertion. □

1.4.2 Normal Superpotential

Problem 1.4.7 Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain, $V := W^1L_\Phi(\Omega, \mathbb{R}^m)$, $Y := L_\Phi(\Gamma, \mathbb{R}^m)$, and $\boldsymbol{f} \in V^*$. Find $\boldsymbol{u} \in V$ such that

$$\langle \mathcal{N}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} + \int_\Gamma j_\nu^0(\mathbf{x}, (\gamma u)_\nu(\mathbf{x}); (\gamma v)_\nu(\mathbf{x}))\, d\Gamma \geq \langle \boldsymbol{f}, \boldsymbol{v}\rangle_{V^*\times V}, \ \forall \boldsymbol{v} \in V, \tag{1.32}$$

where $\mathcal{N}\colon V \to V^*$ is the Navier–Stokes type operator, $\mathcal{N}\boldsymbol{v} = \mathcal{A}\boldsymbol{v} + \mathcal{B}[\boldsymbol{v}]$ with $\mathcal{A}$ being $(\alpha, 2)$-coercive, whereas $(\gamma u)_\nu$ and $(\gamma v)_\nu$ are normal components of traces $\gamma\boldsymbol{u} \in Y$ and $\gamma\boldsymbol{v} \in Y$ of functions $\boldsymbol{u}$ and $\boldsymbol{v}$ on boundary $\partial\Omega = \Gamma$, respectively, and j_ν^0 stands for the generalized directional derivative of $j_\nu(\mathbf{x}, \cdot)$ for $j_\nu\colon \Gamma \times \mathbb{R} \to \mathbb{R}$.

We adopt the following hypotheses on the data.

Hypotheses 1.4.8 Let $\Phi, \Phi^*\colon [0, \infty) \to [0, \infty)$ be a pair of complementary N-functions and let $j_\nu\colon \Gamma \times \mathbb{R} \to \mathbb{R}$ be a Carathéodory function such that

(V4) $\Phi, \Phi^* \in \Delta_2^\infty$ and $\Phi \lll (\Phi_*)^{1-1/d}$ with $d \geq 2$, where Φ_* is the Sobolev conjugate N-function of Φ;
(V5) $j_\nu(\mathbf{x}, \cdot)$ is locally Lipschitz for a.e. $\mathbf{x} \in \Gamma$;
(V6) For some $\kappa > 0$ there exist $a_\kappa \in L^1(\Gamma, [0, \infty))$ and positive constants b_κ and d_κ such that $\Phi^*(|a^*|/d_\kappa) \leq a_\kappa(\mathbf{x}) + b_\kappa\Phi(|a|/\kappa)$ for all $(a, a^*) \in \mathrm{Gr}(\partial j_\nu)$ and a.e. $\mathbf{x} \in \Gamma$.

Remark 1.4.9 By hypothesis (V4), the spaces $V = W^1L_\Phi(\Omega, \mathbb{R}^m)$ and $Y = L_\Phi(\Gamma, \mathbb{R}^m)$ are separable and reflexive satisfying $V \hookrightarrow\hookrightarrow Y$ (cf. (1.2)). Hypotheses (V5) and (V6) follow from [71, 72].

Note that Theorem 1.2.17 on an integral representation of the Clarke subdifferential of locally Lipschitz integral functionals defined on the Orlicz space allows to relate Problems 1.3.1 and 1.4.7 (see Lemmata 1.4.10 and 1.4.12). We proved that every solution of Problem 1.3.1 is also a solution of Problem 1.4.7 (see Lemma 1.4.12). Finally, Lemma 1.4.10 follows from [75]. However, we give a different proof of (n4) here.

Lemma 1.4.10 *Under Hypotheses 1.4.8, if the functional* $J\colon Y \to \mathbb{R}$ *defined by*

$$J(\boldsymbol{y}) = \int_\Gamma j_\nu(\mathbf{x}, y_\nu(\mathbf{x}))\, d\Gamma, \quad \boldsymbol{y} \in Y = L_\Phi(\Gamma, \mathbb{R}^m), \tag{1.33}$$

is finite at least for one $\boldsymbol{y} \in \overline{B}_Y(0, \kappa/2)$, *then*

(n1) J *is Lipschitz on* $\overline{B}_Y(0, \kappa/2)$;
(n2) *for all* $\boldsymbol{y} \in B_Y(0, \kappa/2)$ *and* $\boldsymbol{z} \in Y$, *the following inequality holds:*

$$J^0(\boldsymbol{y}; \boldsymbol{z}) \leq \int_\Gamma j_\nu^0(\mathbf{x}, y_\nu(\mathbf{x}); z_\nu(\mathbf{x}))\, d\Gamma;$$

(n3) $\partial J(\boldsymbol{y}) \subset \mathcal{N}_{(\partial j_\nu)\boldsymbol{\nu}}(y_\nu)$ *for all* $\boldsymbol{y} \in B_Y(0, \kappa/2)$, *where the multivalued superposition operator* $\mathcal{N}_{(\partial j_\nu)\boldsymbol{\nu}} \colon L_\Phi(\Gamma, \mathbb{R}) \multimap Y^*$ *is bounded, which means that if* $\boldsymbol{\zeta} \in \partial J(\boldsymbol{y}) \subset Y^*$ *then*

$$\langle \boldsymbol{\zeta}, \boldsymbol{z} \rangle = \int_\Gamma \boldsymbol{\xi}(\mathbf{x}) \cdot \boldsymbol{z}(\mathbf{x})\, \mathrm{d}\Gamma$$

for all $\boldsymbol{z} \in Y$ *and for some* $\boldsymbol{\xi} \in \mathcal{N}_{(\partial j_\nu)\boldsymbol{\nu}}(y_\nu) = \mathrm{Sel}\, \partial j_\nu(\cdot, y_\nu(\cdot))\boldsymbol{\nu}$, *such that* $\boldsymbol{\xi}(\mathbf{x}) = a(\mathbf{x})\boldsymbol{\nu}$ *and* $a \in \partial j_\nu(\cdot, y_\nu(\cdot)) \subset L_{\Phi^*}(\Gamma, \mathbb{R})$*;*

(n4) $\|\partial J(\boldsymbol{y})\|_{Y^*} \le c_9\, (1 + \|\gamma \boldsymbol{u}\|_Y^\rho)$ *for all* $\boldsymbol{y} \in Y$ *with* $0 < \rho \le 1$, *where* $c_9 := \max\left\{2d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, \frac{2b_\kappa d_\kappa}{\kappa}, 2c_{e1} d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, c_{e1} \frac{2b_\kappa d_\kappa}{\kappa}\right\}$ *with* $b_\kappa, d_\kappa, a_\kappa$ *such as in hypothesis* (V6).

(n5) *If, in addition, either* j_ν *or* $-j_\nu$ *is regular (in the sense of Clarke) at* $y_\nu(\mathbf{x})$ *a.e., then* J *or* $-J$ *is regular (in the sense of Clarke) at* $\boldsymbol{y}$*, respectively.*

(n6) *If, in addition, either* j_ν *or* $-j_\nu$ *is regular (in the sense of Clarke) at* $\boldsymbol{y}(\mathbf{x})$ *a.e., then (n2) and (n3) hold with equalities.*

Proof We apply Theorem 1.2.17. To this end, let $j_1 \colon \Gamma \times \mathbb{R}^m \to \mathbb{R}$ be a function defined by

$$j_1(\mathbf{x}, \mathbf{a}) := j_\nu(\mathbf{x}, \mathrm{a}_\nu), \quad \forall (\mathbf{x}, \mathbf{a}) \in \Gamma \times \mathbb{R}^m. \tag{1.34}$$

Observe that $L \colon \mathbb{R}^m \to \mathbb{R}$ defined by $L\mathbf{a} = \mathrm{a}_\nu$ is a linear continuous operator and its adjoint operator $L^* \colon \mathbb{R} \to \mathbb{R}^m$ is given by $L^* r = r\boldsymbol{\nu}$ for $r \in \mathbb{R}$. Hence,

$$j_1(\mathbf{x}, \mathbf{a}) = j_\nu(\mathbf{x}, L\mathbf{a}) = (j_\nu \circ L)(\mathbf{x}, \mathbf{a}). \tag{1.35}$$

By hypothesis (V5), we infer that j_1 is a Carathéodory function such that $j_1(\mathbf{x}, \cdot)$ is locally Lipschitz for a.e. $\mathbf{x} \in \Gamma$. Hence, by [17, Theorem 2.3.10 and Remark 2.3.11] and (1.35), for $\mathbf{a}, \mathbf{s} \in \mathbb{R}^m$, we obtain

$$j_1^0(\mathbf{x}, \mathbf{a}; \mathbf{s}) = (j_\nu \circ L)^0(\mathbf{x}, \mathbf{a}; \mathbf{s}) \le j_\nu^0(\mathbf{x}, L\mathbf{a}; L\mathbf{s}) = j_\nu^0(\mathbf{x}, \mathrm{a}_\nu; \mathrm{s}_\nu), \tag{1.36}$$

$$\partial j_1(\mathbf{x}, \mathbf{a}) = \partial (j_\nu \circ L)(\mathbf{x}, \mathbf{a}) \subset L^* \partial j_\nu(\mathbf{x}, L\mathbf{a})\boldsymbol{\nu} = \partial j_\nu(\mathbf{x}, \mathrm{a}_\nu)\boldsymbol{\nu}. \tag{1.37}$$

It follows that for $\mathbf{a}^* \in \partial j_1(\mathbf{x}, \mathbf{a})$, we have $\mathbf{a}^* = \mathrm{a}_\nu^* \boldsymbol{\nu}$ and $\mathrm{a}_\nu^* \in \partial j_\nu(\mathbf{x}, \mathrm{a}_\nu)$. Hence, by hypothesis (V6) for some $\kappa > 0$ and for all $\mathrm{a}_\nu^* \in \partial j_\nu(\mathbf{x}, \mathrm{a}_\nu)$, we obtain

$$\Phi^*\left(\left|\mathrm{a}_\nu^*\right| / d_\kappa\right) \le a_\kappa(\mathbf{x}) + b_\kappa \Phi\left(|\mathrm{a}_\nu| / \kappa\right) < a_\kappa(\mathbf{x}) + b_\kappa \Phi\left(\|\mathbf{a}\|_{\mathbb{R}^m} / \kappa\right)$$

for almost all $\mathbf{x} \in \Gamma$ and for all $\mathbf{a} \in \mathbb{R}^m$. The latter follows from the fact that Φ is strictly increasing and $|\mathrm{a}_\nu| = \|\mathbf{a} \cdot \boldsymbol{\nu}\|_{\mathbb{R}^m} \le \|\mathbf{a}\|_{\mathbb{R}^m}$. Thus, for some $\kappa > 0$, there exist $b_\kappa, d_\kappa > 0$, $a_\kappa \in L^1(\Gamma, [0, \infty))$ such that

$$\Phi^*\left(\left\|\mathbf{a}^*\right\|_{\mathbb{R}^m} / d_\kappa\right) \le a_\kappa(\mathbf{x}) + b_\kappa \Phi\left(\|\mathbf{a}\|_{\mathbb{R}^m} / \kappa\right), \quad \forall (\mathbf{a}, \mathbf{a}^*) \in \mathrm{Gr}\, \partial j_1(\mathbf{x}, \cdot) \tag{1.38}$$

for all $\mathbf{a} \in \mathbb{R}^m$ and a.e. $\mathbf{x} \in \Gamma$. It follows that j_1 satisfies the hypotheses of [71, Theorem 4.3], that is, $j_1(\mathbf{x}, \cdot)$ is locally Lipschitz for a.e. $\mathbf{x} \in \Gamma$ and j_1 satisfies (1.38).

Next, we consider the functional $J_1 \colon L_\Phi(\Gamma, \mathbb{R}^m) \to \mathbb{R}$ defined by

$$J_1(\boldsymbol{y}) := \int_\Gamma j_1(\mathbf{x}, \boldsymbol{y}(\mathbf{x}))\, \mathrm{d}\Gamma, \qquad \boldsymbol{y} \in Y = L_\Phi(\Gamma, \mathbb{R}^m), \tag{1.39}$$

where $j_1(\mathbf{x}, \boldsymbol{y}(\mathbf{x})) := j_\nu(\mathbf{x}, y_\nu(\mathbf{x}))$ is given by (1.34).

Since $j_1(\mathbf{x}, \cdot)$ is locally Lipschitz for a.e. $\mathbf{x} \in \Gamma$, from [71, Theorem 4.3], we infer that the functional J_1 is Lipschitz on $\overline{B}_Y(0, \kappa/2)$ and (n1) holds. Therefore,

$$J_1^0(\boldsymbol{y}; \boldsymbol{z}) \le \int_\Gamma j_1^0(\mathbf{x}, \boldsymbol{y}(\mathbf{x}); \boldsymbol{z}(\mathbf{x}))\, \mathrm{d}\Gamma \le \int_\Gamma j_\nu^0(\mathbf{x}, y_\nu(\mathbf{x}); z_\nu(\mathbf{x}))\, \mathrm{d}\Gamma \tag{1.40}$$

for all $\boldsymbol{y} \in B_Y(0, \kappa/2)$ and all $\boldsymbol{z} \in Y$ due to [71, Theorem 4.3] and (1.36). So, (n2) holds.

By [71, Theorem 4.3] and (1.37), we infer that

$$\partial J_1(\boldsymbol{y}) \subset \mathcal{N}_{\partial j_1}(\boldsymbol{y}) \subset \mathcal{N}_{(\partial j_\nu)\boldsymbol{\nu}}(y_\nu) \tag{1.41}$$

for all $\boldsymbol{y} \in B_Y(0, \kappa/2)$. These inclusions mean that if $\boldsymbol{\zeta} \in \partial J_1(\boldsymbol{y}) \subset Y^*$ then

$$\langle \boldsymbol{\zeta}, \boldsymbol{z} \rangle_{Y^* \times Y} = \int_\Gamma \boldsymbol{\xi}(\mathbf{x}) \cdot \boldsymbol{z}(\mathbf{x})\, \mathrm{d}\Gamma, \qquad \forall \boldsymbol{z} \in Y$$

for some $\boldsymbol{\xi} \in \mathcal{N}_{(\partial j_\nu)\boldsymbol{\nu}}(y_\nu) = \mathrm{Sel}\, \partial j_\nu(\cdot, y_\nu(\cdot))\boldsymbol{\nu}$, which has the following properties $\boldsymbol{\xi}(\mathbf{x}) = a(\mathbf{x})\boldsymbol{\nu}$ and $a \in \partial j_\nu(\cdot, y_\nu(\cdot)) \subset L_{\Phi^*}(\Gamma, \mathbb{R})$. Furthermore, the multivalued superposition operator $\mathcal{N}_{(\partial j_\nu)\boldsymbol{\nu}} \colon L_\Phi(\Gamma, \mathbb{R}) \multimap Y^*$ is bounded because $\mathcal{N}_{\partial j_\nu} \colon L_\Phi(\Gamma, \mathbb{R}) \multimap L_{\Phi^*}(\Gamma, \mathbb{R})$ defined by $\mathcal{N}_{\partial j_\nu}(y_\nu) = \mathrm{Sel}\, \partial j_\nu(\cdot, y_\nu(\cdot))$ is bounded, that is, for every $\kappa > 0$ there exists $r(\kappa) > 0$ such that $\|y_\nu\|_{L_\Phi(\Gamma,\mathbb{R})} \le \kappa$ implies $\|\xi\|_{L_{\Phi^*}(\Gamma,\mathbb{R})} \le r(\kappa)$ for $\xi \in \mathcal{N}_{\partial j_\nu}(y_\nu)$. Indeed, fix y_ν and ξ such that $\|y_\nu\|_{L_\Phi(\Gamma,\mathbb{R})} \le \kappa$ and $\xi(\mathbf{x}) \in \partial j_\nu(\mathbf{x}, y_\nu(\mathbf{x}))$ a.e. Hypothesis (V6) yields

$$\begin{aligned} \int_\Gamma \Phi^* \left(|\xi(\mathbf{x})| / d_\kappa\right) \mathrm{d}\Gamma &\le \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + b_\kappa \int_\Gamma \Phi\left(|y_\nu(\mathbf{x})| / \kappa\right) \mathrm{d}\Gamma \\ &\le \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + b_\kappa + 1. \end{aligned}$$

The second inequality holds due to $\int_\Gamma \Phi\left(|y_\nu(\mathbf{x})| / \kappa\right) \mathrm{d}\Gamma \le 1$ by [47, Lemma 3.8.4]. Since Φ^* is convex, we infer that

$$\int_\Gamma \Phi^* \Big(|\xi(\mathbf{x})| / [d_\kappa(\|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + b_\kappa + 1)]\Big)\, \mathrm{d}\Gamma < 1.$$

Therefore, $\|\xi\|_{L_{\Phi^*}(\Gamma,\mathbb{R})} \le r(\kappa) := d_\kappa(\|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + b_\kappa + 1)$. Thus, (n3) holds.

We prove that (n4) is satisfied. Indeed, for every $\boldsymbol{\zeta} \in \partial J_1(\boldsymbol{y})$ and $\boldsymbol{y} \in Y$, in view of (1.40) and Theorem 1.2.14, we infer that

$$|\langle \boldsymbol{\zeta}, z\rangle| \le \int_\Gamma \left| j_\nu^0(\mathbf{x}, y_\nu(\mathbf{x}); z_\nu(\mathbf{x}))\right| \, d\Gamma \le \int_\Gamma \sup_{\boldsymbol{\xi} \in \mathcal{N}_{(\partial j_\nu)\boldsymbol{\nu}}(y_\nu)} \|\boldsymbol{\xi}(\mathbf{x})\|_{\mathbb{R}^d} \, \|z_\nu(\mathbf{x})\|_{\mathbb{R}^d} \, d\Gamma.$$

The Hölder inequality and hypothesis (V3) lead to

$$\begin{aligned} |\langle \boldsymbol{\zeta}, z\rangle| \le {} & 2\, d_\kappa \left\| (\Phi^*)^{-1}(a_\kappa)\right\|_{L_{\Phi^*}(\Gamma,\mathbb{R})} \|z_\nu\|_{L_\Phi(\Gamma,\mathbb{R})} \\ & + 2\tfrac{b_\kappa d_\kappa}{\kappa} \left\| (\Phi^*)^{-1}\Phi(|y_\nu|)\right\|_{L_{\Phi^*}(\Gamma,\mathbb{R})} \|z_\nu\|_{L_\Phi(\Gamma,\mathbb{R})}, \end{aligned} \tag{1.42}$$

where $(\Phi^*)^{-1}$ of Φ^* is defined by $(\Phi^*)^{-1}(s) = \sup\{t \mid \Phi^*(t) \le s\}$ for $s \in [0, \infty)$. From $\Phi^*((\Phi^*)^{-1}(s)) \le s$, we have

$$\left\| (\Phi^*)^{-1}(a_\kappa)\right\|_{L_{\Phi^*}(\Gamma,\mathbb{R})} \le \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, \quad \left\| (\Phi^*)^{-1}\Phi(|y_\nu|)\right\|_{L_{\Phi^*}(\Gamma,\mathbb{R})} \le \|y_\nu\|_{L_\Phi(\Gamma,\mathbb{R})}.$$

Since $\|z_\nu\|_{L_\Phi(\Gamma,\mathbb{R})} \le \|\boldsymbol{z}\|_{L_\Phi(\Gamma,\mathbb{R}^m)}$ and $\|y_\nu\|_{L_\Phi(\Gamma,\mathbb{R})} \le \|\boldsymbol{y}\|_{L_\Phi(\Gamma,\mathbb{R}^m)}$, for the positive constant $\bar{c} := \max\{2d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, \frac{2b_\kappa d_\kappa}{\kappa}\}$, we obtain

$$\|\boldsymbol{\zeta}\|_{Y^*} \le \bar{c}\,(1 + \|\boldsymbol{y}\|_Y), \quad \forall \boldsymbol{\zeta} \in \partial J(\boldsymbol{y}). \tag{1.43}$$

It follows from [38, Theorem 7.3] that for all N-functions A and B such that $A, B \in \Delta_2^\infty$ with $B \prec \Phi \prec A$, we obtain $L_\Phi(\Gamma,\mathbb{R}^m) = (L_B(\Gamma,\mathbb{R}^m))^\theta (L_A(\Gamma,\mathbb{R}^m))^{1-\theta}$ for $0 < \theta < 1$ with equivalent norms. By [38, Proposition 5.1, Remark 5.1], we have $\|\gamma \boldsymbol{u}\|_{L_B(\Gamma,\mathbb{R}^m)} \le 1$ and $\|\gamma \boldsymbol{u}\|_{L_A(\Gamma,\mathbb{R}^m)} \le 1$. Since $Y \hookrightarrow L_B(\Gamma,\mathbb{R}^m)$ with an embedding constant $c_{e1} > 0$, we have

$$\|\gamma \boldsymbol{u}\|_Y \le \|\gamma \boldsymbol{u}\|^\theta_{L_B(\Gamma,\mathbb{R}^m)} \|\gamma \boldsymbol{u}\|^{1-\theta}_{L_A(\Gamma,\mathbb{R}^m)} \le \|\gamma \boldsymbol{u}\|^\theta_{L_B(\Gamma,\mathbb{R}^m)} \le c_{e1}\, \|\gamma \boldsymbol{u}\|^\theta_Y, \tag{1.44}$$

with $0 < \theta < 1$. From (1.43) and (1.44), we obtain

$$\|\boldsymbol{\zeta}\|_{Y^*} \le c_9\,(1 + \|\gamma \boldsymbol{u}\|^\rho_Y) \quad \forall \boldsymbol{\zeta} \in \partial J(\gamma \boldsymbol{u}) \text{ and } 0 < \rho \le 1 \tag{1.45}$$

with $c_9 := \max\{\bar{c}, c_{e1}\bar{c}\} > 0$ and $\rho := \theta$ for $0 < \theta < 1$ or $\rho := 1$ from (1.43).

Furthermore, if either j_ν or $-j_\nu$ is regular (in the sense of Clarke) at $y_\nu(\mathbf{x})$ a.e., we obtain equality in (1.36) due to[17, Theorem 2.3.10] and (1.35). Hence, j_1 and $-j_1$ are regular (in the sense of Clarke) at $\boldsymbol{u}(\mathbf{x})$ a.e. $\mathbf{x} \in \Gamma$. By Theorem 1.2.17, (1.40) and (1.41) become equalities. Therefore, (n5) and (n6) hold. □

The existence and uniqueness result in study of Problem 1.4.7 is the following.

Theorem 1.4.11 *Under Hypotheses 1.4.8, Problem 1.4.7 has a solution* $\boldsymbol{u} \in V$ *provided that* $\alpha > c_{10}$, *and*

(v3) *the solution satisfies the estimate* $\|\boldsymbol{u}\|_V \leq C$ *with* $C := \left(\frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha-c_{10}}\right)^{1/\rho}$;

(v4) *the solution is unique if, in addition, the operator* $\mathcal{A}$ *is strongly monotone with the constant* $m_{\mathcal{A}} > 0$ *and the smallness condition is satisfied:*

$$m_{\mathcal{A}} > 2c_9 \|\gamma\|^2_{V\to Y} + 2c_9 \|\gamma\|^{\rho+2}_{V\to Y} \frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha-c_{10}} + c_4 \left(\frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha-c_{10}}\right)^{1/\rho}$$

where $\rho \in (0, 1]$, α *is the coercivity constant of* $\mathcal{A}$, c_4 *is the continuity constant of the trilinear form b associated with* $\mathcal{B}$, $c_{10} := \max\left\{c_9 \|\gamma\|_{V\to Y}, c_9 \|\gamma\|^{\rho+1}_{V\to Y}\right\}$ *with* $c_9 := \max\left\{2d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, \frac{2b_\kappa d_\kappa}{\kappa}, 2c_{e1}d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, c_{e1}\frac{2b_\kappa d_\kappa}{\kappa}\right\}$, *whereas* $b_\kappa, d_\kappa, a_\kappa$ *such as in hypothesis (V6).*

Proof The proof is complete by Lemmata 1.4.10 and 1.4.12, and Theorem 1.3.4. Note that, because of (1.33) and (n4), we have to modify the proof of Theorem 1.3.4. It is enough to replace Steps 1.3.3, 1.3.5, and 1.3.6 by Steps 1.4.1 to 1.4.3, respectively:

Step 1.4.1 We show that $\|\mathsf{F}(\boldsymbol{v})\|_{Y^*} \leq c_{10}(1+\|\gamma\boldsymbol{v}\|^\rho_V)$ *for all* $\boldsymbol{v} \in V$ *with* $0 < \rho \leq 1$ *and* $c_{10} := \max\{c_9 \|\gamma\|_{V\to Y}, c_9 \|\gamma\|^{\rho+1}_{V\to Y}\}$ *with* c_9 *such as in (n4).* Let $\boldsymbol{\eta} \in \mathsf{F}(\boldsymbol{v})$ for $\boldsymbol{v} \in V$. Thus

$$|\langle\boldsymbol{\eta}, \boldsymbol{v}\rangle_{V^*\times V}| = \left|\langle\gamma^*\boldsymbol{\zeta}, \boldsymbol{v}\rangle_{V^*\times V}\right| \leq \left\|\gamma^*\right\|_{Y^*\to V^*} \|\boldsymbol{\zeta}\|_{Y^*} \|\boldsymbol{v}\|_V, \quad \forall \boldsymbol{v} \in V$$

for $\boldsymbol{\eta} = \gamma^*\boldsymbol{\zeta}$ and $\boldsymbol{\zeta} \in \partial J(\gamma\boldsymbol{v})$. It follows from (n4) that

$$\|\boldsymbol{\eta}\|_{V^*} \leq \left\|\gamma^*\right\|_{Y^*\to V^*} c_9(1 + \|\gamma\boldsymbol{v}\|^\rho_Y) \leq c_{10}(1 + \|\boldsymbol{v}\|^\rho_V).$$

Step 1.4.2 We show that the multivalued operator $\mathsf{G}\colon V \multimap V^*$ *is pseudomonotone and coercive provided* $\alpha > c_{10}$. Since the class of multivalued pseudomonotone operators is closed under addition of mappings (see [18, Proposition 6.3.68]), it follows that G is pseudomonotone due to Lemma 1.2.5 and Step 1.3.4. Furthermore, Lemma 1.2.5 shows that for all $\boldsymbol{v} \in V$ and $\boldsymbol{\eta} \in \mathsf{F}(\boldsymbol{v})$, we have

$$\langle\mathsf{G}(\boldsymbol{v}), \boldsymbol{v}\rangle_{V^*\times V} = \langle\mathcal{N}(\boldsymbol{v}), \boldsymbol{v}\rangle_{V^*\times V} + \langle\boldsymbol{\eta}, \boldsymbol{v}\rangle_{V^*\times V} \geq \alpha\|\boldsymbol{v}\|^2_V + \langle\boldsymbol{\eta}, \boldsymbol{v}\rangle_{V^*\times V}.$$

By Step 1.4.1, we obtain

$$\langle\boldsymbol{v}^*, \boldsymbol{v}\rangle_{V^*\times V} \geq \alpha\|\boldsymbol{v}\|^2_V - c_{10}(1+\|\boldsymbol{v}\|^\rho_V)\|\boldsymbol{v}\|_V = \ell(\|\boldsymbol{v}\|_V)\|\boldsymbol{v}\|_V, \quad \forall(\boldsymbol{v}, \boldsymbol{v}^*) \in \mathrm{Gr}(\mathsf{G}),$$

where $\ell(t) := \alpha t - c_{10}t^\rho - c_{10}$ and $\lim_{t\to\infty}\ell(t) = \infty$ if $0 < \rho < 1$ or $\rho = 1$ provided $\alpha > c_{10}$. Therefore, G is coercive as claimed. In conclusion, by

[8, Theorem 3], the operator G is surjective. Hence, for every $\boldsymbol{f} \in V^*$, there exists $\boldsymbol{u} \in V$ such that $\mathcal{N}\boldsymbol{u} + \mathsf{F}(\boldsymbol{u}) \ni \boldsymbol{f}$. Furthermore, the coercivity of G yields

$$\alpha \, \|\boldsymbol{u}\|_V^2 - (c_{10} \, \|\boldsymbol{u}\|_V^\rho + c_{10}) \, \|\boldsymbol{u}\|_V \le \|\boldsymbol{f}\|_{V^*} \, \|\boldsymbol{u}\|_V \, .$$

Thus, $\alpha \, \|\boldsymbol{u}\|_V \le c_{10} \, \|\boldsymbol{u}\|_V^\rho + c_{10} + \|\boldsymbol{f}\|_{V^*}$ implies the estimate $\|\boldsymbol{u}\|_V \le C$ provided $\alpha > c_{10}$, where

- $C := \left(\frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha-c_{10}}\right)^{1/\rho}$ for $0 < \rho < 1$ and $\|\boldsymbol{u}\|_V \ge 1$;
- $C := \frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha-c_{10}}$ for $0 < \rho < 1$ and $\|\boldsymbol{u}\|_V < 1$;
- $C := \frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha-c_{10}}$ for $\rho = 1$.

Step 1.4.3 It is enough to modify a part of Step 1.3.6. Let $\boldsymbol{u}_1, \boldsymbol{u}_2 \in V$ be two solutions of Problem 1.3.1. By (n4), we have $\left\|\boldsymbol{\zeta}_j\right\|_{Y^*} \le c_9 \, (1 + \left\|\gamma \boldsymbol{u}_j\right\|_Y^\rho)$ for $\boldsymbol{\zeta}_j \in \partial J(\gamma \boldsymbol{u}_j)$ and $j = 1, 2$. Thus

$$\langle \boldsymbol{\eta}_1 - \boldsymbol{\eta}_2, \boldsymbol{u}_1 - \boldsymbol{u}_2 \rangle_{V^* \times V} \ge -2r \, \|\gamma\|_{V \to Y}^2 \, \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V^2$$

with $r := c_9 \left(1 + C^\rho \, \|\gamma\|_{V \to Y}^\rho\right) > 0$. It implies that the solution to Problem 1.3.1 is unique provided $m_{\mathcal{A}} > c_4 C + 2r \, \|\gamma\|_{V \to Y}^2$. □

Lemma 1.4.12 *Every solution of Problem 1.3.1 is also a solution of Problem 1.4.7. The converse is also true if either j or $-j$ is regular (in the sense of Clarke).*

Proof Let $\boldsymbol{u} \in V$ be a solution of Problem 1.3.1. Then, there exist $\boldsymbol{\eta} = \gamma^* \boldsymbol{\zeta} \in V^*$ and $\boldsymbol{\zeta} \in \partial J(\gamma \boldsymbol{u}) \subset Y^*$ such that $\mathcal{N}\boldsymbol{u} + \boldsymbol{\eta} = \boldsymbol{f}$. By [17, Proposition 2.1.5], we have $\langle \boldsymbol{\zeta}, \gamma \boldsymbol{v} \rangle_{Y^* \times Y} \le J^0(\gamma \boldsymbol{u}, \gamma \boldsymbol{v})$ for all $\boldsymbol{v} \in V$. Thus,

$$\begin{aligned} \langle \boldsymbol{f} - \mathcal{N}\boldsymbol{u}, \boldsymbol{v} \rangle_{V^* \times V} &= \langle \boldsymbol{\eta}, \boldsymbol{v} \rangle_{V^* \times V} = \langle \gamma^* \boldsymbol{\zeta}, \boldsymbol{v} \rangle_{V^* \times V} = \langle \boldsymbol{\zeta}, \gamma \boldsymbol{v} \rangle_{Y^* \times Y} \\ &\le J^0(\gamma \boldsymbol{u}; \gamma \boldsymbol{v}) \le \int_\Gamma j_\nu^0(\mathbf{x}, (\gamma u)_\nu(\mathbf{x}); (\gamma v)_\nu(\mathbf{x})) \, \mathrm{d}\Gamma, \quad \forall \boldsymbol{v} \in V, \end{aligned}$$

where the last inequality holds due to (n2) of Lemma 1.4.10. Therefore, $\boldsymbol{u}$ is also a solution of Problem 1.4.7. On the other hand, if either j_ν or $-j_\nu$ is regular (in the sense of Clarke), then every solution of Problem 1.4.7 is a solution of Problem 1.3.1. Indeed, from (1.32) and (n6), we obtain

$$\langle \boldsymbol{f} - \mathcal{N}\boldsymbol{u}, \boldsymbol{v} \rangle_{V^* \times V} \le \int_\Gamma j_\nu^0(\mathbf{x}, (\gamma u)_\nu(\mathbf{x}); (\gamma v)_\nu(\mathbf{x})) \, \mathrm{d}\Gamma = J^0(\gamma \boldsymbol{u}; \gamma \boldsymbol{v}).$$

It follows from [17, Theorem 2.3.10, Corollary p. 47] that

$$J^0(\gamma \boldsymbol{u}; \gamma \boldsymbol{v}) = (J \circ \gamma)^0(\boldsymbol{u}; \boldsymbol{v}) \quad \text{and} \quad \partial(J \circ \gamma)(\boldsymbol{u}) = \gamma^* \partial J(\gamma \boldsymbol{u}).$$

Therefore, $\langle \boldsymbol{f} - \mathcal{N}\boldsymbol{u}, \boldsymbol{v} \rangle_{V^* \times V} \le (J \circ \gamma)^0(\boldsymbol{u}; \boldsymbol{v})$ and $\boldsymbol{f} - \mathcal{N}\boldsymbol{u} \in \partial(J \circ \gamma)(\boldsymbol{u}) = \gamma^* \partial J(\gamma \boldsymbol{u})$. □

1.5 Steady Flows of Non-Newtonian Fluids Under Slip Boundary Conditions of Frictional Type

1.5.1 Existence and Uniqueness

In this section we apply existence and uniqueness results of Sects. 1.3 and 1.4.1 in the study of steady-state flows of isotropic, isothermal, inhomogeneous, viscous, and incompressible non-Newtonian fluids with a multivalued nonmonotone subdifferential frictional boundary condition. The results of this chapter are based on our research paper [66].

1.5.1.1 Setting of the Flow Problem

Let Ω be a bounded simply connected open set in $\mathbb{R}^d$, $d = 2, 3$, with connected boundary Γ of class C^2. We consider the following nonlinear system of equations:

$$-\operatorname{div}\mathbf{S} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} + \nabla\pi = \boldsymbol{f} \quad \text{in } \Omega, \tag{1.46a}$$

$$\operatorname{div}\boldsymbol{u} = 0 \quad \text{in } \Omega. \tag{1.46b}$$

The system describes the steady-state flow of an incompressible non-Newtonian fluid occupying the volume Ω subjected to given volume forces $\boldsymbol{f}$. Here $\boldsymbol{u}\colon \Omega \to \mathbb{R}^d$ denotes the velocity field, $\pi : \Omega \to \mathbb{R}$ the pressure, $\boldsymbol{f}\colon \Omega \to \mathbb{R}^d$ the density of external forces, and $(\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = (\sum_{j=1}^d u_j \frac{\partial u_i}{\partial x_j})_{i=1,\ldots,d}$ the convective term. The symbol $\mathbf{S}\colon \Omega \to \mathbb{S}^d$ denotes the extra (viscous) part of the (Cauchy) stress tensor in the fluid $\boldsymbol{\sigma} = -\pi\mathbf{I} + \mathbf{S}$, where $\mathbf{I}$ is the identity matrix. The extra stress tensor $\mathbf{S}$ is given by a constitutive law $\mathbf{S} = \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))$ in Ω, where $\mathbf{D}(\boldsymbol{u}) = \frac{1}{2}(\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^\top)$ denotes the symmetric part of the velocity gradient. We underline that the extra stress tensor $\mathbf{S}$ is defined on Ω, while the constitutive function $\mathbf{S} = \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))$ is defined on $\Omega \times \mathbb{S}^d$. The divergence free condition in (1.46b) is the equation for law of mass conservation and it states the motion is incompressible. The symbol div denotes the divergence operators for tensor and vector valued functions $\mathbf{S}$ and $\boldsymbol{u}$ defined by $\operatorname{div}\mathbf{S} = (\mathbf{S}_{ij,j})$ and $\operatorname{div}\boldsymbol{u} = (u_{i,i})$, where the index that follows a comma represents the partial derivative with respect to the corresponding component of $\mathbf{x}$.

The most common nonlinear model among rheologists is the power law model, corresponding to the choice $\mathbf{S}(\mathbf{x}, \mathbf{A}) = \nu_0(\kappa_0 + \|\mathbf{A}\|_{\mathbb{S}^d})^{p-2}\mathbf{A}$ for all $\mathbf{A} \in \mathbb{S}^d$ and a.e. $\mathbf{x} \in \Omega$. Here, $\nu_0 \in (0, \infty)$ and $\kappa_0 \in [0, \infty)$ are constants, and $p \in (0, \infty)$ is an exponent which needs to be specified via physical experiments. An extensive list of specific p-values for different fluids can be found in [3].

We complement the system (1.46a) and (1.46b) with boundary conditions. We denote by $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d)$ the unit outward normal vector on the boundary Γ. We also assume that the boundary Γ is composed of two sets $\overline{\Gamma}_0$ and $\overline{\Gamma}_1$, with disjoint relatively open sets Γ_0 and Γ_1 such that Γ_1 has a positive measure. Our main interest lies in the contact and slip frictional boundary conditions on the surface Γ_0.

Problem 1.5.1 Find a velocity field $\boldsymbol{u}\colon \Omega \to \mathbb{R}^d$, an extra stress tensor $\mathbf{S}\colon \Omega \to \mathbb{S}^d$ and a pressure $\pi : \Omega \to \mathbb{R}$ such that

$$-\operatorname{div} \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + \nabla \pi = \boldsymbol{f} \quad \text{in } \Omega, \tag{1.47a}$$

$$\operatorname{div} \boldsymbol{u} = 0 \quad \text{in } \Omega, \tag{1.47b}$$

$$u_\nu = 0 \quad \text{on } \Gamma_0, \tag{1.47c}$$

$$-\mathbf{S}_\tau \in \partial j_\tau(\cdot, \boldsymbol{u}_\tau(\cdot)) \quad \text{on } \Gamma_0, \tag{1.47d}$$

$$\boldsymbol{u} = 0 \quad \text{on } \Gamma_1. \tag{1.47e}$$

Conditions (1.47c) and (1.47d) in Problem 1.5.1 are called the slip boundary conditions of frictional type. On the part Γ_0, the velocity vector is decomposed into the normal and tangential components defined by $u_\nu = \boldsymbol{u} \cdot \boldsymbol{\nu}$ and $\boldsymbol{u}_\tau = \boldsymbol{u} - u_\nu \boldsymbol{\nu}$, respectively. Similarly, for the stress tensor field $\mathbf{S}$, we define its normal and tangential components by $S_\nu = (\mathbf{S}\boldsymbol{\nu}) \cdot \boldsymbol{\nu}$ and $\mathbf{S}_\tau = \mathbf{S}\boldsymbol{\nu} - S_\nu \boldsymbol{\nu}$, respectively. We assume that there is no flux through Γ_0, so that the normal component of the velocity on this part of the boundary satisfies condition (1.47c). The tangential components of the extra stress tensor and the velocity are assumed to satisfy the multivalued friction law (1.47d). In the latter, $j_\tau : \Gamma_0 \times \mathbb{R}^d \to \mathbb{R}$ is called a *superpotential* and denotes the function which is locally Lipschitz in the second variable and ∂j_τ denotes the subdifferential of $j_\tau(\mathbf{x}, \cdot)$ in the sense of Clarke. On the part Γ_1 of the boundary, it is supposed that the fluid adheres to the wall, and therefore, we consider, without loss of generality, the homogeneous Dirichlet condition (1.47e). We underline that in the slip boundary condition (1.47d), the function j_τ is, in general, nonconvex and nondifferentiable. Therefore, it models the nonmonotone slip boundary condition and the weak formulation of Problem 1.5.1 leads to a hemivariational inequality. If the potential j_τ which generates the slip condition is a convex function, then the variational formulation of Problem 1.5.1 is a variational inequality.

We comment on two concrete examples of monotone and nonmonotone slip boundary conditions of the form (1.47d). First, suppose that $\boldsymbol{u}_0 \in \mathbb{R}^d$ is a given velocity of the moving part of boundary Γ_0 and $g \in L^\infty(\Gamma_0, \mathbb{R})$ denotes a nonnegative function called a modulus of friction. Consider the convex potential $j_\tau\colon \Gamma_0 \times \mathbb{R}^d \to \mathbb{R}$ defined by $j_\tau(\mathbf{x}, \boldsymbol{\xi}) = g(\mathbf{x})\|\boldsymbol{\xi} - \boldsymbol{u}_0\|_{\mathbb{R}^d}$ for $\boldsymbol{\xi} \in \mathbb{R}^d$, a.e. $\mathbf{x} \in \Gamma_0$. This choice of j_τ leads to a threshold slip condition considered earlier in [21, 34, 35, 44]. It is easy to see that in this case condition (1.47d) has the following form:

$$\begin{cases} \boldsymbol{u}_\tau = \boldsymbol{u}_0 \implies \|\mathbf{S}_\tau\|_{\mathbb{R}^d} \le g(\mathbf{x}), \\ \boldsymbol{u}_\tau \ne \boldsymbol{u}_0 \implies -\mathbf{S}_\tau = g(\mathbf{x}) \frac{\boldsymbol{u}_\tau - \boldsymbol{u}_0}{\|\boldsymbol{u}_\tau - \boldsymbol{u}_0\|_{\mathbb{R}^d}}. \end{cases}$$

This condition is the well-known Tresca friction law on Γ_0, cf. [67, Section 6.3] for a detailed discussion. The interpretation of the above law is the following. In the case where the velocity of the fluid equals the velocity of the moving boundary, the tangential stress is below a certain threshold value. If the slip between the velocity of the fluid and the velocity of boundary occurs, then the friction force is directed opposite to the slip velocity and its magnitude is determined by the slip rate value according to the function g.

Second, we give an example of condition (1.47d) which leads to a version of a nonmonotone threshold condition. Let $a \in L^\infty(\Gamma_0, \mathbb{R})$ be a prescribed function such that $0 \le a(\mathbf{x}) < 1$ for a.e. $\mathbf{x} \in \Gamma_0$. Consider the nonconvex potential $j_\tau \colon \Gamma_0 \times \mathbb{R}^d \to \mathbb{R}$ defined by

$$j_\tau(\mathbf{x}, \boldsymbol{\xi}) = (a(\mathbf{x}) - 1)e^{-\|\boldsymbol{\xi}\|_{\mathbb{R}^d}} + a(\mathbf{x})\, \|\boldsymbol{\xi}\|_{\mathbb{R}^d} \quad \text{for } \boldsymbol{\xi} \in \mathbb{R}^d, \text{ a.e. } \mathbf{x} \in \Gamma_0. \tag{1.48}$$

The function j_τ in (1.48) is nonconvex in its second argument and its generalized gradient is given by

$$\partial j_\tau(\mathbf{x}, \boldsymbol{\xi}) = \begin{cases} \overline{B}(\mathbf{0}, 1) & \text{if} \quad \boldsymbol{\xi} = 0, \\ \left((1 - a(\mathbf{x}))e^{-\|\boldsymbol{\xi}\|_{\mathbb{R}^d}} + a(\mathbf{x})\right) \dfrac{\boldsymbol{\xi}}{\|\boldsymbol{\xi}\|_{\mathbb{R}^d}} & \text{if} \quad \boldsymbol{\xi} \neq 0 \end{cases}$$

for all $\boldsymbol{\xi} \in \mathbb{R}^d$, a.e. $\mathbf{x} \in \Gamma_0$, where $\overline{B}(\mathbf{0}, 1)$ denotes the closed unit ball in $\mathbb{R}^d$. In this case the condition (1.47d) reduces to the threshold slip law of the form

$$\begin{cases} \|\mathbf{S}_\tau\|_{\mathbb{R}^d} \le 1 & \text{if} \quad \boldsymbol{u}_\tau = 0, \\ -\mathbf{S}_\tau = \left((1 - a(\mathbf{x}))e^{-\|\boldsymbol{u}_\tau\|_{\mathbb{R}^d}} + a(\mathbf{x})\right) \dfrac{\boldsymbol{u}_\tau}{\|\boldsymbol{u}_\tau\|_{\mathbb{R}^d}} & \text{if} \quad \boldsymbol{u}_\tau \neq 0 \end{cases}$$

on Γ_0. We also mention that the multivalued condition (1.47d) incorporates various nonmonotone multivalued relations which are useful in applications, cf. [67, Section 3.3], [69, Section 1.2], and [23].

In the study of Problem 1.5.1 we adopt the following hypotheses (hypotheses (E4) and (E5) follow from [9, 14]).

Hypotheses 1.5.2 Let $\Phi, \Phi^* \colon [0, \infty) \to [0, \infty)$ be a pair of complementary N-functions, and $j_\tau \colon \Gamma \times \mathbb{R}^d \to \mathbb{R}$ and $\mathbf{S} \colon \Omega \times \mathbb{S}^d \to \mathbb{S}^d$ with $\mathbf{S}(\mathbf{x}, \mathbf{0}) = \mathbf{0}$ for a.e. $\mathbf{x} \in \Omega$ be Carathéodory functions such that:

(E1) $\Phi, \Phi^* \in \Delta_2^\infty$ and $A \prec \Phi \lll (\Phi_*)^{1-1/d}$, with Φ_* being the Sobolev conjugate N-function of Φ and $A(t) = t^p$ for $t \in [0, \infty)$ with $p \ge \frac{3d}{d+2}$ and $d = 2, 3$;
(E2) $j_\tau(\mathbf{x}, \cdot)$ is locally Lipschitz for a.e. $\mathbf{x} \in \Gamma$;
(E3) for some $\kappa > 0$ there exist constants $b_\kappa, d_\kappa > 0$ and $a_\kappa \in L^1(\Gamma, [0, \infty))$ such that $\Phi^*\left(\|\mathbf{a}^*\|_{\mathbb{R}^d} / d_\kappa\right) \le a_\kappa(\mathbf{x}) + b_\kappa \Phi\left(\|\mathbf{a}\|_{\mathbb{R}^d} / \kappa\right)$ for all $(\mathbf{a}, \mathbf{a}^*) \in \mathrm{Gr}(\partial j_\tau)$ and a.e. $\mathbf{x} \in \Gamma$;

(E4) there exists $c_{11} > 0$ such that for all $\mathbf{A} \in \mathbb{S}^d$ and a.e. $\mathbf{x} \in \Omega$, it holds

$$\mathbf{S}(\mathbf{x}, \mathbf{A}) : \mathbf{A} \geq c_{11}\Phi(\|\mathbf{A}\|_{\mathbb{S}^d}) + c_{11}\Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{A})\|_{\mathbb{S}^d});$$

(E5) $(\mathbf{S}(\mathbf{x}, \mathbf{A}) - \mathbf{S}(\mathbf{x}, \mathbf{B})) : (\mathbf{A} - \mathbf{B}) \geq 0$ for all $\mathbf{A}, \mathbf{B} \in \mathbb{S}^d$ and a.e. $\mathbf{x} \in \Omega$.

Remark 1.5.3 By hypothesis (E1), the spaces $V = W^1L_\Phi(\Omega, \mathbb{R}^d)$ and $Y = L_\Phi(\Gamma, \mathbb{R}^d)$ are separable and reflexive satisfying $V \hookrightarrow\hookrightarrow Y$, see (1.2). Hypothesis (E3) yields the result on an integral representation of the Clarke subdifferential of locally Lipschitz integral functionals defined on the Orlicz space (see [71, 72]). This representation is required to establish existence and uniqueness results in the study of Problems 1.5.1 and 1.5.7 (see Theorem 1.5.6 and 1.5.10). Note that, for instance, the condition $p \geq \frac{3d}{d+2}$ was introduced by [49], the condition $p \geq \frac{2d}{d+1}$ with $d \geq 2$ was introduced by [29] and [81] (in [81] the limiting case $p = \frac{2d}{d+1}$ is not included), the condition $p > \frac{2d}{d+2}$ with $d \geq 2$ was introduced by [28] by Lipschitz truncation methods; see [28, 1.3 Historical comments] for more details.

Remark 1.5.4 If the extra stress tensor $\mathbf{S}$ has convex potential (vanishing at $\mathbf{0}$), then it is simple to verify hypothesis (E4). To find N-functions Φ and Φ^* we use the following relation $\Phi(t) + \Phi^*(\Phi'(t)) = t\Phi'(t)$ for $t \in [0, \infty)$. It corresponds to the case when the Young inequality for N-functions becomes the equality. Consider, for simplicity, the constitutive function of the form $\mathbf{S}(\mathbf{x}, \mathbf{A}) = 2\mu(\|\mathbf{A}\|^2_{\mathbb{S}^d})\mathbf{A}$ for $\mathbf{A} \in \mathbb{S}^d$ which appears in modeling of the so-called generalized Newtonian fluids. For this $\mathbf{S}$, we choose

$$\Phi(t) = \int_0^{t^2} \mu(s)\,\mathrm{d}s \quad \text{for } t \in [0, \infty).$$

Then, we can show that hypothesis (E4) holds with constant $c_{17} = 1$. For such choice of Φ we only need to verify whether the N-function conditions, i.e. behavior in/near zero and near infinity, are satisfied. The monotonicity of $\mathbf{S}$ follows from the convexity of the potential.

1.5.1.2 Weak Formulation and Main Result

In order to give the variational formulation of Problem 1.5.1, we introduce some notation:

$$\mathcal{W} := \left\{ \boldsymbol{w} \in C^\infty(\overline{\Omega}, \mathbb{R}^d) \,\middle|\, \operatorname{div} \boldsymbol{w} = 0 \text{ in } \Omega,\, w_\nu = 0 \text{ on } \Gamma_0,\, \boldsymbol{w} = 0 \text{ on } \Gamma_1 \right\},$$

$$V := \overline{\mathcal{W}}^{\|\cdot\|_{W^1L_\Phi(\Omega,\mathbb{R}^d)}}, \quad Y := L_\Phi(\Gamma, \mathbb{R}^d), \tag{1.49}$$

where V stands for the closure of $\mathcal{W}$ in the norm of $W^1 L_\Phi(\Omega, \mathbb{R}^d)$. Note that one can introduce on V the norm given by

$$\|\boldsymbol{v}\|_V := \|\mathbf{D}(\boldsymbol{v})\|_{L_\Phi(\Omega, \mathbb{S}^d)} \quad \text{for } \boldsymbol{v} \in V, \tag{1.50}$$

because in virtue of [36, Lemmata 1.2 and 1.12] and [37, p. 55] together with $\Phi^* \in \Delta_2^\infty$ (see hypothesis (E1)), one can admit on V the following norm $\|\boldsymbol{v}\|_V = \|\boldsymbol{v}\|_{W^1 L_\Phi(\Omega, \mathbb{R}^d)}$ for $\boldsymbol{v} \in V$. Owing to [36, Corollary 5.8], it is equivalent to the norm $\|\boldsymbol{v}\|_{\mathring{W}^1 L_\Phi(\Omega, \mathbb{R}^d)} = \|\nabla \boldsymbol{v}\|_{L_\Phi(\Omega, \mathbb{R}^{d\times d})}$ for $\boldsymbol{v} \in V$. Hence, we obtain (1.50) due to $\Phi, \Phi^* \in \Delta_2^\infty$ and Korn's inequality (see Theorem 1.2.6).

Next, multiplying (1.47a) by $\boldsymbol{v} \in V$ and applying Green's formula [37, p. 57], we obtain

$$\int_\Omega \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) : \mathbf{D}(\boldsymbol{v})\, \mathrm{d}\mathbf{x} + \int_\Omega (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} \cdot \boldsymbol{v}\, \mathrm{d}\mathbf{x} + \int_\Omega \nabla \pi \cdot \boldsymbol{v}\, \mathrm{d}\mathbf{x}$$
$$= \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}\, \mathrm{d}\mathbf{x} + \int_\Gamma \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\boldsymbol{\nu} \cdot \boldsymbol{v}\, \mathrm{d}\Gamma.$$

By boundary conditions (1.47c) and (1.47e), as well as the fact that the functions in V are divergence free, we obtain

$$\int_\Omega \nabla \pi \cdot \boldsymbol{v}\, \mathrm{d}\mathbf{x} = -\int_\Omega \pi \operatorname{div} \boldsymbol{v}\, \mathrm{d}\mathbf{x} + \int_\Gamma (\boldsymbol{v} \cdot \boldsymbol{\nu})\pi\, \mathrm{d}\Gamma = \left(\int_{\Gamma_0} v_\nu \pi\, \mathrm{d}\Gamma + \int_{\Gamma_1} v_\nu \pi\, \mathrm{d}\Gamma \right) = 0.$$

Likewise, by (1.47c) and (1.47e), we infer that

$$\int_\Gamma \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\boldsymbol{\nu} \cdot \boldsymbol{v}\, \mathrm{d}\Gamma = \int_{\Gamma_0 \cup \Gamma_1} \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\boldsymbol{\nu} \cdot (\boldsymbol{v}_\tau + v_\nu \boldsymbol{\nu})\, \mathrm{d}\Gamma = \int_{\Gamma_0} \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\boldsymbol{\nu} \cdot \boldsymbol{v}_\tau\, \mathrm{d}\Gamma.$$

Therefore,

$$\int_\Omega \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) : \mathbf{D}(\boldsymbol{v})\, \mathrm{d}\mathbf{x} + \int_\Omega (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} \cdot \boldsymbol{v}\, \mathrm{d}\mathbf{x} + \int_{\Gamma_0} -\mathbf{S}_\tau(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) \cdot \boldsymbol{v}_\tau\, \mathrm{d}\Gamma = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}\, \mathrm{d}\mathbf{x}.$$

The definition of the Clarke subdifferential and the friction law (1.47d) yield

$$\int_{\Gamma_0} -\mathbf{S}_\tau(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) \cdot \boldsymbol{v}_\tau\, \mathrm{d}\Gamma \le \int_{\Gamma_0} j_\tau^0(\boldsymbol{u}_\tau; \boldsymbol{v}_\tau)\, \mathrm{d}\Gamma.$$

Hence, the variational formulation of Problem 1.5.1 has the following form:

$$\int_\Omega \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) : \mathbf{D}(\boldsymbol{v})\, \mathrm{d}\mathbf{x} + \int_\Omega (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} \cdot \boldsymbol{v}\, \mathrm{d}\mathbf{x} + \int_{\Gamma_0} j_\tau^0(\boldsymbol{u}_\tau; \boldsymbol{v}_\tau)\, \mathrm{d}\Gamma \ge \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}\, \mathrm{d}\mathbf{x}, \tag{1.51}$$

for all $\boldsymbol{v} \in V$, where $\mathbf{S} = \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))$ is the extra stress tensor and j_τ^0 stands for the generalized directional derivative of $j_\tau(\mathbf{x}, \cdot)$ (with $j_\tau : \Gamma \times \mathbb{R}^d \to \mathbb{R}$ being a prescribed function). Introducing the operators $\mathcal{A}\colon V \to V^*$ and $\mathcal{B}[\cdot]\colon V \to V^*$ defined by

$$\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} = \int_\Omega \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) : \mathbf{D}(\boldsymbol{v})\, d\mathbf{x}, \quad \boldsymbol{u}, \boldsymbol{v} \in V, \tag{1.52}$$

$$\langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{v}), \boldsymbol{w}\rangle_{V^*\times V} = \int_\Omega (\boldsymbol{u} \cdot \nabla)\boldsymbol{v} \cdot \boldsymbol{w}\, d\mathbf{x}, \quad \mathcal{B}[\boldsymbol{v}] = \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}), \quad \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V, \tag{1.53}$$

we obtain the following variational formulation of Problem 1.5.1.

Problem 1.5.5 Let Ω be a bounded simply connected open set in $\mathbb{R}^d$, $d = 2, 3$, with connected boundary $\partial\Omega = \Gamma$ of class C^2, V and Y be Banach spaces as in (1.49), and $\boldsymbol{f} \in V^*$. Find $\boldsymbol{u} \in V$ such that

$$\langle \mathcal{A}\boldsymbol{u} + \mathcal{B}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} + \int_\Gamma j_\tau^0(\mathbf{x}, \boldsymbol{u}_\tau(\mathbf{x}); \boldsymbol{v}_\tau(\mathbf{x}))\, d\Gamma \ge \langle \boldsymbol{f}, \boldsymbol{v}\rangle_{V^*\times V}, \ \forall \boldsymbol{v} \in V, \tag{1.54}$$

where the operators $\mathcal{A}$ and $\mathcal{B}$ are given by (1.52) and (1.53), respectively, and j_τ^0 stands for the generalized directional derivative of the superpotential $j_\tau(\mathbf{x}, \cdot)$.

The existence and uniqueness result in study of Problem 1.5.5 is the following.

Theorem 1.5.6 *Under Hypotheses 1.5.2, Problem 1.5.5 has a solution $\boldsymbol{u} \in V$ provided $c_{11} > c_6$, and*

(e1) *the solution satisfies the estimate $\|\boldsymbol{u}\|_V \le C$ with $C := \left(\frac{\|\boldsymbol{f}\|_{V^*} + c_5}{c_{11} - c_6}\right)^{1/(\rho-1)}$;*
(e2) *the solution is unique if, in addition, there exists $c_{12} > 0$ such that:*

(E6) $(\mathbf{S}(\mathbf{x}, \mathbf{A}) - \mathbf{S}(\mathbf{x}, \mathbf{B})) : (\mathbf{A} - \mathbf{B}) \ge c_{12}\Phi(\|\mathbf{A} - \mathbf{B}\|_{\mathbb{S}^d})$ *for all* $\mathbf{A}, \mathbf{B} \in \mathbb{S}^d$, *a.e.* $\mathbf{x} \in \Gamma$;
(E7) $c_{12} > 2c_8 \|\gamma\|_{V\to Y}^2 + 2c_7 \frac{\|\boldsymbol{f}\|_{V^*} + c_5}{c_{11} - c_6} \|\gamma\|_{V\to Y}^{\rho+1} + c_4 \left(\frac{\|\boldsymbol{f}\|_{V^*} + c_5}{c_{11} - c_6}\right)^{1/(\rho-1)}$;

where $\rho \in (1, \infty)$, the constant $c_{11} > 0$ follows from hypothesis (S4)*, and c_4 is the continuity constant of the trilinear form b associated with the operator $\mathcal{B}$, whereas $c_5 := c_8 \|\gamma\|_{V\to Y}$ and $c_6 := c_7 \|\gamma\|_{V\to Y}^\rho$ with $c_7 := \frac{2b_\kappa d_\kappa}{\kappa} > 0$ and $c_8 := 2d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + c_7 > 0$ for $b_\kappa, d_\kappa, a_\kappa$ from hypothesis* (S3).

Proof We apply Theorem 1.4.5. To this end, we prove that $\mathcal{A} + \mathcal{B}$ is the Navier–Stokes type operator. The proof is complete by Steps 1.5.1 to 1.5.3.

Step 1.5.1 We show that the operator $\mathcal{A}$ given by (1.52) *satisfies conditions of Definition 1.2.4.* Firstly, we prove that the operator $\mathcal{A}$ pseudomonotone. To this end, by [67, Theorem 3.69], it suffices to show that $\mathcal{A}$ is bounded, monotone, and hemicontinuous.

We show that operator $\mathcal{A}$ is bounded. By hypothesis (E4), for a constant $c_{13} \in (0, 1]$, we have

$$\frac{c_{13}}{2}\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) : \frac{2}{c_{13}}\mathbf{D}(\boldsymbol{u}) \geq c_{11}\Phi(\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d}) + c_{11}\Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d})$$
$$\geq c_{11}\Phi(\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d}) + c_{13}\Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d}).$$

Applying Young's inequality to the left-hand side of the above inequality, we obtain

$$\Phi^*\left(\frac{c_{13}}{2}\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d}\right) + \Phi\left(\frac{2}{c_{13}}\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d}\right)$$
$$\geq c_{11}\Phi(\|\mathbf{D}(\boldsymbol{v})\|_{\mathbb{S}^d}) + c_{13}\Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d}).$$

Since Φ^* is convex, $\Phi(0) = 0$ and $c_{13} \in (0, 1]$, it follows that

$$\Phi\left(\frac{2}{c_{13}}\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d}\right) \geq c_{13}\Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d}) - \Phi^*\left(\frac{c_{13}}{2}\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d}\right)$$
$$\geq c_{13}\Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d}) - \frac{c_{13}}{2}\Phi^*\left(\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d}\right)$$
$$\geq \frac{c_{13}}{2}\Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d}).$$

Having $\Phi \in \Delta_2^\infty$ due to hypothesis (E1), it follows from $\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d} \in L_\Phi(\Omega, \mathbb{R})$ that $\frac{2}{c_{13}}\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d} \in L_\Phi(\Omega, \mathbb{R})$. Thus

$$\int_\Omega \Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d})\, d\mathbf{x} \leq \frac{2}{c_{13}}\int_\Omega \Phi\left(\frac{2}{c_{13}}\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d}\right) d\mathbf{x} < \infty. \tag{1.55}$$

By the Young inequality together with (1.55), we infer that

$$\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} = \int_\Omega \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) : \mathbf{D}(\boldsymbol{v})\, d\mathbf{x}$$
$$\leq \int_\Omega \Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d})\, d\mathbf{x} + \int_\Omega \Phi(\|\mathbf{D}(\boldsymbol{v})\|_{\mathbb{S}^d})\, d\mathbf{x}$$
$$\leq \frac{2}{c_{13}}\int_\Omega \Phi\left(\frac{2}{c_{13}}\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d}\right) d\mathbf{x} + \int_\Omega \Phi(\|\mathbf{D}(\boldsymbol{v})\|_{\mathbb{S}^d})\, d\mathbf{x}.$$

Because of hypothesis (E1) and Corollary 1.2.8, we obtain

$$\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} \leq 2c_1\left(1 + \frac{2^\rho}{c_{13}^\rho}\|\mathbf{D}(\boldsymbol{u})\|^\rho_{L_\Phi(\Omega,\mathbb{S}^d)}\right) + c_1\left(1 + \|\mathbf{D}(\boldsymbol{v})\|^\rho_{L_\Phi(\Omega,\mathbb{S}^d)}\right)$$
$$\leq c_{14} + c_{15}(\|\boldsymbol{u}\|_V^\rho + \|\boldsymbol{v}\|_V^\rho),$$

where $c_{14}, c_{15} > 0$. This implies that the operator $\mathcal{A}\colon V \to V^*$ is well defined and bounded due to estimate

$$\|\mathcal{A}\boldsymbol{u}\|_{V^*} \le c_{16} + c_{15} \|\boldsymbol{u}\|_V^\rho, \quad \forall \boldsymbol{u} \in V, \tag{1.56}$$

where $c_{15}, c_{16} > 0$ and $\rho > 1$.

Next, we show that $\mathcal{A}$ is continuous, which implies that $\mathcal{A}$ is hemicontinuous. Let $\boldsymbol{v}_n \to \boldsymbol{v}$ in V. Hence, $\mathbf{D}(\boldsymbol{v_n}) \to \mathbf{D}(\boldsymbol{v})$ in $L_\Phi(\Omega, \mathbb{S}^d)$. By [43, Theorems IV3.1 and I.6.4], there exists a subsequence $\boldsymbol{v}_{n_k}$ of $\boldsymbol{v}_n$ such that $\mathbf{D}(\boldsymbol{v_{n_k}}) \to \mathbf{D}(\boldsymbol{v})$ for a.e. $\mathbf{x} \in \Omega$. Since $\mathbf{S}$ is continuous by Hypotheses 1.5.2, we infer that $\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{v_{n_k}})) \to \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{v}))$ for a.e. $\mathbf{x} \in \Omega$. Note that $\mathbf{S}(\cdot, \mathbf{D}(\boldsymbol{v_{n_k}})), \mathbf{S}(\cdot, \mathbf{D}(\boldsymbol{v})) \in L_{\Phi^*}(\Omega, \mathbb{S}^d)$ due to hypothesis (E4). Applying the Lebesgue dominated convergence theorem for Orlicz spaces (see [47, p. 159]), we obtain $\mathbf{S}(\cdot, \mathbf{D}(\boldsymbol{v_{n_k}})) \to \mathbf{S}(\cdot, \mathbf{D}(\boldsymbol{v}))$ in $L_{\Phi^*}(\Omega, \mathbb{S}^d)$, where $\Phi^* \in \Delta_2^\infty$ by hypothesis (E1). By the Hölder inequality, we obtain

$$\begin{aligned}
\langle \mathcal{A}\boldsymbol{v}_{n_k} - \mathcal{A}\boldsymbol{v}, \boldsymbol{w}\rangle_{V^*\times V} &= \int_\Omega (\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{v_{n_k}})) - \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{v}))) : \mathbf{D}(\boldsymbol{w})\, \mathrm{d}\mathbf{x} \\
&\le 2 \left\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{v_{n_k}})) - \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{v}))\right\|_{L_{\Phi^*}(\Omega, \mathbb{S}^d)} \|\mathbf{D}(\boldsymbol{w})\|_{L_\Phi(\Omega, \mathbb{S}^d)} \\
&\le 2 \left\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{v_{n_k}})) - \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{v}))\right\|_{L_{\Phi^*}(\Omega, \mathbb{S}^d)} \|\boldsymbol{w}\|_V
\end{aligned}$$

for all $\boldsymbol{w} \in V$. Therefore, $\left\|\mathcal{A}\boldsymbol{v}_{n_k} - \mathcal{A}\boldsymbol{v}\right\|_{V^*} \to 0$, as $k \to \infty$ for any subsequence $\boldsymbol{v}_{n_k}$ of $\boldsymbol{v}_n$. Thus, we have $\mathcal{A}\boldsymbol{v}_n \to \mathcal{A}\boldsymbol{v}$ in V^*. This shows that the operator $\mathcal{A}$ is continuous.

Secondly, it is clear that the operator $\mathcal{A}$ is monotone due to hypothesis (E5). We proceed to show that $\mathcal{A}$ is (α, ρ)-coercive. Hypothesis (E4) and (1.50) yield

$$\begin{aligned}
\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{u}\rangle_{V^*\times V} &= \int_\Omega \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) : \mathbf{D}(\boldsymbol{u})\, \mathrm{d}\mathbf{x} \\
&\ge c_{11} \int_\Omega \Phi(\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d})\, \mathrm{d}\mathbf{x} + c_{11} \int_\Omega \Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\|_{\mathbb{S}^d})\, \mathrm{d}\mathbf{x} \\
&\ge c_{11} \int_\Omega \Phi(\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d})\, \mathrm{d}\mathbf{x} \ge c_{11} \|\mathbf{D}(\boldsymbol{u})\|^\rho_{L_\Phi(\Omega, \mathbb{S}^d)} = c_{11} \|\boldsymbol{u}\|_V^\rho
\end{aligned}$$

for all $\boldsymbol{u} \in V$ with $\rho \in (1, \infty)$. The last inequality holds due to (1.3) and (1.6) of Lemma 1.2.7. So, $\mathcal{A}$ is (α, ρ)-coercive with $\alpha := c_{11}$, which completes the proof.

Step 1.5.2 The operator $\mathcal{B}$ given by (1.53) *satisfies conditions of Definition 1.2.4.* By hypothesis (E1) and [42, Theorem 2.2], we obtain $W^1L_\Phi(\Omega, \mathbb{R}^d) \hookrightarrow W^{1,p}(\Omega, \mathbb{R}^d)$ with $p \ge \frac{3d}{d+2}$ and $d = 2, 3$. In view of (1.50) and the standard estimate

$$\int_\Omega (\boldsymbol{u} \cdot \nabla)\boldsymbol{v} \cdot \boldsymbol{w}\, \mathrm{d}\mathbf{x} \le c\, \|\boldsymbol{u}\|_{W^{1,p}(\Omega, \mathbb{R}^d)} \|\boldsymbol{v}\|_{W^{1,p}(\Omega, \mathbb{R}^d)} \|\boldsymbol{w}\|_{W^{1,p}(\Omega, \mathbb{R}^d)}$$

with $c > 0$ for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in W^{1,p}(\Omega, \mathbb{R}^d)$ with $p \geq \frac{3d}{d+2}$ and $d = 2, 3$, we infer that

$$\begin{aligned} \langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{v}), \boldsymbol{w} \rangle_{V^* \times V} &\leq c \, \|\boldsymbol{u}\|_{W^1 L_\Phi(\Omega, \mathbb{R}^d)} \, \|\boldsymbol{v}\|_{W^1 L_\Phi(\Omega, \mathbb{R}^d)} \, \|\boldsymbol{w}\|_{W^1 L_\Phi(\Omega, \mathbb{R}^d)} \\ &\leq c \, \|\boldsymbol{u}\|_V \, \|\boldsymbol{v}\|_V \, \|\boldsymbol{w}\|_V \end{aligned} \tag{1.57}$$

for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ with a positive constant c. Hence, the operator $\mathcal{B} \colon V \times V \to V^*$ is well defined and continuous.

We are now in position to show that $\langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{v}), \boldsymbol{v} \rangle_{V^* \times V} = 0$. Applying the Green formula to (1.53), we obtain

$$\begin{aligned} \langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{v}), \boldsymbol{v} \rangle_{V^* \times V} &= \int_\Omega \sum_{i,j=1}^d u_j \frac{\partial v_i}{\partial x_j} v_i \, d\mathbf{x} = \frac{1}{2} \int_\Omega \sum_{j=1}^d u_j \frac{\partial}{\partial x_j} \left(\sum_{i=1}^d v_i^2 \right) d\mathbf{x} \\ &= -\frac{1}{2} \int_\Omega \sum_{j=1}^d \frac{\partial u_j}{\partial x_j} \left(\sum_{i=1}^d v_i^2 \right) d\mathbf{x} + \frac{1}{2} \int_\Gamma \sum_{j=1}^d u_j \nu_j \left(\sum_{i=1}^d v_i^2 \right) d\Gamma = 0, \end{aligned}$$

where the last equality holds due to (1.47b) and (1.47c). In consequence

$$\langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{v}), \boldsymbol{w} \rangle_{V^* \times V} = -\langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{w}), \boldsymbol{v} \rangle_{V^* \times V}, \tag{1.58}$$

by $\langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{v}+\boldsymbol{w}), \boldsymbol{v}+\boldsymbol{w} \rangle = \langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{v}), \boldsymbol{v} \rangle + \langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{v}), \boldsymbol{w} \rangle + \langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{w}), \boldsymbol{v} \rangle + \langle \mathcal{B}(\boldsymbol{u}, \boldsymbol{w}), \boldsymbol{w} \rangle$.

It remains to prove that $\mathcal{B}[\cdot] \colon V \to V^*$ is weakly sequentially continuous, that is, $\boldsymbol{v}_n \rightharpoonup \boldsymbol{v}$ in $\sigma(V, V^*)$ implies $\langle \mathcal{B}(\boldsymbol{v}_n, \boldsymbol{v}_n), \boldsymbol{w} \rangle \to \langle \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}), \boldsymbol{w} \rangle$ for all $\boldsymbol{w} \in V$. By (1.58), we have to show that

$$\lim_{n \to \infty} \langle \mathcal{B}(\boldsymbol{v}_n, \boldsymbol{w}), \boldsymbol{v}_n \rangle_{V^* \times V} = \langle \mathcal{B}(\boldsymbol{v}, \boldsymbol{w}), \boldsymbol{v} \rangle_{V^* \times V}, \quad \forall \boldsymbol{w} \in V. \tag{1.59}$$

Indeed, let $\boldsymbol{w} \in \mathcal{W}$ and $\boldsymbol{v}_n \rightharpoonup \boldsymbol{v}$ in $\sigma(V, V^*)$. By (1.1) and (1.49), we infer that $V \subset W^1 L_\Phi(\Omega, \mathbb{R}^d)$ and $\boldsymbol{v}_n \to \boldsymbol{v}$ in $L_\Phi(\Omega, \mathbb{R}^d)$. From (1.57) and

$$\begin{aligned} |\langle \mathcal{B}(\boldsymbol{v}_n, \boldsymbol{w}), \boldsymbol{v}_n \rangle - \langle \mathcal{B}(\boldsymbol{v}, \boldsymbol{w}), \boldsymbol{v} \rangle| &= |\langle \mathcal{B}(\boldsymbol{v}_n, \boldsymbol{w}), \boldsymbol{v}_n - \boldsymbol{v} \rangle - \langle \mathcal{B}(\boldsymbol{v}_n - \boldsymbol{v}, \boldsymbol{w}), \boldsymbol{v} \rangle| \\ &\leq 2c_{\mathcal{B}} (\|\boldsymbol{v}_n\|_{L_\Phi(\Omega, \mathbb{R}^d)} + \|\boldsymbol{v}\|_{L_\Phi(\Omega, \mathbb{R}^d)}) \, \|\boldsymbol{v}_n - \boldsymbol{v}\|_{L_\Phi(\Omega, \mathbb{R}^d)} \, \|\boldsymbol{w}\|_{C^1(\overline{\Omega}, \mathbb{R}^d)}, \end{aligned}$$

we infer that (1.59) holds for all $\boldsymbol{w} \in \mathcal{W}$, and by (1.58) we obtain the desired conclusion $\langle \mathcal{B}(\boldsymbol{v}_n, \boldsymbol{v}_n), \boldsymbol{w} \rangle \to \langle \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}), \boldsymbol{w} \rangle$ for all $\boldsymbol{w} \in \mathcal{W}$.

Now, for $\boldsymbol{w} \in V$ there exists a sequence $\boldsymbol{w}_k \in \mathcal{W}$ such that $\boldsymbol{w}_k \to \boldsymbol{w}$, as $k \to \infty$. Since $\langle \mathcal{B}(\boldsymbol{v}_j, \boldsymbol{v}_j), \boldsymbol{w} \rangle = \langle \mathcal{B}(\boldsymbol{v}_j, \boldsymbol{v}_j), \boldsymbol{w}_k \rangle + \langle \mathcal{B}(\boldsymbol{v}_j, \boldsymbol{v}_j), \boldsymbol{w} - \boldsymbol{w}_k \rangle$, it follows from (1.57) that $\langle \mathcal{B}(\boldsymbol{v}_j, \boldsymbol{v}_j), \boldsymbol{w} - \boldsymbol{w}_k \rangle_{V^* \times V} \le c_{\mathcal{B}} \left\| \boldsymbol{v}_j \right\|_V \left\| \boldsymbol{v}_j \right\|_V \| \boldsymbol{w} - \boldsymbol{w}_k \|_V$. Thus

$$\begin{aligned} &\lim_{k\to\infty} \limsup_{j\to\infty} \left| \langle \mathcal{B}(\boldsymbol{v}_j, \boldsymbol{v}_j), \boldsymbol{w} - \boldsymbol{w}_k \rangle_{V^* \times V} \right| \\ &\quad \le \lim_{k\to\infty} \limsup_{j\to\infty} c_{\mathcal{B}} \left\| \boldsymbol{v}_j \right\|_V \left\| \boldsymbol{v}_j \right\|_V \| \boldsymbol{w} - \boldsymbol{w}_k \|_V = 0, \end{aligned}$$

which yields $\langle \mathcal{B}(\boldsymbol{v}_n, \boldsymbol{v}_n), \boldsymbol{w} \rangle \to \langle \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}), \boldsymbol{w} \rangle$ for all $\boldsymbol{w} \in V$.

Step 1.5.3 The uniqueness of a solution to Problem 1.5.5. To provide the uniqueness of a solution to Problem 1.5.5, it suffices to show that $\mathcal{A}$ is strongly monotone due to Theorem 1.4.5. Hypothesis (E6) and Lemma 1.2.7 lead to

$$\begin{aligned} \langle \mathcal{A}\boldsymbol{u}_1 - \mathcal{A}\boldsymbol{u}_2, \boldsymbol{u}_1 - \boldsymbol{u}_2 \rangle &= \int_\Omega (\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}_1)) - \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}_2)))(\mathbf{D}(\boldsymbol{u}_1) - \mathbf{D}(\boldsymbol{u}_2))\, d\mathbf{x} \\ &\ge c_{12} \int_\Omega \Phi(\|\mathbf{D}(\boldsymbol{u}_1) - \mathbf{D}(\boldsymbol{u}_2)\|_{\mathbb{S}^d})\, d\mathbf{x} \\ &\ge c_{12} \|\mathbf{D}(\boldsymbol{u}_1) - \mathbf{D}(\boldsymbol{u}_2)\|^2_{L_\Phi(\Omega, \mathbb{S}^d)} \ge c_{12} \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V^2, \end{aligned}$$

which means that the operator $\mathcal{A}$ is strongly monotone with $m_{\mathcal{A}} := c_{12} > 0$. □

1.5.2 *Slow Flows*

In various instances of interest in applications, the convective term $(\boldsymbol{u} \cdot \nabla)\boldsymbol{u}$ is negligible in (1.46a), compared with the other terms appearing in this equation. This is the case, for example, if the modulus of the velocity is small. Another situation where the role of the convective term is negligible appears in plastic or pseudo-plastic fluids. Indeed, the convective term accounts for the inner rotation in the fluid flow, and for such fluids the impact of this term is very limited. Dropping the convective term reduces Problem 1.5.1 to the following system. Note that the following slow steady fluid flow problem without friction has been studied extensively by variational methods, see, e.g., [7, 20, 33].

Problem 1.5.7 Find a velocity field $\boldsymbol{u} \colon \Omega \to \mathbb{R}^d$, an extra stress tensor $\mathbf{S} \colon \Omega \to \mathbb{S}^d$ and a pressure $\pi : \Omega \to \mathbb{R}$ such that

$$\begin{aligned} -\operatorname{div} \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) + \nabla\pi &= \boldsymbol{f} \quad \text{in } \Omega, \\ \operatorname{div} \boldsymbol{u} &= 0 \quad \text{in } \Omega, \\ u_\nu &= 0 \quad \text{on } \Gamma_0, \end{aligned}$$

$$-\mathbf{S}_\tau \in \partial j_\tau(\cdot, \boldsymbol{u}_\tau(\cdot)) \quad \text{on } \Gamma_0,$$
$$\boldsymbol{u} = 0 \quad \text{on } \Gamma_1.$$

For the above problem we provide a variational formulation.

Problem 1.5.8 Let Ω be a bounded simply connected open set in $\mathbb{R}^d$, $d = 2, 3$, with connected boundary $\partial\Omega = \Gamma$ of class C^2, V and Y be Banach spaces as in (1.49), and $\boldsymbol{f} \in V^*$. Find $\boldsymbol{u} \in V$ such that

$$\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} + \int_\Gamma j_\tau^0(\mathbf{x}, \boldsymbol{u}_\tau(\mathbf{x}); \boldsymbol{v}_\tau(\mathbf{x}))\, \mathrm{d}\Gamma \geq \langle \boldsymbol{f}, \boldsymbol{v}\rangle_{V^*\times V}, \ \forall \boldsymbol{v} \in V, \tag{1.60}$$

where $\mathcal{A}\colon V \to V^*$ is given by (1.52).

In the study of Problem 1.5.8 we adopt the following hypotheses.

Hypotheses 1.5.9 Let $\Phi, \Phi^*\colon [0, \infty) \to [0, \infty)$ be a pair of complementary N-functions, and $j_\tau\colon \Gamma \times \mathbb{R}^d \to \mathbb{R}$ and $\mathbf{S}\colon \Omega \times \mathbb{S}^d \to \mathbb{S}^d$ with $\mathbf{S}(\mathbf{x}, \mathbf{0}) = \mathbf{0}$ for a.e. $\mathbf{x} \in \Omega$ be Carathéodory functions such that:

(S1) $\Phi, \Phi^* \in \Delta_2^\infty$ and $\Phi \lll (\Phi_*)^{1-1/d}$, with Φ_* being the Sobolev conjugate N-function of Φ and $d = 2, 3$;
(S2) $j_\tau(\mathbf{x}, \cdot)$ is locally Lipschitz continuous for a.e. $\mathbf{x} \in \Gamma$;
(S3) for some $\kappa > 0$ there exist constants $b_\kappa, d_\kappa > 0$ and $a_\kappa \in L^1(\Gamma, [0, \infty))$ such that $\Phi^*\left(\|\mathbf{a}^*\|_{\mathbb{R}^d}/d_\kappa\right) \leq a_\kappa(\mathbf{x}) + b_\kappa \Phi\left(\|\mathbf{a}\|_{\mathbb{R}^d}/\kappa\right)$ for all $(\mathbf{a}, \mathbf{a}^*) \in \mathrm{Gr}(\partial j_\tau)$ and a.e. $\mathbf{x} \in \Gamma$;
(S4) there exists $c_{11} > 0$ such that for all $\mathbf{A} \in \mathbb{S}^d$ and a.e. $\mathbf{x} \in \Omega$, it holds

$$\mathbf{S}(\mathbf{x}, \mathbf{A}) : \mathbf{A} \geq c_{11}\Phi(\|\mathbf{A}\|_{\mathbb{S}^d}) + c_{11}\Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{A})\|_{\mathbb{S}^d});$$

(S5) $(\mathbf{S}(\mathbf{x}, \mathbf{A}) - \mathbf{S}(\mathbf{x}, \mathbf{B})) : (\mathbf{A} - \mathbf{B}) \geq 0$ for all $\mathbf{A}, \mathbf{B} \in \mathbb{S}^d$ and a.e. $\mathbf{x} \in \Omega$.

The existence and uniqueness result in study of Problem 1.5.8 reads as follows.

Theorem 1.5.10 *Under Hypotheses 1.5.9, Problem 1.5.8 has a solution $\boldsymbol{u} \in V$ provided $c_{11} > c_6$, and*

(s1) *the solution satisfies the estimate $\|\boldsymbol{u}\|_V \leq C$ with $C := \left(\frac{\|\boldsymbol{f}\|_{V^*} + c_5}{c_{11} - c_6}\right)^{1/(\rho-1)}$;*
(s2) *the solution is unique if, in addition, there exists $c_{12} > 0$ such that:*

  (S6) $(\mathbf{S}(\mathbf{x}, \mathbf{A}) - \mathbf{S}(\mathbf{x}, \mathbf{B})) : (\mathbf{A} - \mathbf{B}) \geq c_{12}\Phi(\|\mathbf{A} - \mathbf{B}\|_{\mathbb{S}^d})$ *for all* $\mathbf{A}, \mathbf{B} \in \mathbb{S}^d$, *a.e.* $\mathbf{x} \in \Gamma$*;*
  (S7) $c_{12} > 2c_8 \|\gamma\|_{V\to Y}^2 + 2c_7 \frac{\|\boldsymbol{f}\|_{V^*} + c_5}{c_{11} - c_6} \|\gamma\|_{V\to Y}^{\rho+1}$*;*

where $\rho \in (1, \infty)$, the constant $c_{11} > 0$ follows from hypothesis (S4)*, whereas $c_5 := c_8 \|\gamma\|_{V\to Y}$ and $c_6 := c_7 \|\gamma\|_{V\to Y}^\rho$ with $c_7 := \frac{2b_\kappa d_\kappa}{\kappa} > 0$ and $c_8 := 2d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})} + c_7 > 0$ for $b_\kappa, d_\kappa, a_\kappa$ from hypothesis* (S3).

Proof The proof is analogous to the one of Theorem 1.5.6 with the operator $\mathcal{B} \equiv 0$. By Step 1.5.1, the operator $\mathcal{A}$ given by (1.52) satisfies conditions of Definition 1.2.4 and Step 1.5.3. □

1.6 Steady Flows of Newtonian Fluids Under Leak Boundary Conditions of Frictional Type

1.6.1 Existence and Uniqueness

In this section we apply existence and uniqueness results of Sects. 1.3 and 1.4.2 in the study of steady-state flows of isotropic, isothermal, inhomogeneous, viscous, and incompressible Newtonian fluids with a multivalued nonmonotone subdifferential frictional boundary condition. The results of this chapter are based on our research paper [65].

1.6.1.1 Setting of the Flow Problem

Let Ω be a bounded simply connected open set in $\mathbb{R}^d$, $d = 2, 3$, with connected boundary Γ of class C^2. We consider the following system of stationary Navier–Stokes equations:

$$-c_{\text{visc}}\Delta \boldsymbol{u} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + \nabla \pi = \boldsymbol{f} \quad \text{in } \Omega, \tag{1.61a}$$

$$\operatorname{div} \boldsymbol{u} = 0 \quad \text{in } \Omega. \tag{1.61b}$$

The system describes the steady-state flow of an inhomogeneous incompressible viscous fluid occupying the volume Ω subjected to given volume forces $\boldsymbol{f}$. Here $\boldsymbol{u} \colon \Omega \to \mathbb{R}^d$ is the velocity field, $\pi : \Omega \to \mathbb{R}$ is the pressure, $c_{\text{visc}} > 0$ the kinematic viscosity of the fluid ($c_{\text{visc}} = 1/\text{Re}$, where Re is the Reynolds number), $\boldsymbol{f} \colon \Omega \to \mathbb{R}^d$ the density of external forces. Observe that, by the divergence free condition we have $\Delta \boldsymbol{u} = \operatorname{div}(\mathbf{D}(\boldsymbol{u}))$, where $\mathbf{D}(\boldsymbol{u}) = \frac{1}{2}(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^\top)$. Therefore, we can rewrite (1.61a) and (1.61b) in an equivalent form

$$-\operatorname{div} \mathbf{S} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + \nabla \pi = \boldsymbol{f} \quad \text{in } \Omega, \tag{1.62a}$$

$$\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) = c_{\text{visc}} \mathbf{D}(\boldsymbol{u}) \quad \text{in } \Omega, \tag{1.62b}$$

$$\operatorname{div} \boldsymbol{u} = 0 \quad \text{in } \Omega. \tag{1.62c}$$

The symbol $\mathbf{S} \colon \Omega \to \mathbb{S}^d$ denotes the extra (viscous) part of the (Cauchy) stress tensor in the fluid $\boldsymbol{\sigma} = -\pi \mathbf{I} + \mathbf{S}$, where $\mathbf{I}$ stands for the identity matrix. The extra stress tensor $\mathbf{S}$ is given by a constitutive law $\mathbf{S} = \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))$ in Ω, where $\mathbf{S} \colon \Omega \times \mathbb{S}^d \to \mathbb{S}^d$ and $\mathbf{D}(\boldsymbol{u}) = \frac{1}{2}(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^\top)$ denotes the symmetric part of the velocity gradient. Here, (1.62b) represents the linear Stokes' law. Incompressible fluids described by Stokes' law are called Newtonian fluids. However, only fluids with a simple molecular structure such as water, oil, and several gases are governed by this law.

The formulation of the stationary model of the flow of an inhomogeneous, isotropic, and viscous incompressible Newtonian fluid with a nonlinear boundary condition reads as follows.

Problem 1.6.1 Find a velocity field $\boldsymbol{u}\colon \Omega \to \mathbb{R}^d$ and a pressure $\pi\colon \Omega \to \mathbb{R}$ satisfying

$$-\operatorname{div}\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} + \nabla\pi = \boldsymbol{f} \quad \text{in } \Omega, \tag{1.63a}$$

$$\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) = c_{\text{visc}}\mathbf{D}(\boldsymbol{u}) \quad \text{in } \Omega, \tag{1.63b}$$

$$\operatorname{div}\boldsymbol{u} = 0 \quad \text{in } \Omega, \tag{1.63c}$$

$$\boldsymbol{u}_\tau = 0 \quad \text{on } \Gamma, \tag{1.63d}$$

$$-\sigma_\nu \in \partial j_\nu(\cdot, u_\nu(\cdot)) \quad \text{on } \Gamma. \tag{1.63e}$$

Conditions (1.63d) and (1.63e) are called the leak boundary condition of frictional type. In these conditions, we suppose that the velocity vector is decomposed into its normal and tangential components defined by $u_\nu = \boldsymbol{u}\cdot\boldsymbol{\nu}$ and $\boldsymbol{u}_\tau = \boldsymbol{u} - u_\nu\boldsymbol{\nu}$, respectively. Similarly, for the stress tensor field $\boldsymbol{\sigma}$, we define its normal and tangential components by $\sigma_\nu = (\boldsymbol{\sigma}\boldsymbol{\nu})\cdot\boldsymbol{\nu}$ and $\boldsymbol{\sigma}_\tau = \boldsymbol{\sigma}\boldsymbol{\nu} - \sigma_\nu\boldsymbol{\nu}$, respectively. Here, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d)$ is the unit outward normal vector on the boundary Γ. Condition (1.63d) is also called the nonslip condition. It states that the tangential component of the velocity vector is known and without loss of generality we put it equal to zero. As concerns the boundary condition (1.63e), it represents a multivalued subdifferential condition between the normal velocity and the normal stress. It is used to model fluid control problems in which we regulate the normal velocity of the fluid to reduce the pressure on the boundary. In concrete models, one considers the motion of a fluid in a tube or a channel. The fluid is pumped into tube and can leave it at the boundary orifices while a device can change the sizes of the latter, see [26, 60]. The function $j_\nu\colon \Gamma\times\mathbb{R} \to \mathbb{R}$ is called a *superpotential* and stands for a locally Lipschitz function and ∂j_ν denotes the subdifferential of $j_\nu(\mathbf{x}, \cdot)$ in the sense of Clarke. Since $j_\nu(\mathbf{x}, \cdot)$ is not assumed to be convex in general, the variational inequality approach to Problem 1.6.1 is not possible.

We note that when $j_\nu(\mathbf{x}, \cdot)$ is convex, subdifferential boundary conditions were studied for stationary Navier–Stokes equations in [10, 34] and for the Boussinesq equations in [12]. The evolution counterparts can be found in [35, 44, 45]. For problems with nonconvex superpotential $j_\nu(\mathbf{x}, \cdot)$, we refer to [59, 60] for the stationary problems, and to [26, 61] for the evolution case, and the references therein.

We comment on two concrete examples of monotone and nonmonotone leak boundary conditions of the form (1.63e). Let $g \in L^\infty(\Gamma, \mathbb{R})$ be a strictly positive function. First, consider the convex potential $j_\nu : \Gamma \times \mathbb{R} \to \mathbb{R}$ defined by $j_\nu(\mathbf{x}, r) = g(\mathbf{x})\,|r|$ for $r \in \mathbb{R}$, a.e. $\mathbf{x} \in \Gamma$. This choice of j_ν leads to a threshold leak condition considered earlier in [34, 44] and condition (1.63e) has the following equivalent form:

$$\begin{cases} |\sigma_\nu| \le g(\mathbf{x}) \Longrightarrow u_\nu = 0, \\ -\sigma_\nu = g(\mathbf{x})\,\dfrac{u_\nu}{|u_\nu|} \Longrightarrow u_\nu \neq 0. \end{cases}$$

Physically, the function g represents the threshold of the normal stress. If $|\sigma_\nu| \le g$, then (1.63e) entails $u_\nu = 0$, that is, no leak occurs. Otherwise, non-trivial leak takes place. Second, we give an example of a nonconvex function $j_\nu(\mathbf{x}, \cdot)$ in condition (1.63e) which leads to a version of a nonmonotone threshold condition. Consider the function

$$j_\nu(\mathbf{x}, r) = g(\mathbf{x}) \times \begin{cases} 0 & \text{if } r < 0, \\ \frac{1}{8}r^2 + 2r & \text{if } 0 \le r < 4, \\ \frac{1}{8}r^2 + 8 & \text{if } r \ge 4. \end{cases}$$

This function generates a law which is nonmonotone and multivalued. Condition $u_\nu > 0$ is interpreted as the outflow of the fluid through the boundary. We refer to [60, 61] for a detailed discussion of a control problem when the pressure is regulated by a hydraulic control device. For other examples of various nonmonotone and multivalued relations which are useful in applications, cf. [67, Section 3.3] and [69, Section 1.2].

In the study of Problem 1.6.1 we adopt the following hypotheses (hypothesis (N4) follows from [9, 14]).

Hypotheses 1.6.2 Let $\Phi, \Phi^* : [0, \infty) \to [0, \infty)$ be a pair of complementary N-functions and $j_\nu : \Gamma \times \mathbb{R} \to \mathbb{R}$ be a Carathéodory function such that:

(N1) $\Phi, \Phi^* \in \Delta_2^\infty$ and $A \prec \Phi \lll (\Phi_*)^{1-1/d}$, with Φ_* being the Sobolev conjugate N-function of Φ and $A(t) = t^p$ for $t \in [0, \infty)$ with $p \ge \frac{3d}{d+2}$ and $d = 2, 3$;
(N2) $j_\nu(\mathbf{x}, \cdot)$ is locally Lipschitz continuous for a.e. $\mathbf{x} \in \Gamma$;
(N3) for some $\kappa > 0$ there exist $a_\kappa \in L^1(\Gamma, [0, \infty))$ and positive constants b_κ and d_κ such that $\Phi^*\left(|a^*|/d_\kappa\right) \le a_\kappa(\mathbf{x}) + b_\kappa \Phi\left(|a|/\kappa\right)$ for all $(a, a^*) \in \mathrm{Gr}(\partial j_\nu)$ and a.e. $\mathbf{x} \in \Gamma$;
(N4) there exists $c_{17} > 0$ such that for all $\mathbf{A} \in \mathbb{S}^d$ and a.e. $\mathbf{x} \in \Omega$, it holds

$$\mathbf{S}(\mathbf{x}, \mathbf{A}) : \mathbf{A} \ge c_{17}\Phi(\|\mathbf{A}\|_{\mathbb{S}^d}) + c_{17}\Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{A})\|_{\mathbb{S}^d})$$

where $\mathbf{S}$ is given by linear Stokes' law (1.63b).

Remark 1.6.3 By hypothesis (N1), the spaces $V = W^1L_\Phi(\Omega, \mathbb{R}^d)$ and $Y = L_\Phi(\Gamma, \mathbb{R}^d)$ are separable and reflexive satisfying $V \hookrightarrow\hookrightarrow Y$, see (1.2). Hypothesis (S3) yields the result on an integral representation of the Clarke subdifferential of locally Lipschitz integral functionals defined on the Orlicz space (see [71, 72]). This representation is required to establish existence and uniqueness results in the study of Problems 1.6.1 and 1.6.13. Hypothesis (N4) follows from [9, 14].

1.6.1.2 Weak Formulation and Main Result

In order to give the variational formulation of Problem 1.6.1, we introduce some notation:

$$\begin{aligned} \mathcal{W} &:= \left\{ \boldsymbol{w} \in C^\infty(\overline{\Omega}, \mathbb{R}^d) \,\middle|\, \operatorname{div} \boldsymbol{w} = 0 \text{ in } \Omega,\, \boldsymbol{w}_\tau = 0 \text{ on } \Gamma \right\}, \\ V &:= \overline{\mathcal{W}}^{\|\cdot\|_{W^1L_\Phi(\Omega,\mathbb{R}^d)}}, \quad Y := L_\Phi(\Gamma, \mathbb{R}^d), \end{aligned} \tag{1.64}$$

where V stands for the closure of $\mathcal{W}$ in the norm of $W^1L_\Phi(\Omega, \mathbb{R}^d)$. By Korn's inequality in Theorem 1.2.6 and [36, Lemmata 1.2 and 1.12], we can introduce on V the norm given by

$$\|\boldsymbol{v}\|_V := \|\boldsymbol{v}\|_{L_\Phi(\Omega,\mathbb{R}^d)} + \|\mathbf{D}(\boldsymbol{v})\|_{L_\Phi(\Omega,\mathbb{S}^d)} \quad \text{for } \boldsymbol{v} \in V \tag{1.65}$$

which is equivalent to the norm $\|\boldsymbol{v}\|_V := \|\boldsymbol{v}\|_{W^1L_\Phi(\Omega,\mathbb{R}^d)}$ for $\boldsymbol{v} \in V$ due to [16, p. 2315]. Now, we derive a variational formulation of Problem 1.6.1. Assume that $\boldsymbol{u}$ and π are sufficiently smooth functions which satisfy (1.63a) to (1.63e). Multiplying the equation (1.63a) by $\boldsymbol{v} \in V$ and applying Green's formula [37, p. 57], we obtain

$$\begin{aligned} &\int_\Omega \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) : \mathbf{D}(\boldsymbol{v})\,\mathrm{d}\mathbf{x} + \int_\Omega (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} \cdot \boldsymbol{v}\,\mathrm{d}\mathbf{x} + \int_\Omega \nabla\pi \cdot \boldsymbol{v}\,\mathrm{d}\mathbf{x} \\ &\quad = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}\,\mathrm{d}\mathbf{x} + \int_\Gamma \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\boldsymbol{\nu} \cdot \boldsymbol{v}\,\mathrm{d}\Gamma. \end{aligned}$$

Since the functions in V are divergence free, we have

$$\int_\Omega \nabla\pi \cdot \boldsymbol{v}\,\mathrm{d}\mathbf{x} = -\int_\Omega \pi \operatorname{div} \boldsymbol{v}\,\mathrm{d}\mathbf{x} + \int_\Gamma v_\nu \pi\,\mathrm{d}\Gamma = \int_\Gamma v_\nu \pi\,\mathrm{d}\Gamma.$$

It follows from (1.64) that

$$\int_\Gamma \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\boldsymbol{\nu} \cdot \boldsymbol{v}\,\mathrm{d}\Gamma = \int_\Gamma \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\boldsymbol{\nu} \cdot (\boldsymbol{v}_\tau + v_\nu \boldsymbol{\nu})\,\mathrm{d}\Gamma = \int_\Gamma \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))\boldsymbol{\nu} \cdot v_\nu \boldsymbol{\nu}\,\mathrm{d}\Gamma.$$

Therefore,

$$\int_\Omega \mathbf{S}(\mathbf{x},\mathbf{D}(\boldsymbol{u})):\mathbf{D}(\boldsymbol{v})\,\mathrm{d}\mathbf{x}+\int_\Omega (\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\cdot\boldsymbol{v}\,\mathrm{d}\mathbf{x}$$
$$=\int_\Omega \boldsymbol{f}\cdot\boldsymbol{v}\,\mathrm{d}\mathbf{x}+\int_\Gamma \mathbf{S}(\mathbf{x},\mathbf{D}(\boldsymbol{u}))\boldsymbol{\nu}\cdot v_\nu\boldsymbol{\nu}-v_\nu\pi\,\mathrm{d}\Gamma.$$

Note that $\mathbf{S}(\mathbf{x},\mathbf{D}(\boldsymbol{u}))\boldsymbol{\nu}\cdot v_\nu\boldsymbol{\nu}-v_\nu\pi=(\mathbf{S}(\mathbf{x},\mathbf{D}(\boldsymbol{u}))\,|\boldsymbol{\nu}|^2-\pi\mathbf{I})v_\nu=\boldsymbol{\sigma}\boldsymbol{\nu}\cdot\boldsymbol{v}$. Because of $\boldsymbol{\sigma}\boldsymbol{\nu}\cdot\boldsymbol{v}=(\sigma_\nu\boldsymbol{\nu}+\boldsymbol{\sigma}_\tau)\cdot(v_\nu\boldsymbol{\nu}+\boldsymbol{v}_\tau)=\sigma_\nu v_\nu+\boldsymbol{\sigma}_\tau\cdot\boldsymbol{v}_\tau$ (see [67, p. 183]) and (1.64), we obtain

$$\int_\Omega \mathbf{S}(\mathbf{x},\mathbf{D}(\boldsymbol{u})):\mathbf{D}(\boldsymbol{v})\,\mathrm{d}\mathbf{x}+\int_\Omega(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\cdot\boldsymbol{v}\,\mathrm{d}\mathbf{x}+\int_\Gamma -\sigma_\nu v_\nu\,\mathrm{d}\Gamma=\int_\Omega \boldsymbol{f}\cdot\boldsymbol{v}\,\mathrm{d}\mathbf{x}. \tag{1.66}$$

The subdifferential boundary condition (1.63e) yields

$$\int_\Gamma -\sigma_\nu v_\nu\,\mathrm{d}\Gamma\le\int_\Gamma j_\nu^0(\mathbf{x},u_\nu(\mathbf{x});v_\nu(\mathbf{x}))\,\mathrm{d}\Gamma, \tag{1.67}$$

where $j_\nu^0(\mathbf{x},a;s)$ denotes the generalized directional derivative of $j_\nu(\mathbf{x},\cdot)$ at the point $a\in\mathbb{R}$ in the direction $s\in\mathbb{R}$. In view of (1.66) and (1.67) we introduce the operators $\mathcal{A}\colon V\to V^*$ and $\mathcal{B}[\cdot]\colon V\to V^*$ defined by

$$\langle\mathcal{A}\boldsymbol{u},\boldsymbol{v}\rangle_{V^*\times V}=c_{\mathrm{visc}}\int_\Omega \mathbf{D}(\boldsymbol{u}):\mathbf{D}(\boldsymbol{v})\,\mathrm{d}\mathbf{x},\quad \boldsymbol{u},\boldsymbol{v}\in V, \tag{1.68}$$

$$\langle\mathcal{B}(\boldsymbol{u},\boldsymbol{v}),\boldsymbol{w}\rangle_{V^*\times V}=\int_\Omega(\boldsymbol{u}\cdot\nabla)\boldsymbol{v}\cdot\boldsymbol{w}\,\mathrm{d}\mathbf{x},\quad \mathcal{B}[\boldsymbol{v}]=\mathcal{B}(\boldsymbol{v},\boldsymbol{v}),\quad \boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in V. \tag{1.69}$$

Hence, we obtain the following variational formulation of Problem 1.6.1 in the form of a *hemivariational inequality*.

Problem 1.6.4 Let Ω be a bounded simply connected domain in $\mathbb{R}^d$, $d=2,3$, with connected boundary $\partial\Omega=\Gamma$ of class C^2, V and Y be Banach spaces as in (1.64), and $\boldsymbol{f}\in V^*$. Find $\boldsymbol{u}\in V$ such that

$$\langle\mathcal{A}\boldsymbol{u}+\mathcal{B}\boldsymbol{u},\boldsymbol{v}\rangle_{V^*\times V}+\int_\Gamma j_\nu^0(\mathbf{x},u_\nu(\mathbf{x});v_\nu(\mathbf{x}))\,\mathrm{d}\Gamma\ge\langle\boldsymbol{f},\boldsymbol{v}\rangle_{V^*\times V},\quad\forall\boldsymbol{v}\in V, \tag{1.70}$$

where the operators $\mathcal{A}$ and $\mathcal{B}$ are given by (1.68) and (1.69), respectively, and j_ν^0 stands for the generalized directional derivative of the superpotential $j_\nu(\mathbf{x},\cdot)$.

The existence and uniqueness result in study of Problem 1.6.4 is the following.

Theorem 1.6.5 *Under Hypotheses 1.6.2, Problem 1.6.4 has a solution* $\boldsymbol{u} \in V$ *provided that* $\alpha > c_{10}$*, and*

(a1) *the solution satisfies the estimate* $\|\boldsymbol{u}\|_V \leq C$ *with* $C := \left(\frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha-c_{10}}\right)^{1/\rho}$*;*
(a2) *the solution is unique if, in addition, there exists* $c_{18} > 0$ *such that:*

(N5) $(\mathbf{S}(\mathbf{x}, \mathbf{A}) - \mathbf{S}(\mathbf{x}, \mathbf{B})) : (\mathbf{A} - \mathbf{B}) \geq c_{18}\Phi(\|\mathbf{A} - \mathbf{B}\|_{\mathbb{S}^d})$ *for all* $\mathbf{A}, \mathbf{B} \in \mathbb{S}^d$*, a.e.* $\mathbf{x} \in \Gamma$*;*
(N6) $m_{\mathcal{A}} > 2c_9 \|\gamma\|^2_{V\to Y} + 2c_9 \|\gamma\|^{\rho+2}_{V\to Y} \frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha-c_{10}} + c_4 \left(\frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha-c_{10}}\right)^{1/\rho}$*;*

where $\rho \in (0, 1]$*,* $\alpha := \frac{1}{4}\min\left\{\frac{c_{17}}{2}, \frac{c_{17}c_{19}}{2}\right\}$ *with* c_{17} *from hypothesis* (N4) *and* $c_{19} > 0$*,* $m_{\mathcal{A}} := \frac{1}{4}\min\left\{\frac{c_{18}}{2}, \frac{c_{18}c_{19}}{2}\right\}$*,* c_4 *is the continuity constant of the trilinear form* b *associated with the operator* $\mathcal{B}$*, whereas* $c_{10} := \max\left\{c_9 \|\gamma\|_{V\to Y}, c_9 \|\gamma\|^{\rho+1}_{V\to Y}\right\}$ *with* $c_9 := \max\left\{2d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, \frac{2b_\kappa d_\kappa}{\kappa}, 2c_{e1}d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, c_{e1}\frac{2b_\kappa d_\kappa}{\kappa}\right\}$ *and* $b_\kappa, d_\kappa, a_\kappa$ *such as in hypothesis* (N3).

Proof We apply Theorem 1.4.11. To this end, we show that $\mathcal{N}\colon V \to V^*$ given by $\mathcal{N}\boldsymbol{v} = \mathcal{A}\boldsymbol{v} + \mathcal{B}\boldsymbol{v}$ for $\boldsymbol{v} \in V$, which appears in the hemivariational inequality (1.70), is the Navier–Stokes type operator according to Definition 1.2.4.

First, we show that $\mathcal{A}$ is coercive. By hypotheses (N1) and (N4), we obtain

$$\begin{aligned}\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{u}\rangle_{V^*\times V} &\geq c_{17}\int_\Omega \Phi(\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d})\,\mathrm{d}\mathbf{x} + c_{17}\int_\Omega \Phi^*(c_{\text{visc}}\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d})\,\mathrm{d}\mathbf{x}\\ &\geq c_{17}\int_\Omega \Phi(\|\mathbf{D}(\boldsymbol{u})\|_{\mathbb{S}^d})\,\mathrm{d}\mathbf{x} \geq c_{17}\|\mathbf{D}(\boldsymbol{u})\|^2_{L_\Phi(\Omega,\mathbb{S}^d)}\end{aligned} \tag{1.71}$$

for all $\boldsymbol{u} \in V$. The last inequality holds due to (1.3) and 1.6 of Lemma 1.2.7. In view of [16, Proposition 4.2], there exists a positive constant $C = C(\Omega)$ such that $\inf_{z\in\mathcal{R}} \|\boldsymbol{u} - \boldsymbol{z}\|_{L_\Phi(\Omega,\mathbb{R}^d)} \leq C\|\mathbf{D}(\boldsymbol{u})\|_{L_\Phi(\Omega,\mathbb{S}^d)}$ for all $\boldsymbol{u} \in W^1L_\Phi(\Omega, \mathbb{R}^d)$, where $\mathcal{R} := \left\{\boldsymbol{z}\colon \mathbb{R}^d \to \mathbb{R}^d \,\middle|\, \boldsymbol{z}(\mathbf{x}) = \mathbf{a} + Q\mathbf{x}\right\}$ for some $\mathbf{a} \in \mathbb{R}^d$ and $Q \in \mathbb{S}^d$ with $Q = -Q^\top$. So, there exist $c_{19} > 0$ and $\boldsymbol{z} \in L_\Phi(\Omega, \mathbb{R}^d)$ such that for any $\boldsymbol{u} \in V$, we obtain

$$\begin{aligned}\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{u}\rangle_{V^*\times V} &\geq \frac{c_{17}}{2}\left(\|\mathbf{D}(\boldsymbol{u})\|^2_{L_\Phi(\Omega,\mathbb{S}^d)} + \|\mathbf{D}(\boldsymbol{u})\|^2_{L_\Phi(\Omega,\mathbb{S}^d)}\right)\\ &\geq \frac{c_{17}}{2}\left(\|\mathbf{D}(\boldsymbol{u})\|^2_{L_\Phi(\Omega,\mathbb{S}^d)} + \frac{c_{17}c_{19}}{2}\|\boldsymbol{u} - \boldsymbol{z}\|^2_{L_\Phi(\Omega,\mathbb{R}^d)}\right)\\ &\geq \tilde{c}\left(\frac{1}{2}\|\boldsymbol{u}\|^2_V + \left|\|\boldsymbol{u}\|_{L_\Phi(\Omega,\mathbb{R}^d)} - \|\boldsymbol{z}\|_{L_\Phi(\Omega,\mathbb{R}^d)}\right|^2 - \|\boldsymbol{u}\|^2_{L_\Phi(\Omega,\mathbb{R}^d)}\right),\end{aligned} \tag{1.72}$$

where $\tilde{c} := \min\left\{\frac{c_{17}}{2}, \frac{c_{17}c_{19}}{2}\right\}$ due to (1.65) and $|a+b|^2 \le 2(|a|^2+|b|^2)$ for $a, b \in \mathbb{R}$. Hence,

$$
\begin{aligned}
\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{u}\rangle_{V^*\times V} &\ge \tilde{c}\left(\frac{1}{2}\|\boldsymbol{u}\|_V^2 - \|\boldsymbol{u}\|^2_{L_\Phi(\Omega,\mathbb{R}^d)}\left(1-\left|1-\frac{\|\boldsymbol{z}\|_{L_\Phi(\Omega,\mathbb{R}^d)}}{\|\boldsymbol{u}\|_{L_\Phi(\Omega,\mathbb{R}^d)}}\right|^2\right)\right)\\
&\ge \tilde{c}\left(\frac{1}{2}\|\boldsymbol{u}\|_V^2 - \frac{1}{4}\|\boldsymbol{u}\|^2_V\left(1-\left|1-\frac{\|\boldsymbol{z}\|_{L_\Phi(\Omega,\mathbb{R}^d)}}{\|\boldsymbol{u}\|_{L_\Phi(\Omega,\mathbb{R}^d)}}\right|^2\right)\right)\\
&= \frac{\tilde{c}}{4}\|\boldsymbol{u}\|_V^2\left(1+\left|1-\frac{\|\boldsymbol{z}\|_{L_\Phi(\Omega,\mathbb{R}^d)}}{\|\boldsymbol{u}\|_{L_\Phi(\Omega,\mathbb{R}^d)}}\right|^2\right) \ge \frac{\tilde{c}}{4}\|\boldsymbol{u}\|_V^2, \quad \forall \boldsymbol{u}\in V.
\end{aligned}
\tag{1.73}
$$

Thus, $\mathcal{A}$ is $(\alpha, 2)$-coercive with $\alpha = \frac{\tilde{c}}{4}$.

Second, by hypotheses (N1) and (N4) together with Step 1.5.1, the operator $\mathcal{A}$ given by (1.68) is bounded, hemicontinuous, and monotone. Thus, $\mathcal{A}$ is pseudomonotone due to [67, Theorem 3.69]. By Step 1.5.2 and [60, Remark 12], we infer that the operator $\mathcal{B}$ given by (1.69) satisfies conditions of Definition 1.2.4 (cf. [60, p. 206]).

To provide the uniqueness of a solution to Problem 1.6.4, it suffices to show that $\mathcal{A}$ is strongly monotone due to Theorem 1.4.11. By hypothesis (N5), (1.3), and (1.6) of Lemma 1.2.7 and the same arguments such as in (1.71) and (1.73), we obtain

$$
\begin{aligned}
\langle \mathcal{A}\boldsymbol{u}_1 - \mathcal{A}\boldsymbol{u}_2, \boldsymbol{u}_1-\boldsymbol{u}_2\rangle &= \int_\Omega (\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}_1)) - \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}_2))) : (\mathbf{D}(\boldsymbol{u}_1) - \mathbf{D}(\boldsymbol{u}_2))\,\mathrm{d}\mathbf{x}\\
&\ge c_{18}\int_\Omega \Phi(\|\mathbf{D}(\boldsymbol{u}_1) - \mathbf{D}(\boldsymbol{u}_2)\|_{\mathbb{S}^d})\,\mathrm{d}\mathbf{x}\\
&\ge c_{18}\|\mathbf{D}(\boldsymbol{u}_1 - \boldsymbol{u}_2)\|^2_{L_\Phi(\Omega,\mathbb{S}^d)} \ge \frac{1}{4} m_{\mathcal{A}}\|\boldsymbol{u}_1-\boldsymbol{u}_2\|_V^2
\end{aligned}
$$

with $m_{\mathcal{A}} := \frac{1}{4}\min\left\{\frac{c_{18}}{2}, \frac{c_{18}c_{19}}{2}\right\} > 0$ for all $\boldsymbol{u}_1, \boldsymbol{u}_2 \in V$, which implies that the operator $\mathcal{A}$ is strongly monotone. □

Remark 1.6.6 Note that hypothesis (N5) allows to solve the steady Newtonian fluid flow Problem 1.6.1 in $W^1L_\Phi(\Omega, \mathbb{R}^d)$ with $\Phi \succ t^p$ and $p \ge \frac{3d}{d+2}$ for $d = 2, 3$, including the Sobolev space $W^{1,p}(\Omega, \mathbb{R}^d)$ with the same conditions on p (cf. [60] for $W^{1,p}(\Omega, \mathbb{R}^d)$ with $p \ge 2$). For the steady generalized Newtonian fluid flow problem with the Dirichlet boundary condition in the context of Orlicz spaces, we refer to [7, 30], where the authors assumed that $\Phi, \Phi^* \in \Delta_2^{\text{glob}}$ (i.e., Φ, Φ^* satisfy the Δ_2-*condition* for all $t \ge 0$) and $\Phi \in C^2$ has the lower bound $\Phi(t) \ge \frac{1}{2}\Phi''(0)t^2$ for $t \ge 0$ (i.e., Φ has the so-called superquadratic growth). Note that the conditions $\Phi, \Phi^* \in \Delta_2^{\text{glob}}$ are stronger than $\Phi, \Phi^* \in \Delta_2^{\infty}$. In [7, 30], the superquadratic growth of Φ implies the unique solution in $W^1 L_\Phi(\Omega, \mathbb{R}^d)$ with $\Phi \succ t^2$ for all $t \geq 0$, including the solution in the Sobolev space $W^{1,p}(\Omega, \mathbb{R}^d)$ with $p \geq 2$.

1.6.2 Slow Flows

In this part we comment on a particular form of Problem 1.6.1. Consider the flow problem under leak boundary conditions of frictional type in which the convective term $(\boldsymbol{u} \cdot \nabla)\boldsymbol{u}$ in (1.61a) is negligible. In this case, Problem 1.6.1 reduces to the following system.

Problem 1.6.7 Find a velocity field $\boldsymbol{u} \colon \Omega \to \mathbb{R}^d$ and a pressure $\pi \colon \Omega \to \mathbb{R}$ satisfying

$$\begin{aligned}
-\operatorname{div} \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) + \nabla\pi &= \boldsymbol{f} && \text{in } \Omega,\\
\mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u})) &= c_{\text{visc}}\mathbf{D}(\boldsymbol{u}) && \text{in } \Omega,\\
\operatorname{div} \boldsymbol{u} &= 0 && \text{in } \Omega,\\
\boldsymbol{u}_\tau &= 0 && \text{on } \Gamma,\\
-\sigma_\nu &\in \partial j_\nu(\cdot, u_\nu(\cdot)) && \text{on } \Gamma.
\end{aligned}$$

The variational formulation of Problem 1.6.7 reads as follows.

Problem 1.6.8 Let Ω be a bounded simply connected domain in $\mathbb{R}^d$, $d = 2, 3$, with connected boundary $\partial\Omega = \Gamma$ of class C^2, V and Y be Banach spaces as in (1.64), and $\boldsymbol{f} \in V^*$. Find $\boldsymbol{u} \in V$ such that

$$\langle \mathcal{A}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} + \int_\Gamma j_\nu^0(\mathbf{x}, u_\nu(\mathbf{x}); v_\nu(\mathbf{x}))\, \mathrm{d}\Gamma \geq \langle \boldsymbol{f}, \boldsymbol{v}\rangle_{V^*\times V}, \quad \forall \boldsymbol{v} \in V, \tag{1.74}$$

where the operator $\mathcal{A}$ is given by (1.68), and j_ν^0 stands for the generalized directional derivative of the superpotential $j_\nu(\mathbf{x}, \cdot)$.

In the study of Problem 1.6.8 we adopt the following hypotheses.

Hypotheses 1.6.9 Let $\Phi, \Phi^* \colon [0, \infty) \to [0, \infty)$ be a pair of complementary N-functions and $j_\nu \colon \Gamma \times \mathbb{R} \to \mathbb{R}$ be a Carathéodory function such that:

(L1) $\Phi, \Phi^* \in \Delta_2^\infty$ and $\Phi \lll (\Phi_*)^{1-1/d}$, with Φ_* being the Sobolev conjugate N-function of Φ and $d = 2, 3$;
(L2) $j_\nu(\mathbf{x}, \cdot)$ is locally Lipschitz continuous for a.e. $\mathbf{x} \in \Gamma$;
(L3) for some $\kappa > 0$ there exist $a_\kappa \in L^1(\Gamma, [0, \infty))$ and positive constants b_κ and d_κ such that $\Phi^*(|a^*|/d_\kappa) \leq a_\kappa(\mathbf{x}) + b_\kappa \Phi(|a|/\kappa)$ for all $(a, a^*) \in \operatorname{Gr}(\partial j_\nu)$ and a.e. $\mathbf{x} \in \Gamma$;

(L4) there exists $c_{17} > 0$ such that for all $\mathbf{A} \in \mathbb{S}^d$ and a.e. $\mathbf{x} \in \Omega$, it holds

$$\mathbf{S}(\mathbf{x}, \mathbf{A}) : \mathbf{A} \geq c_{17}\Phi(\|\mathbf{A}\|_{\mathbb{S}^d}) + c_{17}\Phi^*(\|\mathbf{S}(\mathbf{x}, \mathbf{A})\|_{\mathbb{S}^d})$$

where $\mathbf{S}$ is given by linear Stokes' law (1.63b).

The existence and uniqueness result in study of Problem 1.6.8 is the following.

Theorem 1.6.10 *Under Hypotheses 1.6.9, Problem 1.6.8 has a solution $\boldsymbol{u} \in V$ provided that $\alpha > c_{10}$, and*

(l1) *the solution satisfies the estimate $\|\boldsymbol{u}\|_V \leq C$ with $C := \left(\frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha - c_{10}}\right)^{1/\rho}$;*
(l2) *the solution is unique if, in addition, there exists $c_{18} > 0$ such that:*

(L5) $(\mathbf{S}(\mathbf{x}, \mathbf{A}) - \mathbf{S}(\mathbf{x}, \mathbf{B})) : (\mathbf{A} - \mathbf{B}) \geq c_{18}\Phi(\|\mathbf{A} - \mathbf{B}\|_{\mathbb{S}^d})$ *for all* $\mathbf{A}, \mathbf{B} \in \mathbb{S}^d$, *a.e.* $\mathbf{x} \in \Gamma$*;*
(L6) $m_{\mathcal{A}} > 2c_9 \|\gamma\|^2_{V\to Y} + 2c_9 \|\gamma\|^{\rho+2}_{V\to Y} \frac{\|\boldsymbol{f}\|_{V^*}+c_{10}}{\alpha - c_{10}}$*;*

where $\alpha := \frac{1}{4}\min\left\{\frac{c_{17}}{2}, \frac{c_{17}c_{19}}{2}\right\}$ *with the constants* $c_{17} > 0$ *from hypothesis* (N4) *and* $c_{19} > 0$, $m_{\mathcal{A}} := \frac{1}{4}\min\left\{\frac{c_{18}}{2}, \frac{c_{18}c_{19}}{2}\right\}$, *whereas* $c_{10} := \max\left\{c_9 \|\gamma\|_{V\to Y}, c_9 \|\gamma\|^{\rho+1}_{V\to Y}\right\}$ *with* $c_9 := \max\left\{2d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, \frac{2b_\kappa d_\kappa}{\kappa}, 2c_{e1}d_\kappa \|a_\kappa\|_{L^1(\Gamma,\mathbb{R})}, c_{e1}\frac{2b_\kappa d_\kappa}{\kappa}\right\}$ *and* $b_\kappa, d_\kappa, a_\kappa$ *such as in hypothesis* (N3).

Proof The proof is analogous to the one of Theorem 1.6.5 with the operator $\mathcal{B} \equiv 0$. □

Remark 1.6.11 Owing to hypothesis (L5), we are able to solve Problem 1.6.7 in the Sobolev space $W^{1,p}(\Omega, \mathbb{R}^d)$ with $1 < p < 2$. For the slow steady fluid flow problem with the Dirichlet boundary condition, we refer to [33, Theorem 3.1.1] in the Sobolev space setting $W^{1,p}(\Omega, \mathbb{R}^d)$ with $p \geq 2$.

1.6.3 Optimal Control Problem

In this section we study the class of optimal control problems for a system described by the hemivariational inequality (1.70) in Problem 1.6.4.

1.6.3.1 Continuous Dependence on External Forces

In this section we study the upper semicontinuous dependence of the solution set of Problem 1.6.4 with respect to density of external forces $\boldsymbol{f}$. The following result is crucial for the optimal control problem.

Theorem 1.6.12 *Assume that Hypotheses 1.6.2 hold and*

(d1) $\boldsymbol{f}_k, \boldsymbol{f} \in V^*$ *and* $\boldsymbol{f}_k \rightharpoonup \boldsymbol{f}$ *in* $\sigma(V^*, V)$*;*
(d2) $0 < \rho \leq 1$ *and* $\alpha > c_{10}$.

If $(\boldsymbol{u}_k) \subset V$ *is a sequence of solutions of Problem 1.6.4 corresponding to* $\boldsymbol{f}_k$*, then there exists a subsequence* $(\boldsymbol{u}_{k_j})$ *of* $(\boldsymbol{u}_k)$ *such that* $\boldsymbol{u}_{k_j} \rightharpoonup \boldsymbol{u}$ *in* $\sigma(V, V^*)$ *and* $\boldsymbol{u} \in V$ *is a solution to Problem 1.6.4 corresponding to* $\boldsymbol{f}$*.*

Proof By Theorem 1.6.5, there exists $(\boldsymbol{u}_k) \subset V$ such that

$$\langle \mathcal{A}\boldsymbol{u}_k + \mathcal{B}\boldsymbol{u}_k, \boldsymbol{v}\rangle_{V^*\times V} + \int_\Gamma j_\nu^0(\mathbf{x}, (u_k)_\nu(\mathbf{x}); v_\nu(\mathbf{x}))\, \mathrm{d}\Gamma \geq \langle \boldsymbol{f}_k, \boldsymbol{v}\rangle_{V^*\times V}, \quad \forall \boldsymbol{v} \in V$$

and $\|\boldsymbol{u}_k\|_V \leq C$ with $C > 0$ independent of $k \in \mathbb{N}$ due to (d2). By Hypotheses 1.6.2 (see (N1)), the sequence $(\boldsymbol{u}_k)$ belongs to a bounded subset of the reflexive Banach space V. Hence, there exists a subsequence $(\boldsymbol{u}_{k_j})$ of $(\boldsymbol{u}_k)$ such that $\boldsymbol{u}_{k_j} \rightharpoonup \boldsymbol{u}$ in $\sigma(V, V^*)$ and $\boldsymbol{u} \in V$. Owing to Theorem 1.6.5, the operators $\mathcal{A}$ and $\mathcal{B}$ given by (1.68) and (1.69) satisfy conditions of Definition 1.2.4. Therefore,

$$\langle \mathcal{A}\boldsymbol{u}_{k_j} + \mathcal{B}\boldsymbol{u}_{k_j}, \boldsymbol{v}\rangle_{V^*\times V} \to \langle \mathcal{A}\boldsymbol{u} + \mathcal{B}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V}, \quad \forall \boldsymbol{v} \in V.$$

From Hypotheses 1.6.2 (see (N1)), we deduce that the compactness of the trace operator γ (see (1.2)) yields $\gamma \boldsymbol{u}_{k_j} \to \gamma \boldsymbol{u}$ in Y and $\gamma \boldsymbol{u} \in Y$. It follows that $(\gamma u_{k_j})_\nu = \gamma \boldsymbol{u}_{k_j} \cdot \boldsymbol{\nu} \to \gamma \boldsymbol{u} \cdot \boldsymbol{\nu} = (\gamma u)_\nu$ in $L_\Phi(\Gamma, \mathbb{R})$. Owing to [43, Theorems IV3.1 and I.6.4], there exists a subsequence $(\boldsymbol{u}_l)$ of $(\boldsymbol{u}_{k_j})$ such that

$$(\gamma u_l)_\nu(\mathbf{x}) \to (\gamma u)_\nu(\mathbf{x}) \ \text{ a.e. } \mathbf{x} \in \Gamma.$$

Since the function $\mathbb{R} \times \mathbb{R} \ni (r, s) \mapsto j^0(\mathbf{x}, r; s) \in \mathbb{R}$ is upper semicontinuous, we have

$$\limsup_{l\to\infty} j_\nu^0(\mathbf{x}, (u_l)_\nu(\mathbf{x}); v_\nu(\mathbf{x})) \leq j_\nu^0(\mathbf{x}, u_\nu(\mathbf{x}); v_\nu(\mathbf{x})) < \infty$$

for all $\boldsymbol{v} \in V$ and a.e. $\mathbf{x} \in \Gamma$. The last inequality holds due to (n4) of Lemma 1.4.10 (see (1.42)). Using the Fatou lemma, we obtain

$$\limsup_{l\to\infty} \int_\Gamma j_\nu^0(\mathbf{x}, (u_l)_\nu(\mathbf{x}); v_\nu(\mathbf{x}))\, \mathrm{d}\Gamma \leq \int_\Gamma j_\nu^0(\mathbf{x}, u_\nu(\mathbf{x}); v_\nu(\mathbf{x}))\, \mathrm{d}\Gamma.$$

From the above and (d1), we infer that

$$\langle \boldsymbol{f}, \boldsymbol{v}\rangle_{V^*\times V} \leq \langle \mathcal{A}\boldsymbol{u} + \mathcal{B}\boldsymbol{u}, \boldsymbol{v}\rangle_{V^*\times V} + \int_\Gamma j_\nu^0(\mathbf{x}, u_\nu(\mathbf{x}); v_\nu(\mathbf{x}))\, \mathrm{d}\Gamma, \quad \forall \boldsymbol{v} \in V.$$

Thus, $\boldsymbol{u} \in V$ is a solution to hemivariational inequality (1.70) in Problem 1.6.4 corresponding to $\boldsymbol{f}$, and the proof is complete. □

1.6.3.2 Optimal Control Problem

We suppose that $U := V^*$ represents the space of controls. We will denote by $S(\boldsymbol{f}) \subset V$ the solution set of Problem 1.6.4 corresponding to $\boldsymbol{f}$. It is a nonempty set for all $\boldsymbol{f} \in U$ due to Hypotheses 1.6.2 of Theorem 1.6.5. The control problem is formulated as follows.

Problem 1.6.13 Let $U_{ad} \subset U$ be a nonempty set representing the set admissible controls and $\mathcal{F}\colon U \times V \to \mathbb{R}$, $\mathcal{F} = \mathcal{F}(\boldsymbol{f}, \boldsymbol{u})$ be an objective functional. Find a control $\boldsymbol{f}^* \in U_{ad}$ and a state $\boldsymbol{u}^* \in S(\boldsymbol{f}^*)$ such that

$$\mathcal{F}(\boldsymbol{f}^*, \boldsymbol{u}^*) = \inf\{\mathcal{F}(\boldsymbol{f}, \boldsymbol{u}) \mid \boldsymbol{f} \in U_{ad}, \boldsymbol{u} \in S(\boldsymbol{f})\}. \tag{1.75}$$

A pair which solves (1.75) is called an *optimal solution*. The existence result for the optimal control problem Problem 1.6.13 is the following.

Theorem 1.6.14 *Assume that Hypotheses 1.6.2 hold and*

(o1) *U_{ad} is a bounded and weakly closed subset of U;*
(o2) *$\mathcal{F}$ is lower semicontinuous with respect to $U \times V$ endowed with weak topology;*
(o3) $0 < \rho \le 1$ *and* $\alpha > c_{10}$.

Then, Problem 1.6.13 has an optimal solution.

Proof Let $(\boldsymbol{f}_k, \boldsymbol{u}_k)$ be a minimizing sequence for Problem 1.6.13. Hence, $\boldsymbol{f}_k \in U_{ad}$, $\boldsymbol{u}_k \in S(\boldsymbol{f}_k)$ and

$$\lim_{k\to\infty} \mathcal{F}(\boldsymbol{f}_k, \boldsymbol{u}_k) = \inf\{\mathcal{F}(\boldsymbol{f}, \boldsymbol{u}) \mid \boldsymbol{f} \in U_{ad}, \boldsymbol{u} \in S(\boldsymbol{f})\} =: m.$$

Since V is separable and reflexive due to Hypotheses 1.6.2 (see (N1)), it follows from (o1) that there exists a subsequence $(\boldsymbol{f}_{k_j})$ of $(\boldsymbol{f}_k)$ and $\boldsymbol{f}^* \in U_{ad}$ such that $\boldsymbol{f}_{k_j} \rightharpoonup \boldsymbol{f}^*$ in $\sigma(V, V^*)$. By Theorem 1.6.12, there exists a subsequence $(\boldsymbol{u}_{k_j})$ of $(\boldsymbol{u}_k)$ such that $\boldsymbol{u}_{k_j} \rightharpoonup \boldsymbol{u}^*$ in $\sigma(V, V^*)$ and $\boldsymbol{u}^* \in S(\boldsymbol{f})$. Owing to (o2), we deduce that $m \le \mathcal{F}(\boldsymbol{f}^*, \boldsymbol{u}^*) \le \liminf_{j\to\infty} \mathcal{F}(\boldsymbol{f}_{k_j}, \boldsymbol{u}_{k_j}) = m$, and the proof is complete. □

1.7 Concluding Remarks

Most of the available results concerning incompressible fluids deal with the constitutive law $\mathbf{S} = \mathbf{S}(\mathbf{x}, \mathbf{D}(\boldsymbol{u}))$ which describes polynomial dependence between the extra stress tensor $\mathbf{S}$ and the symmetric part of the velocity gradient $\mathbf{D}(\boldsymbol{u})$.

The assumptions presented in this chapter capture power-law and Carreau-type models, and more general models of fluids with non-polynomial growth condition in reflexive Orlicz spaces. Power-law and Carreau-type models are very popular among rheologists, in chemical engineering, and colloidal mechanics (see [52] and the references therein). In various instances of interest in applications, results which may yield the loss of integrability can be achieved in the framework of non-reflexive Orlicz spaces. They are flexible enough to study fine properties of measurable functions which are required in models of fluids of Prandtl–Eyring [24], Powell–Eyring [79], and Sutterby [3]. These models are broadly used in geophysics, engineering, and medical applications, for instance in models of glacier ice and blood flow (see, e.g., [1], [53] and the references therein) and many others. For these reasons, we will continue a study on steady/unsteady flows of fluids which are described by non-polynomial growth conditions and nonmonotone frictional boundary conditions, see, e.g., [64, 74].

Acknowledgements This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie Grant Agreement No. 823731 - CONMECH. The first author is also supported by the Natural Science Foundation of Guangxi Grant No. 2018JJA110006, and the Beibu Gulf University Project No. 2018KYQD03.

References

1. N.S. Akbar, Biomathematical study of Sutterby fluid model for blood flow in stenosed arteries. Int. J. Biomath. **8**(6), 1550075 (2015)
2. L.C. Berselli, L. Diening, M. Růžička, Existence of strong solutions for incompressible fluids with shear dependent viscosities. J. Math. Fluid Mech. **12**, 101–132 (2010)
3. R.B. Bird, O. Hassager, *Dynamics of Polymeric Liquids: Fluid Mechanics*. Dynamics of Polymer Liquids, vol. 1, 2nd edn. (Wiley, Hoboken, 1987)
4. D. Breit, Analysis of generalized Navier–Stokes equations for stationary shear thickening flows. Nonlinear Anal. **75**, 5549–5560 (2012)
5. D. Breit, A. Cianchi, Negative Orlicz–Sobolev norms and strongly nonlinear systems in fluid mechanics. J. Differ. Equ. **259**, 48–83 (2015)
6. D. Breit, L. Diening, Sharp conditions for Korn inequalities in Orlicz spaces. J. Math. Fluid Mech. **14**, 565–573 (2012)
7. D. Breit, M. Fuchs, The nonlinear Stokes problem with general potentials having superquadratic growth. J. Math. Fluid Mech. **13**, 371–385 (2011)
8. F.E. Browder, P. Hess, Nonlinear mappings of monotone type in Banach spaces. J. Funct. Anal. **11**, 251–294 (1972)
9. M. Bulíček, P. Gwiazda, J. Málek, A. Świerczewska-Gwiazda, On unsteady flows of implicitly constituted incompressible fluids. SIAM J. Math. Anal. **44**(4), 2756–2801 (2012)
10. A.Yu. Chebotarev, Subdifferential boundary value problems for stationary Navier–Stokes equations. Differ. Uravn. (Differ. Equ.) **28**, 1443–1450 (1992)
11. A.Yu. Chebotarev, Stationary variational inequalities in a model of an inhomogeneous incompressible fluid. Sib. Math J. (Sib. Math. Zh.) **38**(5), 1028–1037 (1997)
12. A.Yu. Chebotarev, Variational inequalities for Navier–Stokes type operators and one-side problems for equations of viscous heat-conducting fluids. Math. Notes (Mat. Zametki) **70**(2), 264–274 (2001)

13. A.Yu. Chebotarev, Modeling of steady flows in a channel by Navier–Stokes variational inequalities. J. Appl. Mech. Tech. Phys. **44**(6), 852–857 (2003)
14. K. Chełmiński, P. Gwiazda, Convergence of coercive approximations for strictly monotone quasistatic models in inelastic deformation theory. Math. Models Methods Appl. Sci. **30**, 1357–1374 (2007)
15. A. Cianchi, A sharp embedding theorem for Orlicz-Sobolev spaces. Indiana Univ. Math. J. **45**, 39–65 (1996)
16. A. Cianchi, Korn type inequalities in Orlicz spaces. J. Funct. Anal. **267**, 2313–2352 (2014)
17. F.H. Clarke, *Optimization and Nonsmooth Analysis*. Classics in Applied Mathematics (SIAM, Philadelphia, 1990)
18. Z. Denkowski, S. Migórski, N.S. Papageorgiou, *An Introduction to Nonlinear Analysis: Theory*, vol. I (Kluwer, Boston, 2003)
19. W. Desch, R. Grimmer, On the wellposedness of constitutive laws involving dissipation potentials. Trans. Am. Math. Soc. **353**(12), 5095–5120 (2001)
20. L. Diening, P. Kaplicky, L^q theory for a generalized Stokes system. Manuscripta Math. **141**, 333–361 (2013)
21. J.K. Djoko, J.M. Lubuma, Analysis of a time implicit scheme for the Oseen model driven by nonlinear slip boundary conditions. J. Math. Fluid Mech. **18**, 717–730 (2016)
22. T.K. Donaldson, N.S. Trudinger, Orlicz–Sobolev spaces and imbedding theorems. J. Funct. Anal. **8**, 52–75 (1971)
23. S. Dudek, P. Kalita, S. Migórski, Stationary flow of non-Newtonian fluid with nonmonotone frictional boundary conditions. Z. Angew. Math. Phys. **66**, 2625–2646 (2015)
24. H.J. Eyring, Viscosity, plasticity, and diffusion as example of absolute reaction rates. J. Chem. Phys. **4**, 283–291 (1936)
25. C. Fang, W. Han, Well-posedness and optimal control of a hemivariational inequality for nonstationary Stokes fluid flow. Discrete Contin. Dyn. Syst. **39**(10), 5369–5386 (2016)
26. C. Fang, W. Han, S. Migórski, M. Sofonea, A class of hemivariational inequalities for nonstationary Navier–Stokes equations. Nonlinear Anal. Real World Appl. **31**, 257–276 (2016)
27. A. Fougères, Théoremès de trace et de prolongement dans les espaces de Sobolev et Sobolev–Orlicz. C.R. Acad. Sci. Paris, Ser. A **274**, 181–184 (1972)
28. J. Freshe, On analysis of steady flows of fluids with shear-dependent viscosity based on the Lipschitz truncation method. SIAM J. Math. Anal. **34**, 1064–1083 (2003)
29. J. Freshe, J. Málek, M. Steinhauer, An existence result for fluids with shear dependent viscosity-steady flows. Nonlinear Anal. **30**, 3041–3049 (1997)
30. M. Fuchs, A note on non-uniformly elliptic Stokes-type systems in two variables. J. Math. Fluid Mech. **12**, 266–279 (2010)
31. M. Fuchs, V. Osmolovskii, Variational integrals on Orlicz-Sobolev spaces. Z. Anal. Anwend. **17**(2), 393–415 (1998)
32. M. Fuchs, G. Seregin, Variational methods for fluids of Prandtl–Eyring type and plastic materials with logarithmic hardening. Math. Methods Appl. Sci. **22**, 317–351 (1999)
33. M. Fuchs, G. Seregin, *Variational Methods for Problems from Plasticity Theory and for Generalized Newtonian Fluids*. Lecture Notes in Mathematics, vol. 1749 (Springer, Berlin, 2000)
34. H. Fujita, A mathematical analysis of motions of viscous incompressible fluid under leak or slip boundary conditions. RIMS Kokyuroku **888**, 199–216 (1994)
35. H. Fujita, Non-stationary Stokes flows under leak boundary conditions of friction type. J. Comput. Appl. Math. **19**, 1–8 (2001)
36. J.-P. Gossez, Nonlinear elliptic boundary value problems for equations with rapidly (or slowly) increasing coefficients. Trans. Am. Math. Soc. **190**, 163–205 (1974)
37. J.-P. Gossez, A remark on strongly nonlinear elliptic boundary value problems. Bol. Soc. Brasil. Mat. **8**, 53–63 (1977)
38. J. Gustavsson, J. Peetre, Interpolation of Orlicz spaces. Studia Math. **60**, 33–59 (1977)
39. P. Gwiazda, A. Świerczewska-Gwiazda, On non-Newtonian fluids with the property of rapid thickening under different stimulus. Math. Models Methods Appl. Sci. **18**, 1073–1092 (2008)

40. P. Gwiazda, A. Świerczewska-Gwiazda, A. Wróblewska, Monotonicity methods in generalized Orlicz spaces for a class of non-Newtonian fluids. Math. Models Methods Appl. Sci. **33**, 125–137 (2010)
41. J. Haslinger, P.D. Panagiotopoulos, Optimal control of hemivariational inequalities, in *Control of Boundaries and Stabilization. Proceedings of the IFIP WG 7.2 Conference, Clermont Ferrand, June 20–23, 1988*, ed. by J. Simon. Lecture Notes in Control and Information Sciences, vol. 125 (Springer, Berlin, 1989), pp. 128–139
42. H. Hudzik, On continuity of the imbedding operation from $W^k_{M_1}(\Omega)$ *into* $W^k_{M_2}(\Omega)$. Funct. Approx. Comment. Math. **6**, 111–118 (1978)
43. L.V. Kantorovich, G.P. Akilov, *Functional Analysis* (Pergamon, Oxford, 1982)
44. T. Kashiwabara, On a strong solution of the non-stationary Navier-Stokes equations under slip or leak boundary conditions of friction type. J. Differ. Equ. **254**, 756–778 (2013)
45. D.S. Konovalova, Subdifferential boundary value problems for the nonstationary Navier–Stokes equations. Differ. Equ. (Differ. Uravn.) **36**(6), 878–885 (2000)
46. M.A. Krasnosel'skiĭ, Ya.B. Rutickiĭ, *Convex Functions and Orlicz Spaces* (P. Noordhoof Ltd., Groningen, 1961)
47. A. Kufner, O. John, S. Fučík, *Function Spaces* (Academia Publishing House of the Czechoslovak Academy of Sciences, Prague, 1977)
48. O.A. Ladyzhenskaya, *The Mathematical Theory of Viscous Incompressible Flow* (Gordon and Breach, New York, 1969)
49. J.-L. Lions, Quelques méthodes de résolution des problèmes aux limites non linéaires (Dunod, Paris, 1969)
50. J. Málek, On weak solutions to a class of non-Newtonian incompressible fluids in bounded three-dimensional domains. The case $p \geq 2$. Adv. Differ. Equ. **6**, 257–302 (2001)
51. J. Málek, J. Nečas, M. Růžička, On the non-Newtonian incompressible fluids. Math. Models Methods Appl. Sci. **3**(1), 35–63 (1993)
52. J. Málek, K.R. Rajagopal, M. Růžička, Existence and regularity of solutions and the stability of the rest state for fluids with shear dependent viscosity. Math. Models Methods Appl. Sci. **6**, 789–812 (1995)
53. J. Málek, J. Nečas, M. Rokyta, M. Růžička, *Weak and Measure-Valued Solutions to Evolutionary PDEs* (Chapman & Hall, London, 1996)
54. L. Maligranda, Indices and interpolation. Diss. Math. **234**, 1–49 (1985)
55. L. Maligranda, *Orlicz Spaces and Interpolation*. Seminars in Mathematics, vol. 5 (Departamento de Matemática, Universidade Estadual de Campinas, Campinas, 1989)
56. M. Miettinen, J. Haslinger, Approximation of optimal control problems of hemivariational inequalities. Numer. Funct. Anal. Optim. **13**, 43–68 (1992)
57. S. Migórski, A note on optimal control problem for a hemivariational inequality modeling fluid flow. Discrete Contin. Dyn. Syst. Suppl. **2013**, 545–554 (1984)
58. S. Migórski, Optimal control of a class of boundary hemivariational inequalities of hyperbolic type. Dyn. Contin. Discret. Impuls. Syst. Ser. B Appl. Alg. **2003**, 159–164 (2003)
59. S. Migórski, Hemivariational inequalities modeling viscous incompressible fluids. J. Nonlinear Convex Anal. **5**, 217–227 (2004)
60. S. Migórski, Hemivariational inequalities for stationary Navier–Stokes equations. J. Math. Anal. Appl. **306**, 197–217 (2005)
61. S. Migórski, Navier–Stokes problems modeled by evolution hemivariational inequalities. Discret. Contin. Dyn. Syst. Suppl. **2007**, 731–740 (2007)
62. S. Migórski, S. Dudek, Evolutionary oseen model for generalized Newtonian fluid with multivalued nonmonotone friction law. J. Math. Fluid Mech. **20**, 1317–1333 (2018)
63. S. Migórski, A. Ochal, Optimal control of parabolic hemivariational inequalities. J. Glob. Optim. **17**, 285–300 (2000)
64. S. Migórski, D. Pączka, Analysis of steady flow of non-Newtonian fluid with leak boundary condition (submitted)
65. S. Migórski, D. Pączka, Hemivariational inequality for Newtonian fluid flow with leak boundary condition (submitted)

66. S. Migórski, D. Pączka, On steady flow of non-Newtonian fluids with frictional boundary conditions in reflexive Orlicz spaces. Nonlinear Anal. Real World Appl. **39**, 337–361 (2018)
67. S. Migórski, A. Ochal, M. Sofonea, *Nonlinear Inclusions and Hemivariational Inequalities. Models and Analysis of Contact Problems*. Advances in Mechanics and Mathematics, vol. 26 (Springer, New York, 2013)
68. J. Musielak, *Orlicz Spaces and Modular Spaces*. Lecture Notes in Mathematics, vol. 1034 (Springer, Berlin, 1983)
69. Z. Naniewicz, P.D. Panagiotopoulos, *Mathematical Theory of Hemivariational Inequalities and Applications* (Dekker, New York, 1995)
70. H.T. Nguyen, The superposition operators in Orlicz spaces of vector functions. Dokl. Akad. Nauk BSSR **31**, 191–200 (1987)
71. H.T. Nguyen, D. Pączka, Existence theorems for the Dirichlet elliptic inclusion involving exponential-growth-type multivalued right-hand side. Bull. Pol. Acad. Sci. Math. **53**, 361–375 (2005)
72. H.T. Nguyen, D. Pączka, Generalized gradients for locally Lipschitz integral functionals on non-L^p-type spaces of measurable functions, in *Function Spaces VIII. Proceedings of the 8th Conference on Function Spaces, Będlewo, 2006 (Warsaw)*, ed. by H. Hudzik, J. Musielak, M. Nowak, L. Skrzypczak, vol. 79 (Banach Center Publications, Warsaw, 2008), pp. 135–156
73. H.T. Nguyen, D. Pączka, Weak and Young measure solutions for hyperbolic initial-boundary value problems of elastodynamics in the Orlicz–Sobolev space setting. SIAM J. Math. Anal. **48**(2), 1297–1331 (2016)
74. D. Pączka, Frictional contact problem for steady flow of electrorheological fluids (submitted)
75. D. Pączka, Elastic contact problem with Coulomb friction and normal compliance in Orlicz spaces. Nonlinear Anal. Real World Appl. **45**, 97–115 (2019)
76. P.D. Panagiotopoulos, *Inequality Problems in Mechanics and Applications. Convex and Nonconvex Energy Functions* (Birkhäuser, Basel, 1985)
77. P.D. Panagiotopoulos, Optimal control of systems governed by hemivariational inequalities. Necessary conditions, in *Free Boundary Value Problems. Proceedings of a Conference held at the Mathematisches Forschungsinstitut, Oberwolfach, July 9–15, 1989*, ed. by K.H. Hoffmann, J. Sprekels. International Series of Numerical Mathematics, vol. 95 (Birkhäuser, Basel, 1990), pp. 207–228
78. P.D. Panagiotopoulos, *Hemivariational Inequalities, Applications in Mechanics and Engineering* (Springer, New York, 1993)
79. R.E. Powell, H.J. Eyring, Mechanism for relaxation theory of viscosity. Nature **154**, 427–428 (1944)
80. M.M. Rao, Z.D. Ren, *Theory of Orlicz Spaces* (Marcel Dekker, New York, 1991)
81. M. Růžička, A note on steady flow of fluids with shear dependent viscosity. Nonlinear Anal. **197**, 3029–3039 (1997)
82. J. Wolf, Existence of weak solutions to the equations of non-stationary motion of non-Newtonian fluids with shear rate dependent viscosity. J. Math. Fluid Mech. **9**, 104–138 (2007)

Chapter 2
Discrete Fourier Transform and Theta Function Identities

R. A. Malekar

Abstract The classical identities of the Jacobi theta functions are obtained from the multiplicities of the eigenvalues i^k and the corresponding eigenvectors of the DFT $\Phi(n)$ expressed in terms of the theta functions. An extended version of the classical Watson addition formula and Riemann's identity on theta functions is derived. Watson addition formula and Riemann's identity are obtained as a particular case. An extensions of some classical identities corresponding to the theta functions $\theta_{a,b}(x, \tau)$ with $a,b \in \frac{1}{3}\mathbb{Z}$ are also derived.

2.1 Introduction

The discrete Fourier transform (DFT) is a well-known computing tool and it is also a source of interesting mathematical problems. The work by Auslander and Tolimeri [1] unified the pure and applied aspects of computing DFT. The DFT is connected with different number of theoretical problems, for example, the trace of the DFT matrix is a well-known quadratic Gauss sum, up to trivial normalization factor. This was used by Schur to present a new method of evaluation of Gauss sums. Recently, instead of the DFT the expression 'quantum Fourier transform'(QFTR) has been frequently used starting with seminal work of Shore [16]. In our work we use more traditional abbreviation DFT for the same object.

DFT is a well-known computing tool, but the eigenfunctions of the DFT were not well known. The first major step in this direction was taken by Mehta [13]. He had constructed eigenfunctions of the DFT in terms of discrete analogue of the Hermite functions. Galetti and Marchiolli [4] expressed the eigenfunctions obtained by Mehta as derivatives of Jacobi theta functions. The properties of Jacobi theta functions and their derivatives as eigenfunctions of the DFT were further studied by Ruzzi [15].

R. A. Malekar (✉)
Department of Mathematics, National Defence Academy, Khadakwasla, Pune, India

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_2

The method of obtaining eigenfunctions of the FT and DFT was generalized by Matveev [10]. He proved beautiful consequences of the fact that DFT and FT are the fourth roots of unity. In addition, eigenfunctions of the DFT were constructed in terms of any absolutely summable series g_n. The particular case of this was eigenfunctions found by Mehta. This was then applied for the case when series arises as summands of a ν-theta functions with characteristics (a,b) denoted by $\theta_{a,b}(x, \tau, \nu)$. (For $\nu = 1$ this reduces to the usual theta function.)

The study of theta functions is central to mathematics. Theta functions obey a bewildering number of identities. This has been emphasized by using the DFT to derive existing classical identities and also new identities of theta functions. This work will place the DFT as an important object of study to derive identities between theta functions. The method is different from the usual method in the literature, which uses the properties of zeros of theta functions and their infinite product representation.

This work shows that the classical identities of Jacobi theta functions can be obtained from the multiplicities of the eigenvalues i^k and corresponding eigenvectors of the DFT, expressed in terms of theta functions. An extended version of the classical Watson addition formula and Riemann's identity on theta functions is derived. The Watson addition formula and Riemann's identity are obtained as a particular case. An extension of classical identities corresponding to the theta functions $\theta_{a,b}(x, \tau)$ with $a,b \in \frac{1}{3}\mathbb{Z}$ is derived.

2.2 Spectral Theory of Discrete Fourier Transform

The multiplicities of the eigenvalues of the DFT $\Phi(n)$ are closely related to the trace of the DFT [1]. The spectral theory of the DFT in this section is developed based on the work of Matveev [10]. He proved a beautiful consequence of the fact that the DFT $\Phi(n)$ is the fourth root of unity, i.e. it satisfies the $\Phi^4 = I$. The spectral theorem gives a spectral decomposition in terms of real symmetric spectral projectors p_k. These projectors are related to the eigenvalues of the DFT by $m_k = Tr\ p_k$. The spectral projectors also lead to a formula for generation of eigenvectors from any vector f. This leads to the construction of eigenfunctions of the DFT in terms of absolutely summable series.

2.2.1 The Discrete Fourier Transform

The matrix $\Phi(n)$ corresponding to the discrete Fourier transform (DFT) of size n is defined by the formula

$$\Phi_{jk}(n) = \frac{1}{\sqrt{n}} q^{jk}, \quad j, k = 0, \ldots, n-1 \quad q = e^{\frac{2\pi i}{n}}. \tag{2.1}$$

Clearly $\Phi_{jk} = \Phi_{kj}$, Also

$$(\Phi\Phi^*)_{jk} = \frac{1}{n}\sum_{r=0}^{n-1} q^{jr}\overline{q^{rk}} = \frac{1}{n}\sum_{r=0}^{n-1} q^{-(j-k)r} = \delta_{jk}.$$

Thus, Φ is unitary and symmetric at the same time. Hence inverse of Φ is obtained by the complex conjugation.

Definition 2.2.1 For $f = (f_0, \ldots, f_{n-1})^t \in C^n$ we define the DFT $\tilde{f} \in C^n$ by $\tilde{f} = \Phi f = (\tilde{f_0}, \tilde{f_1}, \ldots, \tilde{f_{n-1}})$,where

$$\tilde{f_k} = \frac{1}{\sqrt{n}}\sum_{j=0}^{n-1} f_j \, e^{\frac{2\pi ijk}{n}} \;. \tag{2.2}$$

This immediately gives the Parseval's equality, namely

$$\sum_{k=0}^{n-1} |\tilde{f_k}|^2 = ||\tilde{f}||^2 = ||f||^2.$$

The sequence $\tilde{f_j}$ is in fact periodic, i.e. $\tilde{f_{j+n}} = \tilde{f_j}$. The sequence f_j can also be extended periodically with same period n. We will now onwards consider f and $\tilde{f}$ to be extended periodically in this manner. Clearly $\tilde{f} = \Phi f \Rightarrow f = \Phi^{-1}\tilde{f} = \Phi^* f$ that is

$$f_j = \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} \tilde{f_k} e^{\frac{-2\pi ijk}{n}} \quad \text{or} \quad f = \Phi^{-1}\tilde{f}. \tag{2.3}$$

Also,

$$\begin{aligned}(\Phi^2)_{ij} &= \frac{1}{n}\sum_{k=0}^{n-1} q^{(i+j)k},\\ &= 1 \;\text{ if }\; i+j \equiv 0 \pmod{n},\\ &= 0 \;\text{ otherwise.}\end{aligned}$$

This shows that

$$\Phi^2 = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0\\ 0 & 0 & \cdots & 0 & 1\\ 0 & 0 & \cdots & 1 & 0\\ \vdots & \vdots & \cdots & 0 & 0\\ 0 & 1 & 0 & \cdots & 0\end{pmatrix}.$$

In other words we find that

$$\Phi^2_{jm} = \Phi^2_{mj}, \quad (\Phi^2 f)_j = f_{n-j}, \quad \Phi^4 = I, \quad \Phi^3 = \Phi^{-1} = \Phi^*. \tag{2.4}$$

$Tr\ \Phi^2$ takes only two values 1 or 2, according as whether n is odd or even. This can be summarized as

$$Tr\ \Phi^2 = \frac{3 + (-1)^n}{2}. \tag{2.5}$$

The quantity

$$G(n) = \sqrt{n}\ Tr\ \Phi = \sum_{s=0}^{n-1} q^{s^2}, \tag{2.6}$$

is the well-known Gauss sum.

A celebrated result of the Gauss is that

$$G(n) = \frac{1 + (-i)^n}{1 - i}. \tag{2.7}$$

There are many proofs of Gauss sum available in the literature [1]. Schur had given a proof using discrete Fourier transform.

2.2.2 Spectral Decomposition of the Matrix or Operator Roots of Unity

We will need the following result of Matveev [10], the proof follows by sum of the nth roots of unity being zero.

Theorem 2.2.2 (Matveev) *Let U be any bounded operator in a Hilbert space satisfying the relation $U^n = I$ and let $q = e^{\frac{2\pi i}{n}}$, q is a primitive nth root of unity.*

$$U = \sum_{m=0}^{n-1} q^m P_m, \quad where \quad P_m = \frac{1}{n}\sum_{j=0}^{n-1} q^{-jm} U^j. \tag{2.8}$$

The spectral projectors P_j, forming resolution of identity for U, satisfy the relations

$$\sum_{j=0}^{n-1} P_j = I, \qquad U P_j = q^j P_j, \tag{2.9}$$

$$P_j^2 = P_j, \quad P_j P_m = 0 \quad for \quad m \neq j. \tag{2.10}$$

If U is unitary operator, then all spectral projectors P_j are self-adjoint operators.

The multiplicity m_j of eigenvalue q^j can be computed by

$$m_j = Tr\ P_j = \frac{1}{n}\sum_{k=0}^{n-1} q^{-kj} Tr\ U^k\ . \tag{2.11}$$

From (2.11) we can say that the traces of U^k and spectral multiplicities of eigenvalues of U are connected by the DFT.

The spectral decomposition of Φ is obtained from Theorem (2.2.2) by taking $n = 4$.

$$\Phi = \sum_{j=0}^{3} i^j p_j, \qquad p_j = \frac{1}{4}\sum_k (-i)^{jk}\Phi^k, \tag{2.12}$$

$$\sum_{j=0}^{3} p_j = I, \qquad p_j p_k = p_j \delta_{jk}. \tag{2.13}$$

All the projectors p_j are real symmetric matrices. It follows from spectral decomposition that given any vector $f = (f_0, f_1, \ldots, f_{n-1})^t \in C^n$, an eigenvector of Φ can be found as follows. Let $v(k) \in C^n$ be defined by

$$v(k) = \frac{1}{n}\left[I + (-i)^k\Phi + (-i)^{2k}\Phi^2 + (-i)^{3k}\Phi^3\right] f.$$

In particular, this gives that if

$$v_j(k) = f_j + (-i)^k \tilde{f}_j + (-i)^{2k} f_{n-j} + (-i)^{3k} \tilde{f_{-j}}. \tag{2.14}$$

then

$$\Phi v(k) = i^k v(k)\ .$$

Let $m_0 = m(1)$, $m_1 = m(i)$, $m_2 = m(-1)$, $m_3 = m(-i)$ be the multiplicities corresponding to the eigenvalues $1, i, -1, -i$ of Φ. We have the formula

$$m_k = Tr\ P_k.$$

Hence by using (2.11), (2.5) and (2.6) we get for m_k the following explicit expressions.

In particular

$$m_0 = Tr\ P_0 = Tr\ \left(\frac{1}{4}\sum \Phi^k\right) = \frac{1}{4}\ Tr\ \left(I + \Phi + \Phi^2 + \Phi^3\right),$$

However $\Phi^3 = \Phi^{-1} = \Phi^*$. Thus,

$$m_0 = \frac{1}{4}\ Tr\ \left(I + \Phi + \Phi^2 + \Phi^*\right) = \frac{1}{4}\left(n + 2Re\ Tr\ \Phi + \frac{3 + (-1)^n}{2}\right).$$

Using the formula for the Gauss sum $G(n) = \sqrt{n}\ Tr\ \Phi$, this can be simplified. If $n = 4m + k$ for $0 \leq k \leq 3$, then $m_0 = m + 1 = \left[\frac{n}{4}\right] + 1$. Similarly, the expressions for m_1, m_2, m_3 follows.

$$m_1 = \tfrac{1}{4}\left(n + 2Im\ Tr\Phi(n) - \tfrac{3+(-1)^n}{2}\right) = \left[\frac{n+1}{4}\right],$$

$$m_2 = \tfrac{1}{4}\left(n - 2Re\ Tr\Phi(n) + \tfrac{3+(-1)^n}{2}\right) = \left[\frac{n+2}{4}\right],$$

$$m_3 = \tfrac{1}{4}\left(n - 2Im\ Tr\Phi(n) - \tfrac{3+(-1)^n}{2}\right) = \left[\frac{n+3}{4}\right] - 1.$$

where $[x]$ is the largest integer $\leq x$. In a more detailed form the same formulae can be written as follows:

$$\begin{array}{lllll} n = 4m + 2 \Rightarrow & m_1 = m & m_2 = m + 1 & m_3 = m & m_0 = m + 1, \\ n = 4m \Rightarrow & m_1 = m & m_2 = m & m_3 = m - 1 & m_0 = m + 1, \\ n = 4m + 1 \Rightarrow & m_1 = m & m_2 = m & m_3 = m & m_0 = m + 1, \\ n = 4m + 3 \Rightarrow & m_1 = m + 1 & m_2 = m + 1 & m_3 = m & m_0 = m + 1. \end{array}$$

2.2.3 Eigenvectors of $\Phi(n)$

The explicit formula for the spectral projectors p_j provides the complete set of eigenvectors of Φ, as p_j are orthogonal projections on the i^k eigenspace of Φ. Since $\Phi p_k = i^k p_k$, it is clear that each nonzero column of p_k is an eigenvector of Φ. The columns of p_k can be explicitly written down.

Lemma 2.2.3 *Let $v(k, m)$ denote the mth column of p_k then*

$$4v_j(k, m) = \delta_{jm} + (-1)^k \delta_{n-j,m} + (-i)^k \frac{q^{jm}}{\sqrt{n}} + (-i)^{3k} \frac{q^{-jm}}{\sqrt{n}} \tag{2.15}$$

for $0 \leq j \leq n - 1$.

All these eigenvectors have real-valued components, and giving k the values 0, 1, 2, 3 we get the following detailed version of them:

$$4v_j(0,m) = \delta_{jm} + \delta_{n-j,m} + \frac{2}{\sqrt{n}} \cos\left(\frac{2\pi mj}{n}\right),$$

$$4v_j(1,m) = \delta_{jm} - \delta_{n-j,m} + \frac{2}{\sqrt{n}} \sin\left(\frac{2\pi mj}{n}\right),$$

$$4v_j(2,m) = \delta_{jm} + \delta_{n-j,m} - \frac{2}{\sqrt{n}} \cos\left(\frac{2\pi mj}{n}\right),$$

$$4v_j(3,m) = \delta_{jm} - \delta_{n-j,m} - \frac{2}{\sqrt{n}} \sin\left(\frac{2\pi mj}{n}\right).$$

Proof From the formula for

$$p_k = \frac{1}{4}\sum_{r=0}^{3}(i)^{-rk}\Phi^r = \frac{1}{4}\sum_{r=0}^{3}(-i)^{rk}\Phi^r,$$

$$\begin{aligned}
p_k(m,j) &= \frac{1}{4}\sum_{r=0}^{3} i^{-rk}[\Phi^r]_{jm} \\
&= \frac{1}{4}\sum_{r=0}^{3}(-i)^{rk}[\Phi^r]_{jm} \\
&= \frac{1}{4}\left[I + (-i)^k\Phi + (-i)^{2k}\Phi^2 + (-i)^{3k}\Phi^3\right]_{jm} \\
&= \frac{1}{4}\left[I + (-1)^k\Phi^2 + (-i)^k\Phi + (-i)^{3k}\Phi^3\right]_{jm} \\
p_k(m,j) &= \delta_{jm} + (-1)^k\delta_{n-j,m} + (-i)^k\frac{q^{jm}}{\sqrt{n}} + (-i)^{3k}\frac{q^{-jm}}{\sqrt{n}}\ .
\end{aligned}$$

This completes the proof. □

The explicit construction of eigenvectors of the DFT provides the complete set of eigenvectors given by the following theorem.

Theorem 2.2.4 *The vectors*

$$v(0,m), \quad m = 0, 1, 2\ldots, m(1),$$

$$v(1,m), \quad m = 1, 2\ldots, m(i) + 1,$$

$$v(2,m), \quad m = 0, 1, 2\ldots, m(-1),$$
$$v(3,m), \quad m = 1, 2\ldots, m(-i)+1$$

form the basis of C^n.

The proof of the above non-trivial fact is given in McClennan and Parks [11] using the techniques of Chebyshev sets. In terms of spectral projectors we have to take first m_0 columns of p_0 and first nontrivial columns m_1 of p_1, m_2 first columns of p_2 and m_3 first nontrivial columns of p_3 to get it. The completeness follows from the relation that $\sum p_j = I$. The projectors p_j are real symmetric matrices. We illustrate the computation of the spectral projectors for values of $n \leq 4$.

Case (I) $n = 2$
For $n = 2$, $q = e^{\frac{2\pi i}{2}} = -1$, $\Phi^2 = I$.

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

Eigenvalues of Φ are 1 and -1, and there are no imaginary eigenvalues. Hence spectral projectors corresponding to $+i$ and $-i$, i.e. p_1 and p_3 are zero. Also using the formulae for the projectors

$$p_0 + p_2 = \frac{I + \Phi^2}{2} = I, \quad p_1 + p_3 = \frac{I - \Phi^2}{2} = 0,$$

$$p_0 - p_2 = \frac{\Phi + \bar{\Phi}}{2} = \Phi, \quad p_1 - p_3 = \frac{\Phi - \bar{\Phi}}{2i} = 0\,.$$

Thus $p_1 = p_3 = 0$, and

$$p_0 = \frac{I + \Phi}{2} \quad \text{and} \quad p_2 = \frac{I - \Phi}{2}\,.$$

Therefore we have

$$p_0 = \frac{1}{2\sqrt{2}}\begin{pmatrix} \sqrt{2}+1 & 1 \\ 1 & \sqrt{2}-1 \end{pmatrix}.$$

Similarly

$$p_2 = \frac{1}{2\sqrt{2}}\begin{pmatrix} \sqrt{2}-1 & -1 \\ -1 & \sqrt{2}+1 \end{pmatrix}.$$

Case (II) $n = 3$

$$\Phi = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \frac{-1+i\sqrt{3}}{2} & \frac{-1-i\sqrt{3}}{2} \\ 1 & \frac{-1-i\sqrt{3}}{2} & \frac{-1+i\sqrt{3}}{2} \end{pmatrix} .$$

The spectral multiplicities of $\Phi(3)$ are $m_1 = 1$, $m_2 = 1$, $m_3 = 0$, $m_0 = 1$, therefore spectral projector $p_3 = 0$. Using the same formulae for p_0, p_1, p_2 we have

$$p_0 = \frac{1}{2\sqrt{3}} \begin{pmatrix} \sqrt{3}+1 & 1 & 1 \\ 1 & \frac{\sqrt{3}-1}{2} & \frac{\sqrt{3}-1}{2} \\ 1 & \frac{\sqrt{3}-1}{2} & \frac{\sqrt{3}-1}{2} \end{pmatrix},$$

$$p_1 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix},$$

$$p_2 = \frac{1}{2\sqrt{3}} \begin{pmatrix} \sqrt{3}-1 & -1 & -1 \\ -1 & \frac{\sqrt{3}-1}{2} & \frac{\sqrt{3}-1}{2} \\ -1 & \frac{\sqrt{3}-1}{2} & \frac{\sqrt{3}-1}{2} \end{pmatrix} .$$

Case (III) $n = 4$
In case $n = 4$ the Fourier matrix Φ is

$$\Phi = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix} .$$

By the formulae for the spectral multiplicities $m_1 = 1$, $m_2 = 1$, $m_3 = 0$, $m_0 = 2$.

Since $m_3 = 0$ we have spectral projector $p_3 = 0$. Using formulae for the spectral projectors we have

$$p_0 = \frac{1}{4} \begin{pmatrix} 3 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 3 & -1 \\ 1 & 1 & -1 & 1 \end{pmatrix},$$

$$p_1 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix},$$

$$p_2 = \frac{1}{4} \begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{pmatrix}.$$

The complete construction of real orthogonal matrices diagonalizing DFT for some particular values of $n \leq 8$ using the columns of p_k is given in [10]. The complete solution for diagonalization of the DFT $\Phi(n)$ is given in [18] using characters on finite abelian group. The importance of diagonalization in digital signal processing is outlined in [3].

2.2.4 *Eigenvectors of the DFT from Any Absolutely Summable Series*

Let

$$\sum_{m \in Z} g_m$$

be any given absolutely summable series and $q = e^{\frac{2\pi i}{n}}$. Define $g(x)$, a continuous periodic function of period n and a periodic sequence f_j of period n by

$$g(x) \;=\; \sum_{m \in Z} g_m q^{mx} \text{ and } f_j \;= \sum_{m \in Z} g_{mn+j} \; \text{ for } 0 \leq j \leq n-1.$$

Lemma 2.2.5 (Discrete Poisson Summation Formula)

$$\sum_{j=0}^{n-1} q^{jl} \left(\sum_{m \in Z} g_{nm+j} \right) = \sum_{k \in Z} g_k q^{kl} = g(l) \, for \;\; l = 0, 1, 2, \ldots, n-1. \tag{2.16}$$

$$\sum_{m \in Z} g_{nm+j} = \frac{1}{n} \sum_{l=0}^{n-1} g(l) q^{-lj} \;\; for \; j = 0, 1, \ldots, n-1. \tag{2.17}$$

Proof It is clear that (2.17) follows from (2.16). We prove (2.16). Consider the sum

$$\sum_{j=0}^{n-1} q^{jl} \sum_{m \in Z} g_{nm+j}$$

Let $nm + j = k$ and $j = k - nm$ therefore $q^{jl} = q^{kl}$.

LHS $= \sum_k q^{kl} g_k = g(l)$. By applying the inversion of the DFT we get

$$\sum_{m \in Z} g_{nm+j} = \frac{1}{n} \sum_{l=0}^{n-1} g(l) q^{-lj} \quad for\ j = 0, 1, \ldots, n-1.$$

□

Theorem 2.2.6 *Let g_m be any absolutely convergent series. Define for $k \in (0, 1, 2, 3)$ the vector $v(k) \in C^n$ with components given by*

$$[v(k)]_j = \sum_{m \in Z} (g_{nm+j} + (-1)^k g_{nm-j}) + \frac{(-i)^k}{\sqrt{n}} \sum_{m \in Z} (g_m + (-1)^k g_{-m}) e^{\frac{2\pi imj}{n}} \tag{2.18}$$

for $j = 0, 1, 2, \ldots, n-1$.

$v(k)$ is an eigenvector of Φ with eigenvalue i^k, that is

$$\Phi(v(k)) = i^k v(k)\,.$$

All the eigenvectors of the DFT can be constructed by the above formula.

Proof Let $f_j = \sum_{m \in Z} g_{nm+j}$, and

$$s_2 = \frac{(-i)^k}{\sqrt{n}} \sum_{m \in Z} (g_m + (-1)^k g_{-m}) e^{\frac{2\pi imj}{n}}.$$

put $m = kn + r$, then

$$\begin{aligned} s_2 &= \frac{(-i)^k}{\sqrt{n}} \sum_{r=0}^{n-1} f_r e^{\frac{2\pi irj}{n}} + \frac{(-i)^{3k}}{\sqrt{n}} \sum_{r=0}^{n-1} f_r e^{\frac{-2\pi irj}{n}} \\ &= (-i)^k \tilde{f}_j + (-i)^{3k} \tilde{f}_{-j} \\ v_j(k) &= f_j + (-1)^k f_{-j} + (-i)^k \tilde{f}_j + (-i)^{3k} \tilde{f}_{-j}. \end{aligned}$$

Therefore from (2.14) v_k is an eigenvector for Φ with eigenvalue i^k. In order to show that above formula generates all eigenvectors of the DFT we take

$$g_m(l) = \delta_{lm}, \quad l = 0, 1, 2, 3 \ldots n - 1.$$

It is clear that this choice of g_m produces complete system of eigenvectors of Φ. □

The generation of eigenvectors of the DFT in terms of series g_n gives many options to generate eigenvectors of the DFT by a proper choice of summable series g_n. This has been applied for theta functions by Matveev, this leads to new way of deriving identities of theta functions.

2.3 DFT $\Phi(2)$ and Jacobi Theta Function Identities

The connection of the eigenvalues of the DFT to the theta functions was discussed by Auslander and Tolimieri [1]. They related the multiplicity of the eigenvalues of the DFT to certain algebra of theta functions. The eigenfunctions of the DFT are expressed as derivatives of theta functions by Galetti and Marchiolli [4]. These eigenfunctions come from the eigenfunctions constructed by Mehta [13]. Mehta constructed eigenfunctions in terms of discrete analogue of Hermite functions. The eigenfunctions of the DFT in terms of ν-theta functions are constructed by Matveev [10], this reduces to the usual theta functions for $\nu = 1$.

This section is an extension of the work done in [10] where eigenfunctions of the DFT are expressed in terms of theta functions. This will place the DFT as an important object to derive various identities between theta functions and other well-known functions. The well-known fourth order identity between null values of theta functions is derived. An extended Watson addition formula is derived whose particular case is the classical Watson addition formula. A famous fourth order identity of Riemann is derived as a particular case of an extended Riemann's identity. All these identities are derived from the properties of multiplicities of eigenvalue and eigenvectors of the DFT expressed in terms of theta functions. This is a different approach from the classically used techniques which uses properties of zeros of theta functions and their infinite product representations. Liu [5] has also used Fourier series expansion to recover many classical results in theta function due to Jacobi, Ramanujan and others. The recent paper by Srivastava et al. [17] has discussed theta function identities related to the Jacobi triple product identity. There are many well-known and new identities of theta functions discussed in [2, 6]. It will be interesting to derive these identities using the techniques discussed in this chapter.

2.3.1 Jacobi Theta Functions

Theta functions where first studied by Jacobi are central to the theory of elliptic functions [12]. The four classical Jacobi theta functions are defined by

$$\theta_1(x,\tau) = (-i)\sum_{m\in Z}(-1)^m e^{\pi i(m+1/2)^2\tau} e^{2\pi i(m+1/2)x},$$
$$\theta_2(x,\tau) = \sum_{m\in Z} e^{\pi i(m+1/2)^2\tau} e^{2\pi i(m+1/2)x},$$
$$\theta_3(x,\tau) = \sum_{m\in Z} e^{\pi i m^2\tau} e^{2\pi i m x},$$
$$\theta_4(x,\tau) = \sum_{m\in Z}(-1)^m e^{\pi i m^2\tau} e^{2\pi i m(x+1/2)}.$$

Here each θ_i is considered as a function of x, and also depends on the parameter τ which satisfies $\Im(\tau) > 0$. The latter inequality guarantees absolute convergence of all the infinite series for all finite x. These functions are doubly periodic with periods 1 and τ. It follows from the definition that $\theta_1(x)$ is an odd function of x and $\theta_2(x)$, $\theta_3(x)$, $\theta_4(x)$ are even functions of x. The zeros of theta functions are as follows:

$$\theta_1(x,\tau) = 0 \text{ at } x = m + n\tau \quad \text{for } m\ ,\ n\ \in\ Z^+$$
$$\theta_2(x,\tau) = 0 \text{ at } x = \frac{1}{2} + m + n\tau$$
$$\theta_3(x,\tau) = 0 \text{ at } x = \frac{1}{2} + \frac{1}{2}\tau + m + n\tau$$
$$\theta_4(x,\tau) = 0 \text{ at } x = \frac{1}{2}\tau + m + n\tau.$$

Theta functions with characteristics $[a, b]$ are defined by

$$\theta_{a,b}(x,\tau) = \sum_{n\in Z}\exp[\pi i\tau(n+a)^2 + 2\pi i(n+a)(x+b)]. \tag{2.19}$$

$\theta_{a,b}(x,\tau)$ is connected with $\theta_3(x,\tau)$ by (see [20])

$$\theta_{a,b}(x,\tau) = \theta_3(x + a\tau + b,\tau)\exp[\pi i a^2\tau + 2\pi i a(x+b)].$$

We use theta functions $\theta_{a,b}$ with a,$b \in \frac{1}{2}\mathbb{Z}$ to express θ_i in terms of an exponential form (see [14]) used in this chapter. The classical Jacobi theta functions correspond to the discrete lattice $\frac{1}{2}\mathbb{Z} \times \frac{1}{2}\mathbb{Z}$. The extensions and identities of these functions

Table 2.1 Theta function at various arguments

	$x+\frac{1}{2}$	$x+\frac{\tau}{2}$	$x+\frac{1}{2}+\frac{\tau}{2}$	$x+1$	$x+\tau$
θ_1	θ_2	$ia\theta_4$	$a\theta_3$	$-\theta_1$	$-b\theta_1$
θ_2	$-\theta_1$	$a\theta_3$	$-ia\theta_4$	$-\theta_2$	$b\theta_2$
θ_3	θ_4	$a\theta_2$	$ia\theta_1$	θ_3	$b\theta_3$
θ_4	θ_3	$ia\theta_1$	$a\theta_2$	θ_4	$-b\theta_4$

for $\theta_{a,b}$ with $a,b \in \frac{1}{3}\mathbb{Z}$ are discussed in the next chapter.

$$\theta_1(x,\tau) = \theta_{\frac{1}{2},\frac{1}{2}}(x,\tau) = (-1)\sum_{m\in Z}\exp[\pi i\tau(m+1/2)^2 + 2\pi i(m+1/2)(x+1/2)],$$

$$\theta_2(x,\tau) = \theta_{\frac{1}{2},0}(x,\tau) = \sum_{m\in Z}\exp[\pi i\tau(m+1/2)^2 + 2\pi i(m+1/2)x],$$

$$\theta_3(x,\tau) = \theta_{0,0}(x,\tau) = \sum_{m\in Z}\exp(\pi i m^2\tau + 2\pi i m x),$$

$$\theta_4(x,\tau) = \theta_{0,\frac{1}{2}}(x,\tau) = \sum_{m\in Z}\exp[\pi i\tau m^2 + 2\pi i m(x+1/2)].$$

The representation of theta functions in exponential form can be used to evaluate theta functions at the arguments $x+\frac{1}{2}, x+1, x+\frac{1}{2}+\frac{1}{2}\tau$. These results are summarized in Table 2.1.

In Table 2.1 we have used the notations $a = e^{-\pi i x}e^{-\frac{\pi i\tau}{4}}$ and $b = e^{-2\pi i x}q^{-\pi i\tau}$.

2.3.2 *DFT and Theta Functions*

Matveev [10] has defined generalized ν-theta function given by

$$\theta(x,\tau,\nu) = \sum_{m=-\infty}^{\infty} e^{\pi i\tau m^{2\nu}+2\pi i m x}, \qquad \nu\in Z^+, \quad \Im(\tau) > 0. \tag{2.20}$$

It is clear that $\theta(x,\tau,\nu)$ is an entire function of x satisfying the relation

$$\theta(x+1,\tau,\nu) = \theta(x,\tau).$$

$\theta(x,\tau,\nu)$ reduces to the usual theta functions for $\nu = 1$. The ν-theta function satisfies the partial differential equation given by

$$2(2\pi)^{2\nu-1}(-1)^{\nu}\frac{\partial\theta}{\partial\tau} = i\frac{\partial^{2\nu}\theta}{\partial x^{2\nu}}. \tag{2.21}$$

The ν theta function $\theta_{a,b}(x, \tau, \nu)$ with characteristics $[a,b]$ is defined by

$$\theta_{a,b}(x, \tau, \nu) = \sum_{n \in Z} \exp[\pi i \tau (n+a)^{2\nu} + 2\pi i (n+a)(x+b)]. \tag{2.22}$$

We state below the theorem of Matveev which relates $\theta(x, \tau, \nu)$ to the eigenfunctions of the DFT.

Theorem 2.3.1 (Matveev) *For any τ with $\Im(\tau) > 0$ the vector $v(x, \tau, \nu, k)$ with components $v_j(x, \tau, \nu, k)$, $j = 0, 1, 2, \ldots, n-1$ given by*

$$\begin{aligned} v_j(x, \tau, \nu, k) &= \theta_{\frac{j}{n},0}(x, \tau, \nu) + (-1)^k \theta_{-\frac{j}{n},0}(x, \tau, \nu) \\ &+ \frac{1}{\sqrt{n}} \left[(-i)^k \theta \left(\frac{j+x}{n}, \frac{\tau}{n^{2\nu}}, \nu \right) + (-i)^{3k} \theta \left(\frac{x-j}{n}, \frac{\tau}{n^{2\nu}}, \nu \right) \right] \end{aligned} \tag{2.23}$$

is an eigenvector of the DFT $\Phi(n)$ with an eigenvalue i^k:

$$\Phi(n) v(x, \tau, \nu, k) = i^k v(x, \tau, \nu, k).$$

Proof The proof follows from Theorem 2.2.6 by choosing the absolutely summable series g_m. Consider

$$g_m = \exp\left(\frac{\pi i \tau m^{2\nu}}{n^{2\nu}} + \frac{2\pi i m x}{n} \right). \tag{2.24}$$

Then,

$$\begin{aligned} \sum_{m \in Z} g_{mn+j} &= \sum_{m \in Z} \exp\left[\pi i \tau \left(\frac{mn+j}{n} \right)^{2\nu} + 2\pi i \left(\frac{mn+j}{n} \right) x \right) \\ &= \theta_{\frac{j}{n},0}(x, \tau, \nu). \\ \sum_{m \in Z} g_{mn-j} &= \theta_{-\frac{j}{n},0}(x, \tau, \nu). \end{aligned}$$

$$\sum_{m \in Z} (g_m + (-1)^k g_{-m}) e^{\frac{2\pi i m j}{n}} = \left[\theta \left(\frac{j+x}{n}, \frac{\tau}{n^{2\nu}}, \nu \right) + (-1)^k \theta \left(\frac{x-j}{n}, \frac{\tau}{n^{2\nu}}, \nu \right) \right].$$

substituting the above expressions in Theorem (2.2.6) Theorem (2.3.1) follows. □

The ν-theta functions for $\nu = 1$ correspond to the classical Jacobi theta functions. It has been suggested in [10] that this may have applications to determine identities of theta functions. For a given value of n, take eigenvectors of the form (2.23) corresponding to the eigenvalue i^k, with different values of x and τ.

At the most m_k the multiplicities of the eigenvalue i^k of these are linearly independent. Thus minors of the matrix consisting of the eigenvectors $v(x, \tau, k)$ of order greater than m_k vanish. This may lead to new identities among theta functions. This is explored for the DFT $\Phi(2)$ in the following section to derive well-known fourth order identity between null values of theta functions. We obtain an extended Watson addition formula, and its particular case we prove the classical Watson addition formula. The well-known Riemann's fourth order identity is proved using the above technique, and its extension which we call the extended Riemann's identity is derived. Some new cubic identities of Jacobi theta functions are derived.

2.3.3 DFT $\Phi(2)$ and Jacobi Theta Function Identities

In the context of Theorem (2.3.1), classical Jacobi theta function occurs as the components of the eigenvectors of the DFT $\Phi(2)$. We explore this fact to derive the well-known Landen type transformations and the fourth order identity between null values of theta functions.

2.3.4 The Identity $\theta^4(0, \tau) - \theta^4_{0,\frac{1}{2}}(0, \tau) = \theta^4_{\frac{1}{2},0}(0, \tau)$

The DFT $\Phi(2)$ has only two eigenvalues $+1$ and -1. The eigenvector corresponding to eigenvalue $+1$ is given by

$$v(x, \tau, 0) = \begin{bmatrix} 2\theta(x, \tau) + \sqrt{2}\theta\left(\frac{x}{2}, \frac{\tau}{4}\right) \\ 2\theta_{\frac{1}{2},0}(x, \tau) + \sqrt{2}\theta\left(\frac{x+1}{2}, \frac{\tau}{4}\right) \end{bmatrix}. \tag{2.25}$$

The eigenvector corresponding to eigenvalue -1 is given by

$$v(x, \tau, 2) = \begin{bmatrix} 2\theta(x, \tau) - \sqrt{2}\theta\left(\frac{x}{2}, \frac{\tau}{4}\right) \\ 2\theta_{\frac{1}{2},0}(x, \tau) - \sqrt{2}\theta\left(\frac{x+1}{2}, \frac{\tau}{4}\right) \end{bmatrix}. \tag{2.26}$$

We use these eigenvectors to derive the Landen type transformations.

Lemma 2.3.2 (Landen Transformation)

$$\theta^2(x, \tau) + \theta^2_{0,\frac{1}{2}}(x, \tau) = 2\theta(2x, 2\tau)\theta(0, 2\tau) \tag{2.27}$$

$$\theta^2(0, \tau) - \theta^2_{0,\frac{1}{2}}(0, \tau) = 2\theta_{\frac{1}{2},0}(2x, 2\tau)\theta_{\frac{1}{2},0}(0, 2\tau). \tag{2.28}$$

Proof We have

$$\Phi(2)\,[v(x,\tau,0)+v(x,\tau,2)] = v(x,\tau,0) - v(x,\tau,2).$$

This gives the following two identities:

$$\theta(x,\tau) + \theta_{\frac{1}{2},0}(x,\tau) = \theta\left(\frac{x}{2}, \frac{\tau}{2^2}\right), \tag{2.29}$$

$$\theta(x,\tau) - \theta_{\frac{1}{2},0}(x,\tau) = \theta\left(\frac{x+1}{2}, \frac{\tau}{2^2}\right). \tag{2.30}$$

Equations (2.29) and (2.30) are equivalent to

$$\theta(2x,4\tau) + \theta_{\frac{1}{2},0}(2x,4\tau) = \theta(x,\tau), \tag{2.31}$$

$$\theta(2x,4\tau) - \theta_{\frac{1}{2},0}(2x,4\tau) = \theta_{0,\frac{1}{2}}(x,\tau). \tag{2.32}$$

Using (2.25)

$$v\left(x+\frac{\tau}{2},\tau,0\right) = \begin{bmatrix} 2a\theta_{\frac{1}{2},0}(x,\tau) + a\sqrt{2}\theta\left(\frac{x}{2},\frac{\tau}{4}\right) \\ 2a\theta(x,\tau) - a\sqrt{2}\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right) \end{bmatrix} \tag{2.33}$$

where $a = \exp\left(\frac{-\pi i\tau}{4} - \pi i x\right)$.

$$v(x+1,\tau,0) = \begin{bmatrix} 2\theta(x,\tau) + \sqrt{2}\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right) \\ -2\theta_{\frac{1}{2},0}(x,\tau) + \sqrt{2}\theta\left(\frac{x}{2},\frac{\tau}{4}\right) \end{bmatrix}. \tag{2.34}$$

Since $v(x+1,\tau,0)$, $v\left(x+\frac{\tau}{2},\tau,0\right)$ are eigenvectors corresponding to the same eigenvalue 1, which has multiplicity 1 we have

$$\det\left(v(x+1,\tau,0), v\left(x+\frac{\tau}{2},\tau,0\right)\right) = 0.$$

This gives

$$2\theta^2(x,\tau) + 2\theta^2_{\frac{1}{2},0}(x,\tau) = \theta^2\left(\frac{x+1}{2},\frac{\tau}{4}\right) + \theta^2\left(\frac{x}{2},\frac{\tau}{4}\right). \tag{2.35}$$

Equation (2.35) is equivalent to

$$2\theta^2(2x,4\tau) + 2\theta^2_{\frac{1}{2},0}(2x,4\tau) = \theta^2_{0,\frac{1}{2}}(x,\tau) + \theta^2(x,\tau). \tag{2.36}$$

Now consider

$$\theta(2x,2\tau)\theta(0,2\tau)=\sum_{m,n}\exp(\pi i(m^2+n^2)2\tau+2\pi i m2x). \tag{2.37}$$

Let $m+n=n_1$ and $m-n=n_2$, so that n_1 and n_2 are of the same parity. Rewriting (2.37) in terms of n_1 and n_2 we have

$$\begin{aligned}\theta(2x,2\tau)\theta(0,2\tau) &= \sum_{n_1\equiv n_2 \pmod 2}\exp(\pi i(n_1^2+n_2^2)\tau+2\pi i(n_1+n_2)x).\\ &= \sum_{n_1,n_2 \text{ are even}}\exp(\pi i(n_1^2+n_2^2)\tau+2\pi i(n_1+n_2)x)\\ &\quad+\sum_{n_1,n_2 \text{ are odd}}\exp(\pi i(n_1^2+n_2^2)\tau+2\pi i(n_1+n_2)x).\\ &= \theta^2(2x,4\tau)+\theta^2_{\frac{1}{2},0}(2x,4\tau).\end{aligned}$$

From (2.36) we have

$$\theta^2(x,\tau)+\theta^2_{0,\frac{1}{2}}(x,\tau)=2\theta(2x,2\tau)\theta(0,2\tau).$$

This derives Landen type (2.27) transformation. We can derive the other Landen type transformation using a similar method. □

Now the $\det(v(x,\tau,0),v(x+1,\tau,0)=0$. This gives

$$\begin{aligned}&-4\theta(x,\tau)\theta_{\frac{1}{2},0}(x,\tau)+2\sqrt{2}\theta(x,\tau)\theta\left(\frac{x}{2},\frac{\tau}{4}\right)-2\sqrt{2}\theta_{\frac{1}{2},0}(x,\tau)\theta\left(\frac{x}{2},\frac{\tau}{4}\right)\\ &+2\theta^2\left(\frac{x}{2},\frac{\tau}{4}\right)-4\theta(x,\tau)\theta_{\frac{1}{2},0}(x,\tau)-2\sqrt{2}\theta_{\frac{1}{2},0}(x,\tau)\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)\\ &-2\sqrt{2}\theta(x,\tau)\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)-2\theta^2\left(\frac{x+1}{2},\frac{\tau}{4}\right)=0.\end{aligned}$$

Consider the terms with coefficients of $2\sqrt{2}$,

$$\begin{aligned}&\theta(x,\tau)\theta\left(\frac{x}{2},\frac{\tau}{4}\right)\\ &\quad=\theta(x,\tau)\theta(y,\tau)+\theta(x,\tau)\theta_{\frac{1}{2},0}(x,\tau)=A,\\ &\theta_{\frac{1}{2},0}(x,\tau)\theta\left(\frac{x}{2},\frac{\tau}{4}\right)\\ &\quad=\theta_{\frac{1}{2},0}(x,\tau)\theta(x,\tau)+\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(x,\tau)=B,\end{aligned}$$

$$\theta_{\frac{1}{2},0}(x,\tau)\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)$$
$$=\theta_{\frac{1}{2},0}(x,\tau)\theta(x,\tau)-\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(x,\tau)=C,$$
$$\theta(x,\tau)\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)$$
$$=\theta(x,\tau)\theta(x,\tau)-\theta(x,\tau)\theta_{\frac{1}{2},0}(x,\tau)=D.$$

We observe that $A-B-C-D=0$. All the terms with coefficients of $2\sqrt{2}$ cancel each other out. This gives

$$\theta^2\left(\frac{x}{2},\frac{\tau}{4}\right)-\theta^2\left(\frac{x+1}{2},\frac{\tau}{4}\right)=4\theta(x,\tau)\theta_{\frac{1}{2},0}(x,\tau). \tag{2.38}$$

Equation (2.38) is equivalent to

$$\theta^2(x,\tau)-\theta^2_{0,\frac{1}{2}}(x,\tau)=4\theta(2x,4\tau)\theta_{\frac{1}{2},0}(2x,4\tau). \tag{2.39}$$

Applying the same argument as in the derivation of (2.27) we have

$$2\theta(2x,4\tau)\theta_{\frac{1}{2},0}(2x,4\tau)=\theta_{\frac{1}{2},0}(2x,2\tau)\theta_{\frac{1}{2},0}(0,2\tau). \tag{2.40}$$

At $x=0$ we have

$$2\theta(0,4\tau)\theta_{\frac{1}{2},0}(0,4\tau)=\theta^2_{\frac{1}{2},0}(0,2\tau). \tag{2.41}$$

From (2.39) and (2.40) we have

$$\theta^2(x,\tau)-\theta^2_{0,\frac{1}{2}}(x,\tau)=2\theta_{\frac{1}{2},0}(2x,2\tau)\theta_{\frac{1}{2},0}(0,2\tau).$$

This derives (2.28).

Corollary 2.3.3

$$\theta^4(0,\tau)-\theta^4_{0,\frac{1}{2}}(0,\tau)=\theta^4_{\frac{1}{2},0}(0,\tau). \tag{2.42}$$

Proof From (2.27) at $x=0$ we have

$$\theta^2(0,\tau)+\theta^2_{0,\frac{1}{2}}(0,\tau)=2\theta^2(0,2\tau). \tag{2.43}$$

Similarly from (2.28) at $x=0$

$$\theta^2(0,\tau)-\theta^2_{0,\frac{1}{2}}(0,\tau)=2\theta^2_{\frac{1}{2},0}(0,2\tau). \tag{2.44}$$

From (2.43) and (2.44) we have

$$\theta^4(0,\tau) - \theta^4_{0,\frac{1}{2}}(0,\tau) = 4\theta^2(0,2\tau)\theta^2_{\frac{1}{2},0}(0,2\tau). \tag{2.45}$$

Using (2.41) we have

$$\theta^4(0,\tau) - \theta^4_{0,\frac{1}{2}}(0,\tau) = \theta^4_{\frac{1}{2},0}(0,\tau).$$

□

This is a well-known Jacobi identity between the null values of theta functions. Many of the classical identities involving the squares of theta functions (see [19]) can be obtained by the method illustrated above. In the next theorem we illustrate the technique to extend the classical Watson addition formula for theta functions [12].

2.3.5 Extended Watson Addition Formula

The theta functions obey Watson addition formula [12] involving the argument τ and 2τ. We give an extended Watson addition formula involving τ and 4τ [8].

Theorem 2.3.4 (Extended Watson Addition Formula)

$$\begin{aligned}
&\theta_{0,\frac{1}{2}}(x_1+x_2,\tau)\theta(x_1-x_2,\tau) - \theta_{0,\frac{1}{2}}(x_1-x_2,\tau)\theta(x_1+x_2,\tau)\\
&\quad = 2\theta(2x_1+2x_2,4\tau)\theta_{\frac{1}{2},0}(2x_1-2x_2,4\tau)\\
&\qquad -2\theta(2x_1-2x_2,4\tau)\theta_{\frac{1}{2},0}(2x_1+2x_2,4\tau).
\end{aligned} \tag{2.46}$$

Proof Using (2.29), (2.30)

$$\theta\left(\frac{x_1 \pm x_2}{2}, \frac{\tau}{2}\right) = \theta(x_1 \pm x_2, 2\tau) + \theta_{\frac{1}{2},0}(x_1 \pm x_2, 2\tau). \tag{2.47}$$

$$\theta\left(\frac{x_1 \pm x_2 + 1}{2}, \frac{\tau}{2}\right) = \theta(x_1 \pm x_2, 2\tau) - \theta_{\frac{1}{2},0}(x_1 \pm x_2, 2\tau). \tag{2.48}$$

From (2.25) and (2.26), we have

$$v(x_1+x_2,2\tau,0) = \begin{bmatrix} 2\theta(x_1+x_2,2\tau) + \sqrt{2}\theta\left(\frac{x_1+x_2}{2},\frac{\tau}{2}\right) \\ 2\theta_{\frac{1}{2},0}(x_1+x_2,2\tau) + \sqrt{2}\theta\left(\frac{x_1+x_2+1}{2},\frac{\tau}{2}\right) \end{bmatrix},$$

$$v(x_1-x_2,2\tau,0) = \begin{bmatrix} 2\theta(x_1-x_2,2\tau) + \sqrt{2}\theta\left(\frac{x_1-x_2}{2},\frac{\tau}{2}\right) \\ 2\theta_{\frac{1}{2},0}(x_1-x_2,2\tau) + \sqrt{2}\theta\left(\frac{x_1-x_2+1}{2},\frac{\tau}{2}\right) \end{bmatrix}$$

are eigenvectors of $\Phi(2)$ corresponding to eigenvalues $+1$, which has multiplicity 1. Therefore

$$\det(v(x_1+x_2, 2\tau, 0), v(x_1-x_2, 2\tau, 0)) = 0.$$

This gives

$$\begin{aligned}
&4\theta(x_1+x_2, 2\tau)\theta_{\frac{1}{2},0}(x_1-x_2, 2\tau) + 2\sqrt{2}\theta(x_1+x_2, 2\tau)\theta\left(\frac{x_1-x_2+1}{2}, \frac{\tau}{2}\right) \\
&\quad +2\sqrt{2}\theta_{\frac{1}{2},0}(x_1-x_2, 2\tau)\theta\left(\frac{x_1+x_2}{2}, \frac{\tau}{2}\right) \\
&\quad +2\theta\left(\frac{x_1+x_2}{2}, \frac{\tau}{2}\right)\theta\left(\frac{x_1-x_2+1}{2}, \frac{\tau}{2}\right) \\
&\quad -4\theta_{\frac{1}{2},0}(x_1+x_2, 2\tau)\theta(x_1-x_2, 2\tau) \\
&\quad -2\sqrt{2}\theta_{\frac{1}{2},0}(x_1+x_2, 2\tau)\theta\left(\frac{x_1-x_2}{2}, \frac{\tau}{2}\right) \\
&\quad -2\sqrt{2}\theta(x_1-x_2, 2\tau)\theta\left(\frac{x_1+x_2+1}{2}, \frac{\tau}{2}\right) \\
&\quad -2\theta\left(\frac{x_1+x_2+1}{2}, \frac{\tau}{2}\right)\theta\left(\frac{x_1-x_2}{2}, \frac{\tau}{2}\right) = 0.
\end{aligned} \tag{2.49}$$

In this expression consider the terms with $2\sqrt{2}$ as coefficient. Using formulas (2.47) and (2.48),we have

$$\begin{aligned}
A &= \theta(x_1+x_2, 2\tau)\theta\left(\frac{x_1-x_2+1}{2}, \frac{\tau}{2}\right) \\
&= \theta(x_1+x_2, 2\tau)\theta(x_1-x_2, 2\tau) - \theta(x_1+x_2, 2\tau)\theta_{\frac{1}{2},0}(x_1-x_2, 2\tau), \\
B &= \theta_{\frac{1}{2},0}(x_1-x_2, 2\tau)\theta\left(\frac{x_1+x_2}{2}, \frac{\tau}{2}\right) \\
&= \theta_{\frac{1}{2},0}(x_1-x_2, 2\tau)\theta(x_1+x_2, 2\tau) + \theta_{\frac{1}{2},0}(x_1+x_2, 2\tau)\theta_{\frac{1}{2},0}(x_1-x_2, 2\tau), \\
C &= \theta_{\frac{1}{2},0}(x_1+x_2, 2\tau)\theta\left(\frac{x_1-x_2}{2}, \frac{\tau}{2}\right) \\
&= \theta_{\frac{1}{2},0}(x_1+x_2, 2\tau)\theta(x_1-x_2, 2\tau) + \theta_{\frac{1}{2},0}(x_1+x_2, 2\tau)\theta_{\frac{1}{2},0}(x_1-x_2, 2\tau), \\
D &= \theta(x_1-x_2, 2\tau)\theta\left(\frac{x_1+x_2+1}{2}, \frac{\tau}{2}\right) \\
&= \theta(x_1-x_2, 2\tau)\theta(x_1+x_2, 2\tau) - \theta(x_1-x_2, 2\tau)\theta_{\frac{1}{2},0}(x_1+x_2, 2\tau).
\end{aligned}$$

It is clear that $A+B-C-D=0$. Hence all the terms with coefficients $2\sqrt{2}$ cancel each other out. Equation (2.49) becomes

$$\begin{aligned}
&4\theta(x_1+x_2,2\tau)\theta_{\frac{1}{2},0}(x_1-x_2,2\tau)-4\theta_{\frac{1}{2},0}(x_1+x_2,2\tau)\theta(x_1-x_2,2\tau)\\
&\quad=2\theta\left(\frac{x_1+x_2+1}{2},\frac{\tau}{2}\right)\theta\left(\frac{x_1-x_2}{2},\frac{\tau}{2}\right)\\
&\qquad-2\theta\left(\frac{x_1-x_2+1}{2},\frac{\tau}{2}\right)\theta\left(\frac{x_1+x_2}{2},\frac{\tau}{2}\right).
\end{aligned}$$

Replacing x_1, x_2, τ by $2x_1, 2x_2, 2\tau$, respectively, we obtain

$$\begin{aligned}
&2\theta(2x_1+2x_2,4\tau)\theta_{\frac{1}{2},0}(2x_1-2x_2,4\tau)-2\theta_{\frac{1}{2},0}(2x_1+2x_2,4\tau)\theta(2x_1-2x_2,4\tau)\\
&\quad=\theta_{0,\frac{1}{2}}(x_1+x_2,\tau)\theta(x_1-x_2,\tau)\\
&\qquad-\theta_{0,\frac{1}{2}}(x_1-x_2,\tau)\theta(x_1+x_2,\tau).
\end{aligned}$$

This proves (2.46). □

We now show that the classical Watson addition formula is a particular case of (2.46).

Theorem 2.3.5 (Watson Addition Formula)

$$\begin{aligned}
\theta_{\frac{1}{2},\frac{1}{2}}(x_1,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(x_2,\tau)&=\theta(x_1+x_2,2\tau)\theta_{\frac{1}{2},0}(x_1-x_2,2\tau)\\
&\quad-\theta(x_1-x_2,2\tau)\theta_{\frac{1}{2},0}(x_1+x_2,2\tau).
\end{aligned} \tag{2.50}$$

Proof Consider the first term of the left-hand side of (2.46)

$$\begin{aligned}
&\theta_{0,\frac{1}{2}}(x_1+x_2,\tau)\theta(x_1-x_2,\tau)\\
&\quad=\sum\exp\left[\pi i(m^2+n^2)\tau+2\pi i\left((m+n)x_1+(m-n)x_2+\frac{m}{2}\right)\right].
\end{aligned} \tag{2.51}$$

Let $m+n=n_1$, $m-n=n_2$ where n_1 and n_2 are of same parity. Rewriting (2.51) in terms of n_1 and n_2,

$$\begin{aligned}
&=\sum_{n_1\equiv n_2 \pmod 2}\exp\left[\pi i\left(\frac{n_1^2+n_2^2}{2}\right)\tau+2\pi i\left(n_1x_1+n_2x_2+\frac{n_1+n_2}{4}\right)\right]\\
&=\sum_{n_1,n_2\ are\ even}\exp\left[\pi i\left(\frac{n_1^2+n_2^2}{2}\right)\tau+2\pi i\left(n_1x_1+n_2x_2+\frac{n_1+n_2}{4}\right)\right]
\end{aligned}$$

$$+\sum_{n_1,n_2 \ are\ odd} \exp\left[\pi i\left(\frac{n_1^2+n_2^2}{2}\right)\tau + 2\pi i\left(n_1x_1+n_2x_2+\frac{n_1+n_2}{4}\right)\right]$$
$$=\theta_{0,\frac{1}{2}}(2x_1,2\tau)\theta_{0,\frac{1}{2}}(2x_2,2\tau)+\theta_{\frac{1}{2},\frac{1}{2}}(2x_1,2\tau)\theta_{\frac{1}{2},\frac{1}{2}}(2x_2,2\tau).$$

Similarly, using the same argument it is easy to show that,

$$\begin{aligned}\theta_{0,\frac{1}{2}}(x_1-x_2,\tau)\theta(x_1+x_2,\tau) &= \theta_{0,\frac{1}{2}}(2x_1,2\tau)\theta_{0,\frac{1}{2}}(2x_2,2\tau)\\ &\quad -\theta_{\frac{1}{2},\frac{1}{2}}(2x_1,2\tau)\theta_{\frac{1}{2},\frac{1}{2}}(2x_2,2\tau).\end{aligned}$$

By using the above two results in (2.46), we get

$$\begin{aligned}\theta_{\frac{1}{2},\frac{1}{2}}(2x_1,2\tau)\theta_{\frac{1}{2},\frac{1}{2}}(2x_2,2\tau) &= \theta(2x_1+2x_2,4\tau)\theta_{\frac{1}{2},0}(2x_1-2x_2,4\tau)\\ &\quad -\theta(2x_1-2x_2,4\tau)\theta_{\frac{1}{2},0}(2x_1+2x_2,4\tau).\end{aligned}$$

Watson addition formula (2.50) is obtained by replacing x_1, x_2, τ by $\frac{x_1}{2}$, $\frac{x_2}{2}$, $\frac{\tau}{2}$, respectively, in the above equation. □

2.3.6 Riemann's Identity

The Riemann's identity is the well-known fourth order identity of theta functions. The beautiful account of identities derived from Riemann's identity is given in Mumford [14]. We give a version of this identity from determinants of eigenvectors of the DFT $\Phi(2)$ with same eigenvalues [9]. We call it extended Riemann's identity, the particular case of this is the classical Riemann's identity. Liu [7] has given a beautiful and general extension of Jacobi quartic theta function identity using the techniques of Residue theorem.

Theorem 2.3.6 (Extended Riemann's Identity for Theta Functions)

$$\begin{aligned}&4\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)+4\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta(u,\tau)\theta(v,\tau)\\ &\quad+4\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)\theta(x,\tau)\theta(y,\tau)+4\theta(x,\tau)\theta(y,\tau)\theta(u,\tau)\theta(v,\tau)\\ &\quad=\theta\left(\frac{x}{2},\frac{\tau}{4}\right)\theta\left(\frac{y}{2},\frac{\tau}{4}\right)\theta\left(\frac{u}{2},\frac{\tau}{4}\right)\theta\left(\frac{v}{2},\frac{\tau}{4}\right)\\ &\qquad+\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{y+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{u}{2},\frac{\tau}{4}\right)\theta\left(\frac{v}{2},\frac{\tau}{4}\right)\end{aligned}$$

$$+\theta\left(\frac{x}{2},\frac{\tau}{4}\right)\theta\left(\frac{y}{2},\frac{\tau}{4}\right)\theta\left(\frac{u+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{v+1}{2},\frac{\tau}{4}\right)$$
$$+\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{y+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{u+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{v+1}{2},\frac{\tau}{4}\right).$$

Proof Consider the eigenvectors (2.25) at the arguments $x+\frac{\tau}{2}$ and $y+1$, we have

$$v\left(x+\frac{\tau}{2},\tau,0\right)=\begin{bmatrix}2a\theta_{\frac{1}{2},0}(x,\tau)+a\sqrt{2}\theta\left(\frac{x}{2},\frac{\tau}{4}\right)\\ 2a\theta(x,\tau)-a\sqrt{2}\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)\end{bmatrix} \tag{2.52}$$

where $a=\exp\left(\frac{-\pi i\tau}{4}-\pi ix\right)$.

$$v(y+1,\tau,0)=\begin{bmatrix}2\theta(y,\tau)+\sqrt{2}\theta\left(\frac{y+1}{2},\frac{\tau}{4}\right)\\ -2\theta_{\frac{1}{2},0}(y,\tau)+\sqrt{2}\theta\left(\frac{y}{2},\frac{\tau}{4}\right)\end{bmatrix}. \tag{2.53}$$

Since $v(y+1,\tau,0)$, $v\left(x+\frac{\tau}{2},\tau,0\right)$ are eigenvectors corresponding to the same eigenvalue 1, which has multiplicity 1 we have

$$\det\left(v\left(x+\frac{\tau}{2},\tau,0\right),v(y+1,\tau,0)\right)=0.$$

This gives

$$-4\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)+2\sqrt{2}\theta_{\frac{1}{2},0}(x,\tau)\theta\left(\frac{y}{2},\frac{\tau}{4}\right)-2\sqrt{2}\theta_{\frac{1}{2},0}(y,\tau)\theta\left(\frac{x}{2},\frac{\tau}{4}\right)$$
$$+2\theta\left(\frac{x}{2},\frac{\tau}{4}\right)\theta\left(\frac{y}{2},\frac{\tau}{4}\right)-4\theta(x,\tau)\theta(y,\tau)-2\sqrt{2}\theta(x,\tau)\theta\left(\frac{y+1}{2},\frac{\tau}{4}\right)$$
$$+2\sqrt{2}\theta(y,\tau)\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)+2\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{y+1}{2},\frac{\tau}{4}\right)=0. \tag{2.54}$$

Consider in (2.54) the terms with coefficients as $2\sqrt{2}$, using Eqs. (2.29), (2.30) we have,

$$\theta_{\frac{1}{2},0}(x,\tau)\theta\left(\frac{y}{2},\frac{\tau}{4}\right)=\theta_{\frac{1}{2},0}(x,\tau)\theta(y,\tau)$$
$$+\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)=A,$$
$$\theta_{\frac{1}{2},0}(y,\tau)\theta\left(\frac{x}{2},\frac{\tau}{4}\right)=\theta_{\frac{1}{2},0}(y,\tau)\theta(x,\tau)$$
$$+\theta_{\frac{1}{2},0}(y,\tau)\theta_{\frac{1}{2},0}(x,\tau)=B,$$

$$\theta(x,\tau)\theta\left(\frac{y+1}{2},\frac{\tau}{4}\right)=\theta(x,\tau)\theta(y,\tau)$$
$$-\theta(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)=C,$$
$$\theta(y,\tau)\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)=\theta(y,\tau)\theta(x,\tau)$$
$$-\theta(y,\tau)\theta_{\frac{1}{2},0}(x,\tau)=D.$$

Then it is clear that $A-B-C+D=0$. Hence all the terms with coefficients $2\sqrt{2}$ cancel each other out. Equation (2.54) becomes

$$2\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)+2\theta(x,\tau)\theta(y,\tau)=\theta\left(\frac{x}{2},\frac{\tau}{4}\right)\theta\left(\frac{y}{2},\frac{\tau}{4}\right)+\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{y+1}{2},\frac{\tau}{4}\right) \tag{2.55}$$

Similarly by changing the variables x,y to u,v we have

$$2\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)+2\theta(u,\tau)\theta(v,\tau)=\theta\left(\frac{u}{2},\frac{\tau}{4}\right)\theta\left(\frac{v}{2},\frac{\tau}{4}\right)+\theta\left(\frac{u+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{v+1}{2},\frac{\tau}{4}\right) \tag{2.56}$$

Multiplying (2.55), (2.56) gives

$$\begin{aligned}
&4\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)+4\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta(u,\tau)\theta(v,\tau)\\
&\quad+4\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)\theta(x,\tau)\theta(y,\tau)+4\theta(x,\tau)\theta(y,\tau)\theta(u,\tau)\theta(v,\tau)\\
&\quad=\theta\left(\frac{x}{2},\frac{\tau}{4}\right)\theta\left(\frac{y}{2},\frac{\tau}{4}\right)\theta\left(\frac{u}{2},\frac{\tau}{4}\right)\theta\left(\frac{v}{2},\frac{\tau}{4}\right)\\
&\qquad+\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{y+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{u}{2},\frac{\tau}{4}\right)\theta\left(\frac{v}{2},\frac{\tau}{4}\right)\\
&\qquad+\theta\left(\frac{x}{2},\frac{\tau}{4}\right)\theta\left(\frac{y}{2},\frac{\tau}{4}\right)\theta\left(\frac{u+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{v+1}{2},\frac{\tau}{4}\right)\\
&\qquad+\theta\left(\frac{x+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{y+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{u+1}{2},\frac{\tau}{4}\right)\theta\left(\frac{v+1}{2},\frac{\tau}{4}\right)
\end{aligned} \tag{2.57}$$

This proves Riemann's extended identity. □

We now show that the classical Riemann's identity follows from Theorem (2.3.6).

Theorem 2.3.7 (Riemann's Identity)

$$\begin{aligned}
&\theta_{\frac{1}{2},\frac{1}{2}}(x,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(y,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(u,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(v,\tau)\\
&\quad+\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)\\
&\theta_{0,\frac{1}{2}}(u,\tau)\theta_{0,\frac{1}{2}}(v,\tau)\theta_{0,\frac{1}{2}}(x,\tau)\theta_{0,\frac{1}{2}}(y,\tau)\\
&\quad+\theta(x,\tau)\theta(y,\tau)\theta(u,\tau)\theta(v,\tau)\\
&\qquad=2\sum_{m,n,p,q\in\frac{1}{2}Z}\exp\left[\pi i\left(m^2+n^2+p^2+q^2\right)\tau\right.\\
&\qquad\qquad\left.+2\pi i\,(mx+ny+pu+qv)\right].
\end{aligned}\tag{2.58}$$

$m,\ n,\ p,\ q$ are either integers or $m,\ n,\ p,\ q$ are $\in \frac{1}{2}+Z$ and $\sum m = m+n+p+q \in 2Z$.

Proof Consider the right-hand side of (2.57)

$$\begin{aligned}
&=\sum_{m,n,p,q\in Z}\exp\left[\pi i\left(\sum m^2\right)\frac{\tau}{4}+2\pi i\left(\frac{\sum mx}{2}\right)\right]\\
&\quad+\sum_{m,n,p,q\in Z}\exp\left[\pi i(\sum m^2)\frac{\tau}{4}+2\pi i\left(\frac{\sum mx}{2}+\frac{m+n}{2}\right)\right]\\
&\quad+\sum_{m,n,p,q\in Z}\exp\left[\pi i(\sum m^2)\frac{\tau}{4}+2\pi i\left(\frac{\sum mx}{2}+\frac{p+q}{2}\right)\right]\\
&\quad+\sum_{m,n,p,q\in Z}\exp\left[\pi i(\sum m^2)\frac{\tau}{4}+2\pi i\left(\frac{\sum mx}{2}+\frac{\sum m}{2}\right)\right]\\
&=4\sum_{m+n,\,p+q\equiv 0(\text{mod }2)}\exp\left[\pi i\left(\sum m^2\right)\frac{\tau}{4}+2\pi i\left(\frac{\sum mx}{2}\right)\right].
\end{aligned}\tag{2.59}$$

We have from (2.57)

$$\begin{aligned}
&\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)+\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta(u,\tau)\theta(v,\tau)\\
&\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)\theta(x,\tau)\theta(y,\tau)+\theta(x,\tau)\theta(y,\tau)\theta(u,\tau)\theta(v,\tau)\\
&\quad=\sum_{m+n,\,p+q\equiv 0(\text{mod }2)}\exp\left[\pi i\left(\sum m^2\right)\frac{\tau}{4}+2\pi i\left(\frac{\sum mx}{2}\right)\right].
\end{aligned}\tag{2.60}$$

In (2.60) replacing x,y,z,u by $x+\frac{1}{2}$, $y+\frac{1}{2}$, $u+\frac{1}{2}$, $v+\frac{1}{2}$ we get,

$$\begin{aligned}
&\theta_{\frac{1}{2},\frac{1}{2}}(x,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(y,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(u,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(v,\tau)\\
&\quad+\theta_{\frac{1}{2},\frac{1}{2}}(x,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(y,\tau)\theta_{0,\frac{1}{2}}(u,\tau)\theta_{0,\frac{1}{2}}(v,\tau)\\
&\theta_{\frac{1}{2},\frac{1}{2}}(u,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(v,\tau)\theta_{0,\frac{1}{2}}(x,\tau)\theta_{0,\frac{1}{2}}(y,\tau)\\
&\quad+\theta_{0,\frac{1}{2}}(x,\tau)\theta_{0,\frac{1}{2}}(y,\tau)\theta_{0,\frac{1}{2}}(u,\tau)\theta_{0,\frac{1}{2}}(v,\tau)\\
&\qquad=\sum_{m+n,\,p+q\equiv 0 (\mathrm{mod}\ 2)}\exp\left[\pi i\left(\sum m^2\right)\frac{\tau}{4}+2\pi i\left(\frac{\sum m\left(x+\frac{1}{2}\right)}{2}\right)\right].
\end{aligned} \tag{2.61}$$

By adding (2.60), (2.61) we get

$$\begin{aligned}
&\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)+\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta(u,\tau)\theta(v,\tau)\\
&\quad+\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)\theta(x,\tau)\theta(y,\tau)+\theta(x,\tau)\theta(y,\tau)\theta(u,\tau)\theta(v,\tau)\\
&\quad+\theta_{\frac{1}{2},\frac{1}{2}}(x,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(y,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(u,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(v,\tau)\\
&\quad+\theta_{0,\frac{1}{2}}(x,\tau)\theta_{0,\frac{1}{2}}(y,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(u,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(v,\tau)\\
&\quad+\theta_{\frac{1}{2},\frac{1}{2}}(x,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(y,\tau)\theta_{0,\frac{1}{2}}(u,\tau)\theta_{0,\frac{1}{2}}(v,\tau)\\
&\quad+\theta_{0,\frac{1}{2}}(x,\tau)\theta_{0,\frac{1}{2}}(y,\tau)\theta_{0,\frac{1}{2}}(u,\tau)\theta_{0,\frac{1}{2}}(v,\tau)\\
&\qquad=\sum_{m+n,\,p+q\equiv 0 (\mathrm{mod}\ 2)}\exp\left[\pi i\left(\sum m^2\right)\frac{\tau}{4}+2\pi i\left(\frac{\sum mx}{2}\right)\right]\\
&\qquad\quad+\sum_{m+n,\,p+q\equiv 0 (\mathrm{mod}\ 2)}\exp\left[\pi i\left(\sum m^2\right)\frac{\tau}{4}+2\pi i\left(\frac{\sum m\left(x+\frac{1}{2}\right)}{2}\right)\right]
\end{aligned} \tag{2.62}$$

It is clear that (2.62) for $m,\ n,\ p,\ q$ are either integers or $m,\ n,\ p,\ q$ are $\in \frac{1}{2}+Z$ and $\sum m = m+n+p+q \in 2Z$ becomes

$$\begin{aligned}
&\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)+\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta(u,\tau)\theta(v,\tau)\\
&\quad+\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)\theta(x,\tau)\theta(y,\tau)+\theta(x,\tau)\theta(y,\tau)\theta(u,\tau)\theta(v,\tau)\\
&\quad+\theta_{\frac{1}{2},\frac{1}{2}}(x,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(y,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(u,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(v,\tau)\\
&\quad+\theta_{0,\frac{1}{2}}(x,\tau)\theta_{0,\frac{1}{2}}(y,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(u,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(v,\tau)
\end{aligned}$$

$$+\theta_{\frac{1}{2},\frac{1}{2}}(x,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(y,\tau)\theta_{0,\frac{1}{2}}(u,\tau)\theta_{0,\frac{1}{2}}(v,\tau)$$
$$+\theta_{0,\frac{1}{2}}(x,\tau)\theta_{0,\frac{1}{2}}(y,\tau)\theta_{0,\frac{1}{2}}(u,\tau)\theta_{0,\frac{1}{2}}(v,\tau)$$
$$= 2\sum_{m,n,p,q\in\frac{1}{2}Z} \exp\left[\pi i\left(\sum m^2\right)\tau + 2\pi i\left(\sum mx\right)\right]. \tag{2.63}$$

Label the summands on the left-hand side of (2.63) serially as $\sum_{k=1}^{k=8} A_k$. For the notational convenience we denote $m_1 = m + \frac{1}{2}$, $n_1 = n + \frac{1}{2}$, $p_1 = p + \frac{1}{2}$, $q_1 = q + \frac{1}{2}$, then we have

$$A_2 = \theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta(u,\tau)\theta(v,\tau)$$
$$= \sum_{m,n,p,q} \exp\left[\pi i\left(m_1^2+n_1^2+p^2+q^2\right)\tau + 2\pi i\left(m_1x+n_1y+pu+qv\right)\right],$$
$$A_3 = \theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)\theta(x,\tau)\theta(y,\tau)$$
$$= \sum_{m,n,p,q} \exp\left[\pi i\left(m^2+n^2+p_1^2+q_1^2\right)\tau + 2\pi i\left(mx+ny+p_1u+q_1v\right)\right],$$
$$A_6 = \theta_{0,\frac{1}{2}}(x,\tau)\theta_{0,\frac{1}{2}}(y,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(u,\tau)\theta_{\frac{1}{2},frac12}(v,\tau),$$
$$= \sum_{m,n,p,q} \exp\left[\pi i\left(m^2+n^2+p_1^2+q_1^2\right)\tau\right.$$
$$\left. +2\pi i\left(mx+ny+p_1u+q_1v+\frac{\sum m}{2}+\frac{1}{2}\right)\right],$$
$$A_7 = \theta_{\frac{1}{2},\frac{1}{2}}(x,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(y,\tau)\theta_{0,\frac{1}{2}}(u,\tau)\theta_{0,\frac{1}{2}}(v,\tau),$$
$$= \sum_{m,n,p,q} \exp\left[\pi i\left(m_1^2+n_1^2+p^2+q^2\right)\tau\right.$$
$$\left. +2\pi i\left(m_1x+n_1y+pu+qv+\frac{\sum m}{2}+\frac{1}{2}\right)\right], \tag{2.64}$$

Since $\sum_{m,n,p,q} m \in 2Z$, we have $A_2 + A_7 = 0$ and $A_3 + A_6 = 0$. Hence the Riemann's identity follows from (2.63) and (2.64) (see [14]). □

For simplicity if we do a change of the variable as follows:

$$n_1 = \frac{1}{2}(n+m+p+q),\quad x_1 = \frac{1}{2}(x+y+u+v),$$
$$m_1 = \frac{1}{2}(n+m-p-q),\quad y_1 = \frac{1}{2}(x+y-u-v),$$

$$p_1 = \frac{1}{2}(n - m + p - q), \quad u_1 = \frac{1}{2}(x - y + u - v),$$
$$q_1 = \frac{1}{2}(n - m - p + q), \quad v_1 = \frac{1}{2}(x - y - u + v).$$

Then the particular restrictions on parameters n,m,p,q of the summation above exactly means that the resulting n_1, m_1, p_1, q_1 are integers. Also observe that we have the identities: $\sum n^2 = \sum n_1^2$ and $\sum xn = \sum x_1 n_1$. Equation (2.58) becomes

$$\begin{aligned}
&\theta_{\frac{1}{2},\frac{1}{2}}(x,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(y,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(u,\tau)\theta_{\frac{1}{2},\frac{1}{2}}(v,\tau)\\
&\quad+\theta_{\frac{1}{2},0}(x,\tau)\theta_{\frac{1}{2},0}(y,\tau)\theta_{\frac{1}{2},0}(u,\tau)\theta_{\frac{1}{2},0}(v,\tau)\\
&\theta_{0,\frac{1}{2}}(u,\tau)\theta_{0,\frac{1}{2}}(v,\tau)\theta_{0,\frac{1}{2}}(x,\tau)\theta_{0,\frac{1}{2}}(y,\tau)\\
&\quad+\theta(x,\tau)\theta(y,\tau)\theta(u,\tau)\theta(v,\tau)\\
&\quad\quad= 2\theta(x_1,\tau)\theta(y_1,\tau)\theta(u_1,\tau)\theta(v_1,\tau).
\end{aligned}$$

The method presented above indicates that all the identities in [14] derived from the Riemann's identity can be derived from the techniques illustrated above.

2.4 DFT $\Phi(3)$ and Theta Function Identities

This section explores identities of theta functions corresponding to the DFT $\Phi(3)$. Theta functions corresponding to this are $\theta_{a,b}$ for $a,\ b \in \frac{1}{3}\mathbb{Z}$. These functions have not been studied extensively in the literature like classical Jacobi theta functions. This section gives natural extensions of some of the identities obtained in Sect. 2.3. We give an extension of Watson addition formula (2.46) for the DFT $\Phi(3)$. The quadratic identity involving theta functions is given. We give a fourth order extension of Riemann's identity for theta functions $\theta_{a,b}$ for a,$b \in \frac{1}{3}\mathbb{Z}$. This leads to some nontrivial identities of fourth order.

2.4.1 $\theta_{a,b}(x,\tau)$ *with* $a, b \in \frac{1}{3}\mathbb{Z}$ *and* $\Phi(3)$

There are nine theta functions $\theta_{a,b}(x,\tau)$ with $a, b \in \frac{1}{3}\mathbb{Z}$. This can be listed as below

$$\theta_{\frac{1}{3},\frac{1}{3}}(x,\tau) = \sum_{m\in Z} \exp[\pi i\tau(m+1/3)^2 + 2\pi i(m+1/3)(x+1/3)],$$
$$\theta_{\frac{1}{3},0}(x,\tau) = \sum_{m\in Z} \exp[\pi i\tau(m+1/3)^2 + 2\pi i(m+1/3)x],$$

$$\theta_{0,\frac{1}{3}}(x,\tau) = \sum_{m\in Z} \exp[\pi i m^2\tau + 2\pi i m(x+1/3)],$$

$$\theta_{\frac{1}{3},\frac{2}{3}}(x,\tau) = \sum_{m\in Z} \exp[\pi i\tau(m+1/3)^2 + 2\pi i(m+1/3)(x+2/3)],$$

$$\theta_{\frac{2}{3},\frac{1}{3}}(x,\tau) = \sum_{m\in Z} \exp[\pi i\tau(m+2/3)^2 + 2\pi i(m+2/3)(x+1/3)],$$

$$\theta_{\frac{2}{3},\frac{2}{3}}(x,\tau) = \sum_{m\in Z} \exp[\pi i\tau(m+2/3)^2 + 2\pi i(m+2/3)(x+2/3)],$$

$$\theta_{\frac{2}{3},0}(x,\tau) = \sum_{m\in Z} \exp[\pi i\tau(m+2/3)^2 + 2\pi i(m+2/3)x],$$

$$\theta_{0,\frac{2}{3}}(x,\tau) = \sum_{m\in Z} \exp[\pi i\tau m^2 + 2\pi i m(x+2/3)],$$

$$\theta_{0,0} = \sum_{m\in Z} \exp[\pi i\tau m^2 + 2\pi i m^2]$$

The explicit description of zeros of theta functions $\theta_{a,b}$ with $a,b \in \frac{1}{l}\mathbb{Z}$ is given in Mumford [14]. In particular the zeros of $\theta_{a,b}$ with $a,b \in \frac{1}{3}\mathbb{Z}$ are at the points $(a+m+\frac{1}{2})\tau + (b+n+\frac{1}{2})$, for $m,n \in \mathbb{Z}$ and $a,b \in \frac{1}{3}\mathbb{Z}$. The eigenvalues of the DFT $\Phi(3)$ are $+1$, -1 and i with multiplicity one (2.2.2). The eigenvector corresponding to eigenvalue $+1$ is given by using Theorem (2.3.1).

$$v(x,\tau,0) = \begin{bmatrix} 2\theta(x,\tau) + \frac{2}{\sqrt{3}}\theta\left(\frac{x}{3},\frac{\tau}{3^2}\right) \\ \theta_{\frac{1}{3},0}(x,\tau) + \theta_{\frac{2}{3},0}(x,\tau) + \frac{1}{\sqrt{3}}\left[\theta\left(\frac{x+1}{3},\frac{\tau}{3^2}\right) + \theta\left(\frac{x-1}{3},\frac{\tau}{3^2}\right)\right] \\ \theta_{\frac{1}{3},0}(x,\tau) + \theta_{\frac{2}{3},0}(x,\tau) + \frac{1}{\sqrt{3}}\left[\theta\left(\frac{x+2}{3},\frac{\tau}{3^2}\right) + \theta\left(\frac{x-2}{3},\frac{\tau}{3^2}\right)\right] \end{bmatrix}.$$

The eigenvector corresponding to eigenvalue -1 is given by

$$v(x,\tau,2) = \begin{bmatrix} 2\theta(x,\tau) - \frac{2}{\sqrt{3}}\theta\left(\frac{x}{3},\frac{\tau}{3^2}\right) \\ \theta_{\frac{1}{3},0}(x,\tau) + \theta_{\frac{2}{3},0}(x,\tau) - \frac{1}{\sqrt{3}}\left[\theta\left(\frac{x+1}{3},\frac{\tau}{3^2}\right) + \theta\left(\frac{x-1}{3},\frac{\tau}{3^2}\right)\right] \\ \theta_{\frac{1}{3},0}(x,\tau) + \theta_{\frac{2}{3},0}(x,\tau) - \frac{1}{\sqrt{3}}\left[\theta\left(\frac{x+2}{3},\frac{\tau}{3^2}\right) + \theta\left(\frac{x-2}{3},\frac{\tau}{3^2}\right)\right] \end{bmatrix}.$$

Similarly the eigenvector corresponding to eigenvalue i is given by

$$v(x,\tau,1) = \begin{bmatrix} 0 \\ \theta_{\frac{1}{3},0}(x,\tau) - \theta_{\frac{2}{3},0}(x,\tau) - \frac{i}{\sqrt{3}}\left[\theta\left(\frac{x+1}{3},\frac{\tau}{3^2}\right) - \theta\left(\frac{x-1}{3},\frac{\tau}{3^2}\right)\right] \\ \theta_{\frac{2}{3},0}(x,\tau) - \theta_{\frac{1}{3},0}(x,\tau) - \frac{i}{\sqrt{3}}\left[\theta\left[\frac{x+2}{3},\frac{\tau}{3^2}\right) - \theta\left(\frac{x-2}{3},\frac{\tau}{3^2}\right)\right] \end{bmatrix}.$$

We have,

$$\theta(x+1,\tau) = \theta(x,\tau)\ ,\theta_{\frac{1}{3},0}(x+1,\tau) = \omega\theta_{\frac{1}{3},0}(x,\tau)\ ,\theta_{\frac{2}{3},0}(x+1,\tau)$$
$$= \omega^2\theta_{\frac{2}{3},0}(x,\tau),$$

where $\omega = e^{\frac{2\pi i}{3}}$ is the cube root of unity.

$$\Phi(3)\,[v(x,\tau,0)+v(x,\tau,1)+v(x,\tau,2)] = v(x,\tau,0)+iv(x,\tau,1)-v(x,\tau,2).$$

By equating the first component we obtain

$$\theta(x,\tau)+\theta_{\frac{1}{3},0}(x,\tau)+\theta_{\frac{2}{3},0}(x,\tau) = \theta\left(\frac{x}{3},\frac{\tau}{3^2}\right). \tag{2.65}$$

Replacing x by $x+1$ and $x+2$ in (2.65) we get the following identities:

$$\theta(x,\tau)+\omega\theta_{\frac{1}{3},0}(x,\tau)+\omega^2\theta_{\frac{2}{3},0}(x,\tau) = \theta\left(\frac{x+1}{3},\frac{\tau}{3^2}\right). \tag{2.66}$$

$$\theta(x,\tau)+\omega^2\theta_{\frac{1}{3},0}(x,\tau)+\omega\theta_{\frac{2}{3},0}(x,\tau) = \theta\left(\frac{x+2}{3},\frac{\tau}{3^2}\right). \tag{2.67}$$

The above identities are equivalent to

$$\theta(x,\tau) = \theta(3x,9\tau)+\theta_{\frac{1}{3},0}(3x,9\tau)+\theta_{\frac{2}{3},0}(3x,9\tau) \tag{2.68}$$

$$\theta_{0,\frac{1}{3}}(x,\tau) = \theta(3x,9\tau)+\omega\theta_{\frac{1}{3},0}(3x,9\tau)+\omega^2\theta_{\frac{2}{3},0}(3x,9\tau) \tag{2.69}$$

$$\theta_{0,\frac{2}{3}}(x,\tau) = \theta(3x,9\tau)+\omega^2\theta_{\frac{1}{3},0}(3x,9\tau)+\omega\theta_{\frac{2}{3},0}(3x,9\tau). \tag{2.70}$$

Identities (2.65)–(2.70) are extensions of the identities (2.29)–(2.32).

The identities (2.65)–(2.67) show the role that DFT plays in this context. The above identities can be represented as

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{bmatrix} \begin{bmatrix} \theta(x,\tau) \\ \theta_{\frac{1}{3},0}(x,\tau) \\ \theta_{\frac{2}{3},0}(x,\tau) \end{bmatrix} = \begin{bmatrix} \theta\left(\frac{x}{3}\right) \\ \theta\left(\frac{x+1}{3}\right) \\ \theta\left(\frac{x+1}{3}\right) \end{bmatrix}.$$

The above representation shows the general fact stated in the following theorem given in [10, 14]. Theorem 2.4.1 and identity (2.73) have been proved in [10] for generalized ν-theta functions. We reproduce these results for completion corresponding to $\nu = 1$.

Theorem 2.4.1 *The theta functions* $\theta\left(\frac{j+x}{n}, \frac{\tau}{n^2}\right)$ *and* $\theta_{\frac{j}{n},0}(x,\tau)k$, $j = 0, 1, 2, \ldots, n-1$ *are connected by a multiple of the DFT.*

$$\theta_{\frac{j}{n},0}(x,\tau) = \frac{1}{n}\sum_{k=0}^{n-1} q^{-jk}\theta\left(\frac{k+x}{n}, \frac{\tau}{n^2}\right) \tag{2.71}$$

$$\theta\left(\frac{k+x}{n}, \frac{\tau}{n^2}\right) = \sum_{j=0}^{n-1} q^{jk}\theta_{\frac{j}{n},0}(x,\tau)\ . \tag{2.72}$$

In particular, (2.71) says that if

$$v = \left(\theta_{\frac{0}{n},0}(x,\tau), \theta_{\frac{1}{n},0}(x,\tau), \ldots, \theta_{\frac{n-1}{n},0}(x,\tau)\right)^t,$$
$$w = \left(\theta\left(\frac{x}{n}, \frac{\tau}{n^2}\right), \ldots, \theta\left(\frac{x+n-1}{n}, \frac{\tau}{n^2}\right)\right).$$

then $\sqrt{n}\Phi(v) = w$. *Thus* $v = \frac{1}{\sqrt{n}}\Phi^{-1}w$,

Proof

$$\sum_{j=0}^{n-1} q^{jk}\theta_{\frac{j}{n},0}(x,\tau) = \sum_{j=0}^{n-1}\sum_{m\in Z} \exp\left[\frac{\pi i\tau(mn+j)^2}{n^2} + \frac{2\pi i(mn+j)x}{n} + \frac{2\pi ik(mn+j)}{n}\right].$$

Consider $r = mn + j$ then $j = r - mn$ and

$$q^{jk} = q^{rk-mnk} = q^{rk}\,.$$

The double sum in RHS of the above equality can be written as a single sum, by the definition of the theta function,

$$\sum_{r\in Z} \exp\left[\pi i\tau\frac{r^{2\nu}}{n^2} + 2\pi ir\frac{x+k}{n}\right] = \theta\left(\frac{j+x}{n}, \frac{\tau}{n^2}\right)\,.$$

This proves the formula (2.71).

In particular, taking $j = 0$ in (2.71) and $k = 0$ in (2.72) we get the following formulae:

$$\theta(x,\tau) = \frac{1}{n}\sum_{k=0}^{n-1}\theta\left(\frac{k+x}{n},\frac{\tau}{n^2},\right),$$

$$\theta\left(\frac{x}{n},\frac{\tau}{n^2},\right) = \sum_{j=0}^{n-1}\theta_{\frac{j}{n},0}(x,\tau).$$

□

Unitarity of the DFT implies the following relations between theta functions which is Parceval identity,

$$\sum_{k=0}^{k=n-1}\left|\theta\left(\frac{x+k}{n},\frac{\tau}{n^2},\right)\right|^2 = n\sum_{j=0}^{n-1}|\theta_{\frac{j}{n},0}(x,\tau)|^2. \tag{2.73}$$

2.4.2 Extended Watson Addition Formula Corresponding to $\Phi(3)$

Theorem 2.4.2 (Extended Watson Addition Formula Corresponding to $\Phi(3)$)

$$\begin{aligned}
&3\,\theta(3x+3y,9\tau)\theta_{\frac{1}{3},0}(3x-3y,9\tau) + 3\theta(3x+3y,9\tau)\theta_{\frac{2}{3},0}(3x-3y,9\tau)\\
&\quad -3\theta(3x-3y,9\tau)\theta_{\frac{1}{3},0}(3x+3y,9\tau) - 3\theta_{\frac{2}{3},0}(3x+3y,9\tau)\theta(3x-3y,9\tau)\\
&\quad = \theta_{0,\frac{1}{3}}(x+y,\tau)\theta(x-y,\tau) + \theta_{0,\frac{2}{3}}(x+y,\tau)\theta(x-y,\tau)\\
&\qquad -\theta(x+y,\tau)\theta_{0,\frac{1}{3}}(x-y,\tau) - \theta(x+y,\tau)\theta_{0,\frac{2}{3}}(x-y,\tau).
\end{aligned} \tag{2.74}$$

Proof Consider the eigenvectors $v(x+y,3\tau,0)$, $v(x-y,3\tau,0)$.

$$v(x+y,3\tau,0)$$
$$= \begin{bmatrix} 2\theta(x+y,3\tau) + \frac{2}{\sqrt{3}}\theta\left(\frac{x+y}{3},\frac{\tau}{3}\right) \\ \theta_{\frac{1}{3},0}(x+y,3\tau) + \theta_{\frac{2}{3},0}(x+y,3\tau) + \frac{1}{\sqrt{3}}\left[\theta\left(\frac{x+y+1}{3},\frac{\tau}{3}\right) + \theta\left(\frac{x+y-1}{3},\frac{\tau}{3}\right)\right] \\ \theta_{\frac{1}{3},0}(x+y,3\tau) + \theta_{\frac{2}{3},0}(x+y,3\tau) + \frac{1}{\sqrt{3}}\left[\theta\left(\frac{x+y+2}{3},\frac{\tau}{3}\right) + \theta\left(\frac{x+y-2}{3},\frac{\tau}{3}\right)\right] \end{bmatrix},$$

$$v(x-y,3\tau,0)$$
$$=\begin{bmatrix} 2\theta(x-y,3\tau)+\frac{2}{\sqrt{3}}\theta\left(\frac{x-y}{3},\frac{\tau}{3}\right) \\ \theta_{\frac{1}{3},0}(x-y,3\tau)+\theta_{\frac{2}{3},0}(x-y,3\tau)+\frac{1}{\sqrt{3}}\left[\theta\left(\frac{x-y+1}{3},\frac{\tau}{3}\right)+\theta\left(\frac{x-y-1}{3},\frac{\tau}{3}\right)\right] \\ \theta_{\frac{1}{3},0}(x-y,3\tau)+\theta_{\frac{2}{3},0}(x-y,3\tau)+\frac{1}{\sqrt{3}}\left[\theta\left(\frac{x-y+2}{3},\frac{\tau}{3}\right)+\theta\left(\frac{x-y-2}{3},\frac{\tau}{3}\right)\right]\end{bmatrix}.$$

These are eigenvectors corresponding to the same eigenvalue 1 which has multiplicity 1. Therefore any 2×2 minor of $[v(x+y,3\tau,0),\ v(x-y,3\tau,0)]$ is zero. We consider the minor formed by the first two components in the eigenvectors $v(x+y,3\tau,0)$, $v(x-y,3\tau,0)$, i.e.

$$l\begin{vmatrix} v_0(x+y,3\tau,0) & v_0(x-y,3\tau,0) \\ v_1(x+y,3\tau,0) & v_1(x-y,3\tau,0)\end{vmatrix}=0.$$

In the following expansion we have used the temporary shorthand notation:

$$\theta_{\frac{j}{3},0}(x\pm y,3\tau)=\theta_{\frac{j}{3},0}(x\pm y),\ \theta\left(\frac{x\pm y\pm j}{3},\frac{\tau}{3}\right)$$
$$=\theta\left(\frac{x\pm y\pm j}{3}\right)\text{for} j=0,1,2.$$

$$\begin{aligned}
&2\theta(x+y)\theta_{\frac{1}{3},0}(x-y)+2\theta(x+y)\theta_{\frac{2}{3},0}(x-y)\\
&+\frac{2}{\sqrt{3}}\theta(x+y)\theta\left(\frac{x-y+1}{3}\right)+\frac{2}{\sqrt{3}}\theta(x+y)\theta\left(\frac{x-y-1}{3}\right)\\
&+\frac{2}{\sqrt{3}}\theta\left(\frac{x+y}{3}\right)\theta_{\frac{1}{3},0}(x-y)+\frac{2}{\sqrt{3}}\theta\left(\frac{x+y}{3}\right)\theta_{\frac{2}{3},0}(x-y)\\
&+\frac{2}{3}\theta\left(\frac{x+y}{3}\right)\theta\left(\frac{x-y+1}{3}\right)+\frac{2}{3}\theta\left(\frac{x+y}{3}\right)\theta\left(\frac{x-y-1}{3}\right)\\
&-2\theta(x-y)\theta_{\frac{1}{3},0}(x+y)-2\theta(x-y)\theta_{\frac{2}{3},0}(x+y)\\
&-\frac{2}{\sqrt{3}}\theta(x-y)\theta\left(\frac{x+y+1}{3}\right)-\frac{2}{\sqrt{3}}\theta(x-y)\theta\left(\frac{x+y-1}{3}\right)\\
&-\frac{2}{\sqrt{3}}\theta\left(\frac{x-y}{3}\right)\theta_{\frac{1}{3},0}(x+y)-\frac{2}{\sqrt{3}}\theta\left(\frac{x-y}{3}\right)\theta_{\frac{2}{3},0}(x+y)\\
&-\frac{2}{3}\theta\left(\frac{x-y}{3}\right)\theta\left(\frac{x+y+1}{3}\right)-\frac{2}{3}\theta\left(\frac{x-y}{3}\right)\theta\left(\frac{x+y-1}{3}\right)=0.
\end{aligned}\tag{2.75}$$

Label the summands on the left-hand side of (2.75) successively as A_1, A_2, A_3, ..., A_{16}. Then $\sum_{k=1}^{16} A_k = 0$. Consider the terms with coefficients $\frac{2}{\sqrt{3}}$. We use formulas (2.65)–(2.67) in the following formula:

$$
\begin{aligned}
A_3 &= \theta(x+y)\theta\left(\frac{x-y+1}{3}\right)\\
&= \theta(x+y)\left(\theta(x-y) + \omega\theta_{\frac{1}{3},0}\theta(x-y) + \omega^2\theta_{\frac{2}{3},0}(x-y)\right)\\
&= \theta(x+y)\theta(x-y) + \omega\theta(x+y)\theta_{\frac{1}{3},0}(x-y) + \omega^2\theta(x+y)\theta_{\frac{2}{3},0}(x-y),\\
A_4 &= \theta(x+y)\theta\left(\frac{x-y-1}{3}\right)\\
&= \theta(x+y)\left(\theta(x-y) + \omega^2\theta_{\frac{1}{3},0}\theta(x-y) + \omega\theta_{\frac{2}{3},0}(x-y)\right)\\
&= \theta(x+y)\theta(x-y) + \omega^2\theta(x+y)\theta_{\frac{1}{3},0}(x-y) + \omega\theta(x+y)\theta_{\frac{2}{3},0}(x-y),\\
A_5 &= \theta_{\frac{1}{3},0}(x-y)\theta\left(\frac{x+y}{3}\right)\\
&= \theta_{\frac{1}{3},0}(x-y)\theta(x+y) + \theta_{\frac{1}{3},0}(x-y)\theta_{\frac{1}{3},0}(x+y) + \theta_{\frac{1}{3},0}(x-y)\theta_{\frac{2}{3},0}(x+y),\\
A_6 &= \theta_{\frac{2}{3},0}(x-y)\theta\left(\frac{x+y}{3}\right)\\
&= \theta_{\frac{2}{3},0}(x-y)\theta(x+y) + \theta_{\frac{2}{3},0}(x-y)\theta_{\frac{1}{3},0}(x+y) + \theta_{\frac{2}{3},0}(x-y)\theta_{\frac{2}{3},0}(x+y),\\
-A_{11} &= -\theta(x-y)\theta\left(\frac{x+y+1}{3}\right)\\
&= -\theta(x-y)\theta(x+y) - \omega\theta(x-y)\theta_{\frac{1}{3},0}(x+y) - \omega^2\theta(x-y)\theta_{\frac{2}{3},0}(x+y),\\
-A_{12} &= \theta(x-y)\theta\left(\frac{x+y-1}{3}\right)\\
&= -\theta(x-y)\theta(x+y) - \omega^2\theta(x-y)\theta_{\frac{1}{3},0}(x+y) - \omega\theta(x-y)\theta_{\frac{2}{3},0}(x+y),\\
-A_{13} &= -\theta_{\frac{1}{3},0}(x+y)\theta\left(\frac{x-y}{3}\right)\\
&= -\theta_{\frac{1}{3},0}(x+y)\theta(x-y) - \theta_{\frac{1}{3},0}(x+y)\theta_{\frac{1}{3},0}(x-y)\\
&\quad -\theta_{\frac{1}{3},0}(x+y)\theta_{\frac{2}{3},0}(x-y)
\end{aligned}
$$

$$
\begin{aligned}
-A_{14} &= -\theta_{\frac{2}{3},0}(x+y)\theta\left(\frac{x-y}{3}\right)\\
&= -\theta_{\frac{2}{3},0}(x+y)\theta(x-y) - \theta_{\frac{2}{3},0}(x+y)\theta_{\frac{1}{3},0}(x-y)\\
&\quad -\theta_{\frac{2}{3},0}(x+y)\theta_{\frac{2}{3},0}(x-y)
\end{aligned}
$$

We have $A_3+A_4+A_5+A_6+A_{11}+A_{12}+A_{13}+A_{14}=0$. Equation (2.75) becomes

$$
\begin{aligned}
&3\theta(x+y)\theta_{\frac{1}{3},0}(x-y) + 3\theta(x+y)\theta_{\frac{2}{3},0}(x-y)\\
&\quad -3\theta(x-y)\theta_{\frac{1}{3},0}(x+y) - 3\theta(x-y)\theta_{\frac{2}{3},0}(x+y)\\
&\qquad = \theta\left(\frac{x-y}{3}\right)\theta\left(\frac{x+y+1}{3}\right) + \theta\left(\frac{x-y}{3}\right)\theta\left(\frac{x+y+2}{3}\right)\\
&\qquad\quad -\theta\left(\frac{x+y}{3}\right)\theta\left(\frac{x-y+1}{3}\right) - \theta\left(\frac{x+y}{3}\right)\theta\left(\frac{x-y+2}{3}\right).
\end{aligned}
$$

replacing x, y, τ by $3x, 3y, 3\tau$ we get the required formula. □

Corollary 2.4.3

$$
\begin{aligned}
&3\,\theta(3x, 9\tau)\theta_{\frac{1}{3},0}(0, 9\tau) + 3\theta(3x, 9\tau)\theta_{\frac{2}{3},0}(0, 9\tau)\\
&\quad -3\theta(0, 9\tau)\theta_{\frac{1}{3},0}(3x, 9\tau) - 3\theta_{\frac{2}{3},0}(3x, 9\tau)\theta(0, 9\tau)\\
&\qquad = \theta_{0,\frac{1}{3}}(x, \tau)\theta(0, \tau) + \theta_{0,\frac{2}{3}}(x, \tau)\theta(0, \tau)\\
&\qquad\quad -\theta(x, \tau)\theta_{0,\frac{1}{3}}(0, \tau) - \theta(x, \tau)\theta_{0,\frac{2}{3}}(0, \tau).
\end{aligned}
$$

Proof Put $y = x$ in Theorem 2.4.2 and replace x by $\frac{x}{2}$ to get (2.4.3). □

Theorem 2.4.4

$$
\begin{aligned}
&3\theta(x, \tau)\theta(y, \tau) + 3\theta(x, \tau)\theta_{\frac{1}{3},0}(y, \tau) - 3\theta_{\frac{1}{3},0}(x, \tau)\theta_{\frac{2}{3},0}(y, \tau)\\
&\quad -3\theta_{\frac{2}{3},0}(x, \tau)\theta_{\frac{2}{3},0}(y, \tau)\\
&\qquad = \theta\left(\frac{x+1}{3}, \frac{\tau}{9}\right)\theta\left(\frac{y}{3}, \frac{\tau}{9}\right) + \theta\left(\frac{x-1}{3}, \frac{\tau}{9}\right)\theta\left(\frac{y}{3}, \frac{\tau}{9}\right)\\
&\qquad\quad -\omega\theta\left(\frac{x}{3}, \frac{\tau}{9}\right)\theta\left(\frac{y+1}{3}, \frac{\tau}{9}\right) - \omega^2\theta\left(\frac{x}{3}, \frac{\tau}{9}\right)\theta\left(\frac{y-1}{3}, \frac{\tau}{9}\right)
\end{aligned}
$$

Proof The transformation $y \to y + \frac{2\tau}{3}$ transforms $\theta_{\frac{1}{3},0}(y,\tau)$ as follows:

(a)

$$\begin{aligned}
&\theta_{\frac{1}{3},0}\left(y+\frac{2\tau}{3},\tau\right)\\
&=\sum \exp\left[\pi i\left(m+\frac{1}{3}\right)^2\tau + 2\pi i\left(m+\frac{1}{3}\right)\left(y+\frac{2\tau}{3}\right)\right]\\
&=\sum \exp\left[\pi i\left(m^2+\frac{2m}{3}+\frac{1}{9}\right)\tau + 2\pi imy + \frac{4\pi im\tau}{3}+\frac{2\pi iy}{3}+\frac{4\pi i\tau}{9}\right]\\
&=\exp\left[\frac{-4\pi iy}{3}-\frac{4\pi i\tau}{9}\right]\sum\exp\left[\pi i(m+1)^2\tau+2\pi i(m+1)y\right]\\
&=cd\ \ \theta(y,\tau).
\end{aligned}$$

where $c=\exp\left(\frac{-4\pi iy}{3}\right)$, $d=\exp\left(\frac{-4\pi i\tau}{9}\right)$.

(b)

$$\begin{aligned}
&\theta_{\frac{2}{3},0}\left(y+\frac{2\tau}{3},\tau\right)\\
&=\sum \exp\left[\pi i\left(m+\frac{2}{3}\right)^2\tau + 2\pi i\left(m+\frac{2}{3}\right)\left(y+\frac{2\tau}{3}\right)\right]\\
&=\sum \exp\left[\pi i\left(m^2+\frac{4m}{3}+\frac{4}{9}\right)\tau + 2\pi imy + \frac{4\pi im\tau}{3}+\frac{4\pi iy}{3}+\frac{8\pi i\tau}{9}\right]\\
&=\sum \exp\left[\pi i\left(m^2+\frac{8m}{3}+\frac{12}{9}\right)\tau + 2\pi i\left(m+\frac{4}{3}\right)y-\frac{4\pi iy}{3}-\frac{4\pi i\tau}{9}\right]\\
&=cd\ \ \theta_{\frac{1}{3},0}(y,\tau).
\end{aligned}$$

(c)

$$\begin{aligned}
&\theta\left(\frac{y+1+\frac{2\tau}{3}}{3},\frac{\tau}{9}\right)\\
&=\sum\exp\left(\frac{\pi im^2\tau}{9}+2\pi im\left(\frac{y+1+\frac{2\tau}{3}}{3}\right)\right).\\
&=\sum\exp\left(\frac{\pi im^2\tau}{9}+\frac{2\pi imy}{3}+\frac{2\pi im}{3}+\frac{4\pi im\tau}{9}\right).\\
&=\sum\exp\left(\frac{\pi i\tau(m^2+4m+4)}{9}+\frac{2\pi i(m+2)(y+1)}{3}\right)\\
&\quad\times\exp\left(\frac{-4\pi i\tau}{9}+\frac{-4\pi iy}{3}+\frac{-4\pi i}{3}\right).\\
&=cd\omega\theta\left(\frac{y+1}{3},\frac{\tau}{9}\right).
\end{aligned}$$

Similarly,

(d)

$$\begin{aligned}
&\theta\left(\frac{y-1+\frac{2\tau}{3}}{3}, \frac{\tau}{9}\right)\\
&= \sum \exp\left(\frac{\pi i m^2\tau}{9} + 2\pi i m\left(\frac{y-1+\frac{2\tau}{3}}{3}\right)\right).\\
&= \sum \exp\left(\frac{\pi i m^2\tau}{9} + \frac{2\pi i m y}{3} - \frac{2\pi i m}{3} + \frac{4\pi i m\tau}{9}\right).\\
&= \sum \exp\left(\frac{\pi i\tau(m^2+4m+4)}{9} + \frac{2\pi i(m+2)(y-1)}{3}\right)\\
&\quad \times \exp\left(\frac{-4\pi i\tau}{9} + \frac{-4\pi i y}{3} + \frac{4\pi i}{3}\right).\\
&= cd\omega^2\theta\left(\frac{y-1}{3}, \frac{\tau}{9}\right).
\end{aligned}$$

(e)

$$\begin{aligned}
&\theta\left(y + \frac{2\tau}{3}, \tau\right)\\
&= \sum \exp\left(\pi i m^2\tau + 2\pi i m\left(y + \frac{2\tau}{3}\right)\right)\\
&= cd\sum \exp\left(\pi i\tau\left(m^2 + \frac{4m}{3} + \frac{4}{9}\right) + 2\pi i\left(m + \frac{2}{3}\right)\right)y\\
&= cd\theta_{\frac{2}{3},0}(y,\tau).
\end{aligned}$$

The eigenvector corresponding to eigenvalue $+1$ is given by

$$v(x,\tau,0) = \begin{bmatrix} 2\theta(x,\tau) + \frac{2}{\sqrt{3}}\theta\left(\frac{x}{3}, \frac{\tau}{3^2}\right) \\ \theta_{\frac{1}{3},0}(x,\tau) + \theta_{\frac{2}{3},0}(x,\tau) + \frac{1}{\sqrt{3}}\left[\theta\left(\frac{x+1}{3}, \frac{\tau}{3^2}\right) + \theta\left(\frac{x-1}{3}, \frac{\tau}{3^2}\right)\right] \\ \theta_{\frac{1}{3},0}(x,\tau) + \theta_{\frac{2}{3},0}(x,\tau) + \frac{1}{\sqrt{3}}\left[\theta\left(\frac{x+2}{3}, \frac{\tau}{3^2}\right) + \theta\left(\frac{x-2}{3}, \frac{\tau}{3^2}\right)\right] \end{bmatrix}.$$

Now using the transformation $x \to y + \frac{2\tau}{3}$ and using the identities (a)–(e) we get

$$\begin{aligned}
&v(y + \frac{2\tau}{3}, \tau, 0)\\
&= cd\begin{bmatrix} 2\theta_{\frac{2}{3},0}(y,\tau) + \frac{2}{\sqrt{3}}\theta\left(\frac{y}{3}, \frac{\tau}{3^2}\right) \\ \theta(y,\tau) + \theta_{\frac{1}{3},0}(y,\tau) + \frac{1}{\sqrt{3}}\left[\omega\theta\left(\frac{y+1}{3}, \frac{\tau}{3^2}\right) + \omega^2\theta\left(\frac{y-1}{3}, \frac{\tau}{3^2}\right)\right] \\ \theta(y,\tau) + \theta_{\frac{1}{3},0}(y,\tau) + \frac{1}{\sqrt{3}}\left[\omega\theta\left(\frac{y+1}{3}, \frac{\tau}{3^2}\right) + \omega^2\theta\left(\frac{y-1}{3}, \frac{\tau}{3^2}\right)\right] \end{bmatrix}.
\end{aligned}$$

where $c = \exp\left(\frac{-4\pi x}{3}\right)$, $d = \exp\left(\frac{-4\pi\tau}{9}\right)$. We have $v(x,\tau,0)$ and $v\left(y+\frac{2\tau}{3},\tau,0\right)$ are eigenvectors corresponding to eigenvalue $+1$ which has multiplicity one. Therefore any minor of 2×2 order vanishes. We consider the minor formed by the first two components of the eigenvectors $v(x,\tau,0)$,$v\left(y+\frac{2\tau}{3},\tau,0\right)$. This gives

$$
\begin{aligned}
&2\theta(x,\tau)\theta(y,\tau)+2\theta(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)+\frac{2\omega}{\sqrt{3}}\theta(x,\tau)\theta\left(\frac{y+1}{3},\frac{\tau}{9}\right)\\
&\quad+\frac{2\omega^2}{\sqrt{3}}\theta(x,\tau)\theta\left(\frac{y-1}{3},\frac{\tau}{9}\right)+\frac{2}{\sqrt{3}}\theta\left(\frac{x}{3},\frac{\tau}{9}\right)\theta(y,\tau)+\frac{2}{\sqrt{3}}\theta\left(\frac{x}{3},\frac{\tau}{9}\right)\theta_{\frac{1}{3},0}(y,\tau)\\
&\quad+\frac{2\omega}{3}\theta\left(\frac{x}{3},\frac{\tau}{9}\right)\theta\left(\frac{y+1}{3},\frac{\tau}{9}\right)+\frac{2\omega^2}{3}\theta\left(\frac{x}{3},\frac{\tau}{9}\right)\theta\left(\frac{y-1}{3},\frac{\tau}{9}\right)\\
&\quad-2\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)-\frac{2}{\sqrt{3}}\theta_{\frac{1}{3},0}(x,\tau)\theta\left(\frac{y}{3},\frac{\tau}{9}\right)-2\theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)\\
&\quad-\frac{2}{\sqrt{3}}\theta_{\frac{2}{3},0}(x,\tau)\theta\left(\frac{y}{3},\frac{\tau}{9}\right)-\frac{2}{\sqrt{3}}\theta_{\frac{2}{3},0}(y,\tau)\theta\left(\frac{x+1}{3},\frac{\tau}{9}\right)\\
&\quad-\frac{2}{3}\theta\left(\frac{x+1}{3},\frac{\tau}{9}\right)\theta\left(\frac{y}{3},\frac{\tau}{9}\right)\\
&\quad-\frac{2}{\sqrt{3}}\theta_{\frac{2}{3},0}(x,\tau)\theta\left(\frac{x-1}{3},\frac{\tau}{9}\right)-\frac{2}{3}\theta\left(\frac{x-1}{3},\frac{\tau}{9}\right)\theta\left(\frac{y}{3},\frac{\tau}{9}\right)=0
\end{aligned}
$$

We consider all the terms which are coefficients of $\frac{2}{\sqrt{3}}$ in (2.76).

$$
\begin{aligned}
\omega\theta(x,\tau)\theta\left(\frac{y+1}{3},\frac{\tau}{9}\right)&=\omega\theta(x,\tau)\theta(y,\tau)+\omega^2\theta(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)\\
&\quad+\theta(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)=A,\\
\omega^2\theta(x,\tau)\theta\left(\frac{y+2}{3},\frac{\tau}{9}\right)&=\omega^2\theta(x,\tau)\theta(y,\tau)+\omega\theta(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)\\
&\quad+\theta(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)=B,\\
\theta\left(\frac{x}{3},\frac{\tau}{9}\right)\theta(y,\tau)&=\theta(x,\tau)\theta(y,\tau)+\theta_{\frac{1}{3},0}(x,\tau)\theta(y,\tau)\\
&\quad+\theta_{\frac{2}{3},0}(x,\tau)\theta(y,\tau)=C,\\
\theta\left(\frac{x}{3},\frac{\tau}{9}\right)\theta_{\frac{1}{3},0}(y,\tau)&=\theta(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)+\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)\\
&\quad+\theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)=D.\\
\theta_{\frac{1}{3},0}(x,\tau)\theta\left(\frac{y}{3},\frac{\tau}{9}\right)&=\theta_{\frac{1}{3},0}(x,\tau)\theta(y,\tau)+\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)
\end{aligned}
$$

$$+\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau) = E,$$

$$\theta_{\frac{2}{3},0}(x,\tau)\theta\left(\frac{y}{3},\frac{\tau}{9}\right) = \theta_{\frac{2}{3},0}(x,\tau)\theta(y,\tau) + \theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)$$
$$+\theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau) = F$$

$$\theta\left(\frac{x+1}{3},\frac{\tau}{9}\right)\theta_{\frac{2}{3},0}(y,\tau) = \theta(x,\tau)\theta_{\frac{2}{3},0}(y,\tau) + \omega\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)$$
$$+\omega^2\theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau) = G$$

$$\theta\left(\frac{x-1}{3},\frac{\tau}{9}\right)\theta_{\frac{2}{3},0}(y,\tau) = \theta(x,\tau)\theta_{\frac{2}{3},0}(y,\tau) + \omega^2\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)$$
$$+\omega\theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau) = H$$

Then it is clear that $A+B+C+D-E-F-G-H=0$. Hence all the terms with coefficients of $\frac{2}{\sqrt{3}}$ cancel each other out. Thus we obtain the required identity. □

2.4.3 *Extended Riemann's Identity Corresponding to* $\Phi(3)$

In the following identity we give a fourth order extension of extended Riemann's identity for $\theta_{a,b}$ with $a,b \in \frac{1}{3}\mathbb{Z}$.

Theorem 2.4.5

$$9\theta(x,\tau)\theta(y,\tau)\theta(u,\tau)\theta(v,\tau) + 9\theta(x,\tau)\theta(y,\tau)\theta(u,\tau)\theta_{\frac{1}{3},0}(v,\tau)$$
$$-9\theta(x,\tau)\theta(y,\tau)\theta_{\frac{1}{3},0}(u,\tau)\theta_{\frac{2}{3},0}(v,\tau) - 9\theta(x,\tau)\theta(y,\tau)\theta_{\frac{2}{3},0}(u,\tau)\theta_{\frac{2}{3},0}(v,\tau)$$
$$+9\theta(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)\theta(u,\tau)\theta(v,\tau) + 9\theta(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)\theta(u,\tau)\theta_{\frac{1}{3},0}(v,\tau)$$
$$-9\theta(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)\theta_{\frac{1}{3},0}(u,\tau)\theta_{\frac{2}{3},0}(v,\tau)$$
$$-9\theta(x,\tau)\theta_{\frac{1}{3},0}(y,\tau)\theta_{\frac{2}{3},0}(u,\tau)\theta_{\frac{2}{3},0}(v,\tau)$$
$$-9\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)\theta(u,\tau)\theta(v,\tau) - 9\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)\theta(u,\tau)\theta_{\frac{1}{3},0}(v,\tau)$$
$$+9\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)\theta_{\frac{1}{3},0}(u,\tau)\theta_{\frac{2}{3},0}(v,\tau)$$
$$+9\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)\theta_{\frac{2}{3},0}(u,\tau)\theta_{\frac{2}{3},0}(v,\tau)$$
$$-9\theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)\theta(u,\tau)\theta(v,\tau) - 9\theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)\theta(u,\tau)\theta_{\frac{1}{3},0}(v,\tau)$$
$$+9\theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)\theta_{\frac{1}{3},0}(u,\tau)\theta_{\frac{2}{3},0}(v,\tau)$$
$$+9\theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)\theta_{\frac{2}{3},0}(u,\tau)\theta_{\frac{2}{3},0}(v,\tau)$$

$$
\begin{aligned}
&= \\
&\theta\left(\frac{x+1}{3}\right)\theta\left(\frac{y}{3}\right)\theta\left(\frac{u+1}{3}\right)\theta\left(\frac{v}{3}\right)+\theta\left(\frac{x+1}{3}\right)\theta\left(\frac{y}{3}\right)\theta\left(\frac{u-1}{3}\right)\theta\left(\frac{v}{3}\right) \\
&-\omega\theta\left(\frac{x+1}{3}\right)\theta\left(\frac{y}{3}\right)\theta\left(\frac{u}{3}\right)\theta\left(\frac{v+1}{3}\right) \\
&-\omega^2\theta\left(\frac{x+1}{3}\right)\theta\left(\frac{y}{3}\right)\theta\left(\frac{u}{3}\right)\theta\left(\frac{v-1}{3}\right) \\
&+\theta\left(\frac{x-1}{3}\right)\theta\left(\frac{y}{3}\right)\theta\left(\frac{u+1}{3}\right)\theta\left(\frac{v}{3}\right)+\theta\left(\frac{x-1}{3}\right)\theta\left(\frac{y}{3}\right)\theta\left(\frac{u-1}{3}\right)\theta\left(\frac{v}{3}\right) \\
&-\omega\theta\left(\frac{x-1}{3}\right)\theta\left(\frac{y}{3}\right)\theta\left(\frac{u}{3}\right)\theta\left(\frac{v+1}{3}\right) \\
&-\omega^2\theta\left(\frac{x-1}{3}\right)\theta\left(\frac{y}{3}\right)\theta\left(\frac{u}{3}\right)\theta\left(\frac{v-1}{3}\right) \\
&-\omega\theta\left(\frac{x}{3}\right)\theta\left(\frac{y+1}{3}\right)\theta\left(\frac{u+1}{3}\right)\theta\left(\frac{v}{3}\right) \\
&-\omega\theta\left(\frac{x}{3}\right)\theta\left(\frac{y+1}{3}\right)\theta\left(\frac{u-1}{3}\right)\theta\left(\frac{v}{3}\right) \\
&+\omega^2\theta\left(\frac{x}{3}\right)\theta\left(\frac{y+1}{3}\right)\theta\left(\frac{u}{3}\right)\theta\left(\frac{v+1}{3}\right) \\
&+\theta\left(\frac{x}{3}\right)\theta\left(\frac{y+1}{3}\right)\theta\left(\frac{u}{3}\right)\theta\left(\frac{v-1}{3}\right) \\
&-\omega^2\theta\left(\frac{x}{3}\right)\theta\left(\frac{y-1}{3}\right)\theta\left(\frac{u+1}{3}\right)\theta\left(\frac{v}{3}\right) \\
&-\omega^2\theta\left(\frac{x}{3}\right)\theta\left(\frac{y-1}{3}\right)\theta\left(\frac{u-1}{3}\right)\theta\left(\frac{v}{3}\right) \\
&+\theta\left(\frac{x}{3}\right)\theta\left(\frac{y-1}{3}\right)\theta\left(\frac{u}{3}\right)\theta\left(\frac{v+1}{3}\right) \\
&+\omega\theta\left(\frac{x}{3}\right)\theta\left(\frac{y-1}{3}\right)\theta\left(\frac{u}{3}\right)\theta\left(\frac{v-1}{3}\right). \qquad (2.76)
\end{aligned}
$$

where all theta functions on the right-hand side of (2.76) are at the argument $\frac{\tau}{9}$.

Proof Theorem (2.4.4) we have

$$
\begin{aligned}
&3\theta(x,\tau)\theta(y,\tau)+3\theta(x,\tau)\theta_{\frac{1}{3},0}(y,\tau) \\
&\quad -3\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)-3\theta_{\frac{2}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(y,\tau)
\end{aligned}
$$

$$= \theta\left(\frac{x+1}{3}, \frac{\tau}{9}\right)\theta\left(\frac{y}{3}, \frac{\tau}{9}\right) + \theta\left(\frac{x-1}{3}, \frac{\tau}{9}\right)\theta\left(\frac{y}{3}, \frac{\tau}{9}\right)$$
$$-\omega\theta\left(\frac{x}{3}, \frac{\tau}{9}\right)\theta\left(\frac{y+1}{3}, \frac{\tau}{9}\right) - \omega^2\theta\left(\frac{x}{3}, \frac{\tau}{9}\right)\theta\left(\frac{y-1}{3}, \frac{\tau}{9}\right) \quad (2.77)$$

replace variables x, y by u, v, respectively, to get

$$3\theta(u,\tau)\theta(v,\tau) + 3\theta(u,\tau)\theta_{\frac{1}{3},0}(v,\tau) - 3\theta_{\frac{1}{3},0}(u,\tau)\theta_{\frac{2}{3},0}(v,\tau)$$
$$-3\theta_{\frac{2}{3},0}(u,\tau)\theta_{\frac{2}{3},0}(v,\tau)$$
$$= \theta\left(\frac{u+1}{3}, \frac{\tau}{9}\right)\theta\left(\frac{v}{3}, \frac{\tau}{9}\right) + \theta\left(\frac{u-1}{3}, \frac{\tau}{9}\right)\theta\left(\frac{v}{3}, \frac{\tau}{9}\right)$$
$$-\omega\theta\left(\frac{u}{3}, \frac{\tau}{9}\right)\theta\left(\frac{v+1}{3}, \frac{\tau}{9}\right) - \omega^2\theta\left(\frac{u}{3}, \frac{\tau}{9}\right)\theta\left(\frac{v-1}{3}, \frac{\tau}{9}\right) \quad (2.78)$$

Multiplying (2.77) and (2.78) we get the required identity. □

Corollary 2.4.6

$$3\theta^4(x,\tau) + 6\theta^3(x,\tau)\theta_{\frac{1}{3},0}(x,\tau) - 6\theta^2(x,\tau)\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(x,\tau)$$
$$-6\theta^2(x,\tau)\theta^2_{\frac{2}{3},0}(x,\tau) + 3\theta^2(x,\tau)\theta^2_{\frac{1}{3},0}(x,\tau) - 6\theta^2_{\frac{1}{3},0}(x,\tau)\theta(x,\tau)\theta_{\frac{2}{3},0}(x,\tau)$$
$$-6\theta(x,\tau)\theta_{\frac{1}{3},0}(x,\tau)\theta^2_{\frac{2}{3},0}(x,\tau) + 3\theta^2_{\frac{1}{3},0}(x,\tau)\theta^2_{\frac{2}{3},0}(x,\tau) + 6\theta_{\frac{1}{3},0}(x,\tau)\theta^3_{\frac{2}{3},0}(x,\tau)$$
$$+3\theta^4_{\frac{2}{3},0}(x,\tau)$$
$$= 2\theta^2\left(\frac{x}{3}, \frac{\tau}{9}\right)\theta\left(\frac{x+1}{3}, \frac{\tau}{9}\right)\theta\left(\frac{x+2}{3}, \frac{\tau}{9}\right)$$
$$-\omega^2\theta^2\left(\frac{x}{3}, \frac{\tau}{9}\right)\theta^2\left(\frac{x-1}{3}, \frac{\tau}{9}\right)$$
$$-\omega\theta^2\left(\frac{x}{3}, \frac{\tau}{9}\right)\theta^2\left(\frac{x+1}{3}, \frac{\tau}{9}\right)$$

Proof The proof follows by substituting $y = u = v = x$ in (2.76)

$$9\theta^4(x,\tau) + 9\theta^3(x,\tau)\theta_{\frac{1}{3},0}(x,\tau) - 9\theta^2(x,\tau)\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(x,\tau)$$
$$-9\theta^2(x,\tau)\theta^2_{\frac{2}{3},0}(x,\tau) + 9\theta^3(x,\tau)\theta_{\frac{1}{3},0}(x,\tau) + 9\theta^2(x,\tau)\theta^2_{\frac{1}{3},0}(x,\tau)$$
$$-9\theta(x,\tau)\theta_{\frac{2}{3},0}(x,\tau)\theta^2_{\frac{1}{3},0}(x,\tau) - 9\theta(x,\tau)\theta_{\frac{1}{3},0}(x,\tau)\theta^2_{\frac{2}{3},0}(x,\tau)$$

$$
\begin{aligned}
&-9\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(x,\tau)\theta^2(x,\tau)-9\theta^2_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(x,\tau)\theta(x,\tau)\\
&+9\theta^2_{\frac{1}{3},0}(x,\tau)\theta^2_{\frac{2}{3},0}(x,\tau)+9\theta_{\frac{1}{3},0}(x,\tau)\theta^3_{\frac{2}{3},0}(x,\tau)\\
&-9\theta^2_{\frac{2}{3},0}(x,\tau)\theta^2(x,\tau)-9\theta^2_{\frac{2}{3},0}(x,\tau)\theta(x,\tau)\theta_{\frac{1}{3},0}(x,\tau)\\
&+9\theta^3_{\frac{2}{3},0}(x,\tau)\theta_{\frac{1}{3},0}(x,\tau)+9\theta^4_{\frac{2}{3},0}(x,\tau)\\
&\quad=\theta^2\left(\frac{x+1}{3},\frac{\tau}{9}\right)\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\\
&\qquad+\theta\left(\frac{x+1}{3},\frac{\tau}{9}\right)\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+2}{3},\frac{\tau}{9}\right)\\
&\qquad-\omega\theta^2\left(\frac{x+1}{3},\frac{\tau}{9}\right)\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\\
&\qquad-\omega^2\theta\left(\frac{x+1}{3},\frac{\tau}{9}\right)\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+2}{3},\frac{\tau}{9}\right)\\
&\qquad+\theta\left(\frac{x+1}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+2}{3},\frac{\tau}{9}\right)\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)+\theta^2\left(\frac{x+2}{3}\right)\theta^2\left(\frac{x}{3}\right)\\
&\qquad-\omega\theta\left(\frac{x+2}{3}\right)\theta\left(\frac{x+1}{3}\right)\theta^2\left(\frac{x}{3}\right)-\omega^2\theta^2\left(\frac{x+2}{3},\frac{\tau}{9}\right)\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\\
&\qquad-\omega^2\theta^2\left(\frac{x}{3}\right)\theta^2\left(\frac{x+2}{3}\right)-\omega\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\theta^2\left(\frac{x+1}{3},\frac{\tau}{9}\right)\\
&\qquad-\omega\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+1}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+2}{3},\frac{\tau}{9}\right)\\
&\qquad+\omega^2\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\theta^2\left(\frac{x+1}{3},\frac{\tau}{9}\right)\\
&\qquad\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+1}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+2}{3},\frac{\tau}{9}\right)\\
&\qquad-\omega^2\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+2}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+1}{3},\frac{\tau}{9}\right)\\
&\qquad+\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+1}{3},\frac{\tau}{9}\right)\theta\left(\frac{x+2}{3},\frac{\tau}{9}\right)\\
&\qquad+\omega\theta^2\left(\frac{x}{3},\frac{\tau}{9}\right)\theta^2\left(\frac{x+2}{3},\frac{\tau}{9}\right).
\end{aligned}
\tag{2.79}
$$

Collecting the similar terms together in (2.4.3) we get

$$\begin{aligned}
&9\theta^4(x,\tau)+18\theta^3(x,\tau)\theta_{\frac{1}{3},0}(x,\tau)-18\theta^2(x,\tau)\theta_{\frac{1}{3},0}(x,\tau)\theta_{\frac{2}{3},0}(x,\tau)\\
&-18\theta^2(x,\tau)\theta^2_{\frac{2}{3},0}(x,\tau)+9\theta^2(x,\tau)\theta^2_{\frac{1}{3},0}(x,\tau)-18\theta(x,\tau)\theta_{\frac{2}{3},0}(x,\tau)\theta^2_{\frac{1}{3},0}(x,\tau)\\
&+9\theta^2_{\frac{2}{3},0}(x,\tau)\theta^2_{\frac{1}{3},0}(x,\tau)+18\theta_{\frac{1}{3},0}(x,\tau)\theta^3_{\frac{2}{3},0}(x,\tau)+9\theta^4_{\frac{2}{3},0}(x,\tau)\\
&-18\theta_{\frac{1}{3},0}(x,\tau)\theta^2_{\frac{2}{3},0}(x,\tau)\theta(x,\tau)=(1-\omega-\omega+\omega^2)\theta^2\left(\frac{x}{3}\right)\theta^2\left(\frac{x+1}{3}\right)\\
&+(1-\omega^2+1-\omega-\omega+1-\omega^2+1)\theta^2\left(\frac{x}{3}\right)\theta\left(\frac{x+1}{3}\right)\theta\left(\frac{x+2}{3}\right)\\
&+(1-\omega^2-\omega^2+\omega)\theta^2\left(\frac{x}{3}\right)\theta^2\left(\frac{x+2}{3}\right).
\end{aligned}\tag{2.80}$$

simplifying the (2.80) we get the required identity. □

References

1. L. Auslander, R. Tolimieri, Is computing with the finite Fourier transform pure or applied mathematics? Bull. Am. Math. Soc. **I**, 847–97 (1979)
2. S. Cooper, P.C. Toh, Determinant identities for theta functions. J. Math. Anal. Appl. **347**, 1–7 (2008)
3. B.W. Dickinson, K. Steiglitz, Eigenvectors and functions of the discrete Fourier transform. Trans. Acoust. Speech Signal Process. **30**, 25–31 (1982)
4. D. Galetti, M.A. Marchiolli, Discrete coherent states and probability distributions in finite dimensional spaces. Ann. Phys. **249**, 454–480 (1996)
5. Z.-G. Liu, Residue theorem and theta function identities. Ramanujan J. **5**, 129–151, 389–406 (2001)
6. Z.-G. Liu, An addition formula for the Jacobian theta function and its applications. Adv. Math. **212**, 389–406 (2007)
7. Z.-G. Liu, Some inverse relations and theta function identities. Int. J. Number Theory **8**(8), 1977–2002 (2012)
8. R.A. Malekar, H. Bhate, Discrete Fourier transform and Jacobi theta functions identities. J. Math Phys. **51**, 023511 (2010)
9. R.A. Malekar, H. Bhate, Discrete Fourier transform and Riemann's identities for theta functions. Appl. Math. Lett. **25**, 1415–1419 (2012)
10. V.B. Matveev, Intertwining relations between Fourier transform and discrete Fourier transform, the related functional identities and beyond. Inverse Prob. **17**, 633–657 (2001)
11. J.H. McClennan, T.W. Parks, Eigenvalue and eigenvector decomposition of the discrete Fourier transform. Trans. Acoust. Speech Signal Process. **30**, 66–74 (1982)
12. H. Mckean, V. Moll, *Elliptic Curves* (Cambridge University Press, Cambridge, 1999)
13. M.L. Mehta, Eigen values and eigenvectors of finite Fourier transform. J Math. Phys. **28**(4), 781 (1987)
14. D. Mumford, *Tata Lectures on Theta I* (Birkhauser, Basel, 1983)
15. M. Ruzzi, Jacobi θ functions and discrete Fourier transform. J Math. Phys. **47**, 063507 (2006)

16. P.W. Shore, Polynomial time algorithm for prime factorisation and discrete algorithms on quantum computers. SIAM J. Comput. **26**, 1484–509 (1997)
17. H.M. Srivastava, M.P. Choudhary, S. Chaudhary, Some theta-function identities related to Jacobi's triple-product identity. Eur. J. Pure Appl. Math. **11**(1), 1–9 (2018)
18. R. Tolimieri, The construction of orthonormal bases diagonalising the DFT. Adv. Appl. Math. **5**, 56–86 (1984)
19. Z.X. Wang, D.R. Guo, *Special Functions* (World Scientific, Singapore, 1989)
20. E.T. Whittaker, G.N. Watson, *A Course of Modern Analysis*, 4th edn. (Cambridge University Press, Cambridge, 1927)

Chapter 3
On Some Combinatorics of Rogers–Ramanujan Type Identities Using Signed Color Partitions

V. Gupta and M. Rana

Abstract In this work we use combinatorial tools "color partitions," "split color partitions," and "signed partitions" notion to define "signed color partitions" that are further used to derive one hundred Rogers–Ramanujan type identities. The paper lists and provides combinatorial argument using signed color partitions of q-identities listed in Chu–Zhang and Slater's compendium.

Keywords $(m+t)$-color partitions · Split partitions · Signed partitions · Combinatorial interpretations

3.1 Introduction

Informally, a partition [3] of an integer m is a non-increasing sequence of positive integers whose sum is m. Let $2+1+1$ is the partition of 4 and it is denoted by ϑ. The number of parts in the partition ϑ is called the length of the partition and is denoted by $l(\vartheta)$, here $l(\vartheta)=3$. The sum of all parts of a partition ϑ is called the weight of the partitions and is denoted by $|\vartheta|$, here $|\vartheta|=4$. We consider α as the number of distinct parts in a particular partition, here in the above partition $\alpha=2$. In [1], Agarwal and Andrews introduced $(m+t)$-colored partitions in which a part of the size m can come in $(m+t)$, $t \geq 0$, different colors. Note that, zeros are not allowed to repeat and occur only when $t \geq 1$. For $t=0$, the colored partitions corresponding to the above ordinary partitions are $2_1+1_1+1_1$, $2_2+1_1+1_1$, $2_1+1_1+1_1+0_2$, $2_2+1_1+1_1+0_2$. The weighted difference $(W.D.)$ of two consecutive

Supported by NBHM Grant No. 2/48(18)/2016/NBHM(R.P.)/R D II/14983.

V. Gupta · M. Rana (✉)
School of Mathematics, Thapar Institute of Engineering and Technology, Patiala, Punjab, India
e-mail: vasudha@thapar.edu; mrana@thapar.edu

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_3

colored parts a_x and b_y is $a-b-x-y$. Corresponding to the above colored partitions the $W.D.$ are $-1, -2; -2, -2; -1, -2, -2; -2, -2, -2$. In [2], Agarwal and Sood introduced "split $(m+t)$-color partitions" where in the colored partitions the color x is splitted into two parts "the green part (g)" and "the red part (r)" such that $1 \le g \le x, 0 \le r \le x-1$, and $x = g+r$. An $(m+t)$-color partition in which each part splits in this manner is called a split $(m+t)$-color partitions. Also, if $r = 0$, then $a_x = a_{g+0} = a_g$. Corresponding to the above color partitions the split color partitions are $2_1+1_1+1_1, 2_2+1_1+1_1, 2_{1+1}+1_1+1_1, 2_1+1_1+1_1+0_2, 2_1+1_1+1_1+0_{1+1}, 2_2+1_1+1_1+0_2, 2_{1+1}+1_1+1_1+0_2, 2_2+1_1+1_1+0_{1+1}, 2_{1+1}+1_1+1_1+0_{1+1}$. A signed partition δ in [6] of an integer m is a partition pair $(\vartheta^+, \vartheta^-)$ where

$$m = |\vartheta^+| - |\vartheta^-|. \tag{3.1.1}$$

ϑ^+ (resp.ϑ^-) the positive (resp. negative) subpartition δ and $\vartheta_1^+, \vartheta_2^+, \cdots, \vartheta_{l(\vartheta^+)}^+$ (resp. $\vartheta_1^-, \vartheta_2^-, \cdots, \vartheta_{l(\vartheta^-)}^-$) the positive (resp. negative) parts of δ. For example, $\vartheta^+ = 4+2+2, \vartheta^- = 2+1+1$, then $\delta = (4+2+2, 2+1+1)$ and $m = 4+2+2-2-1-1$. Now, we define signed color partition δ, where $\delta = (\vartheta^+, \vartheta^-)$ having $\vartheta^+ = (\vartheta_1)_{\beta_1}^+, (\vartheta_2)_{\beta_2}^+, \cdots, (\vartheta_{l(\vartheta^+)})_{\beta_l}^+$ and $\vartheta^+ = (\vartheta_1)_{\beta_1}^-, (\vartheta_2)_{\beta_2}^-, \cdots, (\vartheta_{l(\vartheta^-)})_{\beta_l}^-$, where $(\vartheta_i)_{\beta_i}^+, (\vartheta_i)_{\beta_i}^-$ are the colored parts. In the sequel the difference condition of ordinary partition is converted to $W.D.$ in a colored partition, also the color β_i can further be split in connection with split colored partition. Hence, when $\beta_i = 1$, then it is a signed partition and when $\beta_i > 1$, it is a signed color partition. It is obvious that there are infinitely many unrestricted signed partitions of any integers. But when we place restrictions on how parts may appear, then the count of relevant partitions is finite. Signed partitions also arise naturally in the study of certain q-series, see [5, 6].

In this paper, we interpret 100 *R*ogers–*R*amanujan type identities(RRTIs) of Chu–Zhang's compendium [4] and Slater's compendium [7] given in Table 3.1 with signed color partition, and the identity no. of each RRTI appearing in [4, 7] is mentioned in the last column of Table 3.1. In the "sum–product" series given in [4, 7], the sum-series is written in the second column of Table 3.1, and for the product-series the reader is referred to [4, 7] for their trivial interpretations. The generating function corresponding to ϑ^+ and ϑ^- is written in the third and fifth column, respectively, and their combinatorial interpretation is given in the fourth and sixth columns, respectively.

Table 3.1 Combinatorial interpretations of RRTIs

Sr. no.	$f_i(m)$	ϑ^+	Interpretation	ϑ^-	Interpretation	Identity no.
1.	$\frac{(-1;q)_m q^{m^2}}{(q;q)_{2m}}$	$\frac{q^{m(3m-1)/2}}{(q;q)_m (q;q^2)_m}$	$W.D. \geq 1$	$\prod_{i=1}^{m-1}(1+q^{-i})$	Distinct parts $< \alpha$	1[4], 6[7]
2.	$\frac{(-q;q)_m q^{m(m+1)}}{(q;q)_m (q;q^2)_{m+1}}$	$\frac{q^{3m(m+1)/2}}{(q;q)_m (q;q^2)_{m+1}}$	$(m+1)$ copies of m, least part is x_{x+1}, $W.D. > 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $< \alpha$	3[4]
3.	$\frac{(q;q^2)_m q^{2m(m+1)}}{(-q;q^2)_{m+1} (q^4,q^4)_m}$	$\frac{q^{3m^2+2m}}{(q;q^2)_{m+1} (q^4;q^4)_m}$	$(m+2)$ copies of m, $a \equiv x(\text{mod } 2)$, $x > 2$, least part is x_{x+2}, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha - 2$	5[4], 27[7]
4.	$\frac{(q;q^2)_m q^{2m^2}}{(-q;q^2)_m (q^4,q^4)_m}$	$\frac{q^{3m^2}}{(q;q^2)_m (q^4;q^4)_m}$	$a \equiv x(\text{mod } 2)$, $x > 2$, least part is $a \equiv x(\text{mod } 4)$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	6[4]
5.	$\frac{(-q;q)_m q^{\frac{m(m+1)}{2}}}{(q;q)_m}$	$\frac{q^{m(m+1)}}{(q;q)_m}$	$a > 1, x = 1, W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct part $\leq \alpha$	7[4], 8[7]
6.	$\frac{(-q;q)_{m+1} q^{\frac{m(m+1)}{2}}}{(q;q)_m}$	$\frac{q^{(m+1)^2}}{(q;q)_m}$	$x = 1$, 1_1 must be a part, $W.D. \geq 0$	$\prod_{i=1}^{m+1}(1+q^{-i})$	Distinct part $\leq \alpha$	8[4]
7.	$\frac{(-1;q)_{2m} q^{m}}{(q;q)_{2m}}$	$\frac{q^{2m^2}}{(q;q)_{2m}}$	$a \equiv x(\text{mod } 2)$, $x > 1$, $W.D. \geq 0$	$\prod_{i=1}^{2m-1}(1+q^{-i})$	Distinct parts $\leq 2\alpha - 1$	9[4]
8.	$\frac{(-q;q)_{2m} q^{m}}{(q;q)_{2m+1}}$	$\frac{q^{2m(m+1)}}{(q;q)_{2m+1}}$	$(m+2)$ copies of m, $a \equiv x(\text{mod } 2)$, least part is x_{x+2} , $x > 1$, $W.D. \geq 2$ or 0	$\prod_{i=1}^{2m}(1+q^{-i})$	Distinct parts $< 2\alpha - 2$	10[4]
9.	$\frac{(-1;q^4)_m q^{m^2}}{(q;q^2)_m (q^4;q^4)_m}$	$\frac{q^{3m^2-2m}}{(q;q^2)_m (q^4;q^4)_m}$	$a \equiv x(\text{mod } 2)$, least part is $a \equiv x(\text{mod } 4)$, $W.D. \geq 4, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m-1}(1+q^{-4i})$	Distinct parts, multiple of 4 and $\leq 4\alpha - 4$	11[4], 66[7]
10.	$\frac{(-1;q^4)_m q^{m(m+2)}}{(q;q^2)_m (q^4;q^4)_m}$	$\frac{q^{3m^2}}{(q;q^2)_m (q^4;q^4)_m}$	$a \equiv x(\text{mod } 2)$, $x > 2$, least part $a \equiv x(\text{mod } 4)$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m-1}(1+q^{-4i})$	Distinct parts, multiple of 4 $\leq 4\alpha - 4$	12[4], 67[7]

(continued)

Table 3.1 (continued)

Sr. no.	$f_i(m)$	ϑ^+	Interpretation	ϑ^-	Interpretation	Identity no.
11.	$\frac{(-q^4;q^4)_m q^{m^2}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$\frac{q^{3m^2+2m}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$(m+2)$ copies of m, $a \equiv x(\text{mod } 2)$, least part is x_{x+2}, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-4i})$	Distinct parts, multiple of $4 \leq 4\alpha - 4$	13[4]
12.	$\frac{(-q;q^2)_m^2 q^{m^2}}{(q;q^2)_m(q^4;q^4)_m}$	$\frac{(-q;q^2)_m q^{2m^2}}{(q;q^2)_m(q^4;q^4)_m}$	$a \equiv x(\text{mod } 2)$, $x > 1$, least part $a \equiv x(\text{mod } 4)$, parts are m_{g+r} $g > 1$, $r = 0, 1$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	14[4]
13.	$\frac{(-q;q^2)_m^2 q^{m(m+2)}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$\frac{(-q;q^2)_m q^{2m(m+1)}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$(m+1)$ copies of m, $a \equiv x+1(\text{mod } 2)$, least part is x_{x+1},part, $g > 0$, $r = 0, 1$, $W.D. \geq 1, \equiv 2(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha - 2$	16[4]
14.	$\frac{(-1)^m(q;q^2)_m q^{3m^2-2m}}{(q^2;q^2)_{2m}}$	$\frac{q^{2m(2m-1)}}{(q^2;q^2)_{2m}}$	a and x are even, least part is $a \equiv x(\text{mod } 4)$ $W.D. \geq 4, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	22[4], 15[7]
15.	$\frac{(-1)^m(q;q^2)_m q^{3m^2}}{(q^2;q^2)_{2m}}$	$\frac{q^{4m^2}}{(q^2;q^2)_{2m}}$	a and x are even, ≥ 4, least part is $a \equiv x(\text{mod } 4)$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	24[4], 19[7]
16.	$\frac{(-1)^m(q;q^2)_m q^{m(m+2)}}{(-q;q^2)_{m+1}(q^4;q^4)_m}$	$\frac{q^{2m(m+1)}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$(m+2)$ copies of m, $a \equiv x(\text{mod } 2)$, $x > 1$, least part is x_{x+2}, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha - 2$	25[4]
17.	$\frac{(-1)^m(q;q^2)_m q^{m(m+2)}}{(-q;q^2)_m(q^4;q^4)_m}$	$\frac{q^{2m(m+1)}}{(q;q^2)_m(q^4;q^4)_m}$	$a \equiv x(\text{mod } 2)$, $x > 3$ least part is $a \equiv x(\text{mod } 4)$, $W.D. \geq -4, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	27[4]
18.	$\frac{(-1)^m(q;q^2)_m q^{m^2}}{(-q;q^2)_m(q^4;q^4)_m}$	$\frac{q^{2m^2}}{(q;q^2)_m(q^4;q^4)_m}$	$a \equiv x(\text{mod } 2)$, $a, x > 1$, least part is $a \equiv x(\text{mod } 4)$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	29[4], 21[7]
19.	$\frac{(-q^2;q^2)_m q^{m(m+1)}}{(-q;q)_{2m}(q^2;q^2)_m}$	$\frac{q^{2m(m+1)}}{(q;q^2)_m(q^4;q^4)_m}$	$a \equiv x(\text{mod } 2)$, $x > 3$, least part is $a \equiv x(\text{mod } 4)$, $W.D. \geq -4, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-2i})$	Distinct parts, multiple of $2 \leq 2\alpha$	32[4]

20.	$\frac{(q;q^2)_m^2 q^{2m^2}}{(q^2;q^2)_{2m}}$	$\frac{(-q;q^2)_m q^{3m^2}}{(q^2;q^2)_{2m}}$	a and x are odd, $a, x > 2$, least part $a \equiv x (\text{mod } 4)$, parts are m_{g+r} $g \geq 3$ and odd $r = 0, 1$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	33[4]
21.	$\frac{(-q;q^2)_m q^{2m(m+1)}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$\frac{q^{3m^2+2m}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$(m+2)$ copies of m, $a \equiv x (\text{mod } 2)$, $x \geq 3$, least part is x_{x+2}, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha - 2$	34[4], 27[7]
22.	$\frac{(-1;q)_{2m} q^m}{(q^2;q^2)_m}$	$\frac{q^{2m^2}}{(q^2;q^2)_m}$	a is even and $x = 2$, $W.D. \geq 0$	$\prod_{i=1}^{2m-1}(1+q^{-i})$	Distinct parts $\leq 2\alpha - 1$	36[4], 24,30[7]
23.	$\frac{(-1;q)_m q^{m^2}}{(q;q)_m(q;q^2)_m}$	$\frac{q^{\frac{m(3m-1)}{2}}}{(q;q)_m(q;q^2)_m}$	$W.D. \geq 1$	$\prod_{i=1}^{m-1}(1+q^{-i})$	Distinct parts $< \alpha$	37[4]
24.	$\frac{(-q;q)_m q^{m^2}}{(q;q)_m(q;q^2)_{m+1}}$	$\frac{q^{\frac{m(3m+1)}{2}}}{(q;q)_m(q;q^2)_{m+1}}$	$m+1$ copies of m, least part is x_{x+1}, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $< \alpha$	38[4], 26[7]
25.	$\frac{(-q;q)_{2m} q^m}{(q^2;q^2)_m}$	$\frac{q^{2m(m+1)}}{(q^2;q^2)_m}$	a is even and $a \geq 4$, $x = 2$, $W.D. \geq 0$	$\prod_{i=1}^{2m}(1+q^{-i})$	Distinct parts $\leq 2\alpha$	39[4]
26.	$\frac{(-q;q)_m q^{m(m+1)}}{(q;q)_m(q;q^2)_{m+1}}$	$\frac{q^{\frac{3m(m+1)}{2}}}{(q;q)_m(q;q^2)_{m+1}}$	$(m+1)$ copies of m, least part is x_{x+1} $W.D. > 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $< \alpha$	40[4], 22[7]
27.	$\frac{(-q;q^2)_m q^{m(m+2)}}{(q;q)_{2m+1}}$	$\frac{q^{2m(m+1)}}{(q;q)_{2m+1}}$	$(m+2)$ copies of m, $a \equiv x (\text{mod } 2)$, $x > 1$,least part is x_{x+2}, $W.D. \geq 0$, even	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha - 2$	41[4]
28.	$\frac{(-q;q^2)_m q^{m^2}}{(q^4;q^4)_m}$	$\frac{q^{2m^2}}{(q^4;q^4)_m}$	a is even and $x = 2$, least part $a \equiv x (\text{mod } 4)$, $W.D. \geq 0 \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	42[4], 25[7]
29.	$\frac{(-q;q^2)_m q^{m(m+2)}}{(q^4;q^4)_m}$	$\frac{q^{2m(m+1)}}{(q^4;q^4)_m}$	$a \equiv 0(\text{mod } 4)$ and $x = 2$, least part $a \equiv x + 2(\text{mod } 4)$, $W.D. \geq 0 \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	43[4],

(continued)

Table 3.1 (continued)

Sr. no.	$f_i(m)$	ϑ^+	Interpretation	ϑ^-	Interpretation	Identity no.
30.	$\frac{(-q;q^2)_m q^{m^2}}{(q;q)_{2m}}$	$\frac{q^{2m^2}}{(q;q)_{2m}}$	$a \equiv x(\text{mod } 2)$, $x > 1$, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	44[4], 29[7]
31.	$\frac{(-1;q^2)_m q^{m(m+1)}}{(q;q)_{2m}}$	$\frac{q^{2m^2}}{(q;q)_{2m}}$	$a \equiv x(\text{mod } 2)$, $x > 1$, $W.D. \geq 0$	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct parts, multiple of $2 \leq 2\alpha - 2$	45[4]
32.	$\frac{(-q^2;q^2)_m q^{m(m+1)}}{(q;q)_{2m+1}}$	$\frac{q^{2m(m+1)}}{(q;q)_{2m+1}}$	$(m+2)$ copies of m, $a \equiv x(\text{mod } 2)$, $x > 1$, least part is x_{x+1}, $W.D. \geq 0$, even	$\prod_{i=1}^{m}(1+q^{-2i})$	Distinct parts, multiple of $2 \leq 2\alpha - 2$	46[4], 28[7]
33.	$\frac{(q;q^2)_m q^{2m^2}}{(q^2;q^2)_{2m}}$	$\frac{q^{3m^2}}{(q^2;q^2)_{2m}}$	a, x are odd and > 2, least part is $a \equiv x(\text{mod } 4)$ $W.D. \geq 0 \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	51[4], 33[7]
34.	$\frac{(q;q^2)_m q^{2m(m+1)}}{(q^2;q^2)_{2m}}$	$\frac{q^{3m^2+2m}}{(q^2;q^2)_{2m}}$	a, x are odd, $a \geq 5$, $x > 2$, least part is $a \equiv x+2(\text{mod } 4)$ $W.D. \geq 0 \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	52[4], 32[7]
35.	$\frac{(q;q^2)_{m+1} q^{2m(m+1)}}{(q^2;q^2)_{2m+1}}$	$\frac{q^{3m^2-2m}}{(q^2;q^2)_{2m-1}}$	$a - x \equiv 1(\text{mod } 2)$, least part is x_x, $W.D. \geq 4 \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	53[4], 31[7]
36.	$\frac{(-q;q^2)_m q^{2m+\binom{m}{2}}}{(q;q)_m (q;q^2)_{m+1}}$	$\frac{q^{\frac{3m(m+1)}{2}}}{(q;q)_m (q;q^2)m+1}$	$(m+2)$-copies of m, least part is i_{i+2}, $W.D \geq 1$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha - 2$	61[4], 35, 106[7]
37.	$\frac{(-q^2;q^2)_m q^{\binom{m+1}{2}}}{(q;q)_m (q;q^2)_{m+1}}$	$\frac{q^{\frac{3m(m+1)}{2}}}{(q;q)_m (q;q^2)m+1}$	$(m+2)$-copies, least part is i_{i+2}, $W.D. \geq 1$	$\prod_{i=1}^{m}(1+q^{-2i})$	Distinct parts, multiple of $2 \leq 2\alpha - 2$	63[4]
38.	$\frac{(-q;q^2)_m q^{\binom{m}{2}}}{(q;q)_m (q;q^2)_m}$	$\frac{q^{\frac{3m(m-1)}{2}}}{(q;q)_m (q;q^2)m}$	$W.D. > 0$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	65[4]

39.	$\frac{(-q;q^2)_m q^{\binom{m+1}{2}}}{(q;q)_m(q;q^2)_{m+1}}$	$\frac{q^{\frac{m(3m+1)}{2}}}{(q;q)_m(q;q^2)m+1}$	$(m+1)$-copies of m, least part is x_{x+1}, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha - 2$	67[4], 37, 105[7]
40.	$\frac{(-1;q^2)_m q^{\binom{m+1}{2}}}{(q;q)_m(q;q^2)_m}$	$\frac{q^{\frac{m(3m-1)}{2}}}{(q;q)_m(q;q^2)m}$	$W.D. > 0$	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct even parts $\leq 2\alpha - 2$	70[4]
41.	$\frac{(-q;q^2)_m q^{m(m+2)}}{(q^2;q^2)_m}$	$\frac{q^{\frac{2m(m+1)}{2}}}{(q^2;q^2)m}$	a, x are even and $a \geq 4$, $x = 2$, $W.D. \geq 0$, even	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	73[4], 34[7]
42.	$\frac{(-q;q^2)_m q^{m^2}}{(q^2;q^2)_m}$	$\frac{q^{\frac{2m^2}{2}}}{(q^2;q^2)m}$	a, x are even and $x = 2$, $W.D. \geq 0$, even	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	74[4], 36[7]
43.	$\frac{(-q;q)_{4m} q^{2m^2}}{(-q^4;q^4)_m(q^4;q^4)_{2m}}$	$\frac{q^{6m^2}}{(q^8;q^8)m(q^2;q^4)_m}$	a, x are even and $a, x \geq 6$, least part satisfies $a \equiv x(\text{mod } 8)$, $W.D. \geq 0, \equiv 0(\text{mod } 8)$	$\prod_{i=1}^{2m}(1+q^{-(2i-1)})$	Distinct odd parts $< 4\alpha$	75[4]
44.	$\frac{(-q;q^2)_{2m} q^{2m(m+2)}}{(q^2;q^2)_{2m+1}(-q^4;q^4)_m}$	$\frac{q^{6m^2+4m}}{(q^8;q^8)m(q^2;q^4)_{m+1}}$	$(m+4)$-copies of m, a, x is even, least part is x_{x+4}, $W.D \geq 0equiv0(\text{mod } 8)$	$\prod_{i=1}^{2m}(1+q^{-(2i-1)})$	Distinct odd part $\leq 4\alpha - 5$	77[4]
45.	$\frac{(-q^2;q^2)_m(-q^3;q^6)_m q^{m(m+1)}}{(q^2;q^2)_{2m+1}(-q;q^2)_m}$	$\frac{(q^3;q^6)_m q^{m(m+1)}}{(q^2;q^2)_{2m+1}(q;q^2)_m}$	$(m+2)$-copies of m, $a \equiv x(\text{mod } 2)$, x is even, least part is x_{x+2}, parts are m_{g+r} $g \geq 0, r = 0, 2,$ $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-2i})$	Distinct even parts $\leq 2\alpha - 2$	92[4]
46.	$\frac{(-q;q)_m q^{3\binom{m+1}{2}}}{(q;q)_{2m+1}}$	$\frac{q^{2m(m+1)}}{(q;q)_{2m+1}}$	$(m+2)$-copies of m, $a \equiv x(\text{mod } 2)$, $a > 1$, x_{x+2} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $< \alpha$	93[4], 44[7]

(continued)

Table 3.1 (continued)

Sr. no.	$f_i(m)$	ϑ^+	Interpretation	ϑ^-	Interpretation	Identity no.
47.	$\frac{(-q;q)_m q^{m^2+\binom{m}{2}}}{(q;q)_{2m}}$	$\frac{q^{2m^2}}{(q;q)_{2m}}$	$a \equiv x \pmod 2$, $a > 1$, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $\leq \alpha$	94[4], 46[7]
48.	$\frac{(-q;q)_m q^{m^2+\binom{m+1}{2}}}{(q;q)_{2m+1}}$	$\frac{q^{m(2m+1)}}{(q;q)_{2m+1}}$	$a \equiv x \pmod 2$, $a \geq 3$, $W.D. \geq 2$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $\leq \alpha$	95[4]
49.	$\frac{(-q;q)_m q^{2m+\binom{m}{2}}}{(q;q)_m (q;q^2)_{m+1}}$	$\frac{q^{m(m+2)}}{(q;q)_m (q;q^2)_{m+1}}$	$(m+2)$-copies of m, x_{x+2} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $< \alpha$	96[4], 43[7]
50.	$\frac{(-q;q)_m q^{\binom{m+1}{2}}}{(q;q)_m (q;q^2)_{m+1}}$	$\frac{q^{m(m+1)}}{(q;q)_m (q;q^2)_{m+1}}$	$(m+1)$-copies of m, x_{x+1} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $< \alpha$	97[4], 45[7]
51.	$\frac{(-1;q)_m q^{\binom{m+1}{2}}}{(q;q)_m (q;q^2)_m}$	$\frac{q^{m^2}}{(q;q)_m (q;q^2)_m}$	$W.D. \geq 0$	$\prod_{i=1}^{m-1}(1+q^{-i})$	Distinct parts $< \alpha$	98[4]
52.	$\frac{(-q;q^2)_m q^{m^2}}{(q;q)_{2m}}$	$\frac{q^{2m^2}}{(q;q)_{2m}}$	$a \equiv x \pmod 2$, $x > 1$, $W.D. \geq 0$ or 2	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	99[4]
53.	$\frac{(-q;q^2)_m q^{m(m+2)}}{(q;q)_{2m+1}}$	$\frac{q^{2m(m+1)}}{(q;q)_{2m+1}}$	$(m+2)$-copies of m, $a \equiv x \pmod 2$, $a > 1$, x_{x+2} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha - 2$	102[4], 50[7]
54.	$\frac{(-q^2;q^2)_m q^{m(m+1)}}{(q;q)_{2m+1}}$	$\frac{q^{2m(m+1)}}{(q;q)_{2m+1}}$	$(m+2)$-copies of m, $a \equiv x \pmod 2$, $a > 1$, x_{x+2} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-2i})$	Distinct parts, multiple of $2 \leq 2\alpha - 2$	103[4]
55.	$\frac{(-q;q^2)_m q^{m(m+1)}}{(q;q)_{2m+1}}$	$\frac{q^{m(2m+1)}}{(q;q)_{2m+1}}$	$a \equiv x \pmod 2$, $W.D. \geq 2$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	104[4], (11,15,64)[7]

56.	$\frac{(-q;q)_{m-1}q^{m^2}}{(q;q)_m(q;q^2)_m}$	$\frac{q^{\frac{m(3m-1)}{2}}}{(q;q)_m(q;q^2)_m}$	$W.D. \geq 1$	$\prod_{i=1}^{m-1}(1+q^{-i})$	Distinct parts $< \alpha$	106[4], 58[7]
57.	$\frac{(q;q^2)_{2m}q^{4m^2}}{(q^4;q^4)_{2m}}$	$\frac{q^{8m^2}}{(q^4;q^4)_{2m}}$	$a \equiv x(\text{mod } 2)$, $a, x \geq 8$, least part $a \equiv x(\text{mod } 8)$ $W.D. \geq 0, \equiv 0(\text{mod } 8)$	$\prod_{i=1}^{2m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha + 2$	109[4], 53[7]
58.	$\frac{(-q^2;q^2)_m q^m}{(q;q)_{2m+1}}$	$\frac{q^{m(m+2)}}{(q;q)_{2m+1}}$	$(m+2)$-copies of m, $a \equiv x(\text{mod } 2)$, x_{x+2} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-2i})$	Distinct parts, multiple of $2 \leq 2\alpha - 2$	110[4]
59.	$\frac{(-q;q^2)_m q^m}{(q;q)_{2m+1}}$	$\frac{q^{m(m+1)}}{(q;q)_{2m+1}}$	$(m+1)$-copies of m, $a \equiv m+1(\text{mod } 2)$, x_{x+1} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha - 2$	111[4]
60.	$\frac{(-1;q^2)_m q^m}{(q;q)_{2m}}$	$\frac{q^{m^2}}{(q;q)_{2m}}$	$a \equiv x(\text{mod } 2)$, $W.D. \geq 0$	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct parts, multiple of $2 \leq 2\alpha - 2$	112[4]
61.	$\frac{(-q;q^2)_m q^{2m(m+1)}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$\frac{q^{3m^2+2m}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$(m+2)$-copies of m, $a \equiv x(\text{mod } 2)$, x_{x+2} is a part, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha - 2$	113[4]
62.	$\frac{(-q;q^2)_m q^{2m(m-1)}}{(q;q^2)_m(q^4;q^4)_m}$	$\frac{q^{3m^2-2m}}{(q;q^2)_m(q^4;q^4)_m}$	$a \equiv x(\text{mod } 2)$, least part $a \equiv x(\text{mod } 4)$, $W.D. \geq 4, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	114[4]
63.	$\frac{(-q;q^2)_{3m} q^{3m^2}}{(q^6;q^6)_{2m}}$	$\frac{q^{12m^2}}{(q^6;q^6)_{2m}}$	$a \equiv x(\text{mod } 2)$, $a, x \geq 12$ least part $a \equiv x(\text{mod } 12)$, $W.D. \geq 0, \equiv 0(\text{mod } 12)$	$\prod_{i=1}^{3m}(1+q^{-(2i-1)})$	Distinct odd parts $\leq 6\alpha - 1$	116[4]
64.	$\frac{(-q^2;q^2)_{3m}(-q^3;q^6)_m q^{3m^2}}{(-q^6;q^6)_m(q^6;q^6)_{2m}}$	$\frac{q^{9m^2}}{(q^{12};q^{12})_m(q^3;q^6)_m}$	$a \equiv x(\text{mod } 2)$, $a, x \geq 9$, least part $a - x \equiv 3(\text{mod } 6)$, $W.D. \geq 0, \equiv 0(\text{mod } 12)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct even parts $\not\equiv 0(\text{mod } 6) \leq$ $6\alpha - 2, 6\alpha - 4$	119[4]
65.	$\frac{(-q^2;q^4)_m q^{m(m+2)}}{(q;q)_{2m+1}(-q^2;q^2)_m}$	$\frac{q^{3m^2+2m}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$(m+2)$ copies of m, $W.D. \geq 0, \equiv 0(\text{mod } 4)$, x_{x+2} is a part, $a \equiv x(\text{mod } 2)$	$\prod_{i=1}^{m}(1+q^{-(4i-2)})$	Distinct part of the form $2(\text{mod } 4) \leq 4\alpha - 6$	127[4], 70[7]

(continued)

Table 3.1 (continued)

Sr. no.	$f_i(m)$	ϑ^+	Interpretation	ϑ^-	Interpretation	Identity no.
66.	$\frac{(-q;q^2)_m^2 q^{m(m+2)}}{(q;q)_{2m+1}(-q^2;q^2)_m}$	$\frac{(-q;q^2)_m q^{2m(m+1)}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$(m+2)$ copies of m, $a-x\equiv 0 \pmod 2$, least part x_{x+2}, $r=0,1$, $W.D.\geq 0, \equiv 0 \pmod 4$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha-2$	128[4]
67.	$\frac{(-q^2;q^2)_{m-1}q^m}{(q^2;q^2)_m}$	$\frac{q^{m^2}}{(q^2;q^2)_m}$	a, x are odd, least part $a\equiv x \pmod 4$ $W.D.\geq 0, \equiv 0 \pmod 4$	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct parts, multiple of 2 $\leq 2\alpha-2$	129[4]
68.	$\frac{(-q^2;q^2)_{m-1}q^{m^2}}{(q;q)_{2m}}$	$\frac{q^{m(2m-1)}}{(q;q)_{2m}}$	$a\equiv x \pmod 2$ $W.D.\geq 2$	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct parts, multiple of 2 $\leq 2\alpha-2$	131[4], 72[7]
69.	$\frac{(-q^2;q^4)_m q^{m^2}}{(q;q)_{2m}(-q^2;q^2)_m}$	$\frac{q^{3m^2}}{(q;q^2)_m(q^4;q^4)_m}$	$a\equiv x \pmod 2$, $x>2$ least part $a\equiv x \pmod 4$ $W.D.\geq 0, \equiv 0 \pmod 4$	$\prod_{i=1}^{m}(1+q^{-(4i-2)})$	Distinct parts $\equiv 2 \pmod 4) \leq 4\alpha-2$	132[4]
70.	$\frac{(q;q^2)_{3m}q^{6m^2}}{(q^6;q^6)_{2m}(q^3;q^6)_m}$	$\frac{q^{12m^2}}{(q^6;q^6)_{2m}}$	$a\equiv x \pmod 2$, $a, x>12$, least part $a\equiv x \pmod{12}$, $W.D.\geq 0, \equiv 0 \pmod{12}$	$\prod_{i=1}^{m}(1+q^{-(6i-5)})(1+q^{-(6i-1)})$	Distinct odd parts $\not\equiv 3 \pmod 6) \leq 6\alpha-5, 6\alpha-1$	135[4]
71.	$\frac{(-q;q^2)_m q^{m(m+2)}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$\frac{q^{2m(m+1)}}{(q;q^2)_{m+1}(q^4;q^4)_m}$,	$(m+2)$-copies of m, $a\equiv x \pmod 2$, $a, x>1$, least part is x_{x+2}, $W.D.\geq 0 \equiv 0 \pmod 4$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha-2$	144[4]
72.	$\frac{(-q;q)_m q^{m^2+\binom{m}{2}}}{(q;q)_{2m}}$	$\frac{q^{2m^2}}{(q;q)_{2m}}$	$a\equiv x \pmod 2$, $a, x>1$, $W.D.\geq 0, 2$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $\leq \alpha$	153[4]
73.	$\frac{(-q;q)_m q^{m^2+\binom{m+1}{2}}}{(q;q)_{2m+1}}$	$\frac{q^{m(2m+1)}}{(q;q)_{2m+1}}$	$a\equiv x \pmod 2$, $W.D.\geq 2$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $\leq \alpha$	154[4], 62[7]

74.	$\frac{(-q;q)_m q^{3\binom{m+1}{2}}}{(q;q)_{2m+1}}$	$\frac{q^{2m(m+1)}}{(q;q)_{2m+1}}$	$(m+2)$-copies of m, $a \equiv x(\text{mod } 2)$, $a, x > 1$, x_{x+2} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $< \alpha$	155[4], 63[7]
75.	$\frac{(-q^3;q^3)_m q^{\binom{m+1}{2}}}{(q;q)_{2m+1}}$	$\frac{q^{2m(m+1)}}{(q;q)_{2m+1}}$	$(m+2)$-copies of m, $a \equiv x(\text{mod } 2)$, $a > 1$, x_{x+2} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-3i})$	Distinct parts, multiple of $3 \leq 3\alpha - 3$	161[4]
76.	$\frac{(-q;q)_m q^{\binom{m+1}{2}}}{(q;q)_{2m}}$	$\frac{q^{m(m+1)}}{(q;q)_{2m}}$	$a - x \equiv 1(\text{mod } 2)$, $a > 1$, x_{x+2}, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $\leq \alpha$	164[4], 81[7]
77.	$\frac{(-q;q)_m q^{\binom{m+1}{2}}}{(q;q)_{2m+1}}$	$\frac{q^{m(m+1)}}{(q;q)_{2m+1}}$	$(m+1)$-copies of m, $a - x \equiv 1(\text{mod } 2)$, x_{x+2} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $< \alpha$	165[4], 80[7]
78.	$\frac{(-q;q)_m q^{2m+\binom{m}{2}}}{(q;q)_{2m+1}}$	$\frac{q^{m(m+2)}}{(q;q)_{2m+1}}$	$(m+2)$-copies of m, $a \equiv x(\text{mod } 2)$, x_{x+2} is a part, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $< \alpha$	166, 82[7]
79.	$\frac{(-q;q^2)_m q^{3m^2}}{(q^2;q^2)_{2m}}$	$\frac{q^{4m^2}}{(q^2;q^2)_{2m}}$	a, x is even, $x \geq 4$, least part $a \equiv x(\text{mod } 4)$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	178[4], 100[7]
80.	$\frac{(-q;q^2)_m q^{3m^2-2m}}{(q^2;q^2)_{2m}}$	$\frac{q^{2m(2m-1)}}{(q^2;q^2)_{2m}}$	a, x is even, $x \geq 2$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	179[4], 95[7]
81.	$\frac{(-q^3;q^6)_m q^{m^2}}{(q^2;q^2)_{2m}}$	$\frac{q^{4m^2}}{(q^2;q^2)_{2m}}$	a, x are even and ≥ 4, least part $a \equiv x(\text{mod } 4)$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(6i-3)})$	Distinct part $\equiv 3(\text{mod } 6) \leq 6\alpha - 3$	183[4]
82.	$\frac{(-q;q^2)_m (q^3;q^6)_m q^{m^2}}{(q^2;q^2)_{2m}(q;q^2)_m}$	$\frac{(q^3;q^6)_m q^{2m^2}}{(q^2;q^2)_{2m}(q;q^2)_m}$	$a \equiv x(\text{mod } 2)$, $a, x > 1$, x is even, least part $a \equiv x(\text{mod } 4)$, parts are m_{g+r}, $g > 0$, $r = 0, 2$, $W.D. \geq 0, \equiv (\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	184[4]

(continued)

Table 3.1 (continued)

Sr. no.	$f_i(m)$	ϑ^+	Interpretation	ϑ^-	Interpretation	Identity no.
83.	$\frac{(-q^2;q^2)_m(-q^3;q^6)_m q^{m(m+1)}}{(q^2;q^2)_{2m}(-q;q^2)_{m+1}}$	$\frac{(q^3;q^6)_m q^{2m(m+1)}}{(q^2;q^2)_{2m}(q;q^2)_{m+1}}$	$(m+2)$-copies of m, $a \equiv x(\text{mod } 2)$, $a, x > 1$, x is even, least part is x_{x+2}, parts are m_{g+r}, $g > 0$, $r = 0, 2$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-2i})$	Distinct parts, multiple of 2 $\leq 2\alpha - 2$	189[4]
84.	$\frac{(-q^2;q^2)_m(-q^3;q^6)_m q^{m(m+1)}}{(q^2;q^2)_{2m+1}(-q;q^2)_m}$	$\frac{(q^3;q^6)_m q^{2m(m+1)}}{(q^2;q^2)_{2m+1}(q;q^2)_m}$	$(m+2)$-copies of m, $a \equiv x(\text{mod } 2)$, $a, x > 1$, x is even, least part is x_{x+2}, parts are m_{g+r}, $g > 0$, $r = 0, 2$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct parts, multiple of 2 $\leq 2\alpha - 2$	190[4], 107[7]
85.	$\frac{(-q^2;q^2)_m(-q^3;q^6)_m q^{m(m+3)}}{(q^2;q^2)_{2m+1}(-q;q^2)_m}$	$\frac{(q^3;q^6)_m q^{2m(m+2)}}{(q;q^2)_m(q^2;q^2)_{2m+1}(q;q^2)_m}$	$(m+4)$-copies of m, $a \equiv x(\text{mod } 2)$, $a, x > 1$, x is even, least part is x_{x+4}, $r = 0, 2$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct parts, multiple of 2 $\leq 2\alpha - 2$	191[4]
86.	$\frac{(-q;q^2)_m q^{m(m+2)}}{(q^2;q^2)_{2m}}$	$\frac{q^{2m(m+1)}}{(q^2;q^2)_{2m}}$	a, x are even, $a \geq 4$, least part $a \equiv x+2(\text{mod } 4)$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	192[4], 118[7]
87.	$\frac{(-q;q^2)_m q^{m^2}}{(q^2;q^2)_{2m}}$	$\frac{q^{2m^2}}{(q^2;q^2)_{2m}}$	a, x are even, $a, x > 2$, least part $a \equiv x(\text{mod } 4)$, $W.D. \geq 0, \equiv 0(\text{mod } 4)$	$\prod_{i=1}^{m}(1+q^{-(2i-1)})$	Distinct odd parts $< 2\alpha$	193[4], 117[7]
88.	$\frac{(-q^2;q^2)_{m-1} q^{m^2}}{(q;q)_{2m}}$	$\frac{q^{m(2m-1)}}{(q;q)_{2m}}$	$a \equiv x(\text{mod } 2)$, $W.D. \geq 2$	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct parts, multiple of 2 $\leq 2\alpha - 2$	195[4], 121[7]
89.	$\frac{(-q^2;q^2)_{m-1}(-q;q)_m q^{\frac{m(m+1)}{2}}}{(q;q)_{2m}}$	$\frac{(-q^2;q^2)_{m-1} q^{m(m+1)}}{(q;q)_{2m}}$	$a \equiv x+1(\text{mod } 2)$, least part x_x, others are m_{g+r}, $g \geq 1$, $r = 2$, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $\leq \alpha$	104[7]

90.	$\frac{(-q;q^2)_m(-q;q)_m q^{\frac{m(m+3)}{2}}}{(q;q)_{2m+1}}$	$\frac{(-q;q^2)_m q^{m(m+2)}}{(q;q)_{2m+1}}$	$(m+2)$-copies of m, $a \equiv x (\text{mod } 2)$, least part is x_{x+2}, parts are m_{g+r}, $g \geq 1, r = 0, 1, W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $< \alpha$	106[7]
91.	$\frac{(-q^2;q^2)_{m-1}(-q;q)_m q^{\frac{m(m-1)}{2}}}{(q;q)_{2m}}$	$\frac{(-q^2;q^2)_{m-1} q^{m^2}}{(q;q)_{2m}}$	$a \equiv x (\text{mod } 2)$, least part x_x, others are m_{g+r}, $g \geq 1, r = 2, W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts $\leq \alpha$	101[7]
92.	$\frac{(q^4;q^4)_{m-1}(-q;q^2)_m q^{m^2}}{(q^2;q^4)_m(q^2;q^2)_m(q^2;q^2)_{m-1}}$	$\frac{(-q;q^2)_m q^{m(2m-1)}}{(q^2;q^4)_m(q^2;q^2)_m}$	$a \equiv x (\text{mod } 2)$, parts are m_{g+r}, $g = \text{odd}$, $r = 0, 1, W.D. \geq 2$	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct parts, multiple of $2 \leq 2\alpha - 2$	72[7]
93.	$\frac{(-1;q^4)_m(-q;q^2)_m q^{m(m+2)}}{(q^2;q^2)_{2m}}$	$\frac{(-q;q^2)_m q^{3m^2}}{(q^2;q^2)_{2m}}$	$a \equiv x (\text{mod } 2)$, $x > 2$, least part is $a \equiv x (\text{mod } 4)$, parts are m_{g+r}, $g = \text{odd}$, $r = 0, 1, W.D. \geq 0, \equiv 0 (\text{mod } 4)$	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct parts, multiple of $4 \leq 4\alpha - 4$	67[7]
94.	$\frac{(-q;q^2)_m q^{m(2m-1)}}{(q^2;q^4)_m(q^2;q^2)_m}$	$\frac{q^{3m^2-m}}{(q^2;q^4)_m(q^2;q^2)_m}$	$a, x \equiv 0 (\text{mod } 2)$, $x \geq 2$, $W.D. \geq 2$	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct odd parts $< 2\alpha$	52[7]
95.	$\frac{(-1;q^2)_m q^{m^2}}{(q;q)_{2m}}$	$\frac{q^{m(2m-1)}}{(q;q)_{2m}}$	$a \equiv x (\text{mod } 2)$, $W.D. \geq 2$	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct parts, multiple of $2 \leq 2\alpha - 2$	47[7]
96.	$\frac{(-1;q^2)_m q^{m(m+1)}}{(q;q)_{2m}}$	$\frac{q^{2m^2}}{(q;q)_{2m}}$	$a \equiv x (\text{mod } 2)$, $x \geq 2$ $W.D. \geq 2$ or 0	$\prod_{i=1}^{m-1}(1+q^{-2i})$	Distinct parts, multiple of $2 \leq 2\alpha - 2$	48[7]
97.	$\frac{(-q;q)_m q^{\frac{m(m-1)}{2}}}{(q;q)_m}$	$\frac{q^{m^2}}{(q;q)_m}$	$x = 1$, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-i})$	Distinct parts, $\leq \alpha$	13[7]
98.	$\frac{(-1;q)_m q^{\frac{m(m+1)}{2}}}{(q;q)_m}$	$\frac{q^{m^2}}{(q;q)_m}$	$x = 1$, $W.D. \geq 0$	$\prod_{i=1}^{m}(1+q^{-(i-1)})$	Distinct parts, $< \alpha$	12[7]
99.	$\frac{(q^3;q^3)_m q^{m(m+1)}}{(q;q)_{2m+1}(q;q)_m}$	$\frac{q^{2m(m+1)}}{(q;q)_{2m+1}}$	$(m+2)$ copies of m, $a \equiv x (\text{mod } 2)$, least part is x_{x+2}, $x > 1$, $W.D. \geq 2$ or 0	$\prod_{i=1}^{m}(1+q^{-(i)}+q^{-(2i)})$	Distinct parts, $< \alpha$, may appear at most twice	147 [4]
100.	$\frac{(-q;q)_{2m} q^{\frac{m(m+1)}{2}}}{(q;q^2)_{m+1}(q^4;q^4)_m}$	$\frac{q^{3m^2-2m}}{(q;q^2)_{m+1}(q^4,q^4)_m}$	$(m+2)$ copies of m, $W.D. \geq 0, \equiv 0 (\text{mod } 4)$, x_{x+2} is a part, $a \equiv x (\text{mod } 2)$ $W.D. \geq 0$	$\prod_{i=1}^{2m}(1+q^{i})$	Distinct parts, $< 2\alpha - 2$	11[7]

3.2 Main Proof

In this section, we give details of generating function corresponding to ϑ^+ and ϑ^- partitions, respectively, and prove the 43rd and 82nd identities given in the table

Theorem 3.1 *For $m \geq 0$, let $f_{43}(m)$ denote the number of signed color partitions $\delta = (\vartheta^+, \vartheta^-)$ of m, where*

(1) ϑ^+ denotes the number of m-color partitions of m such that

  (a) parts and their subscripts are even and ≥ 6,
  (b) the least part a_x satisfy $a - x \equiv 0 (\text{mod } 8)$,
  (c) the weighted difference $(W.D.)$ of any two consecutive parts is non-negative and $\equiv 0 (\text{mod } 8)$,

(2) ϑ^- denotes the number of m-color partitions of m with first copy such that parts are odd and distinct $< 4\alpha$,

where α is the number of distinct parts in ϑ^+. Then

$$f_{43}(m) = \sum_{m=0}^{\infty} \frac{(-q; q)_{4m} q^{2m^2}}{(-q^4; q^4)_m (q^4; q^4)_{2m}}.$$

Example 3.1 Consider $m = 11$, then $f_{43}(m) = 5$, the relevant signed color partitions are:

$12_{12} - 1_1, 14_{14} - 3_1, 14_6 - 3_1, 18_6 + 6_6 - 7_1 - 5_1 - 1_1, 20_8 + 6_6 - 7_1 - 5_1 - 3_1.$

Proof We have

$$\begin{aligned}
\sum_{m=0}^{\infty} A_{43}(m) q^m &= \sum_{m=0}^{\infty} \frac{(-q; q)_{4m} q^{2m^2}}{(-q^4; q^4)_m (q^4; q^4)_{2m}} \\
&= \sum_{m=0}^{\infty} \frac{(-q; q^2)_{2m} q^{2m^2}}{(q^8; q^8)_m (q^2; q^4)_m} \\
&= \sum_{m=0}^{\infty} \frac{q^{2m^2}}{(q^8; q^8)_m (q^2; q^4)_m} \prod_{i=1}^{2m} (1 + q^{(2i-1)}) \\
&= \sum_{m=0}^{\infty} \frac{q^{6m^2}}{(q^8; q^8)_m (q^2; q^4)_m} \prod_{i=1}^{2m} (1 + q^{-(2i-1)}). \qquad (3.2.1)
\end{aligned}$$

In the above, $\sum_{m=0}^{\infty} \frac{q^{6m^2}}{(q^8;q^8)_m(q^2;q^4)_m}$ and $\prod_{i=1}^{2m}(1+q^{-(2i-1)})$ correspond to partitions given by ϑ^+ and ϑ^-, respectively. Now, $\prod_{i=1}^{2m}(1+q^{-(2i-1)})$ denotes the number of n-color partition of m such that parts are odd and distinct $< 4\alpha$ with first copy. Also, $\sum_{m=0}^{\infty} \frac{q^{6m^2}}{(q^8;q^8)_m(q^2;q^4)_m}$ enumerates the partitions denoted by $B_{43}(m)$. Let $B_{43}(l,m)$ denote the number of partitions of m enumerated by $B_{43}(m)$ into l parts. We split the partition into three classes: (i) those do not contain k_k as a part, (ii) those contain 6_6 as a part, and (iii) those contain $k_k (k > 6)$ as a part. With some simple transformation we get the following recurrence relation for the above q-series:

$$B_{43}(l,m) = B_{43}(l,m-8l) + B_{43}(l-1,m-12l+6) + B_{43}(l,m-4l+2) - B_{43}(l,m-12l+2),$$

where $B_{43}(0,0) = 1$ and $B_{43}(l,m) = 0$ for $m < 0$. Further using the fact, $h_{43}(z,q) = \sum_{m=0}^{\infty}\sum_{l=0}^{\infty} B_{43}(l,m)z^l q^m$, where $|q| < 1, |zq| < 1$. We get the q-functional equation

$$h_{43}(z,q) = h_{43}(zq^8,q) + zq^6 h_{43}(zq^{12},q) + q^{-2}h_{43}(zq^4,q) - q^2 h_{43}(zq^{12},q).$$

Setting

$$h_{43}(z,q) = \sum_{m=0}^{\infty} \xi_m(q) z^m \qquad \text{take } \xi_0(q) = 1,$$

we can easily check

$$\xi_m(q) = \frac{q^{6m^2}}{(q^8;q^8)_m(q^2;q^4)_m},$$

and

$$h_{43}(z,q) = \sum_{m=0}^{\infty} \frac{q^{6m^2}}{(q^8;q^8)_m(q^2;q^4)_m} z^m.$$

For $z = 1$, we get

$$h_{43}(1, q) = h_{43}(q) = \sum_{m=0}^{\infty} \frac{q^{6m^2}}{(q^8; q^8)_m (q^2; q^4)_m}.$$

$$\sum_{m=0}^{\infty} B_{43}(m) q^m = \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} B_{43}(m) q^m$$
$$= h_{43}(q).$$

And we get the result. □

Theorem 3.2 *For $m \geq 0$, let $f_{82}(m)$ denote the number of signed color partitions $\delta = (\vartheta^+, \vartheta^-)$ of m, where*

(1) ϑ^+ denotes the number of split m-color partitions of m such that

- *(a) parts and their subscripts have the same parity and the subscript > 1,*
- *(b) the least part a_x satisfy $a \equiv x \pmod 4$,*
- *(c) as $x = g + r$; whenever g is odd, then $r = 2$, g is even, then $r = 0$ or 2,*
- *(d) the weighted difference of any two consecutive parts is non-negative and $\equiv 0 \pmod 4$.*

(2) ϑ^- contains the odd, distinct parts less than 2α with $g = 1$ and $r = 0$, where α is the number of distinct parts in ϑ^+.

Remark 3.1 The conditions (a), (b), and (d) are allowed for the whole subscript x irrespective of green(g) and red(r) parts separately.

Example 3.2 For $m = 6$, $f_{82}(m) = 8$ and the relevant signed color partitions are

$$6_6, 6_2, 6_{4+2}, 7_{5+2}-1_1, 7_{1+2}-1_1, 7_{1+2}+2_2-3_1, 8_4+2_2-3_1-1_1, 8_{2+2}+2_2-3_1-1_1.$$

Proof We have

$$\sum_{m=0}^{\infty} A_{82}(m) q^m = \sum_{m=0}^{\infty} \frac{(-q; q^2)_m (q^3; q^6)_m q^{m^2}}{(q; q^2)_m (q^2; q^2)_{2m}}$$
$$= \sum_{m=0}^{\infty} \frac{(q^3; q^6)_m q^{m^2}}{(q; q^2)_m (q^2; q^2)_{2m}} \prod_{i=1}^{m} (1 + q^{2i-1})$$
$$= \sum_{m=0}^{\infty} \frac{(q^3; q^6)_m q^{2m^2}}{(q; q^2)_m (q^2; q^2)_{2m}} \prod_{i=1}^{m} (1 + q^{-(2i-1)}). \qquad (3.2.2)$$

In the above, $\sum_{m=0}^{\infty} \frac{(q^3;q^6)_m q^{2m^2}}{(q;q^2)_m (q^2;q^2)_{2m}}$ and $\prod_{i=1}^{m}(1+q^{-(2i-1)})$ correspond to partitions given by ϑ^+ and ϑ^-. Now, $\prod_{i=1}^{2m}(1+q^{-(2i-1)})$ denotes the number of split m-color partitions of m such that parts are odd and distinct $< 2\alpha$ with $g = 1, r = 0$. Also, $\sum_{m=0}^{\infty} \frac{(q^3;q^6)_m q^{2m^2}}{(q;q^2)_m (q^2;q^2)_{2m}}$ enumerates the partitions denoted by $B_{82}(m)$. Let $B_{82}(l, m)$ denote the number of partitions of m enumerated by $B_{82}(m)$ into l parts. We split the partition into five classes: (i) those do not contain a part k_k and $k_{\overline{k-2}+2}$, (ii) those contain 2_2 as a part, (iii) those contain 3_{1+2} as a part, (iv) those contain 4_{2+2} as a part, and (v) those contain $k_k (k > 2)$ and $k_{\overline{k-2}+2} (k \geq 5)$. With some simple transformation, we get the recurrence relation, which is reversible.

$$\begin{aligned} B_{82}(l, m) &= B_{82}(l, m-4l) + B_{82}(l-1, m-4l+2) + B_{82}(l-1, m-6l+3) \\ &\quad + B_{82}(l-1, m-8l+4) + B_{82}(l, m-4l+2) - B_{82}(l, m-8l+2), \end{aligned}$$

where $B_{82}(0, 0) = 1$ and $B_{82}(l, m) = 0$ for $m < 0$. Further using the fact, $h_{82}(z, q) = \sum_{m=0}^{\infty}\sum_{l=0}^{\infty} B_{82}(l, m) z^l q^m$, where $|q| < 1, |zq| < 1$. We get the q-functional equation

$$\begin{aligned} h_{82}(z, q) &= h_{82}(zq^4, q) + zq^2 h_{82}(zq^4, q) + zq^3 h_{82}(zq^6, q) \\ &\quad + zq^4 h_{82}(zq^8, q) + q^{-2} h_{82}(zq^4, q) - q^{-2} h_{82}(zq^8, q). \end{aligned}$$

One can further elaborate the proof easily. □

3.3 Conclusion

Signed partitions are an unexplored tool in literature. With the work presented in the paper one is open to many questions such as:

1. Can we find combinatorial interpretation of generalized q-series using signed color partitions?
2. Can there be any possible bijections between signed and other combinatorial tools under some restrictions or in general?

Acknowledgments The authors would like to thank the anonymous referee(s) for their helpful comments that led to a better presentation of the paper.

References

1. A.K. Agarwal, G.E. Andrews, Rogers–Ramanujan identities for partitions with "N copies of N". J. Comb. Theor. Ser. A **45**(1), 40–49 (1987)
2. A.K. Agarwal, G. Sood, Split $(n + t)$-color partitions and Gordon–McIntosh eight order mock theta functions. Electron. J. Comb. **21**(2), P2–P46 (2014)
3. G.E. Andrews, *The Theory of Partitions (Encyclopedia of Mathematics and Its Applications)*, vol. 2 (Addison-Wesley, Boston, 1976)
4. W. Chu, W. Zhang, Bilateral bailey lemma and Rogers–Ramanujan identities. Adv. Appl. Math. **42**(3), 358–391 (2009)
5. W.J. Keith, A bijective toolkit for signed partitions. Ann. Comb. **15**(1), 95–117 (2011)
6. J. McLaughlin, A.V. Sills, Combinatorics of Ramanujan–Slater type identities. Integers **9**(P# 14) (2009)
7. L.J. Slater, Further identities of the Rogers–Ramanujan type. Proc. Lond. Math. Soc. **2**(1), 147–167 (1952)

Chapter 4
Piecewise Continuous Stepanov-Like Almost Automorphic Functions with Applications to Impulsive Systems

Syed Abbas and Lakshman Mahto

Abstract In this chapter, we discuss Stepanov-like almost automorphic function in the framework of impulsive systems. Next, we establish the existence and uniqueness of such solution of a very general class of delayed model of impulsive neural network. The coefficients and forcing term are assumed to be Stepanov-like almost automorphic in nature. Since the solution is no longer continuous, so we introduce the concept of piecewise continuous Stepanov-like almost automorphic function. We establish some basic and important properties of these functions and then prove composition theorem. Composition theorem is an important result from the application point of view. Further, we use composition result and fixed point theorem to investigate existence, uniqueness and stability of solution of the problem under consideration. Finally, we give a numerical example to illustrate our analytical findings.

Keywords Stepanov-like almost automorphic functions · Composition theorem · Impulsive differential equations · Fixed point method · Asymptotic stability

4.1 Introduction

The introduction of almost periodic functions (AP) by H. Bohr [11] in the year 1924–1925 led to various important generalizations of this concept. One important generalization is the concept of almost automorphic function (AA) given by S. Bochner [10]. This concept is further generalized to several other concepts out

S. Abbas (✉)
School of Basic Sciences, Indian Institute of Technology Mandi, Mandi, H.P., India
e-mail: abbas@iitmandi.ac.in

L. Mahto
Department of Mathematics, Indian Institute of Information Technology Dharwad, Hubli, Karnataka, India
e-mail: lakshman@iiitdwd.ac.in

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_4

of which one important generalization is the concept of Stepanov-like almost automorphic function introduced by N'Guérékata and Pankov [21]. Several authors have discussed several classes of almost automorphic functions and their extensions with application to differential equations [13, 14, 20]. It has been observed that one of the natural questions in the field of differential equations is: if the forcing function possesses a special characteristic, then whether the solution possesses the same characteristic or not? Motivated by this many researchers have studied the existence of Stepanov-like almost automorphic solutions of differential equations (see, for example, [14, 20] and the references therein). While studying the behaviour of many physical and biological phenomena, it has been observed that many phenomenons exhibit regularity behaviour which is not exactly periodic. These kind of phenomenons can be modelled by considering more general notions such as almost periodic, almost automorphic, or Stepanov-like almost automorphic. We have the following inclusion $AP \subset AA_u \subset AA \subset BC$, where AA_u stands for uniformly almost automorphic and BC is the space of bounded and continuous functions. If we consider the class of Stepanov-like almost automorphic, then it covers more functions than almost automorphic functions. So, if the underlying behaviour of the systems is not almost automorphic, it may be possible that it is Stepanov-like almost automorphic or it belongs to other more general class of functions. For more work on Stepanov-like almost automorphic and its generalizations, we refer to [2, 4, 15, 16] and the references therein.

Impulsive differential equations involve differential equations on continuous time interval as well as difference equations on discrete set of times. It provides a real framework of modelling the systems, which undergo through abrupt changes like shocks, earthquake, harvesting, etc. Recent years have seen tremendous work in this area due to its applicability in several fields. There are few excellent monographs and literatures on impulsive differential equations [7–9, 19, 24]. As we know that impulses are sudden interruptions in the systems, in neural case, we can say that these abrupt changes are in the neural state. Its effect on humans will depend on the intensity of the change. In signal processing, the faulty elements in the corresponding artificial network may produce sudden changes in the state voltages and thereby affect the normal transient behaviour in processing signals or information. Neural networks have been studied extensively, but the mathematical modelling of dynamical systems with impulses is very recent area of research [1, 3, 5, 6, 25–31].

To the best of our knowledge, the existences, uniqueness and stability of Stepanov-like almost automorphic solution of impulsive differential equations is rarely discussed. In this work, we introduce piecewise continuous Stepanov-like almost automorphic function. We prove composition theorem, which is very important result. As an application we study the existence, uniqueness and stability of Stepanov-like almost automorphic solution of the following impulsive delay differential equations arising from neural network modelling,

$$
\begin{aligned}
\frac{dx_i(t)}{dt} &= \sum_{j=1}^{n} a_{ij}(t)x_j(t) + \sum_{j=1}^{n} \alpha_{ij}(t) f_j(x_j(t)) + \sum_{j=1}^{n} \beta_{ij}(t) f_j(x_j(t-\alpha)) \\
&\quad + \gamma_i(t), \quad t \neq t_k,\ \alpha > 0, \\
\Delta(x(t_k)) &= A_k x(t_k) + I_k(x(t_k)) + \gamma_k, \\
x(t_k - 0) &= x(t_k), \quad x(t_k + 0) = x(t_k) + \Delta x(t_k), \quad k \in \mathbb{Z},\ t \in \mathbb{R}, \\
x(t) &= \Psi_0(t), \quad t \in [-\alpha, 0], \qquad (4.1.1)
\end{aligned}
$$

where $a_{ij}, \alpha_{ij}, \beta_{ij}, f_j, \gamma_i \in \mathcal{C}(\mathbb{R}, \mathbb{R})$ for $i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, n$. The coefficient $A_k \in \mathbb{R}^{n\times n}$, the function $I_k(x) \in \mathcal{C}(\Omega, \mathbb{R}^n)$ and the constant $\gamma_k \in \mathbb{R}^n$. The symbol Ω denotes a domain in $\mathbb{R}^n$ and $\mathcal{C}(X, Y)$ denotes the set of all continuous functions from X to Y.

The organization of this work is as follows: In Sect. 4.2, we give some basic definitions and results. In Sect. 4.3, we establish composition theorem. In Sect. 4.4, we study existence and stability of piecewise continuous Stepanov-like almost automorphic solutions of impulsive differential equations with delay. In Sect. 4.5, we present an example with numerical simulation.

4.2 Preliminaries

Throughout the manuscript, the symbol $\mathbb{R}^n$ denotes the n dimensional space with norm $\|x\| = \max\{|x_i|; i = 1, 2, \cdots, n\}$. We denote $PC(\mathbb{J}, \mathbb{R}^n)$, space of all piecewise continuous functions from $J \subset \mathbb{R}$ to $\mathbb{R}^n$ with points of discontinuity of first kind t_k where it is left continuous.

For smooth reading of the manuscript, we first define the following class of spaces,

- $S^p AA_{pc}(\mathbb{R}, \mathbb{R}^n) = \Big\{\phi \in PC(\mathbb{R}, \mathbb{R}^n) : \phi$ is a piecewise continuous Stepanov-like almost automorphic function $\Big\}$
- $S^p AA_{pc}(\mathbb{R} \times \mathbb{R}^n, \mathbb{R}^n) = \Big\{\phi \in PC(\mathbb{R} \times \mathbb{R}^n, \mathbb{R}^n) : \phi$ is a piecewise continuous Stepanov-like almost automorphic function $\Big\}$
- $S^p AAS(\mathbb{Z}, \mathbb{R}) = \Big\{\phi : \mathbb{Z} \to \mathbb{R} : \phi$ is a Stepanov-like almost automorphic sequence $\Big\}$

Note that the definition of almost automorphic operator is given by N'Guérékata and Pankov [22]. Now we give the following definitions in the framework of impulsive systems motivated by the work of [12, 17, 28].

Definition 4.2.1 ([18]) A function $f \in PC(\mathbb{R}, \mathbb{R}^n)$ is called a PC-almost automorphic if

(i) sequence of impulsive moments $\{t_k\}$ is an almost automorphic sequence,
(ii) for every real sequence (s_n), there exists a subsequence (s_{n_k}) such that $g(t) = \lim_{n\to\infty} f(t+s_{n_k})$ is well defined for each $t \in \mathbb{R}$ and $\lim_{n\to\infty} g(t-s_{n_k}) = f(t)$ for each $t \in \mathbb{R}$.

We denote $AA_{pc}(\mathbb{R}, \mathbb{R}^n)$ the set of all such functions.

Definition 4.2.2 ([18]) A function $f \in PC(\mathbb{R} \times \mathbb{R}^n, \mathbb{R}^n)$ is called PC-almost automorphic in t uniformly for x in compact subsets of X if

(i) sequence of impulsive moments $\{t_k\}$ is an almost automorphic sequence,
(ii) for every compact subset K of X and every real sequence (s_n), there exists a subsequence (s_{n_k}) such that $g(t, x) = \lim_{n\to\infty} f(t + s_{n_k}, x)$ is well defined for each $t \in \mathbb{R}$, $x \in K$ and $\lim_{n\to\infty} g(t - s_{n_k}, x) = f(t, x)$ for each $t \in \mathbb{R}$, $x \in K$.

We denote $AA_{pc}(\mathbb{R} \times \mathbb{R}^n, \mathbb{R}^n)$ the set of all such functions.

Definition 4.2.3 A sequence of continuous functions, $I_k : \mathbb{R}^n \to \mathbb{R}^n$ is almost automorphic, if for integer sequence $\{k_n^{'}\}$, there exist a subsequence $\{k_n\}$ such that $\lim_{n\to\infty} I_{(k+k_n)}(x) = I_k^*(x)$ and $\lim_{n\to\infty} I^*_{(k-k_n)}(x) = I_k(x)$ for each k and $x \in X$.

Definition 4.2.4 A bounded sequence $x : \mathbb{Z}^+ \to \mathbb{R}^n$ is called an almost automorphic sequence, if for every real sequence $(k_n^{'})$, there exists a subsequence (k_n) such that $y(k) = \lim_{n\to\infty} x(k + k_n)$ is well defined for each $m \in \mathbb{Z}$ and $\lim_{n\to\infty} y(k - k_n) = x(k)$ for each $k \in \mathbb{Z}^+$. We denote $AAS(\mathbb{Z}, \mathbb{R}^n)$, the set of all such sequences.

Definition 4.2.5 ([23]) The Bochner transform $f^b(t, s)$, $t \in \mathbb{R}$, $s \in [0, 1]$ of a function $f : \mathbb{R} \to \mathbb{R}^n$ is defined by $f^b(t, s) := f(t + s)$.

Definition 4.2.6 ([23]) Let $p \in [1, \infty)$. The space $BS^p(\mathbb{R}^n)$ of all Stepanov bounded functions, with the exponent p, consists of all measurable functions f on $\mathbb{R}$ with values in $\mathbb{R}^n$ such that $f^b \in L^\infty\big(\mathbb{R}, L^p((0, 1), d\tau)\big)$. This is a Banach space when it is equipped with the norm defined by

$$\|f\|_{S^p} = \|f^b\|_{L^\infty(\mathbb{R},L^p)} = \sup_{t\in\mathbb{R}} \Big(\int_t^{t+1} \|f(\tau)\|^p \, d\tau \Big)^{1/p}.$$

Definition 4.2.7 A bounded piecewise continuous function $f \in PC(\mathbb{R}, \mathbb{R}^n)$ is called a piecewise continuous Stepanov-like almost automorphic if

(i) sequence of impulsive moments $\{t_k\}$ is a Stepanov-like almost automorphic sequence,

(ii) for every real sequence (s'_n), there exists a subsequence (s_n) such that

$$\lim_{n\to\infty}\left(\int_0^1 \|f(t+s_n+s)-g(t+s)\|^p ds\right)^{\frac{1}{p}} = 0$$

is well defined for each $t \in \mathbb{R}$ and

$$\lim_{n\to\infty}\left(\int_0^1 \|g(t-s_n+s)-f(t+s)\|^p ds\right)^{\frac{1}{p}} = 0$$

for each $t \in \mathbb{R}$.

The space of all such functions is denoted by $S^pAA_{pc}(\mathbb{R}, \mathbb{R}^n)$.

Definition 4.2.8 A bounded piecewise continuous function $f \in PC(\mathbb{R}\times\mathbb{R}^n, \mathbb{R}^n)$ is called a piecewise continuous Stepanov-like almost automorphic in t uniformly in x in compact subsets of $\mathbb{R}^n$ if

(i) the sequence of impulsive moments $\{t_k\}$ is a Stepanov-like almost automorphic sequence,
(ii) for every compact subset K of $\mathbb{R}^n$ and every real sequence (s'_n), there exists a subsequence (s_n) such that

$$\lim_{n\to\infty}\left(\int_0^1 \|f(t+s_n+s,x)-g(t+s,x)\|^p ds\right)^{\frac{1}{p}} = 0$$

is well defined for each $t \in \mathbb{R}$ and

$$\lim_{n\to\infty}\left(\int_0^1 \|g(t-s_n+s,x)-f(t+s,x)\|^p ds\right)^{\frac{1}{p}} = 0$$

for each $t \in \mathbb{R}$.

The space of all such functions is denoted by $S^pAA_{pc}(\mathbb{R}\times\mathbb{R}^n, \mathbb{R}^n)$.

Definition 4.2.9 A bounded sequence $x : \mathbb{Z}^+ \to \mathbb{R}^n$ is called Stepanov-like almost automorphic if for every real sequence (k'_n), there exists a subsequence (k_n) and a sequence $y : \mathbb{Z}^+ \to \mathbb{R}^n$ such that

$$\left(\sum_{n=0}^{1} \|x(m+n_k+n)-y(m+n)\|^p\right)^{\frac{1}{p}} \to 0 \text{ as } k \to 0$$

is well defined for each $m \in \mathbb{Z}$ and

$$\left(\sum_{n=0}^{1} \|y(m - n_k + n) - x(m + n)\|^p\right)^{\frac{1}{p}} \to 0 \text{ as } k \to 0$$

for each $m \in \mathbb{Z}^+$.

We denote $S^p AAS(\mathbb{Z}^+, \mathbb{R}^n)$, the set of all such sequences.

We finish this section by defining few examples of Stepanov-like almost automorphic functions below.

(i) Consider $x = (x_n)_{n\in\mathbb{Z}}$, an almost automorphic sequence and the function:

$$a(t) = \begin{cases} x_n, & t \in (n - \epsilon, n + \epsilon), n \in \mathbb{Z}, \\ 0, & \text{otherwise} \end{cases}$$

(ii)

$$b(t) = \begin{cases} \sin\left(\frac{1}{2+\sin(n)+\sin(\sqrt{2}n)}\right), & t \in \left(n - \frac{1}{4}, n + \frac{1}{4}\right), n \in \mathbb{Z}, \\ 0, & \text{otherwise} \end{cases}$$

(iii)

$$c(t) = \begin{cases} \cos\left(\frac{1}{2+\cos(n)+\cos(\sqrt{2}n)}\right), & t \in \left(n - \frac{1}{4}, n + \frac{1}{4}\right), n \in \mathbb{Z}, \\ 0, & \text{otherwise} \end{cases}$$

4.3 Composition Theorem

Lemma 4.3.1 *Let $I_k : \mathbb{R}^n \to \mathbb{R}^n$ be a sequence of Stepanov-like almost automorphic functions and $K \subset \mathbb{R}^n$ be a compact subset. If I_k satisfies Lipschitz condition on $\mathbb{R}^n$, i.e.*

$$\|I_k(x) - I_k(y)\| \leq L\|x - y\|, \forall x, y \in \mathbb{R}^n, \forall k,$$

then the sequence $\{I_k(x) : x \in K\}$ is Stepanov-like almost automorphic.

Proof Since I_k is Lipschitz continuous over a compact set K, its range is also compact. Hence every sequence $I_{k+k_n}(x)$ has a convergent subsequence. So using the fact that I_k is Stepanov almost automorphy, the Stepanov almost automorphy of $I_k(x)$ for $x \in K$ is ensured. □

Lemma 4.3.2 *Let $I_k : \mathbb{R}^n \to \mathbb{R}^n$ be a sequence of Stepanov-like almost automorphic functions and $\phi \in S^p AA_{pc}(\mathbb{R}, \mathbb{R}^n)$. If I_k satisfies Lipschitz condition on $\mathbb{R}^n$, i.e.*

$$\|I_k(x) - I_k(y)\| \leq L\|x - y\|, \forall x, y \in \mathbb{R}^n, \forall k,$$

then the sequence $\{I_k(\phi(t_k))\}$ is Stepanov-like almost automorphic.

Proof Since I_k is a sequence of Stepanov-like almost automorphic functions, there exists I_k^* such that $I_{k+k_n}(x(t_k)) \to I_k^*(x(t_k))$ and $I_{k-k_n}^*(x(t_k)) \to I_k(x(t_k))$. By the above property and Lipschitz continuity of I_k, we obtain

$$\begin{aligned} \|I_{k+k_n}(x(t_{k+k_n})) - I_k^*(x(t_k))\| &\leq \|I_{k+k_n}(x(t_{k+k_n})) - I_{k+k_n}(x(t_k))\| \\ &+ \|I_{k+k_n}(x(t_k)) - I_k^*(x(t_k))\| \\ &\leq L\|x(t_{k+k_n}) - x(t_k)\| \\ &+ \|I_{k+k_n}(x(t_k)) - I_k^*(x(t_k))\|. \end{aligned} \tag{4.3.1}$$

Using Lemma 4.3.1 and the above expression (4.3.1), the sequence $\{I_k(\phi(t_k))\}$ is Stepanov-like almost automorphic. □

Lemma 4.3.3 *If $f, f_1, f_2 \in S^p AA_{pc}(\mathbb{R}, \mathbb{R}^n)$, then the following are true:*

(i) $f_1 + f_2 \in S^p AA_{pc}(\mathbb{R}, \mathbb{R}^n)$,
(ii) $cf \in S^p AA_{pc}(\mathbb{R}, \mathbb{R}^n)$ *for any scalar c,*
(iii) $f_a(t) - f(t + a) \in S^p AA_{pc}(\mathbb{R}, \mathbb{R}^n)$ *for any $a \in \mathbb{R}$,*
(iv) $R_f = \{f(t) : t \in \mathbb{R}\}$ *is relatively compact.*

Proof Proof of $(i), (ii), (iii)$ is obvious from definition of Stepanov-like almost automorphic function. For the proof of (iv) consider a sequence $f(t + s'_n) \in R_f$, then using definition of Stepanov-like almost automorphic function, there exists a function g such that $\lim_{n\to\infty}(\int_0^1 \|f(t + s_n + s, x) - g(t + s, x)\|ds)^{\frac{1}{p}} = 0$. And hence R_f is relatively compact. □

Now we prove our main result of this section.

Lemma 4.3.4 (Composition Theorem) *Let $f \in S^p AA_{pc}(\mathbb{R} \times \mathbb{R}^n, \mathbb{R}^n)$ is uniformly continuous with respect to x on any compact subset of $\mathbb{R}^n$. If $\phi \in S^p AA_{pc}(\mathbb{R}, \mathbb{R}^n)$, then $f(\cdot, \phi(\cdot)) \in S^p AA_{pc}(\mathbb{R}, \mathbb{R}^n)$.*

Proof From the assumption f is uniformly continuous with respect to x on any compact subset of $\mathbb{R}^n$, $i.e.$ for $\epsilon > 0$, there exists $\delta > 0$ such that $\|x - y\| < \delta \Rightarrow \|f(\cdot, x) - f(\cdot, y)\| < \epsilon$.
Also, the range of function ϕ is relatively compact, i.e. $K = \overline{\{\phi(t) : t \in \mathbb{R}\}}$ is compact and hence there exists a finite number of open balls $O_k, k = 1, 2, \cdots, n$

centred at $x_k \in \{\phi(t) : t \in \mathbb{R}\}$ with radius δ such that

$$\{\phi(t) : t \in \mathbb{R}\} \subset \cup_{k=0}^{n} O_k$$

Define B_k such that

$$B_k = \{s \in \mathbb{R} : \phi(s) \in O_k\}, \mathbb{R} = \cup_{k=0}^{n} B_k$$

and set

$$E_1 = B_1, E_k = B_k / \cup_{j=1}^{k-1} B_j$$

Consider a step function $\bar{x} : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\bar{x}(s) = x_k, s \in E_k, \text{ we can see that } \|x(s) - \bar{x}\| \leq \delta.$$

Further using the definition of Stepanov-like almost automorphy of f and ϕ, that is for each sequence $\{s'_n\}$ there exist subsequence $\{s_n\}$ and functions g and ψ such that

$$\int_0^1 \left(\|f(t+s+s_n, x) - g(t+s, x)\|^p ds\right)^{\frac{1}{p}} \to 0, \tag{4.3.2}$$

$$\int_0^1 \left(\|g(t+s-s_n, x) - f(t+s, x)\|^p ds\right)^{\frac{1}{p}} \to 0 \text{ as } n \to \infty \text{ pointwise on } \mathbb{R},$$

and

$$\int_0^1 \left(\|\phi(t+s+s_n) - \psi(t+s)\|^p ds\right)^{\frac{1}{p}} \to 0, \tag{4.3.3}$$

$$\int_0^1 \left(\|\psi(t+s-s_n) - \psi(t+s)\|^p ds\right)^{\frac{1}{p}} \to 0 \text{ as } n \to \infty \text{ pointwise on } \mathbb{R}.$$

Calculating the Stepanov norm of f, we have

$$\begin{aligned}
&\int_0^1 \left(\|f(t+s, x(t+s))\|^p ds\right)^{\frac{1}{p}} \\
&\leq \int_0^1 \left(\|f(t+s, x(t+s)) - f(t+s, \bar{x}(t+s))\|^p ds\right)^{\frac{1}{p}} \\
&+ \int_0^1 \left(\|f(t+s, \bar{x}(t+s))\|^p ds\right)^{\frac{1}{p}}
\end{aligned}$$

$$\leq \int_t^{t+1} \Big(L\|x(s)-\bar{x}(s)\|^p ds\Big)^{\frac{1}{p}} + \int_t^{t+1} \Big(\|f(s,x_k)\|^p ds\Big)^{\frac{1}{p}}$$
$$\leq L\|x(s)-\bar{x}(s)\|_{S^p} + \sum_{k=1}^{n} \int_{E_k \cap [t,t+1]} \Big(\|f(s,x_k)\|^p ds\Big)^{\frac{1}{p}}.$$

Using Eqs. (4.3.2) and (4.3.3), we obtain

$$\int_0^1 \Big(\|f(t+s+s_n,\phi(t+s+s_n)) - g(t+s,\psi(t+s))\|^p ds\Big)^{\frac{1}{p}}$$
$$\leq \int_0^1 \Big(\|f(t+s+s_n,\phi(t+s+s_n)) - f(t+s+s_n,\psi(t+s))\|^p ds\Big)^{\frac{1}{p}}$$
$$+ \int_0^1 \Big(\|f(t+s+s_n,\psi(t+s)) - g(t+s,\psi(t+s))\|^p ds\Big)^{\frac{1}{p}}$$
$$\leq \int_0^1 \Big(L\|\phi(t+s+s_n) - \psi(t+s)\|^p ds\Big)^{\frac{1}{p}}$$
$$+ \int_0^1 \Big(\|f(t+s+s_n,\psi(t+s)) - g(t+s,\psi(t+s))\|^p ds\Big)^{\frac{1}{p}}$$
$$< (L+1)\epsilon.$$

Similarly

$$\int_0^1 \Big(\|g(t+s-s_n,\psi(t+s-s_n)) - f(t+s,\phi(t+s))\|^p ds\Big)^{\frac{1}{p}} < (L+1)\epsilon.$$

Hence $f(\cdot,\phi(\cdot))$ is Stepanov almost automorphic. □

4.4 Impulsive Delay Differential Equations

We can easily see that the Eq. (4.1.1) can be written in the following compact form:

$$\frac{dx(t)}{dt} = A(t)x(t) + f(t,x(t),x(t-\alpha)) \quad t \neq t_k$$
$$\Delta x(t_k) = A_k x(t_k) + I_k(x(t_k)),\ k \in \mathbb{Z},\ t \in \mathbb{R}, \tag{4.4.1}$$

where $A(t) = (a_{ij}(t))_{nxn}, i, j = 1, 2, \cdots, n$, $f = (f_1, f_2, \cdots, f_n)^T$ and

$$f_i(t, x(t), x(t-\alpha)) = \sum_{j=1}^{n} \alpha_{ij}(t) f_j(x_j(t)) + \sum_{j=1}^{n} \beta_{ij}(t) f_j(x_j(t-\alpha)) + \gamma_i(t),$$

for $i = 1, 2, \cdots, n$. In order to prove our results, we need the following assumptions:

(H1) The function $A(t) \in \mathcal{C}(\mathbb{R}, \mathbb{R}^n)$ is a piecewise continuous Stepanov-like almost automorphic function,

(H2) $\det(I + A_k) \neq 0$ and the sequences A_k and t_k are Stepanov-like almost automorphic.

It is well known that if $U_k(t, s)$ is the Cauchy matrix associated with the system

$$\frac{dx(t)}{dt} = A(t)x(t) \quad t_{k-1} \leq t \leq t_k,$$

then the Cauchy matrix of the system (4.4.1) is given by

$$U(t, s) = \begin{cases} U_k(t, s), \quad t_{k-1} \leq t \leq t_k, \\ U_{k+1}(t, t_k + 0)(I + A_k)U_k(t, s), \\ \quad t_{k-1} < s < t_k < t < t_{k+1}, \\ U_{k+1}(t, t_k + 0)(I + A_k)U_k(t_k, t_k + 0) \\ \cdots (I + A_i)U_i(t_i, s), \\ \text{for} \quad t_{i-1} < s \leq t_i < t_k < t < t_{k+1}. \end{cases}$$

For the above Cauchy matrix, the solution of the corresponding homogenous system could be written as $x(t, t_0, x_0) = U(t, t_0)x_0$, where x_0 is the initial condition at the initial point t_0. Let us further assume the followings:

(H3) There exist positive constants K and δ such that $\|U(t, s)\| \leq Ke^{-\delta(t-s)}$, which further implies that $\|U(t + t_{n_k}, s + t_{n_k}) - U(t, s)\| \leq M\epsilon e^{-\frac{\delta}{2}(t-s)}$ for any $\epsilon > 0$ and positive constant M.

(H4) The functions α_{ij}, β_{ij} are Stepanov-like almost automorphic such that

$$-\infty < \alpha_{ij_*} \leq \alpha_{ij}(t) \leq \alpha_{ij}^* < \infty, \qquad -\infty < \beta_{ij_*} \leq \beta_{ij}(t) \leq \beta_{ij}^* < \infty.$$

(H5) The function f_j is Stepanov-like almost automorphic with $0 < \sup_{t \in \mathbb{R}} f_j(t) < \infty$ and satisfies $|f_j(t) - f_j(s)| \leq L_j|t - s|, \quad j = 1, 2, \cdots, n$.

(H6) The function γ_i is Stepanov-like almost automorphic and satisfies $-\infty < \gamma_{i*} \leq \gamma_i(t) \leq \gamma_i^* < \infty$.

(H7) The sequence I_k is Stepanov-like almost automorphic and there exists a positive constant L such that $\|I_k(x) - I_k(y)\| \leq L\|x - y\|$, for $k \in \mathbb{Z}, x, y \in \Omega \subset \mathbb{R}^n$.

Now we have made enough background to prove the main results of this paper, which are presented below.

Lemma 4.4.1 *Under the properties of Cauchy matrix* $U(t, s)$, *the impulsive differential Eq. (4.4.1) is equivalent to the following integral equation:*

$$x(t) = \int_{-\infty}^{t} U(t, s) f(s, x(s), x(s - \alpha))ds + \sum_{t > t_k} U(t, t_k) I_k(x(t_k)). \quad (4.4.2)$$

Proof For $t \in [0, t_1]$, we claim that the following function is the solution of system (4.1.1)

$$x(t) = \int_{-\infty}^{t} U(t, s) f(s, x(s), x(s - \alpha))ds.$$

Differentiating both sides with respect to t, we get

$$\frac{dx(t)}{dt} = \int_{-\infty}^{t} \frac{\partial U(t, s)}{\partial t} f(s, x(s), x(s - \alpha))ds + f(t, x(t), x(t - \alpha)), \ x(0) = \psi_0(0)$$

$$\Leftrightarrow \frac{dx(t)}{dt} = A(t)x(t) + f(t, x(t), x(t - \alpha)), \quad x(0) = \psi_0(0).$$

For $t \in (t_1, t_2]$, define

$$x(t) = \int_{-\infty}^{t} U(t, s) f(s, x(s), x(s - \alpha))ds + U(t, t_1)(I_1 x(t_1))$$

$$\Leftrightarrow x(t) = U(t, t_1)\Big(I_1(x(t_1)) + \int_{-\infty}^{t_1} U(t_1, s) f(s, x(s), x(s - \alpha))ds\Big)$$

$$+ \int_{t_1}^{t} U(t, s) f(s, x(s), x(s - \alpha))ds$$

$$\Leftrightarrow x(t) = U(t, t_1)x(t_1^+) + \int_{t_1}^{t} U(t, s) f(s, x(s), x(s - \alpha))ds,$$

$$x(t_1^+) = I_1(x(t_1)) + \int_{-\infty}^{t_1} U(t_1, s) f(s, x(s), x(s - \alpha))ds.$$

Differentiating both sides of the above relation with respect to t, we obtain

$$\begin{aligned}
\frac{dx(t)}{dt} &= \frac{\partial U(t,t_1)}{\partial t}x(t_1^+) + \int_{t_1}^{t}\frac{\partial U(t,s)}{\partial t}f(s,x(s),x(s-\alpha))ds \\
&\quad + f(t,x(t),x(t-\alpha)), \\
x(t_1^+) &= A_1x(t_1) + I_1(x(t_1)) + \int_{-\infty}^{t_1}U(t_1,s)f(s,x(s),x(s-\alpha))ds \\
\Leftrightarrow \frac{dx(t)}{dt} &= A(t)\Big(U(t,t_1)x(t_1^+) + \int_{t_1}^{t}U(t,s)f(s,x(s),x(s-\alpha))ds\Big) \\
&\quad + f(t,x(t),x(t-\alpha)), \\
\Delta x(t_1) &= A_1x(t_1) + I_1(x(t_1)) \\
\Leftrightarrow \frac{dx(t)}{dt} &= A(t)x(t) + f(t,x(t),x(t-\alpha)), \quad \Delta x(t_1) = A_1x(t_1) + I_1(x(t_1)).
\end{aligned}$$

⋮

For $t \in (t_k, t_{k+1}]$, define

$$\begin{aligned}
x(t) &= \int_{-\infty}^{t}U(t,s)f(s,x(s),x(s-\alpha))ds + U(t,t_k)(I_1x(t_k)) \\
\Leftrightarrow x(t) &= U(t,t_k)(I_k(x(t_k)) + \int_{-\infty}^{t_k}U(t_k,s)f(s,x(s),x(s-\alpha))ds \\
&\quad + \int_{t_k}^{t}U(t,s)f(s,x(s),x(s-\alpha))ds \\
\Leftrightarrow x(t) &= U(t,t_k)(x(t_k^+) + \int_{t_k}^{t}U(t,s)f(s,x(s),x(s-\alpha))ds, \\
x(t_k^+) &= I_k(x(t_k)) + \int_{-\infty}^{t_k}U(t_k,s)f(s,x(s),x(s-\alpha))ds.
\end{aligned}$$

Again differentiating both sides of the above relation with respect to t, we get

$$\begin{aligned}
\frac{dx(t)}{dt} &= \frac{\partial U(t,t_k)}{\partial t}(x(t_k^+) + \int_{t_k}^{t}\frac{\partial U(t,s)}{\partial t}f(s,x(s),x(s-\alpha))ds \\
&\quad + f(t,x(t),x(t-\alpha)), \\
x(t_k^+) &= A_kx(t_k) + I_k(x(t_k)) + \int_{-\infty}^{t_k}U(t_k,s)f(s,x(s),x(s-\alpha))ds \\
\Leftrightarrow \frac{dx(t)}{dt} &= A(t)\Big(U(t,t_k)x(t_k^+) + \int_{t_k}^{t}U(t,s)f(s,x(s))ds\Big) + f(t,x(t),x(t-\alpha))
\end{aligned}$$

$$\Delta x(t_k) = A_k x(t_k) + I_k(x(t_k)),$$

$$\Leftrightarrow \frac{dx(t)}{dt} = A(t)x(t) + f(t, x(t), x(t-\alpha)) \quad \Delta x(t_k) = A_k x(t_k) + I_k(x(t_k)).$$

$\vdots$

Similarly the result holds for any interval $(t_l, t_{l+1}]$. □

Lemma 4.4.2 *If $f : \mathbb{R} \to \mathbb{R}^n$ is a Stepanov-like almost automorphic function, then $\int_{-\infty}^{t} U(t,s)f(s)ds + \sum_{t>t_k} U(t,t_k)I_k(x(t_k))$ is Stepanov-like almost automorphic.*

Proof Since f is Stepanov-like almost automorphic, for each sequence $\{t_n\}$ there exist a subsequence $\{t_{n_k}\}$ and function g such that

$$\lim_{k\to\infty} f(t+t_{n_k}) = g(t), \quad \lim_{k\to\infty} g(t-t_{n_k}) = f(t) \quad \forall t \in \mathbb{R} \text{ in } L^p(\mathbb{R}, \mathbb{R}^n).$$

We define

$$F(t) = \int_{-\infty}^{t} U(t,s)f(s)ds + \sum_{t>t_k} U(t,t_k)I_k(x(t_k))$$

and

$$G(t) = \int_{-\infty}^{t} U(t,s)g(s)ds + \sum_{t>t_k} U(t,t_k)I_k^*(x(t_k)).$$

Using continuity of $U(t,s)$ and Lebesgue's dominated convergence theorem, we obtain

$$\int_{-\infty}^{t} U(t,s)f(s+t_{n_k})ds \to \int_{-\infty}^{t} U(t,s)g(s)ds \text{ in } L^p(\mathbb{R}, \mathbb{R}^n). \tag{4.4.3}$$

Moreover,

$$\sum_{t+t_{n_k}>t_k} U(t+t_{n_k}, t_k)I_k(x(t_k)) = \sum_{t>t_k} U(t+t_{n_k}, t_k+t_{n_k})I_k(x(t_k+t_{n_k}))$$

$$\to \sum_{t>t_k} U(t,t_k)I_k^*(x(t_k)) \text{ in } L^p(\mathbb{R}, \mathbb{R}^n). \tag{4.4.4}$$

Thus using Eqs. (4.4.3) and (4.4.4), we get

$$\lim_{k\to\infty} \left(\int_0^1 \|F(t+t_{n_k}+s) - G(t+s)\|^p ds \right)^{\frac{1}{p}} = 0 \text{ in } \quad \forall t \in \mathbb{R}.$$

Similarly, we can prove that

$$\lim_{k\to\infty}\left(\int_0^1 \|G(t-t_{n_k}+s)-F(t+s)\|^p ds\right)^{\frac{1}{p}}=0 \quad \forall t\in\mathbb{R}.$$

Hence F is piecewise Stepanov-like almost automorphic. □

Theorem 4.4.3 *Under the hypotheses (H1)–(H7), there exists a unique piecewise continuous Stepanov-like almost automorphic solution of Eq. (4.1.1) provided*

$$K\left(\max_i\left(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*\right)L^*(p\delta)^{-\frac{1}{p}} + L(1-e^{-p\delta})^{-\frac{1}{p}}\right) < 1.$$

Proof Define the operator

$$\Lambda\phi(t)=\int_{-\infty}^t U(t,s)f(s,\phi(s),\phi(s-\alpha))ds+\sum_{t>t_k}U(t,t_k)I_k(\phi(t_k)).$$

we denote $B\subset S^pAA_{pc}(\mathbb{R},\mathbb{R}^n)$, the set of all Stepanov-like almost automorphic functions satisfying $\|\phi\|_{S^p}\le K_1$, where $\|\phi\|_{S^p}=\sup_{t\in\mathbb{R}}(\int_t^{t+1}\|\phi(s)\|^p ds)^{\frac{1}{p}}$ and $K_1 = KC\Big((p\delta)^{-\frac{1}{p}}+(1-e^{-p\delta})^{-\frac{1}{p}}\Big)$. Using composition theorem, it is not difficult to see that $\Lambda\phi$ is Stepanov-like almost automorphic as ϕ is Stepanov-like almost automorphic. As the function $f\in S^pAA_{pc}(\mathbb{R}\times\mathbb{R}^n,\mathbb{R}^n)$, define $u(\cdot)=f(\cdot,x(\cdot),x(\cdot-\alpha))$. Again using composition Theorem 4.3.4 and Lemma 4.4.2, we conclude

$$\Lambda_1\phi=\int_{-\infty}^t U(t,s)f(s,\phi(s),\phi(s-\alpha))ds\in S^pAA_{pc}(\mathbb{R},\mathbb{R}^n).$$

Further using Stepanov-like almost automorphy of sequence $I_k\in C(\mathbb{R}^n,\mathbb{R}^n)$, we obtain

$$\begin{aligned}\sum_{t_k<t+t_{n_k}}U(t+t_{n_k},t_k)I_k(\phi(t_k)) &= \sum_{t_k<t}U(t+t_{n_k},t_k+t_{n_k})I_k(\phi(t_k+t_{n_k}))\\ &\to \sum_{t_k<t}U(t,t_k)I_k^*(\phi(t_k)) \text{ in } L^p(\mathbb{R},\mathbb{R}^n).\end{aligned}$$

Similarly

$$\begin{aligned}\sum_{t_k<t-t_{n_k}}U(t-t_{n_k},t_k)(I_k^*\phi(\tau_k)) &= \sum_{t_k<t}U(t-t_{n_k},t_k-t_{n_k})(I_k^*\phi(t_k-t_{n_k}))\\ &\to \sum_{t_k<t}U(t,t_k)(I_k(\phi(t_k)) \text{ in } L^p(\mathbb{R},\mathbb{R}^n).\end{aligned}$$

The above analysis implies $\Lambda\phi \in S^p AA_{pc}(\mathbb{R}, \mathbb{R}^n)$.

Let us denote

$$B \supset B^* = \Big\{\phi \in B : \|\phi\|_{S^p} \leq \frac{rK_1}{1-r}\Big\},$$

where

$$\phi_0(t) = \int_{-\infty}^{t} U(t,s)\gamma(s)ds + \sum_{t_k<t} U(t,t_k)\gamma_k.$$

Now first we calculate the norm of ϕ_0, which is as follows:

$$\begin{aligned}
&\|\phi_0\|_{S^p}\\
&= \sup_{t\in\mathbb{R}}\Big(\int_t^{t+1} \|\int_{-\infty}^{s} U(s,z)\gamma(z)dz\|^p ds\Big)^{\frac{1}{p}} + \sup_{t\in\mathbb{R}}\Big(\int_t^{t+1} \|\sum_{t_k<s} U(s,t_k)\gamma_k\|^p ds\Big)^{\frac{1}{p}}\\
&\leq \sup_{t\in\mathbb{R}}\Big(\int_t^{t+1} \|\int_0^{\infty} U(s,s-z)\gamma(s-z)dz\|^p ds\Big)^{\frac{1}{p}} + \sup_{t\in\mathbb{R}}\Big(\int_t^{t+1} \sum_{t_k<s} \|U(s,t_k)\|^p\\
&\times \|\gamma_k\|^p ds\Big)^{\frac{1}{p}}\\
&\leq K \sup_{t\in\mathbb{R}}\Big(\int_t^{t+1} \int_0^{\infty} e^{-p\delta z}\|\gamma(s-z)dz\|^p ds\Big)^{\frac{1}{p}} + K \sup_{t\in\mathbb{R}}\Big(\int_t^{t+1} \sum_{t_k<s} e^{-p\delta(s-t_k)}\\
&\times \|\gamma_k\|^p ds\Big)^{\frac{1}{p}}\\
&\leq \|\gamma\|_{S^p} K \sup_{t\in\mathbb{R}}\Big(\int_0^{\infty} e^{-p\delta z}dz\Big)^{\frac{1}{p}} + \|\gamma_k\| K \sup_{t\in\mathbb{R}}\Big(\int_t^{t+1} \sum_{t_k<s} e^{-p\delta(s-t_k)}ds\Big)^{\frac{1}{p}}\\
&\leq KC\Big((p\delta)^{-\frac{1}{p}} + (1-e^{-p\delta})^{-\frac{1}{p}}\Big) = K_1.
\end{aligned} \tag{4.4.5}$$

Hence for any $\phi \in B^*$, we get

$$\|\phi\|_{S^p} \leq \|\phi - \phi_0\|_{S^p} + \|\phi_0\|_{S^p} \leq \frac{rK_1}{1-r} + K_1 = \frac{K_1}{1-r}.$$

Our next aim is to prove that Λ maps set B^* to B^*.

In order to achieve this, let us first observe that

$$
\begin{aligned}
&\|\Lambda\phi - \phi_0\|_{S^p} \\
&\quad \le \sup_{t\in\mathbb{R}} \Big(\int_t^{t+1} \| \int_{-\infty}^{s} U(s,z) f(z,\phi(z),\phi(z-\alpha)) dz \|^p ds \Big)^{\frac{1}{p}} \\
&\qquad + \sup_{t\in\mathbb{R}} \Big(\int_t^{t+1} \| \sum_{t_k<s} U(s,t_k) \times I_k(\phi(t_k)) \|^p ds \Big)^{\frac{1}{p}} \\
&\quad \le \sup_{t\in\mathbb{R}} \Big(\int_t^{t+1} \max_i \int_{-\infty}^{s} \|U(s,z)\|^p \sum_{j=1}^{n} \alpha_{ij}^{*} \| f_j(\phi_j(s-z))\|^p dz ds \Big)^{\frac{1}{p}} \\
&\qquad + \sup_{t\in\mathbb{R}} \Big(\int_t^{t+1} \max_i \int_{-\infty}^{s} \|U(s,z)\|^p \sum_{j=1}^{n} \beta_{ij}^{*} \| f_j(\phi_j(s-z-\alpha))\|^p dz ds \Big)^{\frac{1}{p}} \\
&\qquad + \sup_{t\in\mathbb{R}} \Big(\int_t^{t+1} \sum_{t_k<s} \|U(s,t_k)\|^p \|I_k(\phi(t_k))\|^p ds \Big)^{\frac{1}{p}} \\
&\quad \le \sup_{t\in\mathbb{R}} \Big(\int_t^{t+1} \max_i \int_{-\infty}^{s} \|U(s,z)\|^p \sum_{j=1}^{n} \alpha_{ij}^{*} ((L^*)^p \|\phi_j(s-z))\|^p \\
&\qquad + \| f_j(0)\|^p) dz ds \Big)^{\frac{1}{p}} \\
&\qquad + \sup_{t\in\mathbb{R}} \Big(\int_t^{t+1} \max_i \int_{-\infty}^{s} \|U(s,z)\|^p \| f_j(0)\|^p \sum_{j=1}^{n} \beta_{ij}^{*} ((L^*)^p \|\phi_j(s-z-\alpha))\|^p \\
&\qquad + \| f_j(0)\|^p) dz ds \Big)^{\frac{1}{p}} + \sup_{t\in\mathbb{R}} \Big(\int_t^{t+1} \sum_{t_k<s} \|U(s,t_k)\|^p (L^p \|\phi(t_k)\|^p \\
&\qquad\qquad + \|I_k(0)\|^p) ds \Big)^{\frac{1}{p}},
\end{aligned}
$$

where $L^* = \max\{L_i, i = 1, 2, \cdots, n\}$. In order to have zero as an equilibrium solution of the system (4.1.1), we assume that $f_j(0) = I_k(0) = 0$. Thus we have

$$
\begin{aligned}
&\|\Lambda\phi - \phi_0\| \\
&\quad \le K \left(\max_i \left(\sum_{j=1}^{n} \alpha_{ij}^{*} + \sum_{j=1}^{n} \beta_{ij}^{*} \right) L^* \Big(\int_0^{\infty} e^{-p\delta z} dz \Big)^{\frac{1}{p}} \right. \\
&\qquad \left. + L \Big(\sum_{t_k<s} e^{-p\delta(s-t_k)} \Big)^{\frac{1}{p}} \right) \|\phi\|_{S^p}
\end{aligned}
$$

$$\leq K\left(\max_i\left(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*\right) L^*(p\delta)^{-\frac{1}{p}} + L(1-e^{-p\delta})^{-\frac{1}{p}}\right)\|\phi\|_{S^p}$$

$$= r\|\phi\|_{S^p} \leq \frac{rK_1}{1-r}. \tag{4.4.6}$$

Thus we conclude that $\Lambda\phi \in B^*$.

Now we prove that Λ is a contraction. For any $\phi_1, \phi_2 \in B^*$, we obtain

$$\begin{aligned}
&\|\Lambda\phi_1 - \Lambda\phi_2\|_{S^p} \\
&\quad \leq \sup_{t\in\mathbb{R}}\Big(\int_t^{t+1}\|\int_{-\infty}^s U(s,z)(f(z,\phi_1(z),\phi_1(z-\alpha)) \\
&\qquad -f(z,\phi_2(z),\phi_2(z-\alpha)))dz\|^p ds\Big)^{\frac{1}{p}} \\
&\qquad + \sup_{t\in\mathbb{R}}\Big(\int_t^{t+1}\|\sum_{t_k<s} U(s,t_k)(I_k(\phi_1(t_k)) - I_k(\phi_2(t_k))\|^p ds\Big)^{\frac{1}{p}} \\
&\quad \leq K\left(\max_i\left(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*\right) L^*\Big(\int_0^\infty e^{-p\delta z}dz\Big)^{\frac{1}{p}}\right. \\
&\qquad \left. + L\Big(\sum_{t_k<s} e^{-p\delta(s-t_k)}\Big)^{\frac{1}{p}}\right)\|\phi_1 - \phi_2\|_{S^p} \\
&\quad \leq K\left(\max_i\left(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*\right) L^*(p\delta)^{-\frac{1}{p}} + L(1-e^{-p\delta})^{-\frac{1}{p}}\right)\|\phi_1 - \phi_2\|_{S^p} \\
&\quad = r\|\phi_1 - \phi_2\|_{S^p}.
\end{aligned}$$

Using the assumptions, we obtain

$$r = K\left(\max_i\left(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*\right) L^*(p\delta)^{-\frac{1}{p}} + L(1-e^{-p\delta})^{-\frac{1}{p}}\right) < 1.$$

Thus the mapping Λ is a contraction. Hence using Banach contraction principle, we conclude that there exists a unique piecewise continuous Stepanov-like almost automorphic solution of Problem (4.1.1). □

Our next theorem is about asymptotic stability of the system (4.1.1).

Theorem 4.4.4 *Under the hypotheses (H1)–(H7), the solution of the system (4.1.1) is asymptotically stable provided*

$$p^2\delta^2 > 8K^p L^{*p}\left(\max_i\left(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*\right)\right).$$

Proof For any two solutions $x(t)$ and $y(t)$ of the system (4.1.1) with initial values x_0 and y_0, we define $V(t) = x(t) - y(t)$. Using the property $(\|x\| + \|y\|)^p \leq 2^{p-1}(\|x\|^p + \|y\|^p)$ and calculating $p-$th norm of $V(t)$, we obtain

$$\begin{aligned}
\|V(t)\|^p &= \|x(t) - y(t)\|^p \\
&\leq 2^{p-1}\big[2^{p-1}\|U(t,0)\|^p\|x_0 - y_0\|^p \\
&\quad + \int_0^t \|U(t,s)\|^p\|f(s,x(s),x(s-\alpha)) - f(s,y(s),y(s-\alpha))\|^p ds \\
&\quad + 2^{p-1}\sum_{0<t_k<t}\|U(t,t_k)\|^p\|I_k(x(t_k)) - I_k(y(t_k))\|^p\big], \\
&\leq 2^{p-1}\big[2^{p-1}K^p e^{-p\delta t}\|x_0 - y_0\|^p + K^p\int_0^t e^{-\frac{p\delta(t-s)}{2}}ds \\
&\quad \times \int_0^t e^{-\frac{p\delta(t-s)}{2}}\|f(s,x(s),x(s-\alpha)) - f(s,y(s),y(s-\alpha))\|^p ds \\
&\quad + 2^{p-1}\sum_{0<t_k<t}\|U(t,t_k)\|^p\|I_k(x(t_k)) - I_k(y(t_k))\|^p\big] \\
&\leq 2^{p-1}\big[2^{p-1}K^p e^{-p\delta t}\|x_0 - y_0\|^p \\
&\quad + 2\frac{K^p L^{*p}\left(\max_i\left(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*\right)\right)}{p\delta} \\
&\quad \times \int_0^t e^{-\frac{p\delta(t-s)}{2}}\|x(s) - y(s)\|^p ds \\
&\quad + 2^{p-1}\sum_{0<t_k<t}K^p L^p e^{-p\delta(t-t_k)}\|x(t_k) - y(t_k)\|^p\big].
\end{aligned}$$

From the assumption

$$p^2\delta^2 > 2^p K^p L^{*p}\left(\max_i\left(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*\right)\right),$$

there exists an $\epsilon \in (0, \delta)$ such that

$$p\delta\left(\frac{p\delta}{2}-\epsilon\right) > 2^{p-1}K^pL^{*p}\left(\max_i\left(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*\right)\right).$$

We further define $X(t) = \|x(t) - y(t)\|^p e^{\epsilon t}$. Integrating both sides of $X(t)$, we obtain

$$\int_0^\tau X(s)ds \le \frac{2^{2p-2}K^p}{p\delta-\epsilon}X(0) + \frac{2^{p-1}K^pL^{*p}(\max_i(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*)}{p\delta(\frac{p\delta}{2}-\epsilon)} \times$$
$$\times \int_0^\tau X(s)ds + \sum_{0<t_k<\tau} \frac{2^{2p-2}K^pL^p}{p\delta-\epsilon}X(t_k)$$

$$\int_0^\tau X(s)ds \le \frac{p\delta(\frac{p\delta}{2}-\epsilon)}{p\delta(\frac{p\delta}{2}-\epsilon) - 2^{p-1}K^pL^{*p}(\max_i(\sum_{j=1}^n \alpha_{ij}^* + \sum_{j=1}^n \beta_{ij}^*)} \times$$
$$\times \left(\frac{2^{2p-2}K^p}{p\delta-\epsilon} + \left(1+\frac{2^{2p-2}K^pL^p}{p\delta-\epsilon}\right)^{i(0,\tau)}\right)X(0). \tag{4.4.7}$$

Here $i(0, \tau)$ is the number of points t_k in the interval $(0, \tau)$ and the product $\prod_{0<t_k<\tau}\left(1+\frac{2^{2p-2}K^pL^p}{p\delta-\epsilon}\right) = \left(1+\frac{2^{2p-2}K^pL^p}{p\delta-\epsilon}\right)^{i(0,\tau)}$ is convergent because of $\left(1+\frac{2^{2p-2}K^pL^p}{p\delta-\epsilon}\right) \le \left(1+L^p\right)^{\frac{2^{2p-2}K^p}{p\delta-\epsilon}}$.
Since RHS of inequality (4.4.7) is independent of $\tau \in [0, T)$ as well as of T, and hence the LHS integral of inequality (4.4.7) exists in $[0, \infty)$. In particular, we have

$$X(t) \to 0 \text{ as } t \to \infty.$$

Eventually, the Stepanov-like almost automorphic solution is asymptotically stable. □

4.5 Examples

As an example of Problem (4.1.1), consider the following classical model of Hopefield neural network model,

$$\frac{dx_i(t)}{dt} = -a_i(t)x_i(t) + \sum_{j=1}^n \alpha_{ij} f_j(x_j(t)) + \sum_{j=1}^n \beta_{ij} f_j(x_j(t-\alpha)) + \gamma_i(t), t \ne t_k,$$
$$\Delta(x(t_k)) = A_k x(t_k) + I_k(x(t_k)) + \gamma_k,$$

$$x(t_k - 0) = x(t_k), \quad x(t_k + 0) = x(t_k) + \Delta x(t_k), \quad k \in \mathbb{Z},\ t \in \mathbb{R},$$
$$x(t) = \phi_0(t), \quad t \in [-\alpha, 0],\ \alpha > 0, \tag{4.5.1}$$

where $a_i,\ f_j, \gamma_i \in \mathcal{C}(\mathbb{R}, \mathbb{R}),\ \alpha_{ij}, \beta_{ij} \in \mathbb{R}$ for $i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, n$. The coefficient $A_k \in \mathbb{R}^{n\times n}$, the function $I_k(x) \in \mathcal{C}(\Omega, \mathbb{R}^n)$ and the constant $\gamma_k \in \mathbb{R}^n$, where Ω a domain in $\mathbb{R}^n$. In this case our matrix $A(t)$ is a diagonal matrix with diagonal entire $-a_1(t), \cdots, -a_n(t)$. We assume that $a_i(t)$ are Stepanov-like almost automorphic and choose $a_i(t) = 1$ for each $i = 1, 2, \cdots, n$. One can easily verify the hypotheses $(H1)$ and $(H2)$ for this case and we assume the hypothesis $(H3)$. Now under all the conditions of Theorem 4.4.3, there exists a Stepanov-like almost automorphic solution of the Problem (4.5.1).

Let us choose the following set of parameters for the Problem (4.5.1) in $\mathbb{R}^2$:

$$a_1(t) = signum(\cos 2\pi t\theta),\ \beta_{12} = 0.2,\ \gamma_1(t) = 2\sin\sqrt{2t},$$
$$a_2(t) = \cos\left(\frac{1}{2 + \sin(t) + \sin(\sqrt{2}t)}\right),\ \beta_{21} = signum(\cos 2\pi t\theta),\ \gamma_2(t) = c(t),$$
$$A_k = \begin{pmatrix} -0.3 & 0 \\ 0 & -0.3 \end{pmatrix},$$
$$I_k(x) = 0.9|x|,\ x_1(s) = 1 = x_2(s), s \in [-0.1, 0],\ \gamma_k = \begin{pmatrix} 0.25 \\ 0.25 \end{pmatrix}.$$

These parameters clearly satisfy the conditions of our Theorem 4.4.3. The graph of the solution of (4.5.1) corresponding to these parametric values is depicted in Fig. 4.1. It can be easily seen that the nature of the graph is Stepanov almost automorphic.

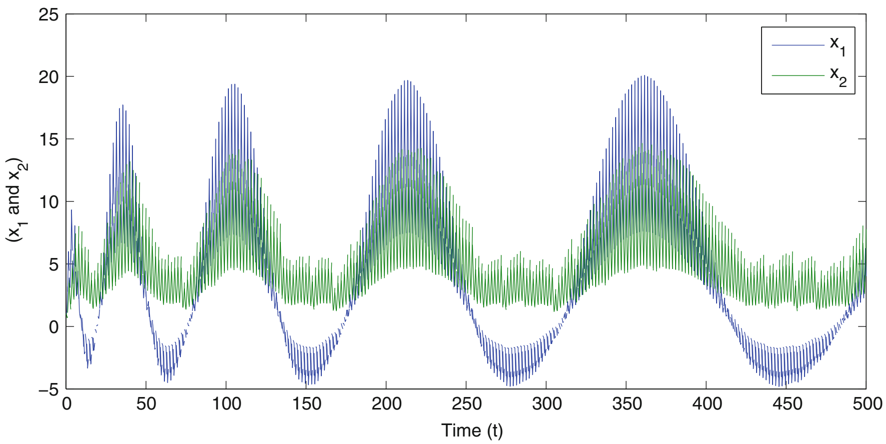

Fig. 4.1 Stepanov-like almost automorphic solution of 4.5.1

4.6 Discussion

The class of Stepanov-like almost automorphic functions covers larger class of functions and hence more complicated behaviour can be expressed in terms of these functions. It already contains the class of almost periodicity, automorphy as a subclass and hence it is more general in nature. One natural question one can always ask in the neural network theory is that what will be the nature of the output when all the parameters are Stepanov-like almost automorphic. In this work, we answered this question under certain condition. The asymptotic stability of solution is also established under certain conditions on the parameters. One can easily see the truth of this claim in the numerical simulation section. The obtained results can be easily generalized to other general class of systems such as neutral system, integro-differential system and systems with deviated arguments.

Acknowledgements We would like to thank the anonymous referee for his/her constructive comments and suggestions.

References

1. S. Abbas, A note on Weyl pseudo almost automorphic functions and their properties. Math. Sci. (Springer), **6**, 5 (2012). Art. 29
2. S. Abbas, Y.K. Chang, M. Hafayed, Stepanov type weighted pseudo almost automorphic sequences and their applications to difference equations. Nonlinear Stud. **21**(1), 99–111 (2014)
3. S. Abbas, L. Mahto, M. Hafayed, A.M. Alimi, Asymptotic almost automorphic solutions of impulsive neural network with almost automorphic coefficients. Neurocomputing **144**, 326–334 (2014)
4. S. Abbas, V. Kavitha, R. Murugesu, Stepanov-like weighted pseudo almost automorphic solutions to fractional order abstract integro-differential equations. Proc. Indian Acad. Sci. Math. Sci. **125**(3), 323–351 (2015)
5. S. Ahmad, I.M. Stamova, Global exponential stability for impulsive cellular neural networks with time-varying delays. Nonlinear Anal. **69**(3), 786–795 (2008)
6. W. Allegretto, D. Papini, M. Forti, Common asymptotic behavior of solutions and almost periodicity for discontinuous, delayed, and impulsive neural networks. IEEE Trans. Neural Netw. **21**(7), 1110–1125 (2010)
7. B. Ammar, F. Cherif, A.M. Alimi, Existence and uniqueness of pseudo almost-periodic solutions of recurrent neural networks with time-varying coefficients and mixed delays. IEEE Trans. Neural Netw. Learning Sys. **23**(1), 109–118 (2012)
8. D.D. Bainov, P.S. Simeonov, *Systems with Impulsive Effects* (Ellis Horwood Limited/John Wiley & Sons, Chichester, 1989)
9. D.D. Bainov, P.S. Simeonov, *Impulsive Differential Equations: Periodic Solutions and Its Applications* (Longman Scientific and Technical Group, England, 1993)
10. S. Bochner, Continuous mappings of almost automorphic and almost periodic functions. Proc. Nat. Acad. Sci. U.S.A. **52**, 907–910 (1964)
11. H. Bohr, *Almost-Periodic Functions* (Chelsea Publishing Company, New York City, 1947)
12. A. Chavez, S. Castiallo, M. Pinto, Discontinuous almost automorphic functions and almost solutions of differential equations with piecewise constant argument. Electron. J. Differ. Equ. **2014**(56), 1–13 (2014)

13. T. Diagana, *Pseudo Almost Periodic Functions in Banach Spaces* (Nova Science, Hauppauge, 2007)
14. T. Diagana, E. Hernndez, J.C. Santos, Existence of asymptotically almost automorphic solutions to some abstract partial neutral integro-differential equations. Nonlinear Anal. **(71)**, 248–257 (2009)
15. V. Kavitha, S. Abbas, R. Murugesu, Existence of Stepanov-like weighted pseudo almost automorphic solutions of fractional integro-differential equations via measure theory. Nonlinear Stud. **24**(4), 825–850 (2017)
16. H.X. Li, L.L. Zhang, Stepanov-like pseudo-almost periodicity and semilinear differential equations with uniform continuity. Results Math. **59**(1–2), 43–61 (2011)
17. J. Liu, C. Zhang, Composition of piecewise pseudo almost periodic functions and applications to abstract impulsive differential equations. Adv. Differ. Equ. **2013**(11), 21 (2013)
18. L. Mahto, S. Abbas, PC-almost automorphic solution of impulsive fractional differential equations. Mediterr. J. Math. **12**(3), 771–790 (2014)
19. L. Mahto, S. Abbas, A. Favini, Analysis of Caputo impulsive fractional order differential equations with applications. Int. J. Differ. Equ. **2013**, 1–11 (2013)
20. G.M. Mophou, G.M. N'Guérékata, On some classes of almost automorphic functions and applications to fractional differential equations. Comput. Math. Appl. **59**, 1310–1317 (2010)
21. G.M. N'Guérékata, *Topics in Almost Automorphy* (Springer, New York, 2005)
22. G.M. N'Guérékata, A. Pankov, Integral operators in spaces of bounded, almost periodic and almost automorphic functions. Differ. Integral Equ. **21**(11–12), 1155–1176 (2008)
23. A. Pankov, *Bounded and Almost Periodic Solutions of Nonlinear Operator Differential Equations* (Kluwer, Dordrecht, 1990)
24. A.M. Samoilenko, N.A. Perestyuk, *Differential Equations with Impulse Effects* (Viska Skola, Kiev, 1987) (in Russian)
25. R. Samidurai, S.M. Anthoni, K. Balachandran, Global exponential stability of neutral-type impulsive neural networks with discrete and distributed delays. Nonlinear Anal. Hybrid Syst. **4**(1), 103–112 (2010)
26. M. Sannay, Exponential stability in Hopfield-type neural networks with impulses. Chaos, Solitons Fractals **32**(2), 456–467 (2007)
27. G.T. Stamov, Impulsive cellular neural networks and almost periodicity. Proc. Jpn. Acad. Sci. 80, Ser. A, **10**, 198–203 (2005)
28. G.T. Stamov, *Almost Periodic Solutions of Impulsive Differential Equations*. Lecture Notes in Mathematics, vol. 2047 (Springer, Heidelberg, 2012), pp. xx+217. ISBN: 978-3-642-27545-6
29. I.M. Stamova, R. Ilarionov, On global exponential stability for impulsive cellular neural networks with time-varying delays. Comput. Math. Appl. **59**(11), 3508–3515 (2010)
30. I.M. Stamova, G.T. Stamov, Impulsive control on global asymptotic stability for a class of impulsive bidirectional associative memory neural networks with distributed delays. Math. Comput. Model. **53**(5–6), 824–831 (2011)
31. G.T. Stamov, I.M. Stamova, J.O. Alzabut, Existence of almost periodic solutions for strongly stable nonlinear impulsive differential-difference equations. Nonlinear Anal. Hybrid Syst. **6**, 818–823 (2012)

Chapter 5
On the Convergence of Secant-Like Methods

I. K. Argyros, M. A. Hernández-Verón, and M. J. Rubio

Abstract In this chapter, our first idea is to improve the speed of convergence of the Secant method by means of iterative processes free of derivatives of the operator in their algorithms. To achieve this, we consider a uniparametric family of Secant-like methods previously constructed. We analyze the semilocal convergence of this uniparametric family of iterative processes by using a technique that consists of a new system of recurrence relations.

Keywords Nonlinear equation · Non-differentiable operator · Divided difference · Iterative method · The Secant method · Local convergence · Semilocal convergence

5.1 Introduction

In this chapter, we deal with the problem of approximating a solution x^* of the equation

$$F(x) = 0, \tag{5.1}$$

where $F : \Omega \subseteq X \rightarrow Y$ is a nonlinear operator defined on a nonempty open and convex domain Ω of a Banach space X with values in a Banach space Y.

This research has been supported in part by the project MTM2014-52016-C2-1-P of the Spanish Ministry of Economy and Competitiveness.

I. K. Argyros
Dept. of Mathematics Sciences, Cameron University, Lawton, OK, USA
e-mail: iargyros@cameron.edu

M. A. Hernández-Verón (✉) · M. J. Rubio
Dept. of Mathematics and Computation, University of La Rioja, Logroño, Spain
e-mail: mahernan@unirioja.es; mjrubio@unirioja.es

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_5

Many scientific and engineering problems can be brought in the form of a nonlinear equation (5.1). Equation (5.1) can be a scalar equation, a system of equations, a differential equation, an integral equation, etc.

The solutions of these equations can be found in the closed form only in some cases. Hence, they are usually approximated by iterative methods of the form

$$\begin{cases} x_{-k}, x_{-k+1}, \ldots, x_{-1}, x_0, \quad \text{given in } \Omega, \\ x_{n+1} = \Phi(x_{n-k}, x_{n-k+1}, \ldots, x_{n-1}, x_n), \quad n \geq 0, \end{cases} \tag{5.2}$$

where $x_{-k}, x_{-k+1}, \ldots, x_{-1}, x_0$ are initial approximations to the solution x^* of (5.1). The most common problem when we use iterative methods is to find initial approximations $x_{-k}, x_{-k+1}, \ldots, x_{-1}, x_0$ is close enough, so that the sequence $\{x_n\}$ converges to x^*.

Three types of studies can be done when we are interested to prove the convergence of the sequence $\{x_n\}$ given by iterative method (5.2): local, semilocal, and global. Firstly, the local study of the convergence is based on demanding conditions to the solution x^*, from certain conditions on the operator F, and provides the so-called ball of convergence of (5.2) that shows the accessibility to x^* from initial approximations $x_{-k}, x_{-k+1}, \ldots, x_{-1}, x_0$ belonging to the ball. Secondly, the semilocal study of the convergence is based on demanding conditions to the initial approximations $x_{-k}, x_{-k+1}, \ldots, x_{-1}, x_0$, from certain conditions on the operator F, and provides the so-called domain of parameters corresponding to the conditions required to the initial approximations that guarantee the convergence of the sequence $\{x_n\}$, given by (5.2), to the solution x^*. Thirdly, the global study of the convergence guarantees, from certain conditions on the operator F, the convergence of the sequence $\{x_n\}$, given by (5.2), to the solution x^* independently of initial approximations.

As we just indicate, the three studies demand conditions on the operator F. However, requirement of conditions to the solution, to the initial approximations, or to none of these determines the different types of studies.

The local study of the convergence has the disadvantage of being able to guarantee that the solution, that is unknown, can satisfy certain conditions. In general, the global study of the convergence is very specific as regards the type of operators to consider, as a consequence of absence of conditions on the initial approximations and on the solution. This study focuses on the semilocal study of the convergence of $\{x_n\}$. Moreover, the last section also includes a local study that is based on a new technique which uses auxiliary points.

The Newton's method [31] is the most used iteration to solve (5.1), as a consequence of its computational efficiency, even though sometimes less speed of convergence is reached. But this method needs the existence of the first Fréchet derivative of the operator F. If we are concerned with approximating a solution x^* of the equation

$$F(x) = G(x) + H(x) = 0, \tag{5.3}$$

where $G, H : \Omega \subseteq X \to Y$, G is a Fréchet differentiable operator and H is a continuous operator but not necessarily differentiable, the Newton's method cannot be applied.

The study of this situation has been considered by several authors. For example, in [7] and [48] it is considered a modification of Newton's method given by

$$x_{n+1} = x_n - \left(G'(x_n)\right)^{-1} \left(G(x_n) + H(x_n)\right), \quad x_0 \in \Omega, \quad n \geq 0. \tag{5.4}$$

In [10], the author considers the iteration

$$x_{n+1} = x_n - (A(x_n))^{-1} \left(G(x_n) + H(x_n)\right), \quad x_0 \in \Omega, \quad n \geq 0, \tag{5.5}$$

where $A(x_n)$ denotes a linear operator which is an approximation of the Fréchet derivative of G evaluated at x_n. In Banach spaces, the divided differences of first order are commonly used as approximations of the Fréchet derivative of an operator. Remember that, if we denote the set of linear and bounded operators from X to Y by $\mathcal{L}(X, Y)$, then if there exists an operator $[x, y; F] \in \mathcal{L}(X, Y)$ such that the condition

$$[x, y; F](x - y) = F(x) - F(y) \tag{5.6}$$

is satisfied, this is a divided differences of first order of F at the points x and y (see [4, 38]). Condition (5.6) does not determine uniquely the divided difference, with the exception of the case when X is one-dimensional. For the existence of divided differences in linear spaces, see [16]. In general, if the operator F is not differentiable, there are several studies (see [1–3, 10, 26, 38, 42, 43]) where the Secant method is considered, with $A(x_n) = [x_{n-1}, x_n; F]$ which is, in (5.5), a divided difference of first order of F on the points $x_{n-1}, x_n \in \Omega$. This method is defined as an iteration which uses new information at two points, so that it is a multipoint method [22]. Their algorithm is given by

$$\begin{cases} x_{-1}, x_0 \in \Omega, \\ x_{n+1} = x_n - [x_{n-1}, x_n; F]^{-1} F(x_n). \end{cases} \tag{5.7}$$

Other well-known methods for solving (5.1) are the Newton-like methods [9, 11, 37]:

$$\begin{cases} x_0 \in \Omega, \\ x_{n+1} = x_n - L_n^{-1} F(x_n), \end{cases} \tag{5.8}$$

where $\{L_n\}$ denotes a sequence of invertible linear operators. These methods are used due to their high efficiency, since the speed of convergence is acceptable and the operational cost is reduced. The study of the convergence of methods (5.8) can be found in [20], where the basic assumption made is that F' is Lipschitz continuous in some ball around the initial iterate. Argyros [8] relaxes this requirement to operators that are only Hölder continuous. Moreover, the Secant method is examined as a particular case of (5.8). As we have written previously, the most inconvenience of Newton's iteration is the evaluation of the first derivative of the operator F at each step. The Secant method, which uses divided differences, is usually applied to solve the previous inconvenience. But, the speed of convergence is reduced. In this chapter, our idea is to improve the speed of convergence of the Secant method [6, 10] by means of iterative methods that maintain the feature of not using derivatives of the operator F in their algorithms [12–14]. We consider a uniparametric family of Secant-like methods constructed in [30], which is given by the following algorithm:

$$\begin{cases} x_{-1},\ x_0 \quad \text{pre-chosen}, \\ y_n = \lambda x_n + (1-\lambda)x_{n-1}, \quad \lambda \in [0,1], \\ x_{n+1} = x_n - [y_n, x_n; F]^{-1} F(x_n), \end{cases} \tag{5.9}$$

that can be considered as a combination of the Secant method ($\lambda = 0$) and Newton's method ($\lambda = 1$). The study of the semilocal convergence of the Secant method is usually made by means of majorizing sequences [8, 20, 37, 38, 42]. In this chapter, we analyze the semilocal convergence of (5.9) by using a technique that consists of a new system of recurrence relations, in the way that Gutiérrez and Hernández analyze the convergence of the Chebyshev method in [25].

As we have written previously, the semilocal convergence studies are based on conditions on the starting points and the operator involved. On the other hand, in the Secant-like methods appearing in (5.9), a divided difference of first order is only used as approximation of a Fréchet derivative of the operator involved, so that the semilocal convergence results are obtained from requiring conditions on such divided difference. As we can see in [27], the conditions depend on the differentiability of the operator. So, in Sect. 5.3, results for Fréchet differentiable operators are included; in Sect. 5.4, results for non-differentiable operators are given; and, in Sect. 5.5, results for any operator are obtained. Finally, the chapter finishes with two new results on the convergence of Secant-like methods (5.9), one local and other semilocal. Both results use auxiliary points and can be applied to any operator (Fréchet differentiable or not).

As already mentioned previously, there is a plethora of choices for the divided difference in the multi-dimensional case. Next, we list some used in this chapter:

1. To solve (5.3), we can use any divided difference satisfying (5.6) and compatible to the sufficient convergence criteria for method (5.4) or (5.5).
2. To apply method (5.9), the divided difference defined in (5.12) is used, which still satisfies (5.6) with x and y having some different component.
3. The divided differences in Sect. 5.4 are chosen to satisfy estimate (5.40).
4. In Sect. 5.5.2.1, the divided difference used is given in (5.12).
5. System (5.47) is solved using a special divided difference given above Table 5.9.

See also the numerical examples, where we use appropriately the various divided differences given in 1–5.

Throughout the chapter, we consider $F : \Omega \subseteq X \to Y$ as a continuous nonlinear operator and Ω a nonempty open convex domain in the Banach space X with values in the Banach space Y. Moreover, we suppose that there exists a divided difference of first order $[z, w; F] \in \mathcal{L}(X, Y)$ for each pair of distinct points $(z, w) \in \Omega \times \Omega$ and denote $\overline{B(x, \varrho)} = \{y \in X;\ \|y - x\| \leq \varrho\}$ and $B(x, \varrho) = \{y \in X;\ \|y - x\| < \varrho\}$, for $\varrho > 0$.

5.2 Preliminaries

The classical Secant method is an efficient algorithm for solving nonlinear operator equation (5.1) [8, 37]. As suggested by the Secant iteration formula (5.7), there are two main elements for applying this method, the smoothness properties of the operator F, and the use of the first order divided difference of the operator F, instead of the first derivative of F. It is well known that for smooth equations, the classical Secant method is superlinearly convergent with R-order at least $(1 + \sqrt{5})/2$ (see [26, 38]). Some Newton-like methods can be considered as generalized Secant methods, since they use only operator values. Considering methods based only on operator values, in this chapter, we consider the Secant-like methods given in (5.9).

Observe that (5.9) is reduced to the Secant method if $\lambda = 0$ and to Newton's method if $\lambda = 1$, since $y_n = x_n$ and $[y_n, x_n; F] = F'(x_n)$ (see [38]). From the geometrical interpretation of the both previous methods, in the real case, it is clear that the closer x_n and y_n are, the higher the speed of the convergence is (see Fig. 5.1). The use of the Secant method is interesting, since the calculation of the first derivative F' is not required and the convergence of the method of successive substitutions is improved, although it is slower than Newton's method. For this, we consider iteration (5.9), whose speed of convergence is closed to that of Newton's iteration, when λ is near 1 (the Newton process).

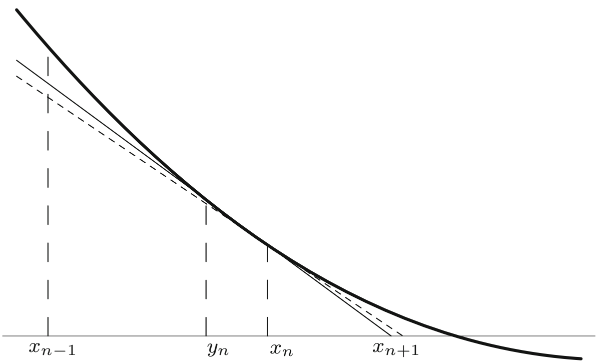

Fig. 5.1 Secant-like methods

Next, we present a numerical example to analyze the speed of convergence of Secant-like methods (5.9). So, we study the solution of a particular nonlinear integral equation of the Hammerstein type [24, 34], which is cited in [47]:

$$x(s) = 1 - \frac{\mu}{2}\int_0^1 \frac{s}{t+s}\frac{1}{x(t)}\,dt, \quad s \in [0,1], \quad \mu \in [0,1] \text{ fixed}, \tag{5.10}$$

in $\Omega = \{x \in C[0,1] : x \text{ is positive}\}$.

To continue, Eq. (5.10) is discretized to replace it with a finite-dimensional problem. For the direct numerical solution of (5.10), we choose $\mu = 1/2$ and introduce the points $t_j = j/m$ $(j = 0, 1, \ldots, m)$, where m is an integer according to the precision required. The composite trapezoidal rule with mesh size $1/m$ is used. A scheme is then designed for the determination of numbers $x(t_j)$. So, we obtain the nonlinear system of equations given by:

$$0 = x(t_j) - 1 + \frac{1}{4m}\left[\frac{1}{2}\frac{t_j}{t_j+t_0}\frac{1}{x(t_0)} + \sum_{k=1}^{m-1}\frac{t_j}{t_j+t_k}\frac{1}{x(t_k)} + \frac{1}{2}\frac{t_j}{t_j+t_m}\frac{1}{x(t_m)}\right]. \tag{5.11}$$

For $\mathbf{u} = (u_1, u_2, \ldots, u_m)^T$, $\mathbf{v} = (v_1, v_2, \ldots, v_m)^T \in \mathbb{R}^m$, with $u_j \neq v_j$ for $j = 1, 2, \ldots, m$, we use the divided difference of first order given by $[\mathbf{u}, \mathbf{v}; F] = ([\mathbf{u}, \mathbf{v}; F]_{ij})_{i,j=1}^m \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m)$, where

$$[\mathbf{u}, \mathbf{v}; F]_{i1} = \begin{cases} \dfrac{F_i(u_1, u_2, \ldots, v_m) - F_i(u_1, v_2, \ldots, v_m)}{u_1 - v_1}, & \text{if } u_2 \neq v_2. \\ 0, \ \text{if } u_1 = v_1. \end{cases}$$

$$[\mathbf{u}, \mathbf{v}; F]_{i2} = \begin{cases} \dfrac{F_i(u_1, u_2, \ldots, v_m) - F_i(u_1, v_2, \ldots, v_m)}{u_2 - v_2}, & \text{if } u_2 \neq v_2. \\ 0, \ \text{if } u_2 = v_2. \end{cases}$$

....

....

$$[\mathbf{u}, \mathbf{v}; F]_{im} = \begin{cases} \dfrac{F_i(u_1, u_2, \ldots, u_m) - F_i(u_1, u_2, \ldots, u_{m-1}, v_m)}{u_m - v_m}, & \text{if } u_m \neq v_m. \\ 0, \ \text{if } u_m = v_m. \end{cases} \tag{5.12}$$

It is easy to prove that $[x, y; F](x - y) = F(x) - F(y)$ with x and y having some different component.

Table 5.1 Numerical solution x^* of (5.11)

j	x_j^*	j	x_j^*	j	x_j^*
0	1.0000000000000000	7	0.8643075347402679	14	0.8141388910219971
1	0.9566527838959406	8	0.8555128396460574	15	0.8095584770363422
2	0.9319739617269809	9	0.8483515334503204	16	0.8121077060370004
3	0.9129724633617881	10	0.8416103236537321	17	0.8083909132037373
4	0.8977438552742314	11	0.8304716661779638	18	0.8049765496475915
5	0.8848953323716495	12	0.8245463960276492	19	0.8017901827237186
6	0.8738883606190559	13	0.8191213272527618	20	0.7988090566367779

Table 5.2 $x_{-1}(t_j) = 1.35$, $x_0(t_j) = 1$ $(j = 0, 1, \ldots, 20)$

n	$\lambda = 0$	$\lambda = 0.6$	$\lambda = 0.85$	$\lambda = 1 - 10^{-5}$
1	1.14120×10^{-2}	8.05623×10^{-3}	6.21652×10^{-3}	4.93910×10^{-3}
2	3.24264×10^{-4}	1.08921×10^{-4}	3.66835×10^{-5}	3.99254×10^{-6}
3	6.08624×10^{-7}	5.91357×10^{-8}	5.82438×10^{-9}	2.61646×10^{-12}
4	3.21652×10^{-11}	4.19664×10^{-13}	5.10703×10^{-15}	0.0

Table 5.3 Error for Newton's method using 16 significant decimal places

n	$\|x^* - z_n\|_\infty$
1	4.93901×10^{-3}
2	3.99072×10^{-6}
3	2.58171×10^{-12}
4	0.0

For the solution of (5.11), we take $x_0(t_j) = 1$ and $x_{-1}(t_j) = 1.35$ $(j = 1, 2, \ldots, m)$.

In Table 5.1 the approximation of the solution x^* of (5.11) is given, using 16 significant decimal places and $m = 20$, when the Secant method is applied to the previous scheme.

Table 5.2 contains the errors $\|x_n - x^*\|_\infty$ for the iterates x_n generated by (5.9) for different values of the parameter λ. The solution x^* is obtained in four steps $(n = 4)$, for a precision of 16 significant decimal places.

As we can see in Table 5.2, the higher λ is, the faster iteration (5.9) converges.

Finally, note that (5.9), where λ is near 1, gives similar approximations to those obtained by Newton's sequence $\{z_n\}$, without using F' (see Table 5.3).

As we can see in the previous numerical example, the speed of convergence of Secant-like methods (5.9) improves the speed of convergence of the Secant method (5.7). Moreover, observe that with $\lambda = 1 - 10^{-5}$ is similar to that of Newton's method but without evaluating the first derivative of the operator F.

To finish, we observe that the use of the Secant method is interesting since the calculation of the first Fréchet derivative F' is not required and the convergence of the successive substitutions method is improved, although it is slower than Newton's method. For this, we consider Secant-like methods (5.9), whose speed of convergence is closed to that of Newton's method when λ is near 1.

As we have indicated previously, to obtain results of semilocal convergence for Secant-like methods (5.9), conditions for the divided difference of first order of the operator F are necessary. So, remember that, if Ω is a open convex domain of X and we suppose that, for each pair of distinct points $x, y \in \Omega$, there exists a divided difference of first order of F at these points. If there exists a non-negative constant k such that

$$\|[x, y; F] - [v, w; F]\| \leq k\,(\|x - v\| + \|y - w\|) \tag{5.13}$$

for all $x, y, v, w \in \Omega$ with $x \neq y$ and $v \neq w$, we say that F has a Lipschitz continuous first order divided difference on Ω.

We can easily generalize this concept. So, If there exists a non-negative constant k such that

$$\|[x, y; F] - [v, w; F]\| \leq k\left(\|x - v\|^p + \|y - w\|^p\right), \quad p \in [0, 1], \tag{5.14}$$

for all $x, y, v, w \in \Omega$ with $x \neq y$ and $v \neq w$, we say that F has a Hölder continuous first order divided difference on Ω. Notice that, if $p = 1$, we obtain that F has a Lipschitz continuous divided difference on Ω.

In the previous case, it is known [8] that the Fréchet derivative of F exists in Ω and satisfies

$$[x, x; F] = F'(x), \quad x \in \Omega. \tag{5.15}$$

In this chapter we relax these previous requirements, (5.13) and (5.14), and we only assume that the divided difference $[x, y; F]$ satisfies

$$\|[x, y; F] - [v, w; F]\| \leq \omega(\|x - v\|, \|y - w\|); \quad x, y, v, w \in \Omega, \tag{5.16}$$

where $\omega : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous non-decreasing function in their components.

In the following lemma we will prove that (5.16) satisfies (5.15) if $\omega(0, 0) = 0$.

Lemma 5.1 *Let Ω be a convex open domain of X and suppose that, for each pair of points $x, y \in \Omega$, there exists a divided difference of first order $[x, y; F] \in \mathcal{L}(X, Y)$ satisfying (5.16) and $\omega(0, 0) = 0$. Then (5.15) is true.*

Proof Let $\{x_n\} \subseteq \Omega$ be so that $\lim_{n\to\infty} x_n = x$. Let us consider $A_n = [x_n, x; F] \in \mathcal{L}(X, Y)$ and it is verified that

$$\|A_n - A_m\| = \|[x_n, x; F] - [x_m, x; F]\| \leq \omega(\|x_n - x_m\|, 0).$$

Since $\{x_n\}$ is convergent, it is evident that $\{A_n\}$ is a Cauchy sequence, and therefore there exists $\lim_{n\to\infty} A_n = \tilde{A} \in \mathcal{L}(X, Y)$. So, we can define $[x, x; F] = \tilde{A} =$

$\lim_{n\to\infty} A_n$. Let us check that $\tilde{A} = F'(x)$:

$$\begin{aligned}
&\|F(x+\Delta x) - F(x) - [x, x; F]\,(\Delta x)\| \\
&\quad = \|[x+\Delta x, x; F]\,(\Delta x) - [x, x; F](\Delta x)\| \\
&\quad\quad = \|\,([x+\Delta x, x; F] - [x, x; F])\ (\Delta x)\| \\
&\quad\quad\quad \leq \|[x+\Delta x, x; F] - [x, x; F]\|\ \|(\Delta x)\| \leq \omega(\|\Delta x\|, 0)\|\Delta x\|
\end{aligned}$$

Then,

$$\lim_{\|\Delta x\|\to 0} \frac{\|F(x+\Delta x) - F(x) - [x, x; F](\Delta x)\|}{\|\Delta x\|} \leq \lim_{\|\Delta x\|\to 0} \omega(\|\Delta x\|, 0) = \omega(0, 0) = 0.$$

Therefore, F is a Fréchet differentiable operator and then verifies (5.15). □

It is easy to see that condition (5.16) generalizes condition (5.14) by only considering $\omega(u_1, u_2) = k(u_1^p + u_2^p)$.

5.3 Convergence of Secant-Like Methods for Fréchet Differentiable Operators

From Lemma 5.1, it is known that when the first order divided difference of operator F is Lipschitz or Hölder continuous, then F is Fréchet differentiable. In this section, we analyze the semilocal convergence for Secant-like methods (5.9) for divided differences of first order Lipschitz or Hölder continuous.

5.3.1 Divided Differences of First Order Lipschitz Continuous

Let X, Y be Banach spaces and $F : \Omega \subseteq X \to Y$ be a nonlinear operator in an open convex domain Ω. Suppose that:

(L1) $\|x_0 - x_{-1}\| = \alpha$,
(L2) there exists $L_0{}^{-1} = [y_0, x_0; F]^{-1}$ such that $\|L_0{}^{-1}\| \leq \beta$,
(L3) $\|L_0{}^{-1}F(x_0)\| \leq \eta$,
(L4) $\|[x, y; F] - [u, v; F]\| \leq k\ (\|x-u\| + \|y-v\|)$, $\quad k \geq 0,\ x, y, u, v \in \Omega;$ $x \neq y,\ u \neq v$.

Under these conditions, we establish a system of recurrence relations from which the convergence of (5.9) is proved. Let us denote

$$a_{-1} = \frac{\eta}{\alpha + \eta}, \quad b_{-1} = \frac{k\beta\alpha^2}{\alpha + \eta}.$$

Define the real sequences

$$a_n = f(a_{n-1})g(a_{n-1})b_{n-1}, \quad b_n = f(a_{n-1})^2 a_{n-1}b_{n-1}, \quad n \geq 0, \tag{5.17}$$

where

$$f(x) = \frac{1}{1-x} \qquad \text{and} \qquad g(x) = (1-\lambda) + (1+\lambda)f(x)x.$$

Note that f and g are increasing in $\mathbb{R} - \{1\}$ and that $f(x) > 1$ in $(0, 1)$.

From the initial hypotheses, it follows that x_1 is well defined, since ${L_0}^{-1}$ exists and

$$\begin{aligned} \|x_1 - x_0\| &= \|{L_0}^{-1}F(x_0)\| \leq \eta = f(a_{-1})a_{-1}\|x_0 - x_{-1}\|, \\ k\|{L_0}^{-1}\|\|x_0 - x_{-1}\| &\leq k\beta\alpha = f(a_{-1})b_{-1}. \end{aligned} \tag{5.18}$$

In [30], the following recurrence relations for $n \geq 1$ are shown by mathematical induction on n:

(i_n) there exists an ${L_n}^{-1} = [y_n, x_n; F]^{-1}$ such that $\|{L_n}^{-1}\| \leq f(a_{n-1})\|{L_{n-1}}^{-1}\|$,
(ii_n) $\|x_{n+1} - x_n\| \leq f(a_{n-1})a_{n-1}\|x_n - x_{n-1}\|$,
(iii_n) $k\|{L_n}^{-1}\|\|x_n - x_{n-1}\| \leq f(a_{n-1})b_{n-1}$.

To study the convergence of the sequence $\{x_n\}$, we analyze the sequences $\{a_n\}$ and $\{b_n\}$ given by (5.17). It is sufficient to see that $\{x_n\}$ is a Cauchy sequence and $a_n < 1$ for all $n \geq 0$.

Firstly, if we denote by $\{\alpha_n\}$ the Fibonacci sequence

$$\alpha_1 = \alpha_2 = 1 \qquad \text{and} \qquad \alpha_{n+2} = \alpha_{n+1} + \alpha_n, \quad n \geq 1, \tag{5.19}$$

and

$$s_n = \alpha_1 + \alpha_2 + \cdots + \alpha_n, \quad n \geq 1. \tag{5.20}$$

Then, the following properties can be proved, again by induction:

(P1) $\alpha_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right] > \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n-1}, \quad n \geq 1,$
(P2) $s_n = \alpha_{n+2} - 1$ and $\beta_n = s_1 + s_2 + \cdots + s_n = \alpha_{n+4} - (n+3), \quad n \geq 1.$

Secondly, some properties for the sequence $\{a_n\}$ and $\{b_n\}$ are proved in the following result.

Lemma 5.2 *Let $\{a_n\}$ and $\{b_n\}$ be the sequences defined in (5.17) and $\lambda \in [0, 1]$ be a fixed element. If $a_{-1} < (3-\sqrt{5})/2$ and $b_{-1} < a_{-1}(1-a_{-1})^2/(1+\lambda(2a_{-1}-1))$, then*

(a) $\{a_n\}$ *and* $\{b_n\}$ *are decreasing,*
(b) $\gamma = b_0/b_{-1} \in (0,1)$ *and* $a_0/(1-a_0) < \gamma$,
(c) $a_n < \gamma^{\alpha_n} a_{n-1}$ *and* $b_n < \gamma^{\alpha_{n+1}} b_{n-1}$, *for* $n \geq 1$,
(d) $a_n < \gamma^{s_n} a_0$, *for* $n \geq 1$.

The following semilocal convergence Theorem [30] shows that the sequence $\{x_n\}$ generated by (5.9) converges to a solution x^* of Eq. (5.1).

Theorem 5.3 *Let* $x_{-1}, x_0 \in \Omega$ *and* $\lambda \in [0,1]$. *Let us suppose that* **(L1)–(L4)** *and the hypotheses of Lemma 5.2 are satisfied. If* $\overline{B(x_0, r_0)} \subseteq \Omega$, *where* $r_0 = \frac{1-a_0}{1-2a_0}\eta$, *then the sequence* $\{x_n\}$ *generated by (5.9) is well defined and converges to a solution* x^* *of (5.3) with R-order of convergence of at least* $(1+\sqrt{5})/2$. *Moreover, it is proved that* $x_n, x^* \in \overline{B(x_0, r_0)}$ *and* x^* *is unique in* $B(x_0, \tau) \cap \Omega$, *where* $\tau = \frac{1}{\beta_k} - r_0 - (1-\lambda)\alpha$. *Furthermore, for all* $n \geq 0$,

$$\|x^* - x_n\| < \frac{\Delta^n}{1-\Delta}\, \eta\, \gamma^{\beta_{n-1}}, \tag{5.21}$$

where $\gamma = b_0/b_{-1}$, $\Delta = \frac{a_0}{1-a_0}$, $\beta_{-1} = 0 = \beta_0$ *and* $\beta_n = s_1 + s_2 + \cdots + s_n$, $n \geq 1$.

5.3.2 Divided Differences of First Order Hölder Continuous

Let X, Y be Banach spaces and $F : \Omega \subseteq X \to Y$ be a nonlinear operator in an open convex domain Ω. Suppose that:

(H1) $\|x_0 - x_{-1}\| = \alpha$,
(H2) there exists $L_0{}^{-1} = [y_0, x_0; F]^{-1}$ such that $\|L_0{}^{-1}\| \leq \beta$,
(H3) $\|L_0{}^{-1} F(x_0)\| \leq \eta$,
(H4) $\|[x, y; F] - F'(z)\| \leq k\,(\|x-z\|^p + \|y-z\|^p)$, $p \in [0,1]$, for all $x, y, z \in \Omega$.

We denote

$$a_{-1} = \frac{\eta}{\alpha + \eta}, \qquad b_{-1} = k\beta\alpha^p,$$

and define the sequences

$$a_n = g(a_{n-1}) b_{n-1}, \quad b_n = f(a_n) f(a_{n-1})^p a_{n-1}^p b_{n-1}, \quad n \geq 0, \tag{5.22}$$

where

$$f(x) = \frac{1}{1-x}, \quad g(x) = (1-\lambda)^p + \frac{2}{p+1}(1+\lambda^p) f(x)^p x^p. \tag{5.23}$$

Note that f and g are increasing, and on the other hand $f(x) > 1$ in $(0,1)$.

As ${L_0}^{-1}$ exists, then x_1 is well defined and, from the initial hypotheses, it follows that

$$\begin{aligned} \|x_1 - x_0\| &\le \eta = f(a_{-1})a_{-1}\|x_0 - x_{-1}\|, \\ k\|{L_0}^{-1}\|\|x_0 - x_{-1}\|^p &\le k\beta\alpha^p = b_{-1}. \end{aligned} \tag{5.24}$$

Then, in [28], by induction on n, the following items are shown for $n \ge 1$:

(i_n) $\exists\, {L_n}^{-1} = [y_n, x_n; F]^{-1}$ such that $\|{L_n}^{-1}\| \le f(a_{n-1})\|{L_{n-1}}^{-1}\|$,
(ii_n) $\|x_{n+1} - x_n\| \le f(a_{n-1})a_{n-1}\|x_n - x_{n-1}\|$,
(iii_n) $k\|{L_n}^{-1}\|\|x_n - x_{n-1}\|^p \le b_{n-1}$.

Next, we study the real sequences defined in (5.22) in order to obtain the convergence of sequence (5.9) in Banach spaces. It will be sufficient that: $a_n < 1$ $(n \ge 0)$ and $\{x_n\}$ is a Cauchy sequence.

Firstly, we provide the following two lemmas on the real sequences given in (5.22).

Lemma 5.4 *Let f and g be the two real functions given in (5.23). If $a_1/a_0 \le b_1/b_0 < 1$, then*

(a) both sequences given in (5.22) are decreasing for $n \ge 0$,
(b) $a_n < \gamma^{\tilde{\alpha}_n} a_{n-1}$ and $b_n < \gamma^{\tilde{\alpha}_{n+1}} b_{n-1}$, for $n \ge 1$, where $\gamma = b_1/b_0 \in (0, 1)$ and $\{\tilde{\alpha}_n\}$ is the Fibonacci generalized sequence:

$$\tilde{\alpha}_1 = \tilde{\alpha}_2 = 1, \quad \tilde{\alpha}_{n+2} = \tilde{\alpha}_{n+1} + p\tilde{\alpha}_n, \quad n \ge 1, \tag{5.25}$$

(c) $a_n < \gamma^{\tilde{s}_n} a_0$, for $n \ge 1$, where $\tilde{s}_n = \tilde{\alpha}_1 + \tilde{\alpha}_2 + \cdots + \tilde{\alpha}_n$.

Next, we provide some properties of (5.25), whose proofs are trivial by applying induction.

Lemma 5.5 *Let $\{\tilde{\alpha}_n\}$ be the sequence defined in (5.25). Then,*

(a) $\tilde{\alpha}_n = \dfrac{1}{\sqrt{1+4p}}\left[\left(\dfrac{1+\sqrt{1+4p}}{2}\right)^n - \left(\dfrac{1-\sqrt{1+4p}}{2}\right)^n\right]$ *and*

$$\tilde{\alpha}_n \ge \frac{1}{\sqrt{1+4p}}\left(\frac{1+\sqrt{1+4p}}{2}\right)^{n-1},$$

(b) $\tilde{s}_n = \tilde{\alpha}_1 + \tilde{\alpha}_2 + \cdots + \tilde{\alpha}_n$ is such that $\tilde{s}_n = (\tilde{\alpha}_{n+2} - 1)/p$ and $\tilde{s}_1 + \tilde{s}_2 + \cdots + \tilde{s}_n = [\tilde{\alpha}_{n+4} - p(n+2) - 1]/p^2, \quad n \ge 1$.

The following semilocal convergence Theorem [28] shows that the sequence $\{x_n\}$ generated by (5.9) converges to a solution x^* of Eq. (5.1).

Theorem 5.6 *Let $x_{-1}, x_0 \in \Omega$ and $\lambda \in [0, 1]$. Let us suppose that* **(H1)**–**(H4)** *and the hypotheses of Lemma 5.4 are satisfied. If $a_0 < 1/2$ and $\overline{B(x_0, r_0)} \subseteq \Omega$, where $r_0 = \frac{1-a_0}{1-2a_0}\eta$, then the sequence $\{x_n\}$ given by (5.9) is well defined and converges* *to a solution* x^* *of (5.3) with at least R-order of convergence* $(1 + \sqrt{1+4p})/2$. *Moreover* $x_n, x^* \in \overline{B(x_0, r_0)}$. *Furthermore,*

$$\|x^* - x_n\| < \frac{\Delta^n}{1-\Delta}\,\eta\,\gamma^{\tilde{\beta}_{n-1}}, \tag{5.26}$$

where $\Delta = \frac{a_0}{1-a_0}$, $\tilde{\beta}_{-1} = 0 = \tilde{\beta}_0$ *and* $\tilde{\beta}_n = \tilde{s}_1 + \tilde{s}_2 + \cdots + \tilde{s}_n$, $n \geq 1$.

5.3.3 Application: A Special Case of Conservative Problems

It is well known that energy is dissipated by the action of any real dynamical system, usually through some form of friction. However, in certain situations this dissipation is so slow that it can be neglected over relatively short periods of time. In such cases we assume the law of conservation of energy, namely, that the sum of the kinetic energy and the potential energy is constant. A system of this kind is said to be conservative.

If ρ and σ are arbitrary functions with the property that $\rho(0) = 0$ and $\sigma(0) = 0$, the general equation

$$m\frac{d^2x(t)}{dt^2} + \sigma\left(\frac{dx(t)}{dt}\right) + \rho(x) = 0, \tag{5.27}$$

can be interpreted as the equation of motion of a mass m under the action of a restoring force $-\rho(x)$ and a damping force $-\sigma(dx/dt)$. In general these forces are nonlinear, and Eq. (5.27) can be regarded as the basic equation of nonlinear mechanics. In this paper we shall consider the special case of a nonlinear conservative system described by the equation

$$m\frac{d^2x(t)}{dt^2} + \rho(x(t)) = 0,$$

in which the damping force is zero and there is consequently no dissipation of energy. Extensive discussions of (5.27), with applications to a variety of physical problems, can be found in classical references [5] and [45].

In this application, we study the existence of a unique solution for a special case of a nonlinear conservative system described by the equation

$$\frac{d^2x(t)}{dt^2} + \Phi(x(t)) = 0, \tag{5.28}$$

with the boundary conditions

$$x(0) = 0 = x(1). \tag{5.29}$$

In order to study the application of (5.9) for the numerical solution of differential equation problems, we illustrate the theory for the case of particular second order ordinary differential equation (5.28) subject to the boundary conditions (5.29).

It is required to find a solution of Problems (5.28) and (5.29) in the interval $0 \leq t \leq 1$. Under suitable restrictions on the function Φ, we will see that a unique solution of (5.28) and (5.29) exists. Moreover the method of discretization is used to project the boundary value problem of second order into a finite-dimensional space. The new class of Secant-like methods are applied to this problem to approximate the solution of the corresponding system of equations.

Firstly, we suppose that Φ is once continuously differentiable and Φ' is Hölder (C, p) continuous. So the operator

$$[F(x)](t) = \frac{d^2x(t)}{dt^2} + \Phi(x(t)) \tag{5.30}$$

is defined from $C^{(2)}[0, 1]$ into $C[0, 1]$ and it is once differentiable.

5.3.3.1 Existence of the Solution

In order to see that a unique solution of Problems (5.28) and (5.29) exists, we apply Theorem 5.6. Then the bounds α, β, η, and k, which appear in the previous section, are necessary. The first derivative of F at $x = x(t)$ is

$$F'(x)y(t) = \frac{d^2y(t)}{dt^2} + \Phi'(x(t))y(t),$$

when it is applied to the function $y(t)$. To start the analysis of the convergence of (5.9) to a solution of Problems (5.28) and (5.29), from the starting functions $x_{-1}(t)$ and $x_0(t)$, we first prove that $L_0^{-1} = [y_0, x_0; F]^{-1}$ exists. Observe that

$$[F'(x) - F'(y)]u(t) = (\Phi'(x) - \Phi'(y))u(t).$$

Then

$$\|F'(x) - F'(y)\| = C\|x - y\|^p,$$

where C is the Hölder constant for Φ'. Since F' exists and is Hölder continuous, it follows that the operator

$$[x, y; F] = \int_0^1 F'(x + \tau(y - x))d\tau$$

is a divided difference at the points $x, y \in C^{(2)}[0, 1]$ and condition **(H4)** is satisfied with $k = C/(1 + p)$.

If x_{-1}, x_0 and $\lambda \in [0,1]$ are now fixed, then $y_0 = \lambda x_0 + (1-\lambda)x_{-1} \in C^{(2)}[0,1]$. Taking into account that

$$L_0 u(t) \equiv [y_0, x_0; F]u(t) = \frac{d^2u(t)}{dt^2} + \int_0^1 \Phi'(y_0(t)+\tau(x_0(t)-y_0(t)))u(t)\,d\tau \equiv v(t),$$

it follows that $u(t) = L_0^{-1}v(t)$ if L_0^{-1} exists.

Next, we consider the linear boundary value problem

$$\begin{aligned} &\frac{d^2u(t)}{dt^2} + \psi(x_0(t), y_0(t))u(t) = v(t) \\ &u(0) = 0 = u(1), \end{aligned} \tag{5.31}$$

where $\psi(x_0(t), y_0(t)) = \int_0^1 \Phi'(y_0(t) + \tau(x_0(t) - y_0(t)))\,d\tau$. It is known, see [36], that Problem (5.31) may be written in the form of the second kind Fredholm equation

$$u(t) = -\int_0^1 K(t,s)v(s)\,ds + [P(u)](t), \quad 0 \le t \le 1,$$

where

$$K(t,s) = \begin{cases} s(1-t), & t \ge s, \\ t(1-s), & t \le s. \end{cases}$$

and

$$[P(u)](t) = \int_0^1 K(t,s)\psi(x_0(s), y_0(s))u(s)\,ds.$$

Thus

$$[(I-P)(u)](t) = -\int_0^1 K(t,s)v(s)\,ds \equiv (Kv)(t).$$

On the other hand, using the max-norm and denoting $S = \sup_{0\le t\le 1} |\psi(x_0(t), y_0(t))|$, we have $\|P\| \le S/8$. Consequently, by the Banach Lemma, $(I-P)^{-1}$ exists if $S < 8$, and then

$$u(t) = (I-P)^{-1}(Kv)(t).$$

Since

$$\|Kv\| \leq \left(\sup_{0\leq t\leq 1} \int_0^1 |K(t,s)|ds \right) \|v\| \leq \|v\|/8,$$

then L_0^{-1} exists, $\|L_0^{-1}\| \leq 1/(8-S)$ and $\|L_0^{-1}F(x_0)\| \leq \|F(x_0)\|/(8-S)$.

We can now establish a result on the existence of the solution of Problems (5.28) and (5.29), whose proof follows as that in Theorem 5.6.

Theorem 5.7 *Following the previous notation, we consider the operator defined in (5.30), where* $F : C^{(2)}[0,1] \to C[0,1]$. *Assume that* $x_{-1}, x_0 \in C^{(2)}[0,1]$, $\lambda \in [0,1]$ *fixed,* $S < 8$, $a_0 < 1/2$ *and* $a_1/a_0 \leq b_1/b_0 < 1$, *where* a_0, a_1, b_0, b_1 *are defined in the previous section with*

$$\alpha = \|x_0 - x_{-1}\|, \quad \beta = \frac{1}{8-S}, \quad \eta = \frac{\|F(x_0)\|}{8-S}, \quad k = \frac{C}{1+p},$$

and C *the Hölder constant for* Φ'. *Then, there exists at least a solution of Problems (5.28) and (5.29) in* $\overline{B(x_0, r_0)}$, *where* $r_0 = \frac{1-a_0}{1-2a_0}\eta$.

Next, we show the application of the previous study to the following boundary value problem:

$$\begin{aligned} &\frac{d^2x(t)}{dt^2} + x(t)^{1+p} + Q = 0, \quad p \in [0,1], \quad Q \in \mathbb{R}, \\ &x(0) = 0 = x(1). \end{aligned} \tag{5.32}$$

in the space $C^{(2)}[0,1]$ of all twice differentiable functions with the max-norm. Now (5.30) can be written in the form

$$[F(x)](t) = \frac{d^2x(t)}{dt^2} + x(t)^{1+p} + Q. \tag{5.33}$$

To obtain the existence and the uniqueness of the solution of (5.32), we first consider

$$\Omega = \{x, y \in C^{(2)}[0,1] : \|\psi(x,y)\| < 8\} \subseteq C^{(2)}[0,1], \tag{5.34}$$

where $\psi(x(t), y(t)) = (1+p)\int_0^1 (y(t) + \tau(x(t) - y(t)))^p \, d\tau$, so that $F : \Omega \to C[0,1]$. Taking into account (5.28), we have $\Phi(x(t)) = x(t)^{1+p} + Q$. Then, by the Banach Lemma, L_0^{-1} exists and $\|L_0^{-1}\| \leq 1/(8-S)$, where

$$S = (1+p) \sup_{0\leq t\leq 1} \left| \int_0^1 (y_0(t) + \tau(x_0(t) - y_0(t)))^p \, d\tau \right|.$$

In addition, we have

$$\|L_0^{-1}F(x_0)\| \leq \frac{\|F(x_0)\|}{8-S}.$$

On the other hand, we get

$$\|[x, y; F] - F'(z)\| \leq \|x-z\|^p + \|y-z\|^p; \quad x, y, z \in \Omega_0, \quad p \in [0, 1].$$

Corollary 8 *Let $F : \Omega \subseteq C^{(2)}[0,1] \to C[0,1]$, where Ω is defined in (5.34) and F in (5.33). Let $x_{-1}, x_0 \in \Omega$, $\lambda \in [0,1]$ fixed. Let us suppose that $a_0 < 1/2$ and $a_1/a_0 \leq b_1/b_0 < 1$, where a_0, a_1, b_0, b_1 are defined in the previous section with*

$$\alpha = \|x_0 - x_{-1}\|, \quad \beta = \frac{1}{8-S}, \quad \eta = \frac{\|F(x_0)\|}{8-S}, \quad k = 1,$$

and $S = (1+p) \sup_{0 \leq t \leq 1} \left| \int_0^1 (y_0(t) + \tau(x_0(t) - y_0(t)))^p \, d\tau \right|$. If $\overline{B(x_0, r_0)} \subseteq \Omega$, where $r_0 = \frac{1-a_0}{1-2a_0}\eta$. Then, a solution of (5.32) exists at least in $\overline{B(x_0, r_0)}$.

5.3.3.2 Location of the Solution

To illustrate the previous result, we consider the Secant method and boundary value Problem (5.32), where $Q = 1/4$ and $p = 1/2$. As the solution would vanish at the endpoints and be positive in the interior, a reasonable choice of initial approximation seems to be $x_{-1}(t) = 0.4 \sin \pi t$. On the other hand, we choose $x_0(t) = 0$ in order to simplify the domain of existence of solution and reduce the operational cost. So,

$$S = \sqrt{0.4}, \quad \alpha = 0.4\pi^2, \quad \beta = 1/(8-S), \quad \eta = 1/(4(8-S)) \quad \text{and} \quad k = 1.$$

As a result,

$$a_0 = 0.303022 < (3-\sqrt{5})/2 < 1/2,$$

$$a_1/a_0 = 0.222462 \leq b_1/b_0 = 0.707029 < 1,$$

and the conditions of Corollary 8 hold. Then, there exists a solution x^* of (5.32) in $\{w \in C^{(2)}[0,1]; \|w\| \leq r_0\}$, where $r_0 = 0.0600328$, see Fig. 5.1.

5.3.3.3 Numerical Solution of the Finite-Difference Equations

To show how boundary value Problem (5.32) can be changed to a system of algebraic equations, we replace the derivative in the differential equation with their finite-difference approximations. The system of algebraic equations can then be solved numerically by (5.9) in order to obtain an approximate solution to boundary value Problem (5.32).

To solve this problem by finite differences, we start by drawing the usual grid line with grid points $t_i = ih$, where $h = 1/n$ and n is an appropriate integer. Note that x_0 and x_n are given by the boundary conditions, then $x_0 = 0 = x_n$, and our work is to find the other x_i $(i = 1, 2, \dots, n-1)$. To do this, we begin by replacing the second derivative $x''(t)$ in the differential equation with its approximation

$$x''(t) \approx [x(t+h) - 2x(t) + x(t-h)]/h^2,$$
$$x''(t_i) = (x_{i+1} - 2x_i + x_{i-1})/h^2, \quad i = 1, 2, \dots, n-1.$$

By substituting this expression into the differential equation, we have the following system of nonlinear equations:

$$\begin{cases} 2x_1 - h^2 x_1^{1+p} - x_2 - h^2 Q = 0, \\ -x_{i-1} + 2x_i - h^2 x_i^{1+p} - x_{i+1} - h^2 Q = 0, \quad i = 2, 3, \dots, n-2, \\ -x_{n-2} + 2x_{n-1} - h^2 x_{n-1}^{1+p} - h^2 Q = 0. \end{cases} \tag{5.35}$$

We therefore have an operator $F : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ such that $F(x) = M(x) - h^2\varphi(x)$, where

$$M = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{pmatrix}, \quad \varphi(x) = \begin{pmatrix} x_1^{1+p} + 1/4 \\ x_2^{1+p} + 1/4 \\ \vdots \\ x_{n-1}^{1+p} + 1/4 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \end{pmatrix}.$$

Thus, we can write

$$F'(x) = M - h^2(1+p)\mathrm{Diag}\{x_1^p, x_2^p, \dots, x_{n-1}^p\}.$$

Let $x \in \mathbb{R}^{n-1}$ and $\|x\| = \max_{1 \le i \le n-1} |x_i|$. The corresponding norm on $A \in \mathbb{R}^{n-1} \times \mathbb{R}^{n-1}$ is

$$\|A\| = \max_{1 \le i \le n-1} \sum_{j=1}^{n-1} |a_{ij}|.$$

It is known that F has a Hölder continuous divided difference at the points $x, y \in \mathbb{R}^{n-1}$, which is defined by the matrix whose entries are given in (5.12) (see [8, 38]).

Table 5.4 Numerical solution x^* of (5.35)

i	x_i^*	i	x_i^*	i	x_i^*
1	0.01141648508671	4	0.03052779202087	7	0.02669609124653
2	0.02032077189279	5	0.03180615401355	8	0.02032077189279
3	0.02669609124653	6	0.03052779202087	9	0.01141648508671

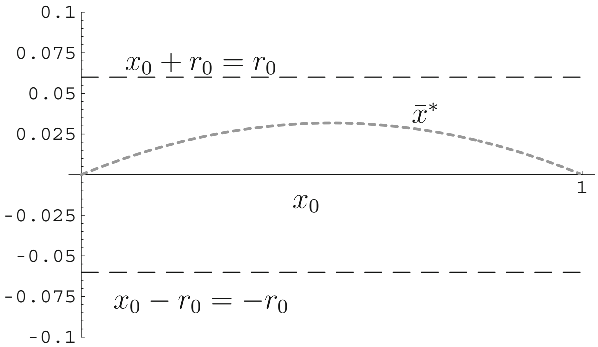

Fig. 5.2 The domain of existence of solution and the approximated solution

If $n = 10$, then (5.35) gives 9 equations. Taking into account the data for the initial iterate are $x_{-1}(t_i) = 0.4 \sin \pi t_i$ and $x_0(t_i) = 0$ for $i = 1, 2, \ldots, 9$. After five iterates, we obtain the vector x^* (see Table 5.4) as the solution of system (5.35).

Finally, we need to relate the solution vector x^* of the system of nonlinear equations (5.35) found in Table 5.4 to an initial point of the method used to solve Eq. (5.32). More precisely, if x^* is now interpolated, the approximation $\bar{x}^*$ to the solution of (5.32) with $p = 1/2$ is that appearing in Fig. 5.2. Notice that the interpolated approximation $\bar{x}^*$ lies within the existence domain of solutions mentioned above.

Hence, according to Fig. 5.2, giving the relationship between x^* and $\bar{x}^*$, we should choose as an initial point (of the method to be used to solve (5.32)) the point $\bar{x}^*$.

5.3.3.4 Final Remark

Finally, we analyze two things. Firstly, we study the domain of the starting points and, secondly, we analyze the speed of convergence of the class of iterative methods given by (5.9). If we now choose $Q = 0$ in (5.32), the corresponding boundary value problem has already been used by other authors as a test problem (see [8, 31, 41]).

We again start using the method of discretization to project this boundary value problem into a finite-dimensional space. Let $n = 10$ and $x_{-1}(t_i) = 135 \sin \pi t_i$ $(i = 1, 2, \ldots, 9)$ be the initial approximation. We choose, as in [8], x_0 by setting $x_0(t_i) = x_{-1}(t_i) - 10^{-5}$, $i = 1, 2, \ldots, 9$. If we apply the Secant method ($\lambda = 0$) to

Table 5.5 Absolute errors for (5.9) and different values of λ

n	$\lambda = 0$	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 0.75$	$\lambda = 0.99$
1	2.24748	2.24748	2.24748	2.24748	2.24748
2	2.60218×10^{-1}	2.08533×10^{-1}	1.53355×10^{-1}	9.43008×10^{-2}	3.35497×10^{-2}
3	3.66518×10^{-3}	2.28181×10^{-3}	1.16525×10^{-3}	3.79963×10^{-4}	1.20966×10^{-5}
4	6.16651×10^{-6}	2.31808×10^{-6}	5.83199×10^{-7}	5.88006×10^{-8}	2.73843×10^{-11}
5	1.47125×10^{-10}	2.58780×10^{-11}	2.20268×10^{-12}	0.0	0.0

Table 5.6 Absolute errors for Newton's method

n	$\|x^* - x_n\|_\infty$
1	2.24749
2	3.09256×10^{-2}
3	2.84217×10^{-13}
4	0.0

the previous points, after two and three iterations we obtain two points x_2 and x_3 in which the conditions required in this paper are satisfied for the Secant method, but the ones required by Argyros in [8] are not. Consequently, we can take x_2 and x_3 as the true starting points.

On the other hand, we obtain the errors $\|x_n - x^*\|_\infty$, which appear in Table 5.5, for the iterates x_n generated by (5.9) for different values of the parameter $\lambda \in [0, 1]$ and starting at x_{-1} and x_0.

The numerical results, using 14 significative decimal figures, indicate that the Secant method is not optimal for approximating the solution x^* of $F(x) = 0$. Moreover, iteration (5.9) converges faster to x^* for increasing values of the parameter $\lambda \in [0, 1]$.

Next, observe that (5.9), where λ is near 1, gives similar approximations, without using F', to the solution x^* of $F(x) = 0$ to Newton's method (see Table 5.6).

5.4 Convergence of Secant-Like Methods for Non-Differentiable Operators

We consider nonlinear integral equations of mixed Hammerstein type

$$x(s) = f(s) + \int_a^b G(s,t)\, H(t, x(t))\, dt, \quad s \in [a, b], \tag{5.36}$$

where $-\infty < a < b < +\infty$, f, G, and H are known functions and x is a solution to be determined. Integral equations of this type appear very often in several applications to real-world problems. For example, in problems of dynamic models of chemical reactors [17], vehicular traffic theory, biology, and queuing theory [19]. The Hammerstein integral equations also appear in the electro-magnetic fluid dynamics and can be reformulated as two-point boundary value problems with certain nonlinear boundary conditions and in multi-dimensional analogues which appear as reformulations of elliptic partial differentiable equations with nonlinear boundary conditions (see [39] and the references given there).

Solving Eq. (5.36) is equivalent to solving $\mathcal{F}(x) = 0$, where $\mathcal{F} : \Omega \subset \mathcal{C}[a, b] \to \mathcal{C}[a, b]$ and

$$[\mathcal{F}(x)](s) = x(s) - f(s) - \int_a^b G(s,t)H(t, x(t))\, dt, \quad s \in [a, b].$$

Examples where the operator $\mathcal{F}$ is differentiable are found in [30].

If we consider (5.36) where G is the Green function in $[a, b] \times [a, b]$, we then use a discretization process to transform Eq. (5.36) into a finite-dimensional problem by approximating the integral of (5.36) by a Gauss–Legendre quadrature formula with m nodes:

$$\int_a^b q(t)\, dt \simeq \sum_{i=1}^{m} w_i q(t_i),$$

where the nodes t_i and the weights w_i are determined.

If we denote the approximations of $x(t_i)$ and $f(t_i)$ by x_i and f_i, respectively, with $i = 1, 2, \ldots, m$, then Eq. (5.36) is equivalent to the following system of nonlinear equations:

$$x_i = f_i + \sum_{j=1}^{m} a_{ij}\, H(t_j, x_j), \quad j = 1, 2, \ldots, m, \tag{5.37}$$

where

$$a_{ij} = w_j G(t_i, t_j) = \begin{cases} w_j \frac{(b-t_i)(t_j-a)}{b-a}, & j \le i, \\ w_j \frac{(b-t_j)(t_i-a)}{b-a}, & j > i. \end{cases}$$

Then, system (5.37) can be written as

$$F(\mathbf{x}) \equiv \mathbf{x} - \mathbf{f} - A\,\mathbf{z} = 0, \qquad F : \mathbb{R}^m \longrightarrow \mathbb{R}^m, \tag{5.38}$$

where

$$\mathbf{x} = (x_1, x_2, \ldots, x_m)^T, \qquad \mathbf{f} = (f_1, f_2, \ldots, f_m)^T, \qquad A = (a_{ij})_{i,j=1}^m,$$

$$\mathbf{z} = (H(t_1, x_1), H(t_2, x_2), \ldots, H(t_m, x_m))^T.$$

As in $\mathbb{R}^m$ we can consider divided difference of first order that does not need that the function F is Fréchet differentiable (see [38]), we then use the divided difference of first order given by (5.12).

If we consider that system of nonlinear Eqs. (5.38) is of the form

$$F(\mathbf{x}) \equiv \mathbf{x} - \mathbf{f} - A\,(\delta \mathbf{v}_\mathbf{x} + \mu \mathbf{w}_\mathbf{x}) = 0, \qquad F : \mathbb{R}^m \longrightarrow \mathbb{R}^m, \tag{5.39}$$

where

$$\mathbf{v}_\mathbf{x} = (x_1^2, x_2^2, \ldots, x_m^2)^T, \qquad \mathbf{w}_\mathbf{x} = (|x_1|, |x_2|, \ldots, |x_m|)^T,$$

$\delta, \mu \in \mathbb{R}$ and $\mu \neq 0$, it is obvious that the function F defined in (5.39) is nonlinear and non-differentiable. Moreover, $[\mathbf{u}, \mathbf{v}; F] = I - (\delta B + \mu C)$, where $B = (b_{ij})_{i,j=1}^m$ with $b_{ij} = a_{ij}(u_j + v_j)$ and $C = (c_{ij})_{i,j=1}^m$ with $c_{ij} = a_{ij}\frac{|u_j| - |v_j|}{u_j - v_j}$. Furthermore,

$$\|[x, y; F] - [u, v; F]\| \leq L + K(\|x - u\| + \|y - v\|) \text{ with } L = 2|\mu|\,\|A\| \text{ and } K = |\delta|\,\|A\|.$$

Observe then that if the divided difference of first order of the function F satisfies a condition of type

$$\begin{aligned} &\|[x, y; F] - [u, v; F]\| \leq L + K(\|x - u\| + \|y - v\|); \\ &L, K \geq 0;\, x, y, u, v \in \Omega;\, x \neq y;\, u \neq v, \end{aligned} \tag{5.40}$$

in $\mathbb{R}^m$, instead of a condition Hölder continuous, from Lemma 5.1 we can solve equations where the function F is non-differentiable, as, for example, Eq. (5.39).

Then, as already discussed above and below Fig. 5.2, the approximate solution $\bar{x}^*$ to Eq. (5.36), related to the solution x^* of Eq. (5.39) by a similar figure, will be chosen to be the initial point to solve Eq. (5.36) using some method.

In view of the above, we present a new semilocal convergence result where the operator F satisfies condition (5.40) that allows us to solve equations where the operator F is non-differentiable ($L \neq 0$).

Next, we present the semilocal convergence [23] result for methods (5.9).

Theorem 5.9 *Let X and Y be two Banach spaces and $F : \Omega \subseteq X \to Y$ be a nonlinear operator defined on a nonempty open convex domain Ω. Suppose that conditions* **(H1)–(H3)** *and (5.40) are satisfied. Once fixed $\lambda \in [0, 1]$, if the equation*

$$t\left(1 - \frac{m}{1 - \beta\,(L + K(2t + (1 - \lambda)\alpha))}\right) - \eta = 0, \tag{5.41}$$

where $m = \max\{\, \beta(L + K((1-\lambda)\alpha + \eta)),\ \beta(L + (2-\lambda)K\eta)\,\}$*, has at least one positive real root and the smallest positive real root, denoted by* R*, satisfies*

$$\beta(L + K(2R + (1-\lambda)\alpha)) < 1 \tag{5.42}$$

and $\overline{B(x_0, R)} \subset \Omega$*, then the sequence* $\{x_n\}$ *defined in (5.9), starting at* x_{-1} *and* x_0*, is well-defined and converges to a solution* x^* *of* $F(x) = 0$*. Moreover, the solution* x^* *and the iterates* x_n *belong to* $\overline{B(x_0, R)}$*, and* x^* *is unique in* $\overline{B(x_0, R)}$*.*

5.4.1 Numerical Example

We illustrate the abovementioned with an example. We consider a non-differentiable system of nonlinear equations of form (5.39) and see that Theorem 5.9 guarantees the semilocal convergence of a method of (5.9).

If in (5.38) we consider $m = 8$, $\delta = \mu = 3/4$, we obtain the non-differentiable system of nonlinear equations

$$F(\mathbf{x}) \equiv \mathbf{x} - \frac{\mathbf{1}}{\mathbf{2}} - \frac{3}{4}A(\mathbf{v_x} + \mathbf{w_x}) = 0, \qquad F : \mathbb{R}^8 \longrightarrow \mathbb{R}^8, \tag{5.43}$$

where

$$\mathbf{x} = (x_1, x_2, \ldots, x_8)^T, \qquad \frac{\mathbf{1}}{\mathbf{2}} = \left(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}\right)^T, \qquad A = (a_{ij})_{i,j=1}^{8}.$$

If we choose the starting points $\mathbf{x_{-1}} = \left(\frac{2}{5}, \frac{2}{5}, \ldots, \frac{2}{5}\right)^T$ and $\mathbf{x_0} = \left(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}\right)^T$, method (5.9) with $\lambda = \frac{3}{4}$ and the max-norm, we obtain $\alpha = 0.1$, $\beta = 1.2170\ldots$, $\eta = 0.0824\ldots$, $L = 0.1853\ldots$, $K = 0.0926\ldots$ and the Eq. (5.41) is reduced to

$$t\left(1 - \frac{(0.2376\ldots)}{1 - (1.2170\ldots)((0.1863\ldots) + (0.0926\ldots)(2t + 0.0250\ldots))}\right) - (0.0824\ldots) = 0.$$

The last equation has two positive real roots and the smallest one, $R = 0.1211\ldots$, satisfies Condition (5.42), since

$$\beta(L + K(2R + (1-\lambda)\alpha)) = 0.2557\ldots < 1.$$

Table 5.7 Numerical solution $\mathbf{x}^*$ of (5.43)

i	x_i^*	i	x_i^*	i	x_i^*	i	x_i^*
1	0.506462...	3	0.561738...	5	0.583219...	7	0.530722...
2	0.530722...	4	0.583219...	6	0.561738...	8	0.506462...

Table 5.8 Absolute errors for (5.9) with $\lambda = \frac{3}{4}$

n	$\|\mathbf{x_n} - \mathbf{x}^*\|$
-1	$1.8321\ldots \times 10^{-1}$
0	$8.3221\ldots \times 10^{-2}$
1	$7.8399\ldots \times 10^{-4}$
2	$1.4181\ldots \times 10^{-6}$
3	$2.3181\ldots \times 10^{-11}$
4	$6.7813\ldots \times 10^{-19}$

Therefore, by Theorem 5.9, we guarantee the semilocal convergence of Method (5.9) with $\lambda = \frac{3}{4}$. After five iterations and using the stopping criterion $\|\mathbf{x_n} - \mathbf{x_{n-1}}\| < 10^{-32}$, we obtain the numerical approximation $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_8^*)^T$ of a solution of (5.43) which is given in Table 5.7. In Table 5.8, we show the errors $\|\mathbf{x_n} - \mathbf{x}^*\|$ obtained with the same stopping criterion. Finally, by Theorem 5.9, the existence and uniqueness of solution is guaranteed in the ball $\overline{B(\mathbf{x_0}, 0.1211\ldots)}$.

5.5 Convergence of Secant-Like Methods Whatever the Operator

To analyze the semilocal convergence of iterative processes that do not use derivatives in their algorithms, the conditions usually required are the Lipschitz or Hölder continuous conditions for the divided difference of first order (see [8, 32]). Notice that, under these conditions, the operator F must be Fréchet differentiable [27]. To generalize the above conditions and even consider situations in which operator F is non-differentiable, we consider ω-continuous for the divided difference (5.16). Moreover, as it is known from Lemma 5.1, if $\omega(0, 0) = 0$, then F is a Fréchet differentiable operator. Therefore, taking into account condition (5.16), we consider the case in which the operator F is non-differentiable; as example, situations where $\omega(0, 0) \neq 0$.

5.5.1 A Semilocal Convergence Result

Let us assume that

(I) $\|x_{-1} - x_0\| = \alpha$,
(II) there exists $L_0{}^{-1} = [y_0, x_0; F]^{-1}$ such that $\|L_0{}^{-1}\| \leq \beta$,

(III) $\|L_0{}^{-1}F(x_0)\| \leq \eta$,
(IV) $\|[x, y; F] - [v, w; F]\| \leq \omega(\|x - v\|, \|y - w\|)$; $x, y, v, w \in \Omega$, where $\omega : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a non-decreasing continuous function in its two arguments.

Next, we can already give a semilocal convergence result.

Theorem 5.10 *Under conditions (I)–(IV), we denote*

$$m = \max\{\beta\omega\left((1 - \lambda)\alpha, \eta\right), \beta\omega\left((1 - \lambda)\eta, \eta\right)\}$$

and assume that the equation

$$u\left(1 - \frac{m}{1 - \beta\omega(u + (1 - \lambda)\alpha, u)}\right) - \eta = 0 \tag{5.44}$$

has at least one positive zero. Let R be the minimum positive one. If $\beta\omega(R + (1 - \lambda)\alpha, R) < 1$, $\dfrac{m}{1 - \beta\omega(R + (1 - \lambda)\alpha, R)} < 1$ and $\overline{B(x_0, R)} \subset \Omega$, then the sequence $\{x_n\}$, given by (5.9), is well defined, remains in $\overline{B(x_0, R)}$, and converges to the unique solution x^* of Eq. (5.1) in $\overline{B(x_0, R)}$.

Remark Note that the operator F is Fréchet differentiable when the divided differences are Lipschitz or (k, p)-Hölder continuous. But, under condition **(IV)**, F is Fréchet differentiable if $\omega(0, 0) = 0$. Therefore, if $\omega(0, 0) \neq 0$, Theorem 5.10 is true for non-differentiable operators.

5.5.2 Applications

We present two kinds of applications. The first one is theoretical and practical for Fréchet differentiable operators, where it is proved the convergence for divided differences that are not Lipschitz or Hölder continuous. Moreover, this application is not usually studied by other authors. The second one is practical for non-differentiable operators, and we compare the methods presented in the paper with other ones given by several authors.

In the first example, a Fréchet differentiable operator is considered, i.e., in (5.3), $F = G$ and $H(x) = 0$. We note that the semilocal convergence conditions required are mild.

5.5.2.1 Example 1

Now we apply the semilocal convergence result given above to the following boundary value problem:

$$\begin{cases} x'' + x^{1+p} + x^2 = 0, \quad p \in [0, 1], \\ x(0) = x(1) = 0. \end{cases} \tag{5.45}$$

To solve this problem by finite differences, we start drawing the usual grid line with the grid points $t_i = ih$, where $h = 1/n$ and n is an appropriate integer. Note that x_0 and x_n are given by the boundary conditions and then $x_0 = 0 = x_n$. We first approximate the second derivative $x''(t)$ by

$$x''(t) \approx [x(t+h) - 2x(t) + x(t-h)]/h^2,$$

$$x''(t_i) = (x_{i+1} - 2x_i + x_{i-1})/h^2, \quad i = 1, 2, \ldots, n-1.$$

Substituting this expression into the differential equation, we have the following system of nonlinear equations:

$$\begin{cases} 2x_1 - h^2 x_1^{1+p} - h^2 x_1^2 - x_2 = 0, \\ -x_{i-1} + 2x_i - h^2 x_i^{1+p} - h^2 x_i^2 - x_{i+1} = 0, \quad i = 2, 3, \ldots, n-2, \\ -x_{n-2} + 2x_{n-1} - h^2 x_{n-1}^{1+p} - h^2 x_{n-1}^2 = 0. \end{cases} \tag{5.46}$$

We therefore have an operator $F : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ such that $F(x) = M(x) - h^2 f(x)$, where

$$f(x) = \left(x_1^{1+p} + x_1^2, x_2^{1+p} + x_2^2, \ldots, x_{n-1}^{1+p} + x_{n-1}^2\right)^t$$

and

$$M = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{pmatrix},$$

Thus,

$$F'(x) = M - h^2(1+p)\text{Diag}\{x_1^p, x_2^p, \ldots, x_{n-1}^p\} - 2h^2\text{Diag}\{x_1, x_2, \ldots, x_{n-1}\}$$

Let $x \in \mathbb{R}^{n-1}$ and choose the norm $\|x\| = \max\limits_{1 \le i \le n-1} |x_i|$. The corresponding norm on $A \in \mathbb{R}^{n-1} \times \mathbb{R}^{n-1}$ is

$$\|A\| = \max_{1 \le i \le n-1} \sum_{j=1}^{n-1} |a_{ij}|.$$

Then F has a divided difference at the points $x, y \in \mathbb{R}^{n-1}$, which is defined by the matrix, whose entries are given in (5.12). Consequently,

$$[x, y; F] = M - h^2 \begin{pmatrix} \frac{x_1^{1+p} - y_1^{1+p} + x_1^2 - y_1^2}{x_1 - y_1} & 0 & \cdots & 0 \\ 0 & \frac{x_2^{1+p} - y_2^{1+p} + x_2^2 - y_2^2}{x_2 - y_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{x_{n-1}^{1+p} - y_{n-1}^{1+p} + x_{n-1}^2 - y_{n-1}^2}{x_{n-1} - y_{n-1}} \end{pmatrix}.$$

In this case, we have that $[x, y; F] = \int_0^1 F'(x + t(y - x))\, dt$. So, we study the value $\|F'(x) - F'(v)\|$ to obtain a bound for $\|[x, y; F] - [v, w; F]\|$.

For all $x, v \in \mathbb{R}^{n-1}$ with $|x_i| > 0, |v_i| > 0, \ (i = 1, 2, \ldots, n-1)$, and taking into account the max-norm, it follows

$$\begin{aligned} \|F'(x) - F'(v)\| &= \|\text{diag}\{h^2(1+p)\ (v_i^p - x_i^p) + 2h^2(v_i - x_i)\}\| \\ &= \max_{1 \le i \le n-1} |h^2(1+p)\ (v_i^p - x_i^p) + 2h^2(v_i - x_i)| \\ &\le (1+p)h^2 \max_{1 \le i \le n-1} |v_i^p - x_i^p| + 2h^2 \max_{1 \le i \le n-1} |v_i - x_i| \\ &\le (1+p)h^2\ [\max_{1 \le i \le n-1} |v_i - x_i|\ ]^p + 2h^2 \|v - x\| \\ &= (1+p)h^2 \|v - x\|^p + 2h^2 \|v - x\|. \end{aligned}$$

Therefore,

$$\begin{aligned} \|[x, y; F] - [v, w; F]\| &\le \int_0^1 \|F'\,(x + t(y - x)) - F'\,(u + t(w - v))\,\|\, dt \\ &\le h^2 \int_0^1 \big((1+p)\|(1-t)(x - v) + t(y - w)\|^p \\ &\quad + 2\|(1-t)(x - v) + t(y - w)\|)\ dt \le h^2(1+p) \\ &\quad \times \int_0^1 \big((1-t)^p \|x - v\|^p + t^p \|y - w\|^p\big)\ dt + 2h^2 \\ &\quad \times \int_0^1 ((1-t)\|x - v\| + t\|y - w\|)\ dt \\ &= h^2 \big(\|x - v\|^p + \|y - w\|^p + \|x - v\| + \|y - w\|\big). \end{aligned}$$

From **(IV)**, we consider the function $\omega(u_1, u_2) = h^2(u_1^p + u_2^p + u_1 + u_2)$.

Next, we apply the Secant method to approximate the solution of $F(x) = 0$. If $n = 10$, then (5.46) gives 9 equations. Since a solution of (5.45) would vanish at the end points and be positive in the interior, a reasonable choice of the initial approximation seems to be $10\sin\pi t$. This approximation gives us the following vector y_{-1}:

$$y_{-1} = \begin{pmatrix} 3.090169943749474 \\ 5.877852522924731 \\ 8.090169943749475 \\ 9.51056516295136 \\ 10.0000000000000 \\ 9.51056516295136 \\ 8.090169943749475 \\ 5.877852522924731 \\ 3.090169943749474 \end{pmatrix}.$$

Choose y_0 by setting $y_0(t_i) = y_{-1}(t_i) - 10^{-5}, i = 1, 2, \ldots, 9$ and using iteration (5.9), ($\lambda = 0$), after two iterations, we obtain y_1 and y_2:

$$y_1 = \begin{pmatrix} 2.453176290658909 \\ 4.812704101582601 \\ 6.8481873135861 \\ 8.252997367741953 \\ 8.75737771678512 \\ 8.252997367741953 \\ 6.8481873135861 \\ 4.812704101582601 \\ 2.453176290658909 \end{pmatrix} \quad \text{and} \quad y_2 = \begin{pmatrix} 2.404324055268407 \\ 4.713971539035271 \\ 6.7003394962933925 \\ 8.066765882171131 \\ 8.556329565792526 \\ 8.066765882171131 \\ 6.7003394962933924 \\ 4.713971539035271 \\ 2.404324055268407 \end{pmatrix}.$$

Taking $x_{-1} = y_1$ and $x_0 = y_2$, we obtain $\alpha = 0.201048$, $\beta = 15.319$, $\eta = 0.0346555$. In this case, the solution of equation (5.44) given in Theorem 5.10 has a minimum positive solution $R = 0.041100361$. Besides, $\beta\omega(\alpha + R, R) = 0.14983 < 1$ and $M = 0.156808 < 1$.

Therefore, the hypotheses of Theorem 5.10 are fulfilled and a unique solution of Eq. (5.3) exists in $\overline{B(x_0, R)}$.

We obtain the vector x^* as the solution of system (5.46), after nine iterations:

$$x^* = \begin{pmatrix} 2.394640794786742 \\ 4.694882371216001 \\ 6.672977546934751 \\ 8.033409358893319 \\ 8.520791423704788 \\ 8.033409358893319 \\ 6.67297754693475 \\ 4.694882371216 \\ 2.394640794786742 \end{pmatrix}.$$

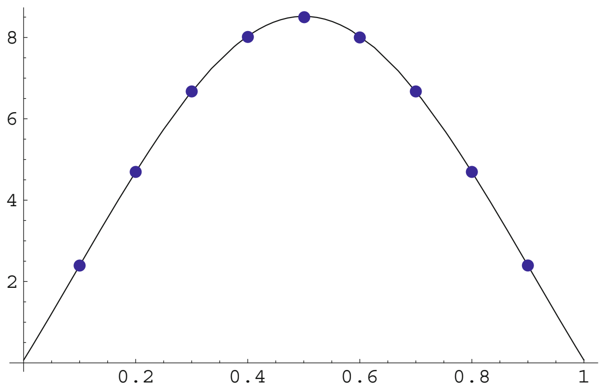

Fig. 5.3 x^* and the approximate solution $\bar{x}^*$

If x^* is now interpolated, its approximation $\bar{x}^*$ to the solution of (5.45) with $p = 1/2$ is that appearing in Fig. 5.3.

Note that, in this example, the convergence cannot be guaranteed from classical studies where divided differences are Lipschitz or Hölder continuous [8, 38], whereas we can do it by the technique presented in this chapter.

5.5.2.2 Example 2

Consider the non-differentiable system of equations

$$\begin{cases} 3x^2y + y^2 - 1 + |x - 1| = 0, \\ x^4 + xy^3 - 1 + |y| = 0. \end{cases} \tag{5.47}$$

We therefore have an operator $F : \mathbb{R}^2 \to \mathbb{R}^2$ such that $F = (F_1, F_2)$. For $x = (x_1, x_2) \in \mathbb{R}^2$, we take $F_1(x_1, x_2) = 3x_1^2x_2+x_2^2-1$, $F_2(x_1, x_2) = x_1^4+x_1x_2^3-1$.

For $v, w \in \mathbb{R}^2$, we take the divided differences as in (5.12) and apply several methods to solve (5.47).

For method (5.4), we have $G = (G_1, G_2)$ where $G_1(x_1, x_2) = 3x_1^2x_2 + x_2^2 - 1$ and $G_2(x_1, x_2) = x_1^4 + x_1x_2^3 - 1$. See Table 5.9 for method (5.4) with $x_0 = (1, 0)$.

Note that the approximated solution is

$$x^* = (0.8946553733346867, 0.3278265117462974).$$

For the Secant method with $x_{-1} = (5, 5)$ and $x_0 = (1, 0)$, see Table 5.10; for method (5.9) with $\lambda = 0.5$, $x_{-1} = (5, 5)$ and $x_0 = (1, 0)$, see Table 5.11; for method (5.9) with $\lambda = 0.99$, $x_{-1} = (5, 5)$ and $x_0 = (1, 0)$, see Table 5.12.

Table 5.9 Method (5.4) with $x_0 = (1, 0)$

n	$x_n^{(1)}$	$x_n^{(2)}$	$\|x^* - x_n\|$
1	1	0.3333333333333333	1.05345×10^{-1}
2	0.9065502183406114	0.3540029112081513	2.61764×10^{-2}
3	0.8853284006634119	0.3380272763613319	1.02008×10^{-2}
4	0.891329556832800	0.3266139765935657	3.32582×10^{-3}
5	0.8952388154638436	0.3264068528436253	1.41967×10^{-3}
6	0.8951546713726346	0.3277303340450432	4.99298×10^{-4}
7	0.8946737434711373	0.3279791543720321	1.52633×10^{-4}
8	0.8945989089774475	0.3278650593487548	5.64644×10^{-5}
9	0.894643228355865	0.3278150392082856	1.2145×10^{-5}
10	0.8946599936156449	0.3278198892648906	6.63248×10^{-6}
11	0.8946576401953287	0.3278267282085600	2.26686×10^{-6}
12	0.8946552195650909	0.3278273518268564	8.30018×10^{-7}
⋮	⋮	⋮	⋮
34	0.8946553733346867	0.3278265217462975	5.55112×10^{-17}

Table 5.10 Secant method, with $x_{-1} = (5, 5)$ and $x_0 = (1, 0)$

n	$x_n^{(1)}$	$x_n^{(2)}$	$\|x^* - x_n\|$
1	0.989800874210782	0.0126274890723652	3.15199×10^{-1}
2	0.9218147654932871	0.3079399161522621	2.71594×10^{-2}
3	0.900073765669214	0.325927010697792	5.41839×10^{-3}
4	0.8949398516241052	0.3277254373962255	2.84478×10^{-4}
5	0.8946584205860127	0.3278253635007827	3.04725×10^{-6}
6	0.8946553750774177	0.3278265210518334	1.74273×10^{-9}
7	0.8946553733346976	0.3278265217462931	1.08802×10^{-14}
8	0.8946553733346867	0.3278265217462976	1.66533×10^{-16}
9	0.8946553733346867	0.3278265217462975	1.11022×10^{-16}

Table 5.11 Method (5.9) with $\lambda = 0.5$, $x_{-1} = (5, 5)$ and $x_0 = (1, 0)$

n	$x_n^{(1)}$	$x_n^{(2)}$	$\|x^* - x_n\|$
1	0.9829778065072182	0.0344753285929756	2.93351×10^{-1}
2	0.9191516755790264	0.3114163466921295	2.44963×10^{-2}
3	0.8976925362896486	0.3267124870002544	3.03037×10^{-3}
4	0.8947380642577267	0.3277957962677528	8.26909×10^{-5}
5	0.8946556314301652	0.3278264207451973	2.58095×10^{-7}
6	0.8946553733563231	0.3278265217375175	2.16364×10^{-11}
7	0.8946553733346867	0.3278265217462975	5.55112×10^{-17}

Therefore, the methods included in (5.9) improve the results given by other authors. Moreover, if the value of the parameter λ is increased, better approximations are obtained.

Table 5.12 Method (5.9) with $\lambda = 0.99$, $x_{-1} = (5, 5)$ and $x_0 = (1, 0)$

n	$x_n^{(1)}$	$x_n^{(2)}$	$\|x^* - x_n\|$
1	0.9228095274055251	0.3269365280425139	2.81542×10^{-2}
2	0.8959888360193688	0.3276958684879607	1.33346×10^{-3}
3	0.8946591561955859	0.3278259055081464	3.78286×10^{-6}
4	0.894655373452723	0.3278265217196517	1.18036×10^{-10}
5	0.8946553733346867	0.3278265217462975	1.11022×10^{-16}
6	0.8946553733346867	0.3278265217462975	5.55112×10^{-17}

5.6 Convergence for the Secant-Like Methods from Auxiliary Points

In this section, we present a local as well as a semilocal convergence analysis for a uniparametric family of Secant-like methods (5.9) in order to approximate a locally unique solution of an equation containing a non-differentiable term in a Banach space setting. In the local convergence case we obtain a wider choice of initial points, tighter error distances, and a more precise uniqueness ball. Moreover, in the semilocal convergence case, the sufficient convergence criteria are more flexible than in the earlier studies. This flexibility allows take a wider convergence domain, tighter error estimates on the distances involved and an at least as precise information on the location of the solution. Numerical examples justify our theoretical results.

The choice of an iterative process for solving (5.1) usually depends on its efficiency [35, 46], which relates the speed of convergence (order of convergence) of the method to its computational cost. However, there is another important aspect that is usually less taken into account: the accessibility of the iterative process. The accessibility of an iterative process shows the domain of starting points from which the sequence $\{x_n\}$ given by a iterative process converges to a solution of the Eq. (5.1). The location of starting approximations, from which the iterative processes converge to a solution of the equation, is a difficult problem to solve. This location is from the study of the convergence that is made of the iterative process: local or semilocal.

To give a greater generality to our study, we are interested in approximating a solution x^* of a nonlinear Eq. (5.1) in Banach spaces, where $F : \Omega \subseteq X \to Y$ is a continuous nonlinear operator, but non-differentiable, and Ω is a nonempty open convex domain in the Banach space X with values in the Banach space Y. If the operator F is not differentiable, there are iterative methods, less studied, that do not use derivatives in their algorithms. This type of methods usually use divided differences instead of derivatives [38].

On the other hand, so far, the study of the local convergence of derivative-free iterative processes shows a small contradiction. Usually, for the known results of local convergence (see [15, 18, 32, 33, 40, 44], and references therein given) the existence of the operator $[F'(x^*)]^{-1}$, forcing the operator F to be Fréchet differentiable. These results therefore study the accessibility of the iterative process for Fréchet differentiable operators. However, in [29], by modifying the hypothesis about the solution x^*, a result of local convergence for (5.9) is obtained, where the operator F is non-differentiable.

Given that we are considering iterative processes to approximate solutions of non-differentiable operators, it seems logical to study the accessibility in this situation. The main aim of this work is to improve the local convergence result obtained in [29] and also extend it to study of the semilocal convergence of Secant-like methods (5.9), where the operator involved F is not differentiable. To do this, we take into account a given result in [15], where a procedure to increase the accessibility of the Secant method is given by combining ω-continuous and ω_0-center-continuous conditions. Notice that, if an operator divided difference is ω-continuous in Ω (5.16), the operator divided difference is ω_0-center-continuous for each pair of distinct points $(a, b) \in \Omega \times \Omega$:

$$\|[x, y; F] - [a, b; F]\| \leq \omega_0(\|x - a\|, \|y - b\|); \quad x, y \in \Omega, \text{for } x \neq y, \tag{5.48}$$

where $\omega_0 : \mathbb{R}_* \times \mathbb{R}_* \longrightarrow \mathbb{R}_*$ is a non-decreasing continuous function in its two variables. We also have that $\omega_0(s, t) \leq \omega(s, t)$ for each $s, t \in \mathbb{R}_*$, and the function $\frac{\omega}{\omega_0}$ can be arbitrarily large [15, 40].

By means of this procedure, we establish new local and semilocal convergence results for (5.9) when F is a non-differentiable operator.

The rest of the chapter is organized as follows. In Sect. 5.6.1, we obtain a local convergence result for the Secant-like iterative processes given by (5.9), including their ball of convergence, for non-differentiable operators. In Sect. 5.6.2, taking into account the result on local convergence obtained previously, we analyze the semilocal convergence of Secant-like iterative processes for non-differentiable operators. Finally, in Sect. 5.6.3, we present an application where we illustrate the results obtained.

5.6.1 Local Convergence Analysis

We shall show the local convergence of method (5.9) based on the following conditions (L):

(L1) There exist $x^* \in \Omega$ with $F(x^*) = 0$, $\delta > 0$ and $\tilde{x} \in \Omega$, with $\|\tilde{x} - x^*\| = \delta$, such that $[x^*, \tilde{x}; F]^{-1} \in \mathcal{L}(Y, X)$.

(L2) $\|[x^*, \tilde{x}; F]^{-1}([x, y; F] - [u, v; F])\| \leq \omega(\|x - u\|, \|y - v\|)$ holds for each $x, y, u, v \in \Omega$, with $x \neq y$ and $u \neq v$, where $\omega : \mathbb{R}_* \times \mathbb{R}_* \longrightarrow \mathbb{R}_*$ is a non-decreasing continuous function in its two variables.

(L2′) $\|[x^*, \tilde{x}; F]^{-1}([x, y; F] - [x^*, \tilde{x}; F])\| \leq \omega_0(\|x - x^*\|, \|y - \tilde{x}\|)$ holds for each $x, y \in \Omega$, with $x \neq y$, where $\omega_0 : \mathbb{R}_* \times \mathbb{R}_* \longrightarrow \mathbb{R}_*$ is a non-decreasing continuous function in its two variables.

(L3) The equation

$$\omega(2(1-\lambda)t, t) + \omega_0(t, \delta + t) - 1 = 0 \tag{5.49}$$

has at least one positive zero. Denote by R the smallest zero.

(L4) $B(x^*, R) \subseteq \Omega$ and $\omega_0(R, \delta + R) < 1$.

Notice that condition $(L2')$ is not additional to $(L2)$, since in practice the computation of function ω involves the computation of function ω_0 as a special case. That is the new results are obtained under the same computational cost as the old ones. However, we have

$$\omega_0(s, t) \leq \omega(s, t) \quad \text{for} \quad \text{each} \quad s, t \in \mathbb{R}_* \tag{5.50}$$

and the function $\frac{\omega}{\omega_0}$ can be arbitrarily large [15, 40].

Next, we present an auxiliary perturbation result on the divided difference of first order for the operator F.

Lemma 5.11 *Suppose that (L) conditions hold. If* $x, y \in B(x^*, R)$*, with* $x \neq y$*, then* $[x, y; F]^{-1} \in \mathcal{L}(Y, X)$ *and*

$$\|[x, y; F]^{-1}[x^*, \tilde{x}; F]\| \leq \frac{1}{1 - \omega_0(\|x - x^*\|, \|y - \tilde{x}\|)} \leq \frac{1}{1 - \omega_0(R, \delta + R)}. \tag{5.51}$$

Proof Using $(L2')$ and $(L4)$, we obtain in turn

$$\begin{aligned}\|[x^*, \tilde{x}; F]^{-1}([x^*, \tilde{x}; F] - [x, y; F])\| &\leq \omega_0(\|x^* - x\|, \|\tilde{x} - x^*\| + \|x^* - y\|) \\ &< \omega_0(R, \delta + R) < 1.\end{aligned}$$

Then, by the Banach Lemma on invertible operators [35], the operator $[x, y; F]^{-1} \in \mathcal{L}(Y, X)$ so that (5.51) is satisfied. □

We can now show the main result of local convergence for method (5.9) using the (L) conditions and the preceding notation.

Theorem 5.12 *Suppose that (L) conditions hold. Then, sequence* $\{x_n\}$ *generated for* $x_{-1}, x_0 \in B(x^*, R)$ *with* $x_{-1} \neq x_0$*, by method (5.9) is well defined, remains in* $B(x^*, R)$ *for each* $n = 0, 1, 2, \ldots$*, and converges to* x^*.

Proof From $\lambda \in [0, 1)$ and the first substep of method (5.9) for $n = 0$, we obtain that $y_0 \neq x_0$ and $y_0 \in B(x^*, R) \subseteq \Omega$. Then, by Lemma 5.11, $[y_0, x_0; F]^{-1} \in \mathcal{L}(Y, X)$. Hence, x_1 is well defined by the second substep of method (5.9) for $n = 0$. We can write from method (5.9) and $(L1)$

$$x_1 - x^* = x_0 - x^* - [y_0, x_0; F]^{-1}F(x_0) + [y_0, x_0; F]^{-1}F(x^*) \tag{5.52}$$

Using $(L2)$, $(L3)$, (5.51), and (5.52), we obtain in turn that

$$\begin{aligned}
\|x_1 - x^*\| &\leq \|[y_0, x_0; F]^{-1}[x^*, \tilde{x}; F]\| \|[x^*, \tilde{x}; F]^{-1}([y_0, x_0; F] \\
&\quad - [x_0, x^*; F])\| \|x_0 - x^*\| \\
&\leq \frac{\omega(\|y_0 - x_0\|, \|x_0 - x^*\|)}{1 - \omega_0(R, \delta + R)} \|x_0 - x^*\| \\
&< \frac{\omega(2(1-\lambda)R, R)}{1 - \omega_0(R, \delta + R)} \|x_0 - x^*\| \\
&= \|x_0 - x^*\|. \qquad (5.53)
\end{aligned}$$

Hence, $\|x_1 - x^*\| < \|x_0 - x^*\| < R$. That is, $x_1 \in B(x^*, R)$ and, therefore by the first substep of method (5.9), $y_1 \in B(x^*, R)$. We also have that $y_1 \neq x_1$, since $\lambda \in [0, 1)$, so $[y_1, x_1; F]^{-1}$ is well defined by Lemma 5.11. By preceding estimates we get an inductive argument that $\|x_{k+1} - x^*\| < \|x_k - x^*\| < R$. Hence, we conclude that $x_{k+1}, y_{k+1} \in B(x^*, R)$ and $\lim_{n \to +\infty} x_k = x^*$. □

Concerning the uniqueness of the solution x^*, we have the following result.

Theorem 5.13 *Under the conditions (L) suppose that there exists $R_1 \geq R$ such that*

$$\omega_0(0, \delta + R_1) < 1. \qquad (5.54)$$

Then, the limit point x^* is the only solution of equation (5.1) in $\overline{B(x^*, R_1)} \cap \Omega$.

Proof Let $y^* \in \overline{B(x^*, R_1)} \cap \Omega$ be such that $F(y^*) = 0$. Define $Q = [x^*, y^*; F]$. Then, using $(L2')$ and (5.54), we get in turn that

$$\begin{aligned}
\|[x^*, \tilde{x}; F]^{-1}([x^*, y^*; F] - [x^*, \tilde{x}; F])\| &\leq \omega_0(\|x^* - x^*\|, \|y^* - \tilde{x}\|) \\
&\leq \omega_0(0, \delta + R_1) < 1.
\end{aligned}$$

Hence, $Q^{-1} \in \mathcal{L}(Y, X)$. From $0 = F(x^*) - F(y^*) = Q(x^* - y^*)$, we deduce that $x^* = y^*$. □

Remarks 1

(1) If $\omega_0 = \omega$, then Lemma 5.11 and Theorem 5.12 reduce to Lemma 1 and Theorem 2 given in [29].
(2) Notice that a uniqueness result was not given in [29].
(3) If strict inequality holds in (5.50), then the following advantages are obtained:

(i) New results are given in affine invariant form whereas the results in [29] are given in nonaffine invariant form. The advantages of affine over nonaffine invariant form are explained in [21, 38].

(ii) New radius of convergence is larger than the old one, see $(L3)$ (for $\omega_0 = \omega$), (5.51) and the last condition in $(L4)$. This advantage allows for a wider choice of initial points.

(iii) New error bounds (see (5.53) for $\omega_0 = \omega$ to get the old ones) are more precise than the old ones. That is fewer steps are needed to obtain the same error tolerance.

5.6.2 Semilocal Convergence Analysis

Now, we present the semilocal convergence analysis of method (5.9). First, we consider the conditions about the divided difference of operator F.

(SL1) There exist $x_0 \in \Omega$, $\mu > 0$ and $\tilde{x} \in \Omega$, with $\|x_0 - \tilde{x}\| = \mu$, such that $[\tilde{x}, x_0; F]^{-1} \in \mathcal{L}(Y, X)$.

(SL2) $\|[\tilde{x}, x_0; F]^{-1}([x, y; F] - [u, v; F])\| \leq \psi(\|x-u\|, \|y-v\|)$ holds for each $x, y, u, v \in \Omega$, with $x \neq y$ and $u \neq v$, where $\psi : \mathbb{R}_* \times \mathbb{R}_* \longrightarrow \mathbb{R}_*$ is a non-decreasing continuous function in its two variables.

As in the case of $(L2)$ it follows from $(SL2)$:

(SL2′) $\|[\tilde{x}, x_0; F]^{-1}([x, y; F] - [\tilde{x}, x_0; F])\| \leq \psi_0(\|x - \tilde{x}\|, \|y - x_0\|)$ holds for each $x, y \in \Omega$, with $x \neq y$, where $\psi_0 : \mathbb{R}_* \times \mathbb{R}_* \longrightarrow \mathbb{R}_*$ is a non-decreasing continuous function in its two variables.

Then, we have again that $\psi_0(s, t) \leq \psi(s, t)$ for each $s, t \in \mathbb{R}_*$, and $\frac{\psi}{\psi_0}$ can be arbitrarily large.

Next, we present a perturbation result of the divided difference of the operator F.

Lemma 5.14 *Under conditions $(SL1)$ and $(SL2')$, if there exists $r \in \mathbb{R}_+$ such that $B(x_0, r) \subseteq \Omega$ and $\psi_0(\mu + r, r) < 1$ holds, then $[x, y; F]^{-1} \in \mathcal{L}(Y, X)$ and*

$$\|[x, y; F]^{-1}[\tilde{x}, x_0; F]\| \leq \frac{1}{1 - \psi_0(\|\tilde{x} - x\|, \|x_0 - y\|)} \leq \frac{1}{1 - \psi_0(\mu + r, r)}, \tag{5.55}$$

for each pair of distinct points $(x, y) \in B(x_0, r) \times B(x_0, r)$.

Proof Using $(SL2')$, we obtain in turn

$$\begin{aligned} \|[\tilde{x}, x_0; F]^{-1}([\tilde{x}, x_0; F] - [x, y; F])\| &\leq \psi_0(\|\tilde{x} - x\|, \|x_0 - y\|) \\ &\leq \psi_0(\|\tilde{x} - x_0\| + \|x_0 - x\|, \|x_0 - y\|) \\ &\leq \psi_0(\mu + r, r) < 1. \end{aligned}$$

Then, by the Banach Lemma on invertible operators, the operator $[x, y; F]^{-1} \in \mathcal{L}(Y, X)$ and (5.55) is satisfied. □

Notice that if $x_{-1} \in \Omega$, with $\|x_{-1} - x_0\| = \alpha > 0$, then, by $\lambda \in [0, 1)$, $y_0 \in \Omega$, $y_0 \neq x_0$ and therefore $[y_0, x_0; F] \in \mathcal{L}(X, Y)$. So, by Lemma 5.14, it follows that $[y_0, x_0; F]^{-1} \in \mathcal{L}(Y, X)$. Suppose $\|[y_0, x_0; F]^{-1}F(x_0)\| \leq \eta$. If moreover $x_{-1} \in B(x_0, r)$, then $y_0 \in B(x_0, r)$.

Taking into account the preceding notation, we shall show the main semilocal convergence result for method (5.9) based on the conditions (*SL*): $(SL1)$, $(SL2)$, $(SL2')$ and

(SL3) The equation

$$(g_0(t) + 1 - g(t))\,\eta - (1 - g(t))t = 0, \tag{5.56}$$

where $g_0(t) = \frac{\psi(\eta+(1-\lambda)\alpha,0)}{1-\psi_0(\mu+t,t)}$, $g(t) = \frac{\psi(\eta+(1-\lambda)\eta,0)}{1-\psi_0(\mu+t,t)}$, has at least one positive zero. Denote by r the smallest such zero.

(SL4) $B(x_0, r) \subseteq \Omega$ and $g_0(r) + g(r) < 1$.

Theorem 5.15 *Suppose that conditions (SL) hold. Then, sequence $\{x_n\}$ generated for x_0, x_{-1} with $\|x_{-1} - x_0\| = \alpha > 0$, by method (5.9) is well defined, remains in $B(x_0, r)$ for each $n = 0, 1, 2, \ldots$, and converges to a solution $x^* \in \overline{B(x_0, r)}$ of Eq. (5.1).*

Proof From $\lambda \in [0, 1)$, as we have just shown previously $[y_0, x_0; F]^{-1} \in \mathcal{L}(Y, X)$, then x_1 is well defined. Moreover, as $\|x_1 - x_0\| \leq \eta$, by (5.56) we get $x_1 \in B(x_0, r)$. On the other hand, by (5.56), we get in turn that

$$\|y_1 - x_0\| = \|\lambda x_1 + (1 - \lambda)x_0 - x_0\| \leq \lambda\|x_1 - x_0\| \leq \lambda\eta < r,$$

so $y_1 \in B(x_0, r)$.

We can write

$$F(x_1) = F(x_1) - F(x_0) - [y_0, x_0; F](x_1 - x_0) = ([x_1, x_0; F] - [y_0, x_0; F])(x_1 - x_0). \tag{5.57}$$

Then, we have by (5.55) and (5.57) that

$$\begin{aligned}
\|x_2 - x_1\| &= \|([y_1, x_1; F]^{-1}[\tilde{x}, x_0; F])[\tilde{x}, x_0; F]^{-1}([x_1, x_0; F] \\
&\quad -[y_0, x_0; F])(x_1 - x_0)\| \leq \|[y_1, x_1; F]^{-1}[\tilde{x}, x_0; F]\|\|[\tilde{x}, x_0; F]^{-1} \\
&\quad \times ([x_1, x_0; F] - [y_0, x_0; F])\|\|x_1 - x_0\| \\
&\leq \frac{\psi(\|x_1 - y_0\|, 0)}{1 - \psi_0(\|\tilde{x} - y_1\|, \|x_0 - x_1\|)}\|x_1 - x_0\|
\end{aligned}$$

$$\leq \frac{\psi(\|x_1 - x_0\| + \|x_0 - y_0\|, 0)}{1 - \psi_0(\|\tilde{x} - x_0\| + \|x_0 - y_1\|, \|x_0 - x_1\|\|)} \|x_1 - x_0\|$$
$$< \frac{\psi(\eta + (1-\lambda)\alpha, 0)}{1 - \psi_0(\mu + r, r)} \|x_1 - x_0\|$$
$$= g_0(r)\|x_1 - x_0\|. \tag{5.58}$$

As $g_0(r) < 1$ by $(SL4)$, from (5.56) and (5.58), we get that

$$\|x_2 - x_0\| \leq \|x_2 - x_1\| + \|x_1 - x_0\| < (g_0(r) + 1)\|x_1 - x_0\| < r,$$

and $x_2 \in B(x_0, r)$. Obviously, we get

$$\|y_2 - x_0\| = \|\lambda x_2 + (1-\lambda)x_1 - x_0\| \leq \lambda\|x_2 - x_0\| + (1-\lambda)\|x_1 - x_0\| < r, \tag{5.59}$$

and $y_2 \in B(x_0, r)$.

Similarly, we have

$$\|x_3 - x_2\| \leq \|[y_2, x_2; F]^{-1}[\tilde{x}, x_0; F]\|\|[\tilde{x}, x_0; F]^{-1}([x_2, x_1; F]$$
$$- ([y_1, x_1; F])\|\|x_2 - x_1\|$$
$$\leq \frac{\psi(\|x_2 - y_1\|, 0)}{1 - \psi_0(\|\tilde{x} - y_2\|, \|x_0 - x_2\|)} \|x_2 - x_1\|$$
$$\leq \frac{\psi(\|x_2 - x_1\| + \|x_1 - y_1\|, 0)}{1 - \psi_0(\|\tilde{x} - x_0\| + \|x_0 - y_2\|, \|x_0 - x_2\|\|)} \|x_2 - x_1\|$$
$$< \frac{\psi(\eta + (1-\lambda)\eta, 0)}{1 - \psi_0(\mu + r, r)} \|x_2 - x_1\|$$
$$= g(r)\|x_2 - x_1\|.$$

Then, as $g(r) < 1$ by $(SL4)$, from (5.56) and (5.58), we get

$$\|x_3 - x_0\| \leq \|x_3 - x_2\| + \|x_2 - x_1\| + \|x_1 - x_0\| < (g(r) + 1)\|x_2 - x_1\|$$
$$+ \|x_1 - x_0\| < [(g(r) + 1)g_0(r) + 1]\eta < r.$$

Therefore, $x_3 \in B(x_0, r)$ and, as before, $y_3 \in B(x_0, r)$:

$$\|y_3 - x_0\| = \|\lambda x_3 + (1-\lambda)x_2 - x_0\| \leq \lambda\|x_3 - x_0\| + (1-\lambda)\|x_2 - x_0\| < r.$$

We prove that the following four items are satisfied, for $j \geq 2$, by the sequence $\{x_n\}$:

(I) $F(x_j) = ([x_j, x_{j-1}; F] - [y_{j-1}, x_{j-1}; F])(x_j - x_{j-1})$,
(II) $\|x_{j+1} - x_j\| \leq g(r)\|x_j - x_{j-1}\| \leq g(r)^{j-1}\|x_2 - x_1\| \leq g(r)^{j-1} g_0(r)\|x_1 - x_0\| < \eta$,
(III) $\|x_{j+1} - x_0\| \leq \|x_{j+1} - x_j\| + \|x_j - x_{j-1}\| + \cdots + \|x_1 - x_0\|$

$< (g(r)^{j-1} + \ldots + g(r) + 1)\|x_2 - x_1\| + \|x_1 - x_0\| < \frac{1}{1-g(r)}\|x_2 - x_1\| + \|x_1 - x_0\|$

$< (\frac{g_0(r)}{1-g(r)} + 1)\,\eta = r$
(IV) $x_{j+1}, y_{j+1} \in B(x_0, r)$.

We have that items (I)–(IV) hold for $j = 2$. If we now suppose that (I)–(IV) are true for some $j = k$, by induction, we prove that (I)-(IV) hold for $k + 1$. It follows, by analogy to the case where $j = 2$, that (I) holds for $k + 1$. Next, we prove (II).

$$
\begin{aligned}
\|x_{k+2} - x_{k+1}\| &= \|[y_{k+1}, x_{k+1}; F]^{-1} F(x_{k+1})\| \\
&\leq \|[y_{k+1}, x_{k+1}; F]^{-1}[\tilde{x}, x_0; F]\| \|[\tilde{x}, x_0; F]^{-1}([x_{k+1}, x_k; F] \\
&\quad -[y_k, x_k; F])(x_{k+1} - x_k)\| \\
&\leq \frac{\psi(\|x_{k+1} - y_k\|, 0)}{1 - \psi_0(\|y_{k+1} - \tilde{x}\|, \|x_{k+1} - x_0\|)}\|x_{k+1} - x_k\| \\
&\leq \frac{\psi(\|x_{k+1} - x_k\| + \|x_k - y_k\|, 0)}{1 - \psi_0(\|y_{k+1} - x_0\| + \|x_0 - \tilde{x}\|, \|x_{k+1} - x_0\|)}\|x_{k+1} - x_k\| \\
&< \frac{\psi(\eta + (1-\lambda)\|x_k - x_{k-1}\|, 0)}{1 - \psi_0(\mu + r, r)}\|x_{k+1} - x_k\| \\
&< \frac{\psi(\eta + (1-\lambda)\eta, 0)}{1 - \psi_0(\alpha + r, r)}\|x_{k+1} - x_k\| \\
&= g(r)\|x_{k+1} - x_k\| \leq g(r)^k g_0(r)\|x_1 - x_0\|.
\end{aligned}
$$

Moreover, by the induction hypothesis and (5.56), we get

$$
\begin{aligned}
\|x_{k+2} - x_0\| &\leq \|x_{k+2} - x_{k+1}\| + \|x_{k+1} - x_k\| + \cdots + \|x_2 - x_1\| + \|x_1 - x_0\| \\
&\leq (1 + g(r) + \cdots + g(r)^k) g_0(r)\|x_1 - x_0\| + \|x_1 - x_0\| \\
&< \left(\frac{g_0(r)}{1 - g(r)} + 1\right)\|x_1 - x_0\| \leq \left(\frac{g_0(r)}{1 - g(r)} + 1\right)\eta = r,
\end{aligned}
$$

so, $x_{k+2} \in B(x_0, r)$ and obviously, as in (5.59), $y_{k+2} \in B(x_0, r)$. Therefore, the items (III) and (IV) are proved.

From (5.58) and (II), we have

$$\|x_{k+2} - x_{k+1}\| = \|[y_{k+1}, x_{k+1}; F]^{-1} F(x_{k+1})\| \leq g(r)^{k+1} g_0(r) \|x_1 - x_0\|.$$

It follows that $\{x_n\}$ is a complete sequence in a Banach space X and as such it converges to some $x^* \in \overline{B(x_0, r)}$.

We showed the estimate

$$\begin{aligned}\|[\tilde{x}, x_0; F]^{-1} F(x_{k+1})\| &= \|[\tilde{x}, x_0; F]^{-1}([x_{k+1}, x_k; F] - [y_k, x_k; F])(x_{k+1} - x_k)\| \\ &\leq \psi(\eta + (1 - \lambda)\eta, 0)\|x_{k+1} - x_k\|.\end{aligned}$$

By letting $k \to +\infty$, we conclude that $F(x^*) = 0$. □

Concerning the uniqueness of the solution x^*, we obtain the following result.

Theorem 5.16 *Under the conditions (SL), we suppose that there exists $r_1 \geq r$ such that*

$$\psi_0(\mu + r, r_1) < 1. \tag{5.60}$$

Then, the limit point x^* is the only solution of Eq. (5.1) in $\overline{B(x_0, r_1)} \cap \Omega$.

Proof As in Theorem 5.13, but using $(SL2')$ instead of $(L2')$, we get by (5.60) that

$$\begin{aligned}\|[\tilde{x}, x_0; F]^{-1}([x^*, y^*; F] - [\tilde{x}, x_0; F])\| &\leq \psi_0(\|x^* - \tilde{x}\|, \|y^* - x_0\|) \\ &\leq \psi_0(\mu + r, r_1) < 1.\end{aligned}$$

Hence, we have that $Q^{-1} \in \mathcal{L}(Y, X)$. □

Remarks 2

(1) Note that Theorem 5.15 and Theorem 5.16 reduce to the semilocal convergence result (Theorem 3.2) given in [27] where $\psi_0 = \psi$, $\tilde{x} = y_0$ and

$$m = \max\{\psi(\eta + (1 - \lambda)\alpha, 0), \psi(\eta + (1 - \lambda)\eta, 0)\}.$$

(2) In the result of semilocal convergence given in Theorem 5.15 not only the error bounds but the information on the convergence domain (i. e. r) and conditions $(SL3)$ and $(SL4)$ can be relaxed as follows: In view of $(SL2)$,

(SL2$''$) there exists a non-decreasing continuous function in its two variables $\widetilde{\psi} : \mathbb{R}_* \times \mathbb{R}_* \longrightarrow \mathbb{R}_*$ such that $\|[\tilde{x}, x_0; F]^{-1}([x, y; F] - [y_0, x_0; F])\| \leq \widetilde{\psi}(\|x - y_0\|, \|y - x_0\|)$ holds for each pair of distinct points $(x, y) \in \Omega \times \Omega$.

We have that

$$\widetilde{\psi}(s,t) \leq \psi(s,t), \quad \text{for} \quad \text{each } s,t \in \mathbb{R}_*. \tag{5.61}$$

Define function $\widetilde{g}_0 : \mathbb{R}_* \longrightarrow \mathbb{R}_*$ by

$$\widetilde{g}_0(t) = \frac{\widetilde{\psi}(\eta + (1-\lambda)\alpha, 0)}{1 - \psi_0(\mu + t, t)}. \tag{5.62}$$

Then, according to the proof of Theorem 5.15 (see (5.58)), conditions $(SL3)$ and $(SL4)$ can be replaced, respectively, by

(SL3′) The equation

$$(\widetilde{g}_0(t) + 1 - g(t))\,\eta - (1 - g(t))t = 0$$

has at least one positive zero, where $\widetilde{g}_0$ is defined by (5.62) and g is given in $(SL3)$. Denote by $\widetilde{r}$ the smallest zero.

(SL4′) $B(x_0, \widetilde{r}) \subseteq \Omega$ and $\widetilde{g}_0(\widetilde{r}) + g(\widetilde{r}) < 1$.

Then, the conclusions of Theorem 5.15 hold in this weaker setting (provided that strict inequality holds in (5.61)). Notice also that in this case we have that $\widetilde{g}_0(t) < g_0(t)$.

5.6.3 Numerical Example

In this section, we apply the local and semilocal convergence results given above to solve a nonlinear system. Let $F : \Omega \subseteq \mathbb{R}^3 \to \mathbb{R}^3$ be defined by

$$F(x_1, x_2, x_3) = (x_1 + 0.0125|x_1|, x_2^2 + x_2 + 0.0125|x_2|, e^{x_3} - 1).$$

We consider $\Omega = B(0, 1)$. Obviously, a solution of $F(\mathbf{x}) = 0$ is $\mathbf{x}^* = (0, 0, 0)$.

For $\mathbf{u} = (u_1, u_2, u_3)^T$, $\mathbf{v} = (v_1, v_2, v_3)^T \in \mathbb{R}^3$, with $u_j \neq v_j$ for $j = 1, 2, 3$, we use the divided difference of first order given in (5.12).

It is easy to prove that $[x, y; F](x - y) = F(x) - F(y)$ with x and y having a some different component. Moreover, it is clear that $x_n = x_{n+1} = x^*$, where x^* is the solution of the problem, if the three components coincide for two terms, $x_n = x_{n+1}$, when applying the Secant-like method, and then no more iterations are done. So, the Secant-like method works correctly computationally with this divided difference of first order.

So, it follows

$$\|[\mathbf{x}, \mathbf{u}; F] - [\mathbf{y}, \mathbf{v}; F]\| \leq \frac{\mathrm{e}}{2}(\|\mathbf{x} - \mathbf{y}\| + \|\mathbf{u} - \mathbf{v}\|) + 0.025,$$

$$\|[\mathbf{x}, \mathbf{u}; F] - [\mathbf{x}^*, \tilde{\mathbf{x}}; F]\| \leq (\|\mathbf{x} - \mathbf{x}^*\| + \|\mathbf{u} - \tilde{\mathbf{x}}\|) + 0.025,$$

for $\mathbf{x} \neq \mathbf{u}$ and $\mathbf{y} \neq \mathbf{v}$.

Then, we obtain the functions

$$\omega(s, t) = \gamma\left(\frac{e}{2}(s + t) + 0.025\right), \quad \omega_0(s, t) = \gamma((s + t) + 0.025),$$

where $\gamma = \|[x^*, \tilde{x}; F]^{-1}\|$.

We choose $\tilde{\mathbf{x}} = (0.2, 0.2, 0.1)$ and $\lambda = 0$. We obtain using the notation of the local result:

$$\delta = 0.2, \quad \|[x^*, \tilde{x}; F]^{-1}\| = 0.987654, \quad R = 0.125464,$$

$$\omega_0(R, \delta + R) = 0.470053 < 1, \quad R_1 = 0.78749.$$

Therefore, the hypotheses of Theorem 5.12 are fulfilled and the sequence $\{x_n\}$, given by (5.9) with $\lambda = 0$, is well defined, converges to $\mathbf{x}^* = \mathbf{0}$, and remains in $B(x^*, 0.125464) \subseteq \Omega$. Moreover, the solution $\mathbf{x}^*$ is unique in $\overline{B(x^*, 0.78749)} \subseteq \Omega$, since $\omega_0(0, \delta + R_1) < 1$.

Note that the radius $R = 0.125464$ is larger than the old radius $R = 0.101634$ obtained for $\omega_0 = \omega$ in [29]. Thus, just as it happened in [29], for greater lambda, a greater radio is obtained. As it is easy to check, for $\lambda = 0.5$, we obtain $R_{\lambda=0.5} = 0.161605$ and for $\lambda = 0.9$, $R_{\lambda=0.9} = 0.209999$.

Next, we study the semilocal convergence. We consider $\tilde{\mathbf{x}} = (0.02, 0.02, 0)$. Then, we obtain

$$\|[\mathbf{x}, \mathbf{u}; F] - [\tilde{\mathbf{x}}, \mathbf{x_0}; F]\| \leq (\|\mathbf{x} - \tilde{\mathbf{x}}\| + \|\mathbf{u} - \mathbf{x_0}\|) + 0.025, \text{ for } \mathbf{x} \neq \mathbf{u}.$$

Therefore, we have the functions

$$\psi(s, t) = \beta\left(\frac{e}{2}(s + t) + 0.025\right), \quad \psi_0(s, t) = \beta((s + t) + 0.025),$$

where $\beta = \|[\tilde{x}, x_0; F]^{-1}\|$.

We choose $\mathbf{x_0} = (0.1, 0.1, 0.01)$, $\mathbf{x_{-1}} = \mathbf{x_0} + 10^{-2}$ and $\lambda = 0$. Using the notation of Theorem 5.15, we obtain

$$\alpha = 0.01, \quad \mu = 0.08, \quad \beta = 0.995008, \quad r = 0.163106, \quad g_0(r) + g(r) = 0.822036 < 1.$$

Therefore, the hypotheses of Theorem 5.15 are fulfilled and the sequence $\{x_n\}$ given by (5.9) for $\lambda = 0$ is well defined, converges, and remains in $B(x_0, r) \subseteq \Omega$. Moreover, the uniqueness ball is $\overline{B(x_0, 0.736785)} \subseteq \Omega$, since $\psi_0(\mu + r, r_1) < 1$.

References

1. V. Alarcón, S. Amat, S. Busquier, D.J. López, A Steffensen's type method in Banach spaces with applications on boundary-value problems. J. Comput. Appl. Math. **216**, 243–250 (2008)
2. S. Amat, S. Busquier, On a higher order secant method. Appl. Math. Comput. **141**, 321–329 (2003)
3. S. Amat, S. Busquier, A modified secant method for semismooth equations. Appl. Math. Lett. **16**, 877–881 (2003)
4. S. Amat, M. Grau-Sánchez, M. Noguera, On the approximation of derivatives using divided differences operators preserving the local convergence order of iterative methods. J. Comput. Appl. Math. **237**, 363–372 (2013)
5. A.A. Andronow, C.E. Chaikin, *Theory of Oscillations* (Princeton University Press, Princeton, 1949)
6. I.K. Argyros, The Secant method and fixed points of nonlinear operators. Monatshefte für Mathematik **106**, 85–94 (1988)
7. I.K. Argyros, The Newton-Kantorovich Method under mild differentiability conditions and the Ptâk error estimates. Monatshefte für Mathematik **109**, 175–193 (1990)
8. I.K. Argyros, On the Secant method. Publ. Math. Debrecen **43**(3–4), 223–238 (1993)
9. I.K. Argyros, Some methods for finding error bounds for Newton-like methods under mild differentiability conditions. Acta Math. Hung. **61**, 183–194 (1993)
10. I.K. Argyros, A convergence theorem for Newton-like methods under generalized Chen-Yamamoto-type assumptions. Appl. Math. Comput. **61**, 25–37 (1994)
11. I.K. Argyros, Newton-like methods under mild differentiability conditions with error analysis. Bull. Austral. Math. Soc. **37**, 131–147 (1998)
12. I.K. Argyros, S.K. Khattri, On the secant method. J. Complex. **29**(6), 454–471 (2013)
13. I.K. Argyros, Á.A. Magreñán, A unified convergence analysis for secant-type methods. J. Korean Math. Soc. **51**(6), 1155–1175 (2014)
14. I.K. Argyros, Á.A. Magreñán, Relaxed secant-type methods. Nonlinear Stud. **21**(3), 485–503 (2014)
15. I.K. Argyros, H. Ren, On an improved local convergence analysis for the Secant method. Numer. Algor. **52**, 257–271 (2009)
16. M. Balazs, G. Goldner, On existence of divided differences in linear spaces. Rev. Anal. Numer. Theor. Approx. **2**, 3–6 (1973)
17. D.D. Bruns, J.E. Bailey, Nonlinear feedback control for operating a nonisothermal CSTR near an unstable steady state. Chem. Eng. Sci. **32**, 257–264 (1977)
18. J. Chen, Z. Shen, Convergence analysis of the secant type methods. Appl. Math. Comput. **188**, 514–524 (2007)
19. K. Deimling, *Nonlinear Functional Analysis* (Springer, Berlin, 1985)
20. J.E. Dennis, Toward a unified convergence theory for Newton-like methods, in *Article in Nonlinear Functional Analysis and Applications*, ed. by L.B. Rall (Academic Press, New York, 1970), pp. 425–472
21. P. Deuflhard, G. Heindl, Affine invariant convergence theorems for Newton's method and extensions to related methods. Siam J. Numer. Anal. **16**, 1–10 (1979)
22. J.A. Ezquerro, M.A. Hernández, Multipoint super-Halley type approximation algorithms in Banach spaces. Numer. Funct. Anal. Optimiz. **21**, 845–858 (2000)

23. J.A. Ezquerro, M.A. Hernández, A.I. Velasco, An analysis of the semilocal convergence for Secant-like methods. Appl. Math. Comput. **266**, 883–892 (2015)
24. M. Ganesh, M.C. Joshi, Numerical solvability of Hammerstein integral equations of mixed type. IMA J. Numer. Anal. **11**, 21–31 (1991)
25. J.M. Gutiérrez, M.A. Hernández, New recurrence relations for Chebyshev method. Appl. Math. Lett. **10**, 63–65 (1997)
26. M.A. Hernández, M.J. Rubio, A new type of recurrence relations for the Secant method. Int. J. Comput. Math. **72**, 477–490 (1999)
27. M.A. Hernández, M.J. Rubio, A uniparametric family of iterative processes for solving nondifferentiable equations. J. Math. Anal. Appl. **275**, 821–834 (2002)
28. M.A. Hernández, M.J. Rubio, Solving a special case of conservative problems by secant-like methods. Appl. Math. Comput. **169**(2), 926–942 (2005)
29. M.A. Hernández, M.J. Rubio, On the ball of convergence of Secant-like methods for non-differentiable operators. Appl. Math. Comput. **273**, 506–512 (2016)
30. M.A. Hernández, M.J. Rubio, J.A. Ezquerro, Secant-like methods for solving nonlinear integral equations of the Hammerstein type. J. Comput. Appl. Math. **115**, 245–254 (2000)
31. M.A. Hernández, M.A. Salanova, Chebyshev method and convexity. Appl. Math. Comput. **95**, 51–62 (1998)
32. R. Hongmin, W. Qingiao, The convergence ball of the Secant method under Hölder continuous divided differences. J. Comput. Appl. Math. **194**, 284–293 (2006)
33. L. Kewei, Homocentric convergence ball of the Secant method. Appl. Math. J. Chinese Univ. Ser. B **22**(3), 353–365 (2007)
34. S. Kumar, H. Sloan, A new collocation-type method for Hammerstein integral equations. Math. Comput. **48**, 585–593 (1987)
35. A.M. Ostrowski, *Solutions of Equations and System of Equations* (Academic Press, New York, 1960)
36. D. Porter, D.S.G. Stirling, *Integral Equations* (Cambridge University Press, Cambridge, 1990)
37. F.A. Potra, On a modified Secant method. Nal. Numer. Theor. Approx. **8**, 203–214 (1979)
38. F.A. Potra, V. Pták, *Nondiscrete induction and iterative methods* (Pitman Publishing Limited, London, 1984)
39. J. Rashidinia, M. Zarebnia, New approach for numerical solution of Hammerstein integral equations. Appl. Math. Comput. **185**, 147–154 (2007)
40. H. Ren, I.K. Argyros, Local convergence of efficient Secant-type methods for solving nonlinear equations. Appl. Math. Comput. **218**, 7655–7664 (2012)
41. J. Rokne, Newton's method under mild differentiability conditions with error analysis. Numer. Math. **18**, 401–412 (1972)
42. J.W. Schmidt, Regula-falsi Verfahren mit konsistenter Steigung und Majoranten prinzip. Period. Math. Hung. **5**, 187–193 (1974)
43. A. Sergeev, On the method of chords. Sibirsk. Mat. Ž. **2**, 282–289 (1961)
44. S. Shakhno, On the Secant method under generalized Lipschitz conditions for the divided difference operator. PAMM-Proc. Appl. Math. Mech. **7**, 2060083–2060084 (2007)
45. J.J. Stoker, *Nonlinear Vibrations* (Interscience-Wiley, New York, 1950)
46. J.F. Traub, *Iterative Methods for the Solution of Equations* (Prentice-Hall, Englewood Cliffs, 1964)
47. W. Werner, Newton-like methods for the computation of fixed points. Comp. & Maths. Appls. **10**, 77–86 (1984)
48. T. Yamamoto, A note on a posteriori error bound of Zabrejko and Nguen for Zicenko's iteration. Numer. Funct. Anal. Optimiz. **9**, 987–994 (1987)

Chapter 6
Spacetimes as Topological Spaces, and the Need to Take Methods of General Topology More Seriously

Kyriakos Papadopoulos and Fabio Scardigli

Abstract Why is the manifold topology in a spacetime taken for granted? Why do we prefer to use Riemann open balls as basic-open sets, while there also exists a Lorentz metric? Which topology is a best candidate for a spacetime: a topology sufficient for the description of spacetime singularities or a topology which incorporates the causal structure? Or both? Is it more preferable to consider a topology with as many physical properties as possible, whose description might be complicated and counterintuitive, or a topology which can be described via a countable basis but misses some important information? These are just a few from the questions that we ask in this chapter, which serves as a critical review of the terrain and contains a survey with remarks, corrections and open questions.

Keywords Zeeman-Göbel topologies · Topologising a spacetime · Spacetime singularities · Causal topologies · Manifold topology

6.1 Introduction

6.1.1 The Manifold Topology vs. Finer or Incomparable Topologies

In [16], the author supports that the manifold topology in a curved spacetime is the best possible and most natural choice, against the class of topologies that was suggested by Zeeman and Göbel (see [38] and [13], respectively). His main focus

K. Papadopoulos (✉)
Department of Mathematics, Kuwait University, Safat, Kuwait

F. Scardigli
Dipartimento di Matematica, Politecnico di Milano, Milano, Italy

Institute-Lorentz for Theoretical Physics, Leiden University, Leiden, The Netherlands
e-mail: fabio@phys.ntu.edu.tw

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_6

lies on topologies finer than the manifold one, but there is a misjudgement here: there are topologies in the class $\mathfrak{Z}$ of Zeeman-Göbel topologies (as we shall see in paragraph 4) that are neither finer nor equal nor coarser than the manifold topology.

Instead of asking whether we need a finer topology for a sufficient mathematical description of a spacetime, we bring the topologisation question into a different level: why should one prefer a topology which describes spacetime singularities against a topology which hides singularities but incorporates the causal structure of the spacetime? As we shall see, the singularity theorems were proven under the frame of the manifold topology, while there are topologies in the class $\mathfrak{Z}$ where the limit curve theorem (LCT for abbreviation) fails to hold, and thus sufficient conditions for the formation of singularities as we understand them in the presence of Riemannian basic-open balls fail as well.

Zeeman's main arguments against the Euclidean $\mathbb{R}^4$ topology for Minkowski spacetime M (extended by Göbel for curved spacetimes) can be summarised as follows:

1. The 4-dimensional Euclidean topology is locally homogeneous, whereas M is not; every point has associated with it a light cone, separating space vectors from time vectors.
2. The group of all homeomorphisms of 4-dimensional Euclidean space is vast, and of no physical significance.

Heathcote's antilogue belongs to (sic) a realist view of spacetime topology as against the instrumentalist position. A realist point of view divides the space into structural levels, such as metric tensor field, affine connection, conformal structure, differentiable manifold and topology. Heathcote highlights that the manifold topology is present as long as the structure of manifold is present, and there are two "untenable" possibilities for a replacement of the manifold topology, in both cases by finer topologies (see [16], page 255, for more details). Heathcote's arguments miss here that there are topologies in the class $\mathfrak{Z}$ that are neither finer nor coarser than the manifold topology, as we shall see, but our disagreement does not lie only on this ground; we believe that the answer to the question "what comes first, the metric or the topology" cannot be a definite answer in favour of the metric (see paragraph 5). There is a Lorentz metric which is ignored by the Riemann open balls that serve as basic-open sets for the manifold topology. In addition, there are topologies different (and not finer, coarser or equal) than the manifold topology, which incorporate the causal structure of the spacetime and they could be considered as natural topologies for a spacetime, as well.

In the recent articles (see [5] and [26] and paragraph 6 here) the authors talked about topologies in the class $\mathfrak{Z}$, in the sense that general relativity generically leads to spacetime singularities where it breaks down as a physical theory; a particular topology in $\mathfrak{Z}$ (that we have called Z), different than the manifold one, was proven to be the most natural one for this frame. It is in those papers that the authors left as an open question a different approach to the topologisation of spacetime: the definition of a dynamical evolution of the spacetime given specific causal and topological conditions (see paragraph 7). It is conjectured that the challenges or even contradictions that arise in the study and understanding of spacetime geometry are due to the "static" nature of a topological structure; indeed, there is a specific fixed topology in the background and the topological properties arising from this topology affect the spacetime as a whole. This rigidity in the study of spacetime geometry might be cured if one develops a topological space with an evolving topology, which will incorporate quantum and relativistic frames within the spacetime and outside it (Planck time and length, singularities, etc.). One of the aims of this critical survey is to open such a discussion as well.

The different approaches in the study of spacetime geometry, and in the topologisation of a spacetime in our case, are due to different cultural backgrounds; Penrose states a similar argument in [32]. Those who come from QFT (quantum field theory), for example, and those from Einstein's general relativity seem to view things in a different way (here we should add those who come from a purely mathematical background, as well). Those from QFT, according to Penrose, would tend to take renormalisability or, better, finiteness, as the primary aim of the union of relativity and quantum theories. Those having a relativistic background would take the deep conceptual conflicts (determinism, causality, background independence) between the principles of quantum mechanics and those of general relativity to be the centrally important issues that needed to be resolved, and from whose resolution we should expect to move forward to a new physics of the future. Those from a purely (theoretical) mathematical background, coming straight from the Platonic world of Penrose, would love to see a spacetime as an integrated mathematical entity, a structure with physical properties coinciding harmonically with the mathematical formulation.

It should be said that the description of fluctuating topologies, or topological transitions, has become a debated topic in the theoretical physics since the visionary introduction of the concept of *Spacetime Foam*, by John Wheeler in the 1950s [36]. In string theory these ideas have been explored in the early 1990s, among others, by Greene (see Refs. [6]), and innumerable have been the applications of the concept of spacetime foam to different problems (see, e.g., [33]). In recent years, research ideas emerged that aim to derive the concept of spacetime itself from quantum entanglement. The seminal paper of Raamsdonk [35] paved the way to the more recent works of Susskind and Maldacena [20]—for a readable review see New Scientist [22]—(authors who, by the way, are all building upon two fundamental, only apparently disconnected, papers written by Einstein in 1935, the so-called E.R. and E.P.R. papers [10] and [9]). On the other hand, already in a model of spacetime as simple as a lattice (see, for example, [17]) we see how the actual topology of spacetime can deeply affect the formulation of the fundamental structures of physical theories (in that case, the definition of the fundamental commutator of QM is deformed by the lattice structure of the underlying spacetime).

The opinions that are presented in this chapter can be considered as opinions stemming from the family of pure mathematicians (plus a theoretical physicist) and it is expected that they will not easily drag the attention of a large number of physicists: it is in our beliefs though that a spacetime as an integrated mathematical entity, a spacetime studied as a topological space, would play a significant role to the search for a theory of quantum gravity. In a few words, the methods of general topology should be taken more seriously from those working in QFT as well as those in general relativity, at least.

6.1.2 On Name-Giving and Notation

In the geometry of spacetime we introduce three relations: the chronological order $\ll$, the causal order $\prec$ and the relation horismos $\rightarrow$. These relations can be extended to any *event space* $(M, \ll, \prec, \rightarrow)$ having no metric (see [18] and [31]).

In particular, we say that x chronologically precedes an event y -written $x \ll y$- if y lies inside the future null cone of x. x *causally precedes* y -written $x \prec y$- if y lies inside or on the future null cone of x. Last, but not least, x is at *horismos* with y -written $x \rightarrow y$- if y lies on the future null cone of x. The order $\ll$ is irreflexive, the order $\prec$ is reflexive and the relation $\rightarrow$ is reflexive, too.

In addition, the chronological future of an event x is denoted by $I^+(x) = \{y \in M : x \ll y\}$ while its causal future by $J^+(x) = \{y \in M : x \prec y\}$ (with a minus instead of a plus sign, dually, for the pasts in each case, respectively). The future null cone of x is denoted by $\mathcal{N}^+(x) \equiv \partial J^+(x) = \{y \in M : x \rightarrow y\}$ and, dually, we put a minus for the null past of x. The chronological past and future of an event x determine its *time cone*, its causal past and future its *causal cone* and its null past and future its *light cone*.

When physicists refer to the *null cone* of an event x they actually mean the light cone. Zeeman, as a working topologist, preferred to substitute the term null cone by three terms, for working with the interior, closure, boundary and exterior of it (see paragraph 3 of [38]).

We should now mention a few problems in name-giving that arise from when one corresponds order-theoretic and topological notions from the classical theory of ordered sets and lattices to a spacetime manifold. Following the construction of the *interval topology* (see [12]), it seems natural to say that a subset $A \subset X$ is a *past set* if $A = I^-(A)$ and a *future set* if $A = I^+(A)$. One then would expect that the *future topology* $\mathcal{T}^+$ is generated by the subbase $\mathcal{S}^+ = \{X \setminus I^-(x) : x \in X\}$ and the *past topology* $\mathcal{T}^-$ by $\mathcal{S}^- = \{X \setminus I^+(x) : x \in X\}$. Then, the *interval topology* $\mathcal{T}_{in}$ on M would consist of basic sets which are finite intersections of subbasic-open sets of the past and the future topologies.

First of all, the names "future topology" and "past topology" are due to the lack of inspiration for other names for such topological analogues in a spacetime, but here we should have in mind that when one considers the chronological relation and identifies $\downarrow \{x\}$ with $I^-(x)$, then obviously $x \notin I^-(x)$. On the contrary, things follow the pattern of the construction in [12] when one considers the causal order $\prec$. Furthermore, $M \setminus I^-(x)$ will not be a future set with $\ll$, according to the definition that a future set satisfies $X =\uparrow X$. All these are not real problems at all, when it comes to our target to describe particular topologies which incorporate the causal structure of a spacetime (see the section Topologies Different than the Manifold

Topology, below, and the corresponding references in it); the problem is sort of corresponding more appropriate names to these topologies, as well as developing a more systematic and simplified notation. We believe that this is not a difficult task to achieve in the near future.

One more point, regarding the appropriateness of a name, the Minkowski space in particular (and spacetimes in general) is not up-complete, and a topology $\mathcal{T}_{in}$ for a spacetime belongs actually to a coarser topology than the interval topology of [12]. So we will treat the interval topology of [12] as a special case referring to up-complete sets, and our $\mathcal{T}_{in}$ spacetime topologies belonging to a more general case where up-completeness is not a necessary condition. It is worth mentioning though that for the particular case of 2-dimensional Minkowski spacetime, $\mathcal{T}_{in}$ under $\prec$ is the interval topology that one defines using [12].

Finally, we would also like to highlight the distinction between the interval topology $\mathcal{T}_{in}$ from the "interval topology" of A.P. Alexandrov (see [31], page 29 and the succeeding section here). $\mathcal{T}_{in}$ is of a more general nature, and it can be defined via any relation, while the Alexandrov topology is restricted to the chronological order. These two topologies are different in nature, as well as in definition, so we propose the use of "interval topology" for $\mathcal{T}_{in}$ exclusively, and not for the Alexandrov topology.

6.2 Topologies Coarser Than or Equal to the Manifold Topology

In the literature, starting from the first modern singularity theorem by Penrose (see [30]) till recent accounts on singularities such as [34], there is no explicit mentioning of the topology of a spacetime M, while Riemann metric and Riemann basic-open balls can be used whenever there is a need, for example, for the proof of the limit curve theorem (LCT) and the convergence of causal curves (for a detailed exposition see [8] and [21]). In parallel to the manifold topology $\mathcal{M}$, one can consider the Alexandrov topology $\mathcal{A}$ which has basic-open sets known as "diamonds" and are simply the intersections of future and past time cones, of two distinct events respectively. This topology incorporates the causal structure of a spacetime, but equals the manifold topology only in the following case (see [31]) and, in other cases, it is coarser than $\mathcal{M}$.

Theorem 6.2.1 *On a spacetime M, the following are equivalent:*

1. *M is strongly causal.*
2. *$\mathcal{A}$ agrees with $\mathcal{M}$.*
3. *$\mathcal{A}$ is Hausdorff.*

So, the main contribution of the topology $\mathcal{A}$ is a characterisation of strong causality, as soon as $\mathcal{A}$ is Hausdorff. Adding the fact that it incorporates the causality (in particular the chronology) of a spacetime by the construction of open diamonds,

$\mathcal{A}$ looks like a great candidate for a spacetime topology when it is Hausdorff but, following Zeeman's arguments, its group of homeomorphisms is vast and of no physical meaning, both in the Minkowski spacetime and in curved spacetimes.

The existence of a Lorentz metric in a spacetime is enough to make us conclude that neither the manifold topology $\mathcal{M}$ nor the Alexandrov topology $\mathcal{A}$ "in its best", that is when Theorem 6.2.1 is satisfied, can fully describe a spacetime topologically. The manifold topology is a natural topology for a manifold, but not such a natural one for a spacetime manifold!

6.3 The Class $\mathfrak{Z}$ of Zeeman-Göbel Topologies

The class $\mathfrak{Z}$ of Zeeman topologies on a spacetime manifold M consists of topologies which have the property that they induce the 1-dimensional manifold topology on every time axis and the 3-dimensional manifold topology on every space axes. This class was first introduced in [38], in the special case of Minkowski spacetime, and it was generalised in [13] for any curved spacetime. In particular, paper [38] is the natural continuation of [37], where Zeeman proved that causality in Minkowski spacetime implies the Lorentz group. He then showed that the group of all homeomorphisms of the finest topology in $\mathfrak{Z}$, which is coarser than the discrete topology, is generated by the inhomogeneous Lorentz group and dilatations. In addition, unlike the topology of $\mathbb{R}^4$, this fine topology F is not locally homogeneous and the light cone through any point can be deduced by F. There is also a quite interesting lemma; the topology on a light ray induced from F is discrete. Discreteness of light, according to Göbel, describes well its physical behaviour: there is no geometric information along a light ray. Here one should not confuse topological discreteness (every set is open) with discreteness in the sense of (finite or infinite) countability. Apart from the group of homeomorphisms of M under F and its physical interpretation, the topological boundary of the null cone has the maximum number of open sets: there is definitely a connection here with the maximum speed, that of light.

Zeeman mentioned three other alternative topologies in $\mathfrak{Z}$ different than F, that we will consider in Sect. 6.4, as well as their analogues for curved spacetimes.

Göbel found that the analogue of F in a curved spacetime has the property that the group of all homeomorphisms under this topology is isomorphic to the group of all homothetic transformations. In a few words, under the relativistic analogue of F, a homeomorphism is an isometry.

A problem that was noticed first by Zeeman himself is that F is technically difficult, as it does not admit a countable base and so it is not the best tool for a working physicist. This was one of the arguments of the authors of [15] and [19] as well, but we object that this is not an attractive reason for avoiding a topology which is much more natural in a spacetime from the manifold topology. Natural in the sense that it incorporates the differential, causal and conformal structures and the group of homeomorphisms of the spacetime is not vast and it has physical meaning.

So, the argument that F has "too many open sets" and does not admit a countable base should be reconsidered. Since we are dealing with both the Lorentz metric and the Riemann metric in a spacetime manifold, a natural topology which will describe the properties of the spacetime should be compatible with every possible structure which is defined on the spacetime. F is such a topology.

For some reason the supporters of the manifold topology, like Heathcote, believed that all the topologies in $\mathfrak{Z}$ are strictly finer than $\mathcal{M}$, but actually this is not true. Three alternative topologies that Zeeman introduces in [38] are linked in their construction to topologies that we mention in the next paragraph, each of which belongs to the class $\mathfrak{Z}$ but is incomparable to $\mathcal{M}$ (see [21, 27] and [29]). Göbel [13, page 297, (C)] actually states that there are other topologies in $\mathfrak{Z}$, but without a clear reference that there are topologies that are not necessarily finer or coarser or equal to the manifold topology. This is important, since the criticism against the class $\mathfrak{Z}$ bases many of its arguments against the term *finer* topology. Let us now look at a sample of topologies in $\mathfrak{Z}$, which are not finer than the manifold topology $\mathcal{M}$.

6.4 Topologies Different Than the Manifold Topology

In [28] we remark that the Path topology $\mathcal{P}$ of Hawking–King–McCarthy (see [15]) is the general relativistic analogue of the topology introduced in Example 1 of [38] (page 169). Low showed in [19] that under this topology $\mathcal{P}$ (that we name Z^T for consistency of notation) the limit curve theorem (LCT) fails to hold. In [28] we introduced three (among others) more topologies that the LCT fails to hold, all incorporating the differential, causal and conformal structure of the spacetime manifold. In particular, we stated the following theorem.

Theorem 6.4.1 *There are three distinct topologies in a spacetime manifold which admit a countable basis, they incorporate the causal and conformal structures and the LCT fails with each one of them, respectively. These are the interval topologies* $T_{in}^{\rightarrow}$, $T_{in}^{\leq}$ *and* $T_{in}^{\ll=}$, *which are all in the class* $\mathfrak{Z}$.

All these topologies are not finer (neither equal nor coarser) than the manifold topology and singularity theorems, under each one of them, respectively, cannot be formed in the way that are described via the manifold topology. These three topologies, together with the manifold topology, give the intersection topologies Z, Z^T, Z^S, which are finer than the manifold topology, where Z is coarser than the fine topology F and Z^T (the Path topology of [15]) and Z^S are incomparable to F.

Low, in [19], supports in his conclusion that LCT failing in the $\mathcal{P}$ (which also fails in the extra five topologies that we suggest in Theorem 6.4.1) makes the manifold topology remaining both technically easier to work with and fruitful. We have some objections. All the six topologies of [28] and in particular those in Theorem 6.4.1 are technically easy to work with (they all have a countable base of open sets) and they are fruitful, as they belong to $\mathfrak{Z}$ and are all behaving like order topologies, in the sense that they satisfy the orderability problem (or weaker versions of it, referring to non-linear orders; see [14, 23, 25] and [24]). Each one of them is induced either from the causal or chronological orders or from (the irreflexive) horismos, with the exception of Z^S which is induced by a particular spacelike non-causal order that we describe in [29]. More specifically, Z, Z^T and Z^S have open sets bounded by Riemann open balls centred at an event x, intersected with the timecone union spacecone of x in the case of Z, the timecone of x in the case of Z^T and the spacecone of x in the case of Z^S, respectively. The rest three topologies have unbounded open sets which are timecone union spacecone in the case of $T_{in}^{\rightarrow}$, timecone in the case of $T_{in}^{\leq}$ and spacecone in the case of $T_{in}^{\ll^=}$, respectively, at an event x. For a more detailed treatment we refer to [28].

On the other hand, the manifold topology misses the Lorentz metric and so the causal structure of the spacetime as well, so we conclude the following.

Corollary 6.4.1 *The manifold topology $\mathcal{M}$, on a spacetime M, is based on the Riemann metric and is sufficient for describing spacetime singularities, but does not incorporate the Lorentz metric, while each of the topologies in Theorem 6.4.1 fail to describe singularities that appear under $\mathcal{M}$, but incorporate the Lorentz metric.*

Corollary 6.4.2 The Fine Topology F is the Best Possible Candidate for a Spacetime *M, as it is strictly finer than $\mathcal{M}$, strictly coarser than the discrete topology and, simultaneously, finer than the topologies introduced in Theorem 6.4.1. In addition, the group of homeomorphisms of M under F is isomorphic to the Lorentz group and dilatations, in the case of special relativity, and to the group of homothetic symmetries in the case of general relativity, while under the manifold topology the group of homeomorphisms of M is vast and of no physical significance. Last, but not least, the LCT holds under F, while it might fail in coarser topologies to F.*

The discussion about F would be incomplete, if we did not mention the comment of Göbel about F in [13], pages 290–291: unphysical world lines, like "bad trips" are avoided if one interprets continuity of worldlines with respect to F. Under F basic assumptions for a kinetic theory in general relativity are satisfied and one can incorporate electromagnetic fields into such a result, if one allows F to depend on a gravitational field as well as on the Maxwell field (and derive corresponding results on orbits of freely falling test particles for charged particles).

6.5 In the Beginning Was the Metric…or the Topology?

This is a more important question as it seems to be. Speaking about spacetime manifolds as mathematical objects, it is vital that a natural topology will incorporate all the mathematical structures appearing in the manifold, including the Lorentz metric as well as the Riemann metric. In this sense (a "Platonic mathematical" sense in the view of Penrose, which is projected to the physical world [32]) the manifold topology is not a natural topology in a spacetime manifold, even if it is defined via the Riemann metric. The metric tensor field, the affine connection and the conformal structure, the differentiable manifold with its topology, are all important constituents of the spacetime manifold, but what about the Lorentz metric and the structure of the null cone? Having mentioned this, we believe that "in the beginning was the topology", in a spacetime manifold. A topology like F, where the group of homeomorphisms of M under F has a physical meaning and which incorporates all the metric structures in the manifold.

6.6 Ambient Cosmology: A Failure Due to a Topological Misconception

In [5] we described the motivation for a 5-dimensional "ambient space", where our 4-dimensional spacetime is its conformally related ambient boundary at infinity, by linking it to the singularity problem in general relativity. In cosmology the infinities that are inherent in the spacetime metric according to the singularity theorems indicate the necessity of a conformal geometry of metrics to absorb them, not a breakdown of general relativity. The construction of this model in ambient cosmology can be found in [2–4] and [7], where the authors started from the construction of the metric, leaving the topological problem at the end. As we observed in [26], it is the topology succeeding (and, unfortunately, not preceding or at least being constructed simultaneously with) the metric that showed a failure in the construction and in results concerning the convergence of causal curves; it is the Path topology $\mathcal{P} = Z^T$ or Z^S or Z where the LCT fails and not in F. Furthermore, why should one bother to add an extra dimension while a 4-dimensional spacetime under a topology like $\mathcal{P}$ has already the properties of the ambient boundary, and while the structure of the ambient boundary is totally unknown to us (we lack knowledge even for basic results on causality: see, for example, [27] for an important correction on [5]). Here we should also mention that a finer topology than the manifold one will contain manifold-open sets; this does not guarantee that the LCT holds. For example, $\mathcal{P}$ is finer than $\mathcal{M}$ and, simultaneously, LCT fails under $\mathcal{P}$.

6.7 Towards an Evolving Topology and a Quantum Theory of Gravity

If the main problem for a working physicist is that F is not an easy topology to work with, due to the lack of a countable basis of open sets, or if topologies like $\mathcal{M}$ and those topologies mentioned in paragraph 4 are missing something important from the spacetime structure, then we believe that there is something deeper behind all this and this certainly is of a topological nature. We have already expressed in [27] an idea of an evolving topology with respect to the class $\mathfrak{Z}$, so that different topologies of this class are assigned to each stage of the evolution as well as where the spacetime itself is subjected to singularities. It could be, for example, that the interval topology from horismos $\rightarrow$ (see [27]) could give a sufficient description to the transition from/to the Planck time and objects like black holes, while other topologies (where the LCT theorem holds, for example) could explain the phase transition from locality to nonlocality. Topologies like Z^T are linked to a discrete space while Z^S to a discrete time, while Z to a discrete light (these are actually remarks of Zeeman in [38], for their special relativistic analogues). By evolution we do not necessarily mean (and this is not our desire at all) to consider kinematically that the spacetimes of our interest are foliated manifolds where leaves of foliation have open sets which vary over time. The question is different: how does a spacetime manifold appear from a functional space? An answer to such a question which refers to the transition from nonlocality to locality seems to need a richer topological background, a background that the class $\mathfrak{Z}$ could possibly provide. Possible tools can be also derived from articles like [1, 11].

6.8 The Need to Take Methods of General Topology More Seriously

We believe that the concerns against a "finer" topology, as expressed in [16], are reasonable. Reasonable are similar concerns expressed in [19], when we restrict ourselves to the validity of general relativity. The problem is that even though the manifold topology $\mathcal{M}$ has somehow worked nicely in the last century or so, it is problematic in describing fully properties of a spacetime in a sufficient way; it lacks important information, as we have seen in the previous paragraphs. F is a finer topology which resolves, at least in a mathematical way, all such issues, and—at the moment—there is no other candidate topology to compete with.

Criticism (in oral communication with physicists) against F, and against topologies like those mentioned in paragraph 4, highlights that there is a value of considering these alternative topologies in $\mathfrak{Z}$ since they may, for example, lead to a new physical theory, or they may allow one to extract new, physically interesting, predictions from the old theory. According to those who criticize alternatives to the manifold topology, there is a point of diminishing returns; that, eventually, further treatment of these topologies, in the abstract, can no longer be justified. At some point, there should be a result of a genuine physical interest; no such a result is in sight and, therefore, that we have reached that point.

The problem of such a criticism is that the main points of [38] and [13] have not been understood, and this is quite disappointing. There is a prejudice against general topology, only the reference to it is enough to discourage working mathematical physicists and theoretical physicists to read carefully a related article. The labyrinth that we seem to be when talking about string theory and quantum theory of gravity, for example, is not only related to the need for an extra physical input, but for an extra mathematical input as well. The authors wish that this chapter contributes to the reopening of a discussion in this serious and fascinating subject.

Acknowledgements The co-author K.P. wishes to thank Robert Low for his remarks on the Interval Topology in [28], some of which we incorporate in the introductory section here, as well as Spiros Cotsakis for introducing him [38] and related literature. He also wishes to thank Nikolaos Kalogeropoulos for discussions on quantum theory of gravity and B.K. Papadopoulos for having taught him lattice topologies. Last, the authors would like to thank Ljubisa Kocinac and Hemen Dutta for their important remarks towards the improvement of the text.

References

1. R.P. Agarwal, C. Cuevas, H. Soto, M. El-Gebeily, Asymptotic periodicity for some evolution equations in Banach spaces. Nonlinear Anal. Theory Methods Appl. **74**(5), 1769–1798 (2011)
2. I. Antoniadis, S. Cotsakis, The large-scale structure of the ambient boundary, to appear in MG14 Proceedings; arXiv:1512.0916
3. I. Antoniadis, S. Cotsakis, Ambient cosmology and spacetime singularities. Eur. Phys. J. C. **75**(35), 1–12 (2015)
4. I. Antoniadis, S. Cotsakis, Topology of the ambient boundary and the convergence of causal curves. Mod. Phys. Lett. A **30**(30), 1550161 (2015)
5. I. Antoniadis, S. Cotsakis, K. Papadopoulos, The causal order on the ambient boundary. Mod. Phys. Lett. A **31**(20), 1650122 (2016)
6. P.S. Aspinwall, B.R.Greene, D.R. Morrison, Multiple mirror manifolds and topology change in string theory. Phys. Lett. B **303**, 249–259 (1993)B.R. Greene, K. Schalm, G. Shiu, Dynamical topology change in M theory. J. Math. Phys. **42**, 3171–3187 (2001)
7. S. Cotsakis, Cosmological Singularities, Springer LNP Proceedings of the First Aegean Summer School of Cosmology, Samos, September 21–29 (2001)
8. S. Cotsakis, *Talking about Singularities, The Eleventh Marcel Grossmann Meeting: On Recent Developments in Theoretical and Experimental General Relativity*. Gravitation and Relativistic Field Theories (In 3 Volumes), 758–777 (2008)
9. A. Einstein, B. Podolsky, N. Rosen, Can quantum mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777 (1935)
10. A. Einstein, N. Rosen, The particle problem in the general theory of relativity. Phys. Rev. **48**, 73 (1935)
11. S. Ferry, K. Mischaikow, V. Nanda, Reconstructing functions from random variables. J. Comput. Dynam. **1**(2), 233–248 (2014)
12. G. Gierz, K.H. Hofmann, K. Keimel, J.D. Lawson, M.W. Mislove, D.S. Scott, *A Compendium of Continuous Lattices* (Springer, Berlin, 1980)
13. Gobel, Zeeman topologies on space-times of general relativity theory. Comm. Math. Phys. **46**, 289–307 (1976)
14. C. Good, K. Papadopoulos, A topological characterization of ordinals: van Dalen and Wattel revisited. Topology Appl. **159**, 1565–1572 (2012)
15. S.W. Hawking, A.R. King, P.J. McCarthy, A new topology for curved space time which incorporates the causal, differential, and conformal structures. J. Math. Phy. **17**(2), 174–181 (1976)
16. A. Heathcote, Zeeman Göbel topologies. Br. J. Phil. Sci. **39**, 247–261 (1988)
17. P. Jizba, H. Kleinert, F. Scardigli, Uncertainty relation on world crystal and its applications to micro black holes. Phys. Rev. D **81**, 084030 (2010), [arXiv:0912.2253 [hep-th]]
18. E.H. Kronheimer, R. Penrose, On the structure of causal spaces. Proc. Camb. Phil. Soc. **63**, 481 (1967)

19. R.J. Low, Spaces of paths and the path topology. J. Math. Phys. **57**, 092503 (2016)
20. J. Maldacena, L. Susskind, Cool horizons for entangled black holes. Fortsch. Phys. **61**, 781 (2013), [arXiv:1306.0533 [hep-th]].L. Susskind, Dear Qubitzers, GR=QM, arXiv:1708.03040 [hep-th].
21. E. Minguzzi, Limit curve theorems in Lorentzian geometry. J. Math. Phys. **49**, 092501 (2008)
22. New Scientist Cover Story, Entangled Universe, 7 November 2015.
23. K. Papadopoulos, On the orderability problem and the interval topology, Chapter in the Volume "Topics in Mathematical Analysis and Applications", in *Optimization and Its Applications Springer Series*, ed. by T. Rassias, L. Toth (Springer, Berlin, 2014)
24. K. Papadopoulos, On properties of nests: some answers and questions. Questions Answers Gen. Topology **33**(2), 71–91 (2015)
25. K. Papadopoulos, Nests, and their role in the orderability problem. *Mathematical Analysis, Approximation Theory and Their Applications* ed. by Th. M. Rassias, V. Gupta (Springer, Berlin, 2016), pp. 517–533
26. K. Papadopoulos, On the possibility of singularities on the ambient boundary. Int. J. Geom. Meth. Mod. Phys. **14**(10), 1750184 (2017)
27. K. Papadopoulos, S. Acharjee, B.K. Papadopoulos, The order on the light cone and its induced topology. Int. J. Geom. Meth. Mod. Phys. **15**(05), 1850069 (2018)
28. K. Papadopoulos, B.K. Papadopoulos, Spacetime Singularities vs. Topologies of Zeeman-Göbel Class, arXiv: 1712.03270
29. K. Papadopoulos, B.K. Papadopoulos, On two topologies that were suggested by Zeeman. Math. Methods Appl. Sci. **41**, 7742–7747 (2018)
30. R. Penrose, Gravitational collapse and space-time singularities. Phys. Rev. Lett. **14**, 57 (1965)
31. R. Penrose, Techniques of Differential Topology in Relativity, CBMS-NSF Regional Conference Series in Applied Mathematics (1972)
32. R. Penrose, *The Road to Reality: A Complete Guide to the Laws of the Universe* (Vintage Books, 2007)
33. F. Scardigli, Black hole entropy: a space-time foam approach. Class. Quant. Grav. **14**, 1781 (1997) [gr-qc/9706030]
34. J.M.M. Senovilla, D. Garfinkle, The 1965 Penrose singularity theorem. Class. Quant. Grav. **32**, 12 (2015)
35. M. Van Raamsdonk, Building up spacetime with quantum entanglement. Gen. Rel. Grav. **42**, 2323 (2010) [Int. J. Mod. Phys. D **19** (2010) 2429] [arXiv:1005.3035 [hep-th]]
36. J.A. Wheeler, Geons. Phys. Rev. **97**, 511 (1955)
37. E.C. Zeeman, Causality implies the Lorentz group. J. Math. Phys. **5**, 490–493 (1964)
38. E.C. Zeeman, The topology of Minkowski space. Topology **6**, 161–170 (1967)

Chapter 7
Analysis of Generalized BBM Equations: Symmetry Groups and Conservation Laws

M. S. Bruzón, T. M. Garrido, and R. de la Rosa

Abstract In this work we study a generalized BBM equation from the point of view of the theory of symmetry reductions in partial differential equations. We obtain the Lie symmetries, then we use the transformation groups to reduce the equations into ordinary differential equations. Physical interpretation of these reductions and some exact solutions are also provided.

Local conservation laws are continuity equations that provide conserved quantities of physical importance for all solutions of a particular equation. In addition, the existence of an infinite hierarchy of local conservation laws of a partial differential equation is a strong indicator of its integrability. For any particular partial differential equation, a complete classification of all local low-order conservation laws can be derived by using the multiplier method.

Keywords Partial differential equations · Lie symmetries · Exact solutions · Conservation laws

7.1 Introduction

Benjamin et al. [6] proposed the regularized long wave (RLW) equation, or Benjamin–Bona–Mahony equation (BBM),

$$u_t + u_x + uu_x - u_{xxt} = 0, \tag{7.1}$$

as an alternative model to the Korteweg–de Vries (KdV) equation for the long wave motion in nonlinear dispersive systems. Equation (7.1) can be used to describe the unidirectional propagation of small-amplitude long waves on the surface of water in a channel. These authors argued that both equations are valid at the same level of

M. S. Bruzón (✉) · T. M. Garrido · R. de la Rosa
Departamento de Matemáticas, Universidad de Cádiz, Cádiz, Spain
e-mail: m.bruzon@uca.es; tamara.garrido@uca.es; rafael.delarosa@uca.es

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_7

approximation, but BBM has some advantages over the KdV from a computational and mathematical viewpoint. Several forms of BBM equation (7.1) have been widely studied, some of them are collected here.

The damped externally excited Benjamin–Bona–Mahony equation

$$u_t + u_x + 2buu_x - cu_{xx} - du - au_{xxt} = \eta \cos k(x + \lambda t), \tag{7.2}$$

where c and d are nonnegative constants that are proportional to the strength of the damping effect, was considered by Eloe and Usman [23]. Equation (7.2) was introduced to model long waves in nonlinear dispersive systems, and it could be checked that BBM equation is a special case of it.

Gandarias and Khalique [24] studied a generalization of (7.2) of the form

$$u_t + u_x + 2buu_x - cu_{xx} - du - au_{xxt} = f(x, t), \tag{7.3}$$

where f is an arbitrary function of the variables x and t. They proved that (7.3) is nonlinearly self-adjoint and determined some exact solutions by using Lie symmetries.

There are also equations that are a combination of the KdV and the BBM equation. For example, the third-order KdV–BBM equation

$$u_t + u_x + \frac{3}{2}uu_x + \nu u_{xxx} - \left(\frac{1}{6} - \nu\right) u_{txx} = 0, \tag{7.4}$$

models long-crested water waves which travel mostly unidirectionally. Travelling wave solutions and conservation laws of Eq. (7.4) were obtained in [34] through Lie symmetry analysis along with the Jacobi elliptic function expansion and Kudryashov methods.

In [32], the following one-dimensional BBM equation with time-dependent coefficients was studied

$$u_t + f(t)u_x + g(t)uu_x + h(t)u_{xxt} = 0, \tag{7.5}$$

where $f(t)$, $g(t)$, and $h(t)$ are nonzero arbitrary functions. These kind of models are more general since in realistic physical systems, no media is homogeneous due to long distance of propagation and the existence of some nonuniformity due to many factors. Molati et al. obtained that the functional forms of the functions $f(t)$, $g(t)$, and $h(t)$ for which Eq. (7.5) admits point symmetries were of a linear, power, exponential, and logarithmic type. Furthermore, by using these symmetries, they derived some exact travelling wave solutions.

On the other hand, we found the one-dimensional modified B-BBM equation with power law nonlinearity and time-dependent coefficients given by

$$u_t + f(t)u^q u_x + g(t)u_{xx} + h(t)u_{xxt} = 0, \tag{7.6}$$

with $f(t)$, $g(t)$, and $h(t)$ nonzero arbitrary functions. Equation (7.6) was derived to encompass the dissipation of practical problems. In [31], a classification of point symmetries depending on the functional forms of the time-dependent variables was obtained as well as some symmetry reductions were taking into account.

Furthermore, in order to understand the role of nonlinear dispersion in the formation of patterns in an undular bore, Yalong [38] introduced and studied a family of BBM-like equations with nonlinear dispersion, $\mathrm{B}(m, n)$ equations

$$u_t + (u^m)_x - (u^n)_{xxt} = 0, \qquad m, n > 1.$$

In [38], exact solitary-wave solutions with compact support and exact special solutions with solitary patterns of the equations were derived.

In [36] the authors introduced the family of BBM equation with strong nonlinear dispersive $\mathrm{B}(m, n)$ equation:

$$u_t + u_x + a\left(u^m\right)_x + \left(u^n\right)_{xxt} = 0. \tag{7.7}$$

By using an algebraic method, solitary pattern solutions of equation (7.7) were obtained. The case $n = 1$ and $m = 2$ corresponds to the BBM equation (7.1).

Clarkson [18] showed that the similarity reduction of Eq. (7.7) for $m = 3, n = 1$, and $a = \frac{1}{3}$ obtained by using the classical Lie group method reduces the partial differential equation (PDE) to an ordinary differential equation (ODE) of Painlevé type, whereas the PDE does not possess the Painlevé property for PDEs defined by Weiss et al. [37]. The author proved that the only non-constant similarity reductions of this equation obtainable by using both the classical Lie method and the direct method [19] are travelling wave type reductions.

In [10] the authors made a classification of symmetry reductions of the family of BBM equations,

$$u_t + bu_x + a\left(u^m\right)_x + \left(u^n\right)_{xxt} = 0, \tag{7.8}$$

where $a, b \neq 0, m \neq 0, n \neq 0$ are constants, m and n are not simultaneously equal to 1, depending on the values of the constants. They constructed all the invariant solutions with regard to one-dimensional subalgebras. Besides travelling wave reductions, they found new similarity reductions for this family of equations. They constructed all nonequivalent ODEs to which (7.8) could be reduced. They obtained, for special values of the parameters of this equation, many exact solutions expressed by various single and combined nondegenerative Jacobi elliptic function solutions and their degenerative solutions.

In [13], the nonlocal symmetries of BBM equations (7.8) were studied. A PDE written as a conservation law can be transformed into an equivalent system by introducing a suitable potential. The nonlocal symmetry group generators of the original PDE were obtained through their equivalent system. The authors proved that the nonclassical method applied to this system leads to new symmetries, which are not solutions arising from potential symmetries of the BBM (7.8) equations.

In [14] the authors considered Eq. (7.8) for $n = 1$ and $m = 2$. By using the $\frac{G'}{G}$-expansion method they obtained three types of travelling wave solutions and also obtained for special values of the parameters of this equation, many exact solutions expressed by various single and combined nondegenerative Jacobi elliptic function solutions and their degenerative solutions (soliton, kink, and compactons). Moreover, by using the concept of self-adjoint equation, introduced by Ibragimov in [28–30], the authors found the general class of self-adjoint equations of (7.10). By using Ibragimov's theorem on conservation laws, they derived conservation laws for this equation.

In [11], the generalized Benjamin–Bona–Mahony–Burgers (BBMB) equation was considered

$$u_t - u_{xxt} - \alpha u_{xx} + \beta u_x + (g(u))_x = 0, \tag{7.9}$$

where $u(x, t)$ represents the fluid velocity in the horizontal direction x, α is a positive constant, $\beta \in \mathbb{R}$ and $g(u)$ is a C^2-smooth nonlinear function which was given in [1]. Classical symmetries of this equation were considered. The functional forms, for which the BBMB equation (7.9) can be reduced to ODEs by using point symmetries, were obtained. Moreover, symmetry reductions and exact solutions were given. In particular, a set of new solitons, kinks, anti-kinks, compactons, and Wadati solitons were derived.

In [12] the authors considered the combined dispersive-dissipative entity

$$u_t + bu_x + a\left(u^m\right)_x + \left(u^n\right)_{txx} + c\left(u^k\right)_{xx} = 0, \tag{7.10}$$

where $a \neq 0$, b, $c \neq 0$, are arbitrary constants, k, m, and n positive constants. They obtained the complete Lie group classification of this equation. Next, from the optimal system, the similarity variables and the similarity solutions were determined and used to reduce Eq. (7.10) into an ODE. Finally, they derived travelling wave solutions of Eq. (7.10).

In [15], the authors showed that this equation is not self-adjoint and we determined the subclasses of Eq. (7.10) which are nonlinearly self-adjoint. Furthermore, by using the multipliers method of Anco and Bluman, we obtained some non-trivial conservation laws. In the case that $n = 1$, $k = 2$, and $m = 3$, taking into account the modified simplest equation method, three types of travelling wave solutions of this equation were obtained.

In [35] the following damped, forced generalized BBM equation was considered

$$u_t - u_{txx} - \partial_x(a(x)\partial_x u) + (g(u))_x = f \tag{7.11}$$

where a, f, and g are arbitrary functions.

In the present paper, we focus on two different cases of Eq. (7.11)

$$u_t - u_{txx} - \partial_x(a(x)\partial_x u) + (g(u))_x = f(t), \tag{7.12}$$

$$u_t - u_{txx} - \partial_x(a(x)\partial_x u) + (g(u))_x = f(x). \tag{7.13}$$

For both equations, Lie point symmetry method is applied. Symmetry analysis is probably the most powerful tool to study linear and nonlinear ODEs and PDEs. Specifically, Lie method allows us to determine the symmetries of a given ODE or PDE through the one-parameter group of transformations that leaves invariant the equation and transforms the solution space into itself. Thus, by using Lie symmetries it is possible to reduce the order of an ODE, reduce the number of dependent variables of a PDE, classify the equation, or get new solutions, among others.

Many books have been written about Lie's method and its theory. A detail description can be found in [7–9, 27, 33], and nowadays it is applied to several differential equations in several fields, such as [16, 17, 20–22, 25, 26].

A conservation law is a space–time divergence expression whose importance lies in its physical meaning. It shows us a property of the physical system model that does not change as the system evolves over time. Furthermore, these laws can be used to assess the accuracy and stability of numerical methods for the solutions of PDEs and the existence of a large number of conservation laws is linked with the integrability of a differential equation. Information about conservation laws and the multiplier method to obtain them can be found in [2, 4, 5].

The aim of this work is to analyze both damped forced generalized BBM equations (7.12)–(7.13) from the point of view of Lie symmetries. Furthermore, once the Lie algebra is obtained, we will reduce these equations to ODEs and obtain travelling wave solutions. We will also construct conservation laws by using the multiplier method. Finally, we study the nonlocal symmetries of the equations and Lie symmetries of the integrated equation.

7.2 Lie Point Symmetries

Before determining Lie point symmetries of Eqs. (7.12) and (7.13), we have determined the equivalence group of these equations. It is remarkable that equivalence transformations facilitate a complete classification of point symmetries. We notice that Eqs. (7.12) and (7.13) are conserved under the equivalence transformation

$$\tilde{t} \longrightarrow t + t_0, \quad \tilde{u} \longrightarrow u + u_0, \quad t_0, u_0 \text{ constants.}$$

Thus, we can simplify the results obtained on point symmetries.

To apply Lie method to Eqs. (7.12) and (7.13) with $g''(u) \neq 0$, we consider the one-parameter Lie group of infinitesimal transformations in (x, t, u) given by

$$\begin{aligned} x^* &= x + \epsilon \xi(x, t, u) + O(\epsilon^2), \\ t^* &= t + \epsilon \tau(x, t, u) + O(\epsilon^2), \\ u^* &= u + \epsilon \eta(x, t, u) + O(\epsilon^2), \end{aligned}$$

where ϵ is the group parameter. We require that this transformation leaves invariant the set of solutions of (7.12) and (7.13), respectively. This yields an overdetermined, linear system of equations for the infinitesimals $\xi(x, t, u)$, $\tau(x, t, u)$, and $\eta(x, t, u)$. The associated Lie algebra of infinitesimal symmetries is the set of vector fields of the form

$$\mathbf{v} = \xi(x, t, u)\partial_x + \tau(x, t, u)\partial_t + \eta(x, t, u)\partial_u. \tag{7.14}$$

7.2.1 Lie Point Symmetries of Eq. (7.12)

Invariance of Eq. (7.12) under a Lie group of point transformations with infinitesimal generator (7.14) leads to a set of 28 determining equations. By simplifying this system we obtain that $\tau = \tau(t)$, $\xi = \xi(x)$, $\eta = \eta(t, x, u)$, $a(x)$, $f(t)$, and $g(u)$ are related by the following conditions:

$$\begin{aligned} &\eta_{uu} = 0, \\ &2\eta_{ux} - \xi_{xx} = 0, \\ &\eta_{uxx} - 2\xi_x = 0, \\ &a\tau_t + a_x\xi + \eta_{ut} = 0, \\ &\eta_{xxt} + a\eta_{xx} - f\eta_u + (a_x - g_u)\,\eta_x - \eta_t + f_t\eta + 2f\xi_x + f\tau_t = 0, \\ &2\eta_{uxt} + 2a\eta_{ux} - \eta g_{uu} - a\xi_{xx} + (a_x - g_u)\,\xi_x + a_{xx}\xi + (a_x - g_u)\,\tau_t = 0. \end{aligned} \tag{7.15}$$

By solving system (7.15), we obtain the following result:

- *Case 1.1.* For $a(x)$ and $g(u)$ arbitrary functions, $f(t)$ constant, the infinitesimal generator is

$$\mathbf{v}_1 = \partial_t.$$

- *Case 1.2.* For $g(u)$ and $f(t)$ arbitrary functions, $a(x)$ constant, the infinitesimal generator is

$$\mathbf{v}_2 = \partial_x.$$

- *Case 1.3.* For $g(u)$ arbitrary function, $f(t)$ and $a(x)$ constants, the infinitesimal generators are

$$\mathbf{v}_1 = \partial_t, \quad \mathbf{v}_2 = \partial_x.$$

- *Case 1.4.* For $a(x) = a_2 \exp(a_1 x)$, $f(t) = f_2 t^{f_1}$, $g(u) = g_1 u^{\frac{f_1}{f_1+1}} + g_2$, a_1, $a_2 \neq 0$, $f_1 \neq -1$, f_2, $g_1 \neq 0$, g_2 constants, the infinitesimal generator is

$$\mathbf{v}_3 = -\partial_x + a_1 t \partial_t + a_1 (f_1 + 1) u \partial_u.$$

  Moreover, if $a_2 = 0$ or $f_2 = 0$ besides $\mathbf{v}_3$ we obtain $\mathbf{v}_2$ or $\mathbf{v}_1$, respectively.
- *Case 1.5.* For $a(x) = a_2 \exp(a_1 x)$, $f(t) = \frac{f_2}{t}$, $g(u) = g_2 \exp(g_1 u) + g_3$, a_1, $a_2 \neq 0$, f_2, $g_1 \neq 0$, $g_2 \neq 0$, g_3 constants, the infinitesimal generator is

$$\mathbf{v}_4 = g_1 \partial_x - a_1 g_1 t \partial_t + a_1 \partial_u.$$

  Moreover, if $a_2 = 0$ or $f_2 = 0$ besides $\mathbf{v}_4$ we obtain $\mathbf{v}_2$ or $\mathbf{v}_1$, respectively.
- *Case 1.6.* For $a(x) = a_2 \exp(a_1 x)$, $f(t) = f_2$, $g(u) = g_1 \ln(u) + g_2$, a_1, $a_2 \neq 0$, f_2, $g_1 \neq 0$, g_2 constants, we obtain besides $\mathbf{v}_1$ the infinitesimal generator

$$\mathbf{v}_5 = -\partial_x + a_1 t \partial_t + a_1 u \partial_u.$$

  Moreover, if $a_2 = f_2 = 0$ besides $\mathbf{v}_1$ and $\mathbf{v}_5$, we obtain $\mathbf{v}_2$.
- *Case 1.7.* For $a(x) = a_2 \exp(a_1 x)$, $f(t) = 0$, $g(u) = g_2 u^{g_1+1} + g_3$, a_1, a_2, $g_1 \neq 0$, $g_2 \neq 0$, g_3 constants, the infinitesimal generators are

$$\mathbf{v}_1, \quad \mathbf{v}_6 = -g_1 \partial_x + a_1 g_1 t \partial_t - a_1 u \partial_u.$$

  Moreover, if $a_2 = 0$ besides $\mathbf{v}_1$ and $\mathbf{v}_6$, we obtain $\mathbf{v}_2$.
- *Case 1.8.* For $a(x) = 0$, $f(t)$ arbitrary, $g(u) = g_2 \exp(g_1 u) + g_3$, $g_1 \neq 0$, g_2, g_3 constants, the infinitesimal generators are

$$\mathbf{v}_2,\ \mathbf{v}_7 = \exp\left(-g_1 \int f(t)\, dt\right) (\partial_t + f(t)\partial_u),\ \mathbf{v}_8 = F(t)\partial_t + (f(t)F(t) - 1)\partial_u,$$

  where

$$F(t) = g_1 \exp\left(-g_1 \int f(t)\, dt\right) \int \exp\left(g_1 \int f(t)\, dt\right) dt.$$

7.2.2 Lie Point Symmetries of Eq. (7.13)

Invariance of Eq. (7.13) under a Lie group of point transformations with infinitesimal generator (7.14) leads to a set of 28 determining equations. By simplifying this system we obtain that $\tau = \tau(t)$, $\xi = \xi(x)$, $\eta = \eta(t, x, u)$, $a(x)$, $f(x)$, and $g(u)$ are related by the following conditions:

$$\begin{aligned}
&\eta_{uu} = 0,\\
&2\eta_{ux} - \xi_{xx} = 0,\\
&\eta_{uxx} - 2\xi_x = 0,\\
&a\tau_t + a_x\xi + \eta_{ut} = 0,\\
&\eta_{xxt} + a\eta_{xx} - f\eta_u + (a_x - g_u)\,\eta_x - \eta_t + f_x\xi + 2f\xi_x + f\tau_t = 0,\\
&2\eta_{uxt} + 2a\eta_{ux} - \eta g_{uu} - a\xi_{xx} + (a_x - g_u)\,\xi_x + a_{xx}\xi + (a_x - g_u)\,\tau_t = 0.
\end{aligned} \tag{7.16}$$

By solving system (7.16), we obtain the following result:

- *Case 2.1.* For $g(u)$, $a(x)$, and $f(x)$ arbitrary functions the infinitesimal generator is $\mathbf{v}_1$.
- *Case 2.2.* For $g(u)$ arbitrary functions, $a(x)$ and $f(x)$ constants, the infinitesimal generators are $\mathbf{v}_1$ and $\mathbf{v}_2$.
- *Case 2.3.* For $a(x) = a_2\exp(a_1x)$, $f(x) = f_2\exp(f_1x)$, $g(u) = g_1u^{\frac{f_1}{f_1-a_1}} + g_2$, $a_1 \neq f_1, a_2, f_1 \neq 0, f_2, g_1 \neq 0, g_2$ constants, the infinitesimal generators are

$$\mathbf{v}_1, \quad \mathbf{v}_9 = -\partial_x + a_1t\partial_t + (a_1 - f_1)u\partial_u.$$

- *Case 2.4.* For $a(x) = a_2\exp(a_1x)$, $f(x) = 0$, $g(u) = g_2u^{g_1+1} + g_3$, $a_1, a_2, g_1 \neq 0, g_2 \neq 0, g_3$ constants, the infinitesimal generators are $\mathbf{v}_1$ and $\mathbf{v}_6$. Moreover, if $a_2 = 0$ besides $\mathbf{v}_1$ and $\mathbf{v}_6$, we obtain $\mathbf{v}_2$.
- *Case 2.5.* For $a(x) = a_2\exp(f_1x)$, $f(x) = f_2\exp(f_1x)$, $g(u) = g_2\exp(g_1u) + g_3$, a_2, f_1, f_2, $g_1 \neq 0$, $g_2 \neq 0$, g_3 constants, the infinitesimal generators are $\mathbf{v}_1$ and $\mathbf{v}_4$. Moreover, if $a_2 = f_2 = 0$ besides $\mathbf{v}_1$ and $\mathbf{v}_4$, we obtain $\mathbf{v}_2$.
- *Case 2.6.* For $a(x) = a_2\exp(a_1x)$, $f(x) = f_2$, $g(u) = g_1\ln(u) + g_2$, a_1, a_2, f_2, $g_1 \neq 0$, g_2 constants, we obtain $\mathbf{v}_1$ and $\mathbf{v}_5$. Moreover, if $a_2 = f_2 = 0$ besides $\mathbf{v}_1$ and $\mathbf{v}_5$, we obtain $\mathbf{v}_2$.
- *Case 2.7.* For $a(x) = 0$, $f(x) = f_2\exp(f_1x)$, $g(u) = g_2u^{g_1+1} + g_3$, f_1, f_2, $g_1 \neq 0$, $g_2 \neq 0$, g_3 constants, the infinitesimal generators are

$$\mathbf{v}_1, \quad \mathbf{v}_{10} = -(g_1 + 1)\partial_x + f_1g_1t\partial_t - f_1u\partial_u.$$

- *Case 2.8.* For $a(x) = 0$, $f(x) = f_2 \exp(f_1 x)$, $g(u) = g_1 \ln(u) + g_2$, f_1, f_2, $g_1 \neq 0$, g_2 constants, the infinitesimal generators are

$$\mathbf{v}_1, \quad \mathbf{v}_{11} = t\partial_t + u\partial_u.$$

  Moreover, if $f_1 = 0$ besides $\mathbf{v}_1$ and $\mathbf{v}_{11}$, we obtain $\mathbf{v}_2$.
- *Case 2.9.* For $a(x) = 0$, $f(x) = f_2$, $g(u) = g_2 \exp(g_1 u) + g_3$, f_2, $g_1 \neq 0$, g_2, g_3 constants, the infinitesimal generators are

$$\mathbf{v}_1, \quad \mathbf{v}_2, \quad \mathbf{v}_{12} = \exp(-f_2 g_1 t)\partial_t + f_2 \exp(-f_2 g_1 t)\partial_u.$$

7.3 Optimal System

In order to determine solutions of Eqs. (7.12) and (7.13) that are not equivalent by the action of an element of the Lie symmetry group, we must calculate the optimal system.

Next, by using the Lie bracket operation $[\mathbf{v}_i, \mathbf{v}_j] = \mathbf{v}_i(\mathbf{v}_j) - \mathbf{v}_j(\mathbf{v}_i)$ we have constructed the Lie commutator table. Moreover, by summing the Lie series, the adjoint representation table shows the separate adjoint actions of each element as it acts over all the other elements.

7.3.1 Optimal System for Eq. (7.12)

For the infinitesimal generators of case 1.3, we have the following:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_2$
$\mathbf{v}_1$	0	0
$\mathbf{v}_2$	0	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_2$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_2$
$\mathbf{v}_2$	$\mathbf{v}_1$	$\mathbf{v}_2$

In case 1.4 and considering $a_2 = 0$, the corresponding commutator and adjoint tables are given by:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_2$	$\mathbf{v}_3$
$\mathbf{v}_2$	0	0
$\mathbf{v}_3$	0	0

Ad	$\mathbf{v}_2$	$\mathbf{v}_3$
$\mathbf{v}_2$	$\mathbf{v}_2$	$\mathbf{v}_3$
$\mathbf{v}_3$	$\mathbf{v}_2$	$\mathbf{v}_3$

Moreover, in case 1.4 for $f_2 = 0$, we have the following:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_3$
$\mathbf{v}_1$	0	$a_1\mathbf{v}_1$
$\mathbf{v}_3$	$-a_1\mathbf{v}_1$	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_3$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_3 - \varepsilon a_1\mathbf{v}_1$
$\mathbf{v}_3$	$e^{\varepsilon a_1}\mathbf{v}_1$	$\mathbf{v}_3$

Considering $a_2 = 0$ in case 1.5, the corresponding commutator and adjoint tables are given by:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_2$	$\mathbf{v}_4$
$\mathbf{v}_2$	0	0
$\mathbf{v}_4$	0	0

Ad	$\mathbf{v}_2$	$\mathbf{v}_4$
$\mathbf{v}_2$	$\mathbf{v}_2$	$\mathbf{v}_4$
$\mathbf{v}_4$	$\mathbf{v}_2$	$\mathbf{v}_4$

The commutator and adjoint tables of case 1.5, with $a_2 \neq 0$, $f_2 = 0$, follow:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_4$
$\mathbf{v}_1$	0	$-a_1 g_1 \mathbf{v}_1$
$\mathbf{v}_4$	$a_1 g_1 \mathbf{v}_1$	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_4$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_4 + \varepsilon a_1 g_1 \mathbf{v}_1$
$\mathbf{v}_4$	$e^{-\varepsilon a_1 g_1} \mathbf{v}_1$	$\mathbf{v}_4$

For the infinitesimal generators of case 1.6 with a_2, $f_2 \neq 0$, we have the following tables:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_5$
$\mathbf{v}_1$	0	$a_1 \mathbf{v}_1$
$\mathbf{v}_5$	$-a_1 \mathbf{v}_1$	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_5$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_5 - \varepsilon a_1 \mathbf{v}_1$
$\mathbf{v}_5$	$e^{\varepsilon a_1} \mathbf{v}_1$	$\mathbf{v}_5$

In case 1.6 considering $a_2 = f_2 = 0$ commutator and adjoint tables are the following:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_5$
$\mathbf{v}_1$	0	0	$a_1 \mathbf{v}_1$
$\mathbf{v}_2$	0	0	0
$\mathbf{v}_5$	$-a_1 \mathbf{v}_1$	0	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_5$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_5 - \varepsilon a_1 \mathbf{v}_1$
$\mathbf{v}_2$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_5$
$\mathbf{v}_5$	$e^{\varepsilon a_1} \mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_5$

Considering $a_2 \neq 0$ in case 1.7, the corresponding commutator and adjoint tables are

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_6$
$\mathbf{v}_1$	0	$a_1 g_1 \mathbf{v}_1$
$\mathbf{v}_6$	$-a_1 g_1 \mathbf{v}_1$	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_6$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_6 - \varepsilon a_1 g_1 \mathbf{v}_1$
$\mathbf{v}_6$	$e^{\varepsilon a_1 g_1} \mathbf{v}_1$	$\mathbf{v}_6$

For case 1.7 with $a_2 = 0$, we obtain

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_6$
$\mathbf{v}_1$	0	0	$a_1 g_1 \mathbf{v}_1$
$\mathbf{v}_2$	0	0	0
$\mathbf{v}_6$	$-a_1 g_1 \mathbf{v}_1$	0	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_6$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_6 - \varepsilon a_1 g_1 \mathbf{v}_1$
$\mathbf{v}_2$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_6$
$\mathbf{v}_6$	$e^{\varepsilon a_1 g_1} \mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_6$

In case 1.8 the commutator and adjoint tables are the following:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_5$
$\mathbf{v}_2$	0	0	0
$\mathbf{v}_7$	0	0	$g_1\mathbf{v}_7$
$\mathbf{v}_8$	0	$-g_1\mathbf{v}_7$	0

Ad	$\mathbf{v}_2$	$\mathbf{v}_7$	$\mathbf{v}_8$
$\mathbf{v}_2$	$\mathbf{v}_2$	$\mathbf{v}_7$	$\mathbf{v}_8$
$\mathbf{v}_7$	$\mathbf{v}_2$	$\mathbf{v}_7$	$\mathbf{v}_8 - \varepsilon g_1\mathbf{v}_7$
$\mathbf{v}_8$	$\mathbf{v}_2$	$\mathrm{e}^{\varepsilon g_1}\mathbf{v}_7$	$\mathbf{v}_8$

Thus, for the generalized BBM equation (7.12), an optimal system of 1-dimensional subalgebras is given by:

- *Case 1.1*: $\{\mathbf{v}_1\}$.
- *Case 1.2*: $\{\mathbf{v}_2\}$.
- *Case 1.3*: $\{b_1\mathbf{v}_1 + b_2\mathbf{v}_2\}$, b_1, b_2 constants.
- *Case 1.4*: If $a_2, f_2 \neq 0$, we obtain $\{\mathbf{v}_3\}$. If $a_2 = 0, f_2 \neq 0$, we obtain $\{b_1\mathbf{v}_2 + b_2\mathbf{v}_3\}$, b_1, b_2 constants. And if $a_2 \neq 0$, $f_2 = 0$, we obtain $\{\mathbf{v}_1, \mathbf{v}_3\}$.
- *Case 1.5*: If $a_2, f_2 \neq 0$, we obtain $\{\mathbf{v}_4\}$. If $a_2 = 0, f_2 \neq 0$, we obtain $\{b_1\mathbf{v}_2 + b_2\mathbf{v}_4\}$, b_1, b_2 constants. And if $a_2 \neq 0$, $f_2 = 0$, we obtain $\{\mathbf{v}_1, \mathbf{v}_4\}$.
- *Case 1.6*: If $a_2, f_2 \neq 0$, we obtain $\{\mathbf{v}_1, \mathbf{v}_5\}$. If $a_2 = f_2 = 0$, $\{b_1\mathbf{v}_1 + b_2\mathbf{v}_2, b_3\mathbf{v}_2 + b_4\mathbf{v}_5\}$, b_1, b_2, b_3, b_4 constants.
- *Case 1.7*: If $a_2 \neq 0$, we obtain $\{\mathbf{v}_1, \mathbf{v}_6\}$. Otherwise, $\{b_1\mathbf{v}_1 + b_2\mathbf{v}_2, b_3\mathbf{v}_2 + b_4\mathbf{v}_6\}$, b_1, b_2, b_3, b_4 constants.
- *Case 1.8*: $\{b_1\mathbf{v}_2 + b_2\mathbf{v}_7, b_3\mathbf{v}_2 + b_4\mathbf{v}_8\}$, b_1, b_2, b_3, b_4 constants.

7.3.2 Optimal System for Eq. (7.13)

For this case, just the new commutator and adjoint tables are shown. Using the infinitesimal generators of case 2.3, we have the following:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_9$
$\mathbf{v}_1$	0	$a_1\mathbf{v}_1$
$\mathbf{v}_9$	$-a_1\mathbf{v}_1$	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_9$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_9 - \varepsilon a_1\mathbf{v}_1$
$\mathbf{v}_9$	$\mathrm{e}^{\varepsilon a_1}\mathbf{v}_1$	$\mathbf{v}_9$

In case 2.4, the corresponding commutator and adjoint tables are given by:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_6$
$\mathbf{v}_1$	0	$a_1g_1\mathbf{v}_1$
$\mathbf{v}_6$	$-a_1g_1\mathbf{v}_1$	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_6$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_6 - \varepsilon a_1g_1\mathbf{v}_1$
$\mathbf{v}_6$	$\mathrm{e}^{\varepsilon a_1g_1}\mathbf{v}_1$	$\mathbf{v}_6$

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_6$
$\mathbf{v}_1$	0	0	$a_1g_1\mathbf{v}_1$
$\mathbf{v}_2$	0	0	0
$\mathbf{v}_6$	$-a_1g_1\mathbf{v}_1$	0	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_6$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_6 - \varepsilon a_1g_1\mathbf{v}_1$
$\mathbf{v}_2$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_6$
$\mathbf{v}_6$	$\mathrm{e}^{\varepsilon a_1g_1}\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_6$

Considering $a_2,\ f_2 = 0$ in case 2.5 we have the following tables:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_4$
$\mathbf{v}_1$	0	0	$-a_1 g_1 \mathbf{v}_1$
$\mathbf{v}_2$	0	0	0
$\mathbf{v}_4$	$a_1 g_1 \mathbf{v}_1$	0	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_4$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_4 - \varepsilon a_1 g_1 \mathbf{v}_1$
$\mathbf{v}_2$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_4$
$\mathbf{v}_4$	$\mathrm{e}^{\varepsilon a_1 g_1} \mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_4$

We have the following for the infinitesimal generators of case 2.7:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_{10}$
$\mathbf{v}_1$	0	$f_1 g_1 \mathbf{v}_1$
$\mathbf{v}_{10}$	$-f_1 g_1 \mathbf{v}_1$	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_{10}$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_{10} - \varepsilon f_1 g_1 \mathbf{v}_1$
$\mathbf{v}_{10}$	$\mathrm{e}^{\varepsilon f_1 g_1} \mathbf{v}_1$	$\mathbf{v}_{10}$

For case 2.8 with $f_1 \neq 0$, the commutator and adjoint tables are

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_{11}$
$\mathbf{v}_1$	0	$\mathbf{v}_1$
$\mathbf{v}_{11}$	$-\mathbf{v}_1$	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_{11}$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_{11} - \varepsilon \mathbf{v}_1$
$\mathbf{v}_{11}$	$\mathrm{e}^{\varepsilon} \mathbf{v}_1$	$\mathbf{v}_{11}$

And considering case 2.8 with $f_1 = 0$, we have the following:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_{11}$
$\mathbf{v}_1$	0	0	$\mathbf{v}_1$
$\mathbf{v}_2$	0	0	0
$\mathbf{v}_{11}$	$-\mathbf{v}_1$	0	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_{11}$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_{11} - \varepsilon \mathbf{v}_1$
$\mathbf{v}_2$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_{11}$
$\mathbf{v}_{11}$	$\mathrm{e}^{\varepsilon} \mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_{11}$

Finally, the commutator and adjoint tables of case 2.9:

$[\mathbf{v}_i, \mathbf{v}_j]$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_{12}$
$\mathbf{v}_1$	0	0	$-f_2 g_1 \mathbf{v}_{12}$
$\mathbf{v}_2$	0	0	0
$\mathbf{v}_{12}$	$f_2 g_1 \mathbf{v}_{12}$	0	0

Ad	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_{12}$
$\mathbf{v}_1$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathrm{e}^{\varepsilon f_2 g_1} \mathbf{v}_{12}$
$\mathbf{v}_2$	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_{12}$
$\mathbf{v}_{12}$	$\mathbf{v}_1 - \varepsilon f_2 g_1 \mathbf{v}_{12}$	$\mathbf{v}_2$	$\mathbf{v}_{12}$

Hence, for the generalized BBM equation (7.13), an optimal system of 1-dimensional subalgebras is given by:

- *Case 2.1*: $\{\mathbf{v}_1\}$.
- *Case 2.2*: $\{b_1\mathbf{v}_1 + b_2\mathbf{v}_2\}$, b_1, b_2 constants.
- *Case 2.3*: $\{\mathbf{v}_1, \mathbf{v}_9\}$.
- *Case 2.4*: If $a_2 \neq 0$, we obtain $\{\mathbf{v}_1, \mathbf{v}_6\}$. Otherwise, $\{b_1\mathbf{v}_1 + b_2\mathbf{v}_2, b_3\mathbf{v}_2 + b_4\mathbf{v}_6\}$, b_1, b_2, b_3, b_4 constants.
- *Case 2.5*: If $a_2,\ f_2 \neq 0$, we obtain $\{\mathbf{v}_1, \mathbf{v}_4\}$. Otherwise, $\{b_1\mathbf{v}_1 + b_2\mathbf{v}_2, b_3\mathbf{v}_2 + b_4\mathbf{v}_4\}$, b_1, b_2, b_3, b_4 constants.

- *Case 2.6*: If $a_2, f_2 \neq 0$, we obtain $\{\mathbf{v}_1, \mathbf{v}_5\}$. Otherwise, $\{b_1\mathbf{v}_1 + b_2\mathbf{v}_2, b_3\mathbf{v}_2 + b_4\mathbf{v}_5\}$, b_1, b_2, b_3, b_4 constants.
- *Case 2.7*: $\{\mathbf{v}_1, \mathbf{v}_{10}\}$.
- *Case 2.8*: If $f_1 \neq 0$, we obtain $\{\mathbf{v}_1, \mathbf{v}_{11}\}$. Otherwise, $\{b_1\mathbf{v}_1 + b_2\mathbf{v}_2, b_3\mathbf{v}_2 + b_4\mathbf{v}_{11}\}$, b_1, b_2, b_3, b_4 constants.
- *Case 2.9*: $\{b_1\mathbf{v}_1 + b_2\mathbf{v}_2, b_3\mathbf{v}_2 + b_4\mathbf{v}_{12}\}$, b_1, b_2, b_3, b_4 constants.

7.4 Reductions and Exact Solutions

We use the method of characteristics to determine the invariants and reduced ODEs corresponding to each generator given in Sect. 7.3.

7.4.1 Reductions for Eq. (7.12)

Case 1.3 For $g(u)$ arbitrary function, $f(t) = f_2$ and $a(x) = a_2$ constants, by using the generator $b_1\mathbf{v}_1 + b_2\mathbf{v}_2$, we obtain the similarity variable and the similarity solution

$$z = b_1 x - b_2 t, \qquad u(x,t) = h(z),$$

and the reduced equation

$$b_1^2 b_2\, h''' - a_2 b_1^2\, h'' + b_1 g_h\, h' - b_2\, h' - f_2 = 0.$$

Case 1.4 For $a(x) = a_2 \exp(a_1 x)$, $f(t) = f_2 t_1^{f}$, $g(u) = g_1 u^{\frac{f_1}{f_1+1}} + g_2$, by using the generator $\mathbf{v}_3$, we obtain the similarity variable and the similarity solution

$$z = t\exp(a_1 x), \qquad u(x,t) = h(z) t^{f_1+1},$$

and the reduced equation

$$\begin{aligned}
&a_1^2(f_1+1)z^3\, hh''' + a_1^2(f_1+1)(a_2 z + f_1 + 4)z^2\, hh'' + 2a_1^2 a_2(f_1+1)z^2\, hh' \\
&\quad - a_1 f_1 g_1 z h^{\frac{f_1}{f_1+1}} h' + \left(a_1^2 f_1^2 + (3a_1^2 - 1) f_1 + 2a_1^2 - 1\right) z\, hh' \\
&\quad - (f_1+1)^2 h^2 + f_2(f_1+1)h = 0.
\end{aligned}$$

Moreover, in the case that $a_2 = 0$ and $f_2 \neq 0$ we consider $b_1\mathbf{v}_2+b_2\mathbf{v}_3$. Therefore, we obtain

$$z = t \exp\left(\frac{a_1 x}{b_2 - b_1}\right), \qquad u(x,t) = h(z)t^{f_1+1},$$

and the following ODE:

$$a_1^2(f_1+1)z^3\, hh''' + a_1^2(f_1+1)(f_1+4)z^2\, hh'' - a_1(b_2-b_1)f_1 g_1 z h^{\frac{f_1}{f_1+1}} h'$$
$$+ \left(a_1^2(f_1^2+3f_1+2) - (b_2-b_1)^2(f_1+1)\right) zhh' - (b_2-b_1)^2(f_1+1)^2 h^2$$
$$+ f_2(b_2-b_1)^2(f_1+1)h = 0.$$

Case 1.5 For $a(x) = a_2 \exp(a_1 x)$, $f(t) = \frac{f_2}{t}$, $g(u) = g_2 \exp(g_1 u) + g_3$, by using the generator $\mathbf{v}_4$, we get the similarity variable and the similarity solution

$$z = t \exp(a_1 x), \qquad u(x,t) = h(z) + \frac{a_1}{g_1} x,$$

and the reduced equation

$$a_1^2 g_1 z^3\, h''' + a_1^2 g_1(a_2 z + 3)z^2\, h'' - a_1 g_1^2 g_2 z^2 e^{g_1 h} h'$$
$$+ g_1\left(2a_1^2 a_2 z + a_1^2 - 1\right) z\, h' - a_1 g_1 g_2 e^{g_1 h} z + a_1^2 a_2 z + f_2 g_1 = 0.$$

Furthermore, in the case that $a_2 = 0$ and $f_2 \neq 0$ we consider $b_1\mathbf{v}_2 + b_2\mathbf{v}_4$. Thus, we obtain

$$z = t \exp\left(\frac{a_1 b_2 g_1 x}{b_2 g_1 + b_1}\right), \qquad u(x,t) = h(z) + \frac{a_1 b_2}{b_2 g_1 + b_1} x,$$

and the following ODE:

$$- a_1^2 b_2^2 g_1^2 z^3\, h''' - 3a_1^2 b_2^2 g_1^2 z^2\, h'' + a_1 b_2 g_1^2(b_2 g_1 + b_1)g_2 z^2 e^{g_1 h} h'$$
$$- (a_1 b_2 g_1 - b_2 g_1 - b_1)(a_1 b_2 g_1 + b_2 g_1 + b_1) z\, h' + a_1 b_2 g_1(b_2 g_1 + b_1) g_2 z\, e^{g_1 h}$$
$$- f_2(b_2 g_1 + b_1)^2 = 0.$$

Case 1.6 For $a(x) = a_2 \exp(a_1 x)$, $f(t) = f_2$, $g(u) = g_1 \ln(u) + g_2$, by using the generator $\mathbf{v}_5$, we get the similarity variable and the similarity solution

$$z = t \exp(a_1 x), \qquad u(x,t) = t\, h(z),$$

and the reduced equation

$$a_1^2 z^3\, hh''' + a_1^2(a_2 z + 4)z^2\, hh'' + (2a_1^2 a_2 z + 2a_1^2 - 1)z\, hh' \\ - a_1 g_1 z h' - h^2 + f_2 h = 0.$$

Furthermore, in the case that $a_2 = f_2 = 0$ we consider the following elements of the optimal system $b_1\mathbf{v}_1 + b_2\mathbf{v}_2$ and $b_3\mathbf{v}_2 + b_4\mathbf{v}_5$. First, for $b_1\mathbf{v}_1 + b_2\mathbf{v}_2$ we obtain

$$z = b_1 x - b_2 t, \qquad u(x,t) = h(z),$$

and the following ODE:

$$b_1^2\, h''' - h' = 0.$$

Finally, by using $b_3\mathbf{v}_2 + b_4\mathbf{v}_5$ we obtain

$$z = t \exp\left(\frac{a_1 b_4 x}{b_4 - b_3}\right), \qquad u(x,t) = t\, h(z),$$

and the following ODE:

$$- a_1^2 b_4^2 z^3\, hh''' - 4a_1^2 b_4^2 z^2\, hh'' + \left((b_4 - b_3)^2 - 2a_1^2 b_4^2\right) z\, hh' \\ + a_1 b_4 (b_4 - b_3) g_1 z\, h' + (b_4 - b_3)^2 h^2 = 0.$$

Case 1.7 For $a(x) = a_2 \exp(a_1 x)$, $f(t) = 0$, $g(u) = g_2 u^{g_1+1} + g_3$, by using the generator $\mathbf{v}_6$, we get the similarity variable and the similarity solution

$$z = t \exp(a_1 x), \qquad u(x,t) = t^{-\frac{1}{g_1}}\, h(z),$$

and the reduced equation

$$a_1^2 g_1 z^3\, h''' + a_1^2(a_2 g_1 z + 3g_1 - 1)z^2\, h'' - a_1 g_1 (g_1 + 1) g_2 z h^{g_1} h' \\ + \left(2a_1^2 a_2 g_1 z + a_1^2 g_1 - a_1^2 - g_1\right) z\, h' + h = 0.$$

Moreover, in the case that $a_2 = 0$ we consider the following elements of the optimal system $b_1\mathbf{v}_1 + b_2\mathbf{v}_2$ and $b_3\mathbf{v}_2 + b_4\mathbf{v}_6$. To begin with, for $b_1\mathbf{v}_1 + b_2\mathbf{v}_2$ we obtain

$$z = b_1 x - b_2 t, \qquad u(x,t) = h(z),$$

and the following ODE:

$$b_1^2 b_2\, h''' - b_1(g_1+1)g_2 h^{g_1} h' - b_2 h' = 0.$$

Finally, by using $b_3\mathbf{v}_2 + b_4\mathbf{v}_6$ we obtain

$$z = t \exp\left(\frac{a_1 g_1 x}{b_4 g_1 - b_3}\right), \qquad u(x,t) = t^{\frac{-1}{g_1}}\, h(z),$$

and the following ODE:

$$a_1^2 g_1^3 z^3\, h''' + a_1^2 g_1^2 (3g_1 - 1) z^2\, h'' - a_1 g_1^2 (g_1+1)(b_4 g_1 - b_3) g_2 z h^{g_1} h'$$
$$- g_1 \left((b_4 g_1 - b_3)^2 - a_1^2 g_1 (g_1 - 1)\right) z\, h' + (b_4 g_1 - b_3)^2 h = 0.$$

Case 1.8 For $a(x) = 0$, $f(t)$ arbitrary, $g(u) = g_2 \exp(g_1 u) + g_3$, by using the generator $b_1\mathbf{v}_2 + b_2\mathbf{v}_7$, we get the similarity variable and the similarity solution

$$z = x - \frac{b_1}{b_2}\int \exp\left(g_1 \int f(t)\,dt\right) dt, \qquad u(x,t) = \int f(t)\,dt\; + h(z),$$

and the reduced equation

$$b_1\, h''' + b_2 g_1 g_2 e^{g_1 h}\, h' - b_1\, h' = 0.$$

Moreover, we consider the following element of the optimal system $b_3\mathbf{v}_2 + b_4\mathbf{v}_8$. In this case, we obtain

$$z = x - \frac{b_3}{b_4 g_1} \log\left(\int \exp\left(g_1 \int f(t)\,dt\right) dt\right),$$
$$u(x,t) = \int f(t)\,dt - \frac{1}{g_1}\log\left(\int \exp\left(g_1 \int f(t)\,dt\right) dt\right) + h(z),$$

and the reduced equation

$$b_3\, h''' + b_4 g_1^2 g_2 e^{g_1 h}\, h' - b_3\, h' - b_4 = 0.$$

7.4.2 Reductions for Eq. (7.13)

Now we will determine the reductions for Eq. (7.13) from the point symmetries obtained. We would like to point out that we will only show those cases in which $f = f(x)$ is not constant. The cases with f constant are included in Sect. 7.4.1.

Case 2.3 For $a(x) = a_2 \exp(a_1 x)$, $f(x) = f_2 \exp(f_1 x)$, $g(u) = g_1 u^{\frac{f_1}{f_1 - a_1}} + g_2$, by using the generator $\mathbf{v}_9$, we obtain the similarity variable and the similarity solution

$$z = t \exp(a_1 x), \qquad u(x,t) = t^{\frac{a_1 - f_1}{f_1}} h(z),$$

and the reduced equation

$$\begin{aligned} &a_1^3 (f_1 - a_1) z^3 \, hh''' + a_1^2 (f_1 - a_1)(a_1 a_2 z - f_1 + 4a_1) z^2 \, hh'' - a_1^2 f_1 g_1 z \, h^{\frac{f_1}{f_1 - a_1}} h' \\ &\quad + a_1 (f_1 - a_1)(2a_1^2 a_2 z - a_1 f_1 + 2a_1^2 - 1) z \, hh' + (f_1 - a_1)^2 h^2 \\ &\quad + a_1 (f_1 - a_1) f_2 z^{\frac{f_1}{a_1}} h = 0. \end{aligned}$$

Case 2.7 For $a(x) = 0$, $f(x) = f_2 \exp(f_1 x)$, $g(u) = g_2 u^{g_1 + 1} + g_3$, by using the generator $\mathbf{v}_{10}$, we obtain the similarity variable and the similarity solution

$$z = t \exp\left(\frac{f_1 g_1 x}{g_1 + 1}\right), \qquad u(x,t) = t^{-\frac{1}{g_1}} h(z),$$

and the reduced equation

$$\begin{aligned} &f_1^2 g_1^3 z^3 \, h''' + f_1^2 g_1^2 (3g_1 - 1) z^2 \, h'' - f_1 g_1^3 (g_1 + 1)^2 z \, h^{g_1} h' + g_1 (f_1^2 g_1 (g_1 - 1) \\ &\quad - (g_1 + 1)^2) z \, h' + (g_1 + 1)^2 h + f_2 g_1 (g_1 + 1)^2 z^{\frac{g_1 + 1}{g_1}} = 0. \end{aligned}$$

Case 2.8 For $a(x) = 0$, $f(x) = f_2 \exp(f_1 x)$, $g(u) = g_1 \ln u + g_2$, by using the generator $\mathbf{v}_{11}$, we obtain the similarity variable and the similarity solution

$$z = x, \qquad u(x,t) = t \, h(z),$$

and the reduced equation

$$hh'' - g_1 \, h' - h^2 + f_2 e^{f_1 z} \, h = 0.$$

7.5 Travelling Waves

If $g(u)$ is an arbitrary function, $f = c$ and $a = k$ are constants, by using the generator $\lambda \mathbf{v}_1 + \mu \mathbf{v}_2$, we obtain the similarity variable and the similarity solution

$$z = \mu x - \lambda t, \qquad u(x,t) = h(z), \tag{7.17}$$

and the reduced equation

$$\lambda\mu^2 h''' - k\mu^2 h'' + \mu g_h h' - \lambda h' - c = 0. \tag{7.18}$$

Consequently the corresponding solutions of (7.18) are travelling wave solutions. Taking $c = 0$ and integrating with respect to z, Eq. (7.18) can be reduced into the equation

$$\lambda\mu^2 h'' - k\mu^2 h' + \mu g - \lambda h + c_0 = 0, \tag{7.19}$$

where c_0 is an integrating constant. In [11], for Eq. (7.19) with $c_0 = 0$, many exact solutions including solitons, kinks, anti-kinks, compactons and Wadati solitons were obtained.

Next, we present some exact solutions of Eq. (7.18) with $c \neq 0$. As the derivative of trigonometric, hyperbolic, and exponential functions can be expressed in terms of themselves, we can choose g as an algebraic function of h, so that Eq. (7.18) admits the trigonometric functions ($p \sin^q z$, $p \cos^q z$, $p \tan^q z$, $p \sinh^q z$, $p \cosh^q z$, $p \tanh^q z$), hyperbolic functions ($p\, \mathrm{sn}^q (z|m)$, $p\, \mathrm{cn}^q (z|m)$, $p\, \mathrm{dn}^q (z|m)$), and exponential function ($\exp(z)$), as solutions.

- One can be easily check that

$$h(z) = p \sin^q(z),$$

is a solution of Eq. (7.18) with

$$\begin{aligned} g_h(h) = &-\frac{\lambda}{\mu}\left(h^{-2q^{-1}}\mu^4(-1+q)(q-2)\,p^{2q^{-1}} - \mu^4 q^2 - 1\right) \\ &- \frac{1}{\mu^2 q\sqrt{\sqrt[q]{p} - \sqrt[q]{h}}\sqrt{\sqrt[q]{p} + \sqrt[q]{h}}} \\ &\times \left(c h^{\frac{1-q}{q}} + kq\left(h^{-q^{-1}}(-1+q)\,p^{2q^{-1}} - q\sqrt[q]{h}\right)\mu^4\right). \end{aligned} \tag{7.20}$$

Consequently, an exact solution of Eq. (7.12), where $g(u)$ is obtained substituting h by u in (7.20), is

$$u(x,t) = p \sin^q(\mu x - \lambda t). \tag{7.21}$$

For $\mu = \lambda = \frac{k}{2}$, $k = \sqrt{\frac{5}{12}}$, $p = 1$, $q = 2$, the solution

$$u(x,t) = \begin{cases} \sin^2(\mu x - \lambda t) & |x - t| \le \frac{2\pi}{k}, \\ 0 & |x - t| > \frac{2\pi}{k}, \end{cases}$$

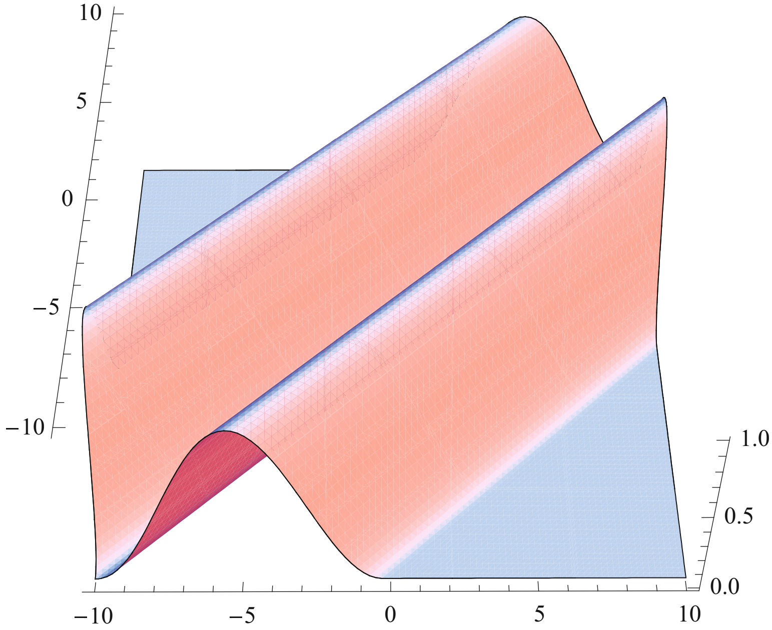

Fig. 7.1 Solution (7.21) for $\mu = \lambda = \frac{k}{2}$, $k = \sqrt{\frac{5}{12}}$, $p = 1$, and $q = 2$

is a sine-type double compacton (that is, solution which has two peaks, see Fig. 7.1)

- For

$$g_h(h) = -\frac{\lambda}{\mu}\left(h^{-2q^{-1}}\mu^4(-1+q)(q-2)\,p^{2q^{-1}} - \mu^4 q^2 - 1\right) - \frac{1}{\mu^2 q\sqrt{\sqrt[q]{p}-\sqrt[q]{h}}\sqrt{\sqrt[q]{p}+\sqrt[q]{h}}} \times \left(ch^{\frac{1-q}{q}} + q\left(h^{-q^{-1}}(-1+q)\,p^{2q^{-1}} - q\sqrt[q]{h}\right)\mu^4 k\right), \tag{7.22}$$

a solution of (7.18) is

$$h(z) = p\,\cos^q(z).$$

So an exact solution of Eq. (7.12) is

$$u(x,t) = p\cos^q(\mu x - \lambda t), \tag{7.23}$$

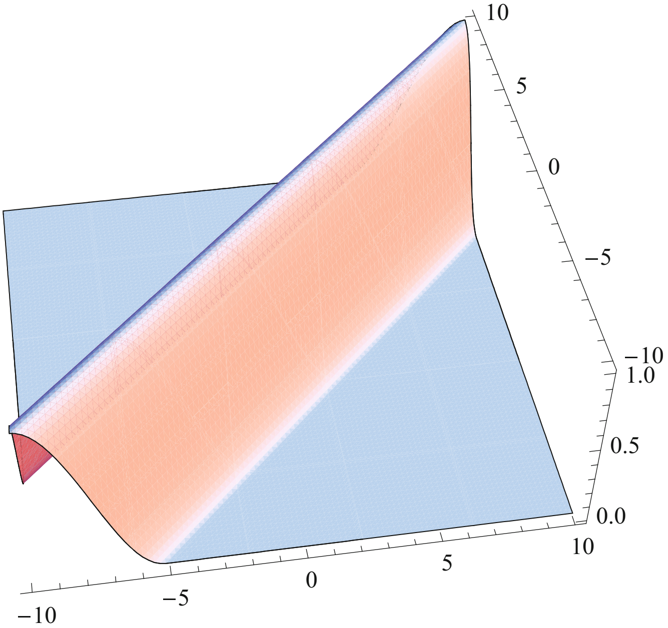

Fig. 7.2 Solution (7.23) for $\mu = \lambda = \frac{k}{2}$, $k = \sqrt{\frac{5}{12}}$, $p = 1$, and $q = 2$

where $g(u)$ is obtained substituting h by u in (7.22). For $\mu = \lambda = \frac{k}{2}$, $k = \sqrt{\frac{5}{12}}$, $p = 1$ and $q = 2$, the solution

$$u(x,t) = \begin{cases} \cos^2(\mu x - \lambda t) & |x - t| \leq \frac{\pi}{k}, \\ 0 & |x - t| > \frac{\pi}{k}, \end{cases}$$

is a compacton solution with a single peak (see Fig. 7.2).

- For

$$\begin{aligned} g_h(h) = &-\frac{1}{\mu^2 q \left(h^{2q^{-1}} + p^{2q^{-1}}\right)} \left(-h^{\frac{1-q}{q}} \sqrt[q]{p} c - \left(2 \left(\left(-3/2\, q^2 l - 3/2\, q\lambda - \lambda\right) \mu^4\right.\right.\right. \\ &+ \lambda/2)\, h^{2q^{-1}} - 3\lambda \left(-1/3 + \left(q^2 - q + 2/3\right) \mu^4\right) p^{2q^{-1}} \\ &\times 2 \left(-1/2\, p^{4q^{-1}} \lambda \mu\, (-1+q)(q-2)\, h^{-2q^{-1}}\right. \\ &- 1/2\, p^{-2q^{-1}} \lambda \mu\, (2+q)(1+q)\, h^{4q^{-1}} \\ &+ k \left(1/2\, p^{3q^{-1}} (-1+q)\, h^{-q^{-1}}\right. \\ &\left.\left.\left.\left.+ 1/2\, p^{-q^{-1}} (1+q)\, h^{3q^{-1}} + \sqrt[q]{h} \sqrt[q]{p} q\right)\right) \mu^3\right) q\mu\right), \end{aligned} \tag{7.24}$$

a solution of (7.18) is

$$h(z) = p\ \tan^q(z).$$

So an exact solution of Eq. (7.12), where $g(u)$ is obtained substituting h by u in (7.24), is

$$u(x,t) = p\tan^q(\mu x - \lambda t). \tag{7.25}$$

- For

$$g_h(h) = -\frac{\lambda}{\mu}\left(h^{-2q^{-1}}\mu^4(-1+q)(q-2)p^{2q^{-1}} - 1 + \left(-q^2+6q-4\right)\mu^4\right) - \frac{1}{\mu^2 q\sqrt{\sqrt[q]{p}-\sqrt[q]{h}}\sqrt{\sqrt[q]{p}+\sqrt[q]{h}}} \times \left(ch^{\frac{1-q}{q}} - q\mu^4 k\left(h^{-q^{-1}}(-1+q)p^{2q^{-1}} - \sqrt[q]{h}(q-2)\right)\right), \tag{7.26}$$

a solution of Eq. (7.18) is

$$h(z) = p\sinh^q(z).$$

So an exact solution of Eq. (7.12), where $g(u)$ is obtained substituting h by u in (7.26), is

$$u(x,t) = p\sinh^q(\mu x - \lambda t). \tag{7.27}$$

- For

$$g_h(h) = -\frac{\lambda}{\mu}\left(h^{-2q^{-1}}\mu^2(-1+q)(q-2)p^{2q^{-1}} - 1 + \left(-q^2+6q-4\right)\mu^2\right) + \frac{1}{\mu q\sqrt{\sqrt[q]{p}-\sqrt[q]{h}}\sqrt{\sqrt[q]{p}+\sqrt[q]{h}}} \times \left(ch^{\frac{1-q}{q}} + q\mu^2 k\left(h^{-q^{-1}}(-1+q)p^{2q^{-1}} - \sqrt[q]{h}(q-2)\right)\right) \tag{7.28}$$

a solution of Eq. (7.18) is

$$h(z) = p\cosh^q(z).$$

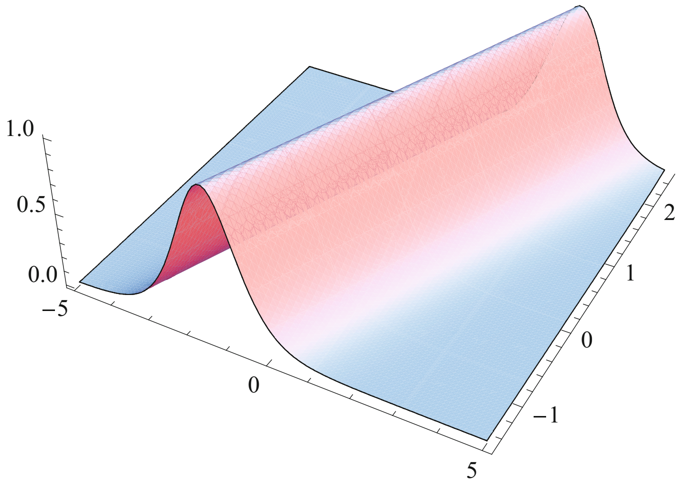

Fig. 7.3 Solution (7.29) for $\lambda = \mu = 1$, $p = 1$, and $q = -2$

Consequently, an exact solution of Eq. (7.12), where $g(u)$ is obtained substituting h by u in (7.28), is

$$u(x,t) = p\cosh^q(\mu x - \lambda t). \tag{7.29}$$

For $\lambda = \mu = 1$, $p = 1$, and $q = -2$ the solution

$$u(x,t) = \operatorname{sech}^2(x - t),$$

describes a soliton moving along a line with constant velocity (see Fig. 7.3).

- For

$$\begin{aligned} g_h(h) = &-\frac{1}{q\mu\left(h^{2q^{-1}} - p^{2q^{-1}}\right)}\Bigg(h^{\frac{1-q}{q}}\sqrt[q]{p}c - \Big(3\left(q^2\mu^2 + \mu^2 q\right. \\ &\left. + 2/3\,\mu^2 + 1/3\right)\lambda\, h^{2q^{-1}} - 3\left(q^2\mu^2 - \mu^2 q + 2/3\,\mu^2 + 1/3\right)\lambda\, p^{2q^{-1}} \\ &+ \mu^2\Big(p^{4q^{-1}}\lambda(-1+q)(q-2)\,h^{-2q^{-1}} - p^{-2q^{-1}}\lambda\,(2+q)(1+q)\,h^{4q^{-1}} \\ &- \left(p^{3q^{-1}}(-1+q)\,h^{-q^{-1}} + p^{-q^{-1}}(1+q)\,h^{3q^{-1}} - 2q\sqrt[q]{h}\sqrt[q]{p}\right)k\Big)\Big)q\Bigg), \end{aligned} \tag{7.30}$$

a solution of Eq. (7.18) is

$$h(z) = p\tanh^q(z).$$

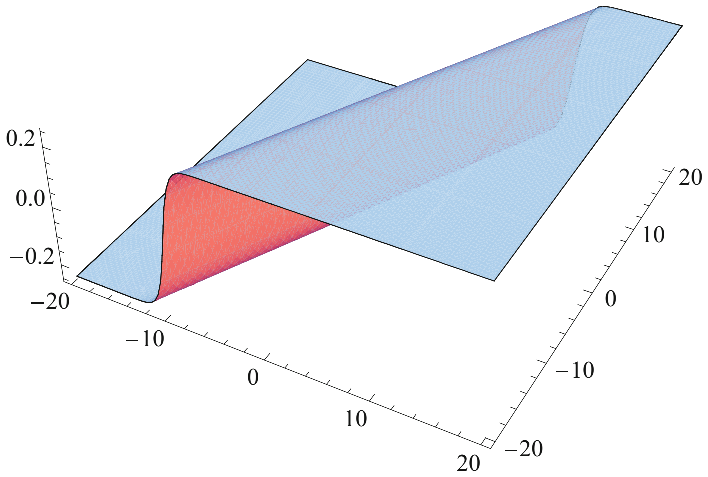

Fig. 7.4 Solution (7.31) for $\mu = 1$, $\lambda = \frac{1}{2}$, $p = \frac{1}{4}$, and $q = 1$

Consequently, an exact solution of Eq. (7.12), where $g(u)$ is obtained substituting h by u in (7.30), is

$$u(x,t) = p \tanh^q(\mu x - \lambda t). \tag{7.31}$$

For $\mu = 1$, $\lambda = \frac{1}{2}$, $p = \frac{1}{4}$, and $q = 1$ the solution

$$u(x,t) = \frac{1}{4}\tanh\left(x - \frac{t}{2}\right),$$

describes a kink solution (see Fig. 7.4).
For $\mu = 1$, $\lambda = \frac{1}{2}$, $p = 1$, and $q = 3$ the solution

$$u(x,t) = \tanh^3\left(x - \frac{t}{2}\right),$$

describes an anti-kink solution (see Fig. 7.5).

- For

$$g_h(h) = \frac{-\lambda\, h\, \mu^2\, q^3 + h\, k\, \mu^2\, q^2 + \lambda\, h\, q + c}{h\, \mu\, q}, \tag{7.32}$$

a solution of (7.18) is

$$h(z) = p \exp(qz).$$

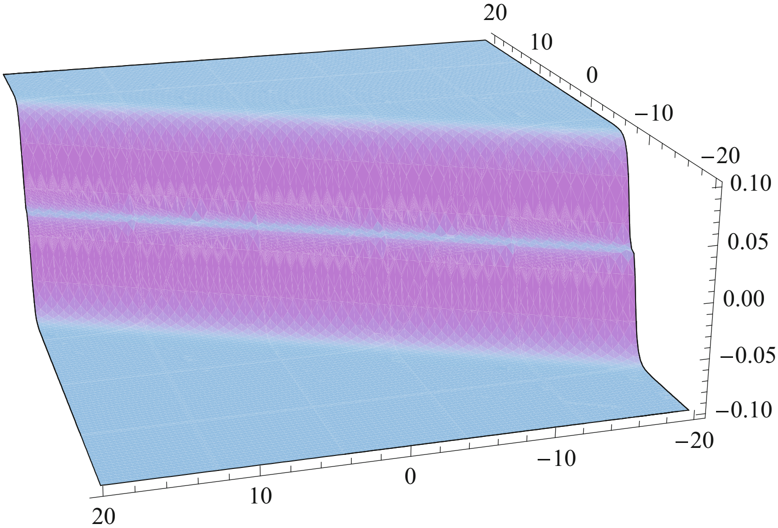

Fig. 7.5 Solution (7.31) for $\mu = 1$, $\lambda = \frac{1}{2}$, $p = 1$, and $q = 3$

Consequently, an exact solution of Eq. (7.12), where $g(u)$ is obtained substituting h by u in (7.32), is

$$u(x,t) = p\exp[q(\mu x - \lambda t)]. \tag{7.33}$$

- For

$$\begin{aligned} g_h(h) = {} & \frac{1}{4\mu\, n^2\sqrt{p^{n/4}-h^{n/4}}\sqrt{p^{n/4}+h^{n/4}}\sqrt{p^{n/2}-h^{n/2}m}} \\ & \times \left(-8\sqrt{p^{n/4}+h^{n/4}}\left(p^{n/2}\mu^2\,(n-2)\,(n-4)\,h^{-n/2}\right.\right. \\ & + p^{-n/2}m\mu^2\,(n+4)\,(n+2)\,h^{n/2} \\ & \left. + (-8\,m-8)\,\mu^2 - 1/2\,n^2\right)\lambda\sqrt{p^{n/4}-h^{n/4}}\sqrt{p^{n/2}-h^{n/2}m} \\ & + \left(cp^{n/4}h^{n/4-1}n^2 - 4\,k\left(4\,h^{n/4}\,(m+1)\,p^{n/4} + p^{3/4\,n}\,(n-4)\,h^{-n/4}\right.\right. \\ & \left.\left.\left. - h^{3/4\,n}p^{-n/4}m\,(n+4)\right)\mu^2\right)n\right), \end{aligned} \tag{7.34}$$

a solution of Eq. (7.18) is

$$h(z) = p\,\mathrm{sn}^q(z|m),$$

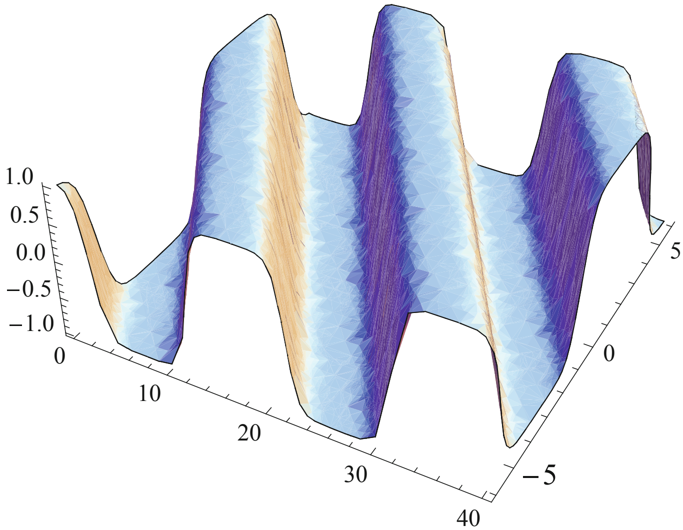

Fig. 7.6 Solution (7.35) for $\mu = \lambda = p = q = 1$ and $m = 0.996$

with $q = \frac{4}{n}$. Consequently, an exact solution of Eq. (7.12), where $g(u)$ is obtained substituting h by u in (7.34), is

$$u(x,t) = p\,\mathrm{sn}^q(\mu x - \lambda t|m). \tag{7.35}$$

For $\mu = \lambda = p = q = 1$ and $m = 0.996$ the solution

$$u(x,t) = \mathrm{sn}(x - t|0.996),$$

shows a stable nonlinear nonharmonic oscillatory periodic wave (see Fig. 7.6).

- For

$$\begin{aligned} g_h(h) = {} & \frac{1}{4\,\mu\,n^2 p^{n/2}\sqrt{p^{n/4}-h^{n/4}}\sqrt{p^{n/4}+h^{n/4}}\sqrt{-mp^{n/2}+p^{n/2}+h^{n/2}m}} \\ & \times \left(-8\sqrt{p^{n/4}+h^{n/4}}\lambda\,\left(p^{n/2}\mu^2\,(n-2)\,(n-4)\,(m-1)\,h^{-n/2}\right.\right. \\ & \times p^{-n/2}m\mu^2\,(n+4)\,(n+2)\,h^{n/2} \\ & \left. + (-16\,m+8)\,\mu^2 + 1/2\,n^2\right)\sqrt{p^{n/4}-h^{n/4}}\sqrt{-mp^{n/2}+p^{n/2}+h^{n/2}m} \\ & + \left(cp^{n/4}h^{n/4-1}n^2 + 4\,k\,\left(p^{3/4\,n}\,(n-4)\,(m-1)\,h^{-n/4}\right.\right. \\ & \left.\left.\left. + \ (8\,m-4)\,h^{n/4}p^{n/4} - h^{3/4\,n}\,p^{-n/4}m\,(n+4)\right)\mu^2\right)n\right), \end{aligned} \tag{7.36}$$

a solution of Eq. (7.18) is

$$h(z) = p\,\mathrm{cn}^q(z|m).$$

Consequently, an exact solution of Eq. (7.12), where $g(u)$ is obtained substituting h by u in (7.36), is

$$u(x,t) = p\,\mathrm{cn}^q(\mu x - \lambda t|m).$$

In the above cases, p and q represent arbitrary constants.

7.6 Conservation Laws

Conservation laws play a significant role in the resolution of equations with a physical background, especially in problems where certain physical properties may not change over time.

We consider T^t and T^x functions of t, x, u and derivatives of u, that represent the conserved density and associated flux, respectively.

A conservation law for Eq. (7.12) (or (7.13)) is a space–time divergence such that

$$D_t T^t(t,x,u,u_t,u_x,\dots) + D_x T^x(t,x,u,u_t,u_x,\dots) = 0, \tag{7.37}$$

on all solutions $u(t,x)$ of Eq. (7.12) (or Eq. (7.13)), where D_t, D_x denote the total derivative operators with respect to t and x, respectively.

If a conserved density is a total x derivatives, $T = D_x\Psi$, when it is restricted to the solution space, then the conservation law (7.37) holds trivially, with the flux being a total t derivative, $X = -D_t\Psi$, when it is restricted to the solution space. Any two conservation laws that differ by such a trivial conservation law are considered to be physically equivalent. The set of all admitted conservation laws forms a vector space on which there is a natural action by the symmetry group of (7.12) (or Eq. (7.13)) [2].

Every local conservation law (7.37) has an equivalent characteristic form in which it has been eliminated u_t and its differential consequences from T^t and T^x by using the equation

$$\widehat{T^t} = T^t\,|_{u_{txx}=\Delta} = T^t - \Phi,$$
$$\widehat{T^x} = T^x\,|_{u_{txx}=\Delta} = T^x - \Psi,$$

where Δ is the result of isolating u_t from Eq. (7.12) (likewise for Eq. (7.13)).

Hence, conservation law (7.37) for (7.12) can be expressed by using its characteristic form

$$D_t\widehat{T^t}(t,x,u,u_t,u_x,u_{tx},\ldots)+D_x\widehat{T^x}(t,x,u,u_t,u_x,u_{tx},\ldots)$$
$$=(u_t-u_{txx}-\partial_x(a(x)\partial_x u)+(g(u))_x-f(t))\,Q(t,x,u,u_t,u_x,\ldots), \tag{7.38}$$

where $Q(x,t,u,u_x,u_t,\ldots)$ is called a multiplier. The characteristic form for Eq. (7.13) is obtained analogously.

A function $Q(x,t,u,u_x,u_t,\ldots)$ is a multiplier if it is non-singular on the set of solutions $u(x,t)$ of Eq. (7.12) (likewise for Eq. (7.13)), and if its product with the equation is a divergence expression with respect to t, x. For this paper we consider multipliers of the form $Q(t,x,u,u_t,u_x,u_{tt})$.

For (7.12), multipliers Q are obtained by means of divergence condition

$$\frac{\delta}{\delta u}\left((u_t-u_{txx}-\partial_x(a(x)\partial_x u)+(g(u))_x-f(t))\,Q\right)=0, \tag{7.39}$$

where $\frac{\delta}{\delta u}=\partial_u-D_x\partial_{u_x}-D_t\partial_{u_t}+D_xD_t\partial_{u_{xt}}+D_x^2\partial_{u_{xx}}+\ldots$, denotes the variational derivative, which has the property of annihilating total derivatives. The divergence condition for Eq. (7.13) can be obtained similarly.

Divergence condition (7.39) gives a multiplier determining equation that splits with respect to u_t, u_{txx} and their differential consequences, yielding an overdetermined system of equations for Q together with the arbitrary functions for Eqs. (7.12) and (7.13), respectively. This system can be solved by the same algorithmic method used to solve the determining equation for infinitesimal symmetries.

We obtain the following multipliers for Eq. (7.12):

1. For $a(x)$, $f(t)$, and $g(u)$ arbitrary functions

$$Q(t,x,u,u_t,u_x,u_{tt})=1 \tag{7.40}$$

2. For $g(u)=c_1u+c_2$ and $a(x)$ and $f(t)$ arbitrary functions $Q(t,x,u,u_t,u_x,u_{tt})$ must satisfy the equations

$$\begin{aligned}\frac{\partial^3}{\partial x^2\partial t}Q\,(t,x,u,u_t,u_x,u_{tt})=&\left(\frac{\partial^2}{\partial x^2}Q\,(t,x,u,u_t,u_x,u_{tt})\right)a\,(x)\\&+\left(\frac{\partial}{\partial x}Q\,(t,x,u,u_t,u_x,u_{tt})\right)\frac{\mathrm{d}}{\mathrm{d}x}a\,(x)\\&-\left(\frac{\partial}{\partial x}Q\,(t,x,u,u_t,u_x,u_{tt})\right)c_1\\&+\frac{\partial^3}{\partial x^3}Q\,(t,x,u,u_t,u_x,u_{tt})\end{aligned}$$

$$+\frac{\partial}{\partial t}Q\left(t,x,u,u_t,u_x,u_{tt}\right),$$
$$\frac{\partial}{\partial u}Q\left(t,x,u,u_1,u_x,u_{tt}\right)=0,$$
$$\frac{\partial}{\partial u_t}Q\left(t,x,u,u_t,u_x,u_{tt}\right)=0,$$
$$\frac{\partial}{\partial u_x}Q\left(t,x,u,u_t,u_x,u_{tt}\right)=0,$$
$$\frac{\partial}{\partial u_{tt}}Q\left(t,x,u,u_t,u_x,u_{tt}\right)=0$$

For Eq. (7.13) we obtain the same multipliers.

There is a one-to-one correspondence between non-trivial multipliers and non-trivial conservation laws in characteristic form [3–5, 33]. Each multiplier determines a corresponding conserved density and flux from the characteristic Eq. (7.38) by splitting it with respect to all derivatives of u that do not appear in the multiplier function. This yields a linear system of equations, which can be straightforwardly integrated to obtain T^t and T^x.

For Eq. (7.12) and multiplier (7.40), the corresponding conserved densities and fluxes are

- For $a(x)$, $f(t)$, and $g(u)$ arbitrary functions
  $Q\ =1,$
  $T^t\ =u,$
  $T^x=u_{xx}-u_{tx}-a\left(x\right)u_x-g\left(u\right).$

For Eq. (7.13) and multiplier (7.40), the corresponding conservation law is

- For $a(x)$, $f(x)$, and $g(u)$ arbitrary functions
  $Q\ =1,$
  $T^t\ =u,$
  $T^x=u_{xx}-u_{tx}-a\left(x\right)u_x-g\left(u\right).$

7.7 Potential Symmetries

In order to find potential symmetries of (7.12) and (7.13) we write the equation in a conserved form by using the conservation law obtained in the previous Sect. 7.6. Hence, the associated auxiliary system is given by

$$\begin{cases} v_x=u, \\ v_t=u_{tx}+a(x)u_x-g(u). \end{cases} \tag{7.41}$$

If $(u(x),v(x))$ satisfies (7.41), then $u(x)$ solves the generalized BBM equations (7.12) and (7.13), with $f(t)=0$ and $f(x)=0$, respectively.

The basic idea for obtaining potential symmetries is to require that the infinitesimal generator

$$X = \xi(x,t,u,v)\frac{\partial}{\partial x} + \tau(x,t,u,v)\frac{\partial}{\partial t} + \phi_1(x,t,u,v)\frac{\partial}{\partial u} + \phi_2(x,t,u,v)\frac{\partial}{\partial v} \tag{7.42}$$

leaves invariant the set of solutions of (7.41).

We obtain classical potential symmetries if

$$(\xi_v)^2 + (\tau_v)^2 + \left(\phi_{1,v}\right)^2 \neq 0. \tag{7.43}$$

Consequently, if the projection of a point symmetry of system (7.41) has essential dependence on v, then the resulting symmetry of Eqs. (7.12) and (7.13), with $f(t) = 0$ and $f(x) = 0$, respectively, will be a nonlocal symmetry.

The condition for the vector field (7.42) to generate a point symmetry of system (7.41) is given by

$$\begin{aligned} &\mathrm{pr}^{(1)}X(v_x - u) = 0 \\ &\mathrm{pr}^{(2)}X(v_t - u_{tx} - au_x + g(u)) = 0 \end{aligned} \tag{7.44}$$

on the solution space of the system (7.41), where pr$X$ denotes the prolongation of the vector field (7.42).

Equations (7.44) split with respect to the x and t derivatives of u and v, yielding an overdetermined, linear system of equations for the infinitesimals $\xi(t,x,u,v)$, $\tau(t,x,u,v)$, $\phi_1(t,x,u,v)$, $\phi_2(t,x,u,v)$ together with the functions $a(x)$ and $g(u)$. We derive and solve this system by using the Maple software.

Each admitted point symmetry can be used to reduce system (7.41) to a system of ordinary differential equations whose solutions correspond to invariant solutions $(u(t,x), v(t,x))$ of system (7.41) under the point symmetry. These invariant solutions are naturally expressed in terms of similarity variables which are found by solving the invariance conditions

$$\phi_1(t,x,u,v) - \tau(t,x,u,v)u_t - \xi(t,x,u,v)u_x = 0,$$
$$\phi_2(t,x,u,v) - \tau(t,x,u,v)u_t - \xi(t,x,u,v)u_x = 0.$$

We obtain the symmetries and we observe that the condition (7.43) is not satisfied, then the Lie method applied to (7.41) leads to the Lie symmetries.

On the other hand, the potential system (7.41) yields a further potential subsystem given by eliminating u in terms of v_x, we obtain the integrated equation

$$u_t - u_{txx} - a\,(x)\,u_{xx} + g\,(u_x) = 0. \tag{7.45}$$

We now classify all point symmetries of this equation. A point symmetry of Eq. (7.45) is a one-parameter Lie group of transformations on (t, x, u) generated by a vector field of the form

$$\mathbf{X} = \tau(t, x, u)\partial_t + \xi(t, x, u)\partial_x + \phi(t, x, u)\partial_u \tag{7.46}$$

which is required to leave invariant the solution space of system (7.45). Invariance of Eq. (7.45) under a Lie group of point transformations with infinitesimal generator (7.46) leads to the determining equations. The solutions of this system depend on the functions of the equation:

- *Case 1*. If $a(x)$ and $g(u_x)$ are arbitrary functions, the only symmetries admitted by (7.12) are the group of space and time translations, which are defined by the infinitesimal generators

$$\mathbf{V}_1 = \frac{\partial}{\partial t}, \quad \mathbf{V}_2 = \frac{\partial}{\partial u}.$$

- *Case 2*. If $a\,(x) = \mathrm{e}^{a_1 x} a_2$ and

$$g(u_x) = \frac{{g_1}^{2-n}\,(u_x + g_2)^n\,(n-1)^{1-n}}{n} + g_3$$

  besides $\mathbf{V}_1$ and $\mathbf{V}_2$ we obtain

$$\mathbf{V}_3^1 = -a_1 t\partial_t + \partial_x + \frac{a_1\,(g_3\,nt + xg_2 + u)}{n-1}\partial_u$$

7.8 Conclusions

In this paper, a damped generalized Benjamin–Bona–Mahony equation with a forcing term has been studied. The equation considered involves three arbitrary functions. In particular, we have considered two different cases. First, the case with a forcing term depending on t and then, a forcing term depending on x. We have obtained all the Lie point symmetries that these equations admit. Point symmetries can be used to reduce the PDE into ODEs. To achieve this goal, we have constructed the optimal system of one-dimensional subalgebras of the Lie symmetry algebra of these equations. By using the optimal systems, we have determined all nonequivalent group-invariant solutions of these equations. Furthermore, some travelling wave solutions have been determined. Moreover, for special values of the parameters of this equation, we obtain many exact solutions expressed by various single and combined nondegenerative Jacobi elliptic function solutions and their degenerative solutions (soliton, kink, and compactons).

On the other hand, by using the multipliers method by Anco and Bluman, we have obtained non-trivial conservation laws. From the conservation laws, we have written the equations under study in a conserved form, and we have proved that the potential symmetries are projected into the Lie symmetries. Finally, we have determined Lie symmetries of integrated equation.

Acknowledgement The authors acknowledge the financial support from Junta de Andalucía group FQM-201, and they express their sincere gratitude to the Plan Propio de Investigación de la Universidad de Cádiz.

References

1. K. Al-Khaled, S. Momani, A. Alawneh, Approximate wave solutions for generalized Benjamin–Bona–Mahony–Burgers equations. Appl. Math. Comput. **171**, 281–292 (2005)
2. S.C. Anco, Generalization of Noether's theorem in modern form to non-variational partial differential equations, in *Recent Progress and Modern Challenges in Applied Mathematics, Modelling and Computational Science*, vol. 79 (Springer, Berlin, 2017), pp. 119–182
3. S.C. Anco, G.W. Bluman, Direct construction of conservation laws from field equations. Phys. Rev. Lett. **78**, 2869–2942 (1997)
4. S.C. Anco, G.W. Bluman, Direct construction method for conservation laws of partial differential equations. Part I: examples of conservation law classifications. Eur. J. Appl. Math. **13**, 545–566 (2002)
5. S.C. Anco, G.W. Bluman, Direct construction method for conservation laws of partial differential equations. Part II: general treatment. Eur. J. Appl. Math. **41**, 567–585 (2002)
6. T.B. Benjamin, J.L. Bona, J.J. Mahony, Model equations for long waves in non-linear dispersive systems. Philos. Trans. R. Soc. A **272**, 47–78 (1972)
7. G.W. Bluman, S.C. Anco, *Symmetry and Integration Methods for Differential Equations* (Springer, New York, 2002)
8. G.W. Bluman, A. Cheviakov, S.C. Anco, *Applications of Symmetry Methods to Partial Differential Equations* (Springer, New York, 2010)
9. G.W. Bluman, S. Kumei, *Symmetries and Differential Equations in Applied Mathematical Sciences*, vol. 81 (Springer, New York, 1989)
10. M.S. Bruzón, M.L. Gandarias, J.C. Camacho, Symmetry for a family of BBM equations. J. Nonlinear Math. Phys. **15**, 81–90 (2008)
11. M.S. Bruzón, M.L. Gandarias, Travelling wave solutions for a generalized Benjamin–Bona–Mahony–Burgers equation. Int. J. Math. Models Methods Appl. Sci. **2**, 103–108 (2008)
12. M.S. Bruzón, M.L. Gandarias, Exact solutions for a generalized Benjamin–Bona–Mahony–Burgers equation. J. Nonlinear Syst. Appl. 151–154 (2009)
13. M.S. Bruzón, M.L. Gandarias, Nonlocal symmetries for a family Benjamin–Bona–Mahony–Burgers equations. Some exact solutions. Int. J. Appl. Math. Inform. **5**, 180–187 (2011)
14. M.S. Bruzón, M.L. Gandarias, On the group classification and conservation laws of the self–adjoint of a family Benjamin–Bona–Mahony equations. Int. J. Math. Model Methods Appl. Sci. **6**, 527–534 (2012)
15. M.S. Bruzón, T.M. Garrido, R. de la Rosa, Conservation laws and exact solutions of a generalized Benjamin–Bona–Mahony–Burgers equation. Chaos, Solitons and Fractals **89**, 578–583 (2016)
16. M.S. Bruzón, T.M. Garrido, R. de la Rosa, Symmetry reductions for a generalized fifth order KdV equation. Appl. Math. Nonlinear Sci. **2**(2), 485–494 (2017)

17. M.S. Bruzón, E. Recio, T.M. Garrido, A.P. Márquez, Conservation laws, classical symmetries and exact solutions of the generalized KdV-Burgers-Kuramoto equation. Open Phys. **15**, 433–439 (2017)
18. P.A. Clarkson, New similarity reductions and Painlevé analysis for the symmetric regularised long wave and modified Benjamin–Bona–Mahoney equations. J. Phys. A Math. Gen. **22**, 3281–3848 (1989)
19. P.A. Clarkson, M.D. Kruskal M D, New similarity solutions of the Boussinesq equation. J. Math. Phys. **30**, 2201–2213 (1989)
20. R. de la Rosa, M.L. Gandarias, M.S. Bruzón, On symmetries and conservation laws of a Gardner equation involving arbitrary functions. Appl. Math. Comput. **290**, 125–134 (2016)
21. R. de la Rosa, M.L. Gandarias, M.S. Bruzón, Symmetry group analysis of a fifth-order KdV equation with variable coefficients. J. Comput. Theor. Transp. **45**, 275–289 (2016)
22. R. de la Rosa, E. Recio, T.M. Garrido, M.S. Bruzón, On a generalized variable-coefficient Gardner equation with linear damping and dissipative terms. Math. Methods Appl. Sci. **41**, 7158–7169 (2018)
23. P. Eloe, M. Usman, Bifurcations in steady state solutions of a class of nonlinear dispersive wave equations. Nonlinear Stud. **19**(2), 215–224 (2012)
24. M.L. Gandarias, C.M. Khalique, Nonlinearly self-adjoint, conservation laws and solutions for a forced BBM equation. Abstr. Appl. Anal. **2014**, 1–5 (2014)
25. T.M. Garrido, M.S. Bruzón, Lie point symmetries and travelling wave solutions for the generalized Drinfeld–Sokolov system. J. Comput. Theor. Transp. **45**(4), 290–298 (2016)
26. T.M. Garrido, A.A. Kasatkin, M.S. Bruzón, R.K. Gazizov, Lie symmetries and equivalence transformations for the Barenblatt–Gilman model. J. Comput. Appl. Math. **318**, 253–258 (2017)
27. N.H. Ibragimov, *CRC Handbook of Lie Group Analysis of Differential Equations*, vols. 1–3 (CRC Press, Boca Raton, 1994–1996)
28. N.H. Ibragimov, The answer to the question put to me by L.V. Ovsyannikov 33 years ago. Arch. ALGA **3**, 53–80 (2006)
29. N.H. Ibragimov, A new conservation theorem. J. Math. Anal. Appl. **333**, 311–328 (2007)
30. N.H. Ibragimov, Quasi self-adjoint differential equations. Arch. ALGA **4**, 55–60 (2007)
31. M. Molati, C.M. Khalique, Lie symmetry analysis of the time-variable coefficient B-BBM equation. Adv. Differ. Equ. **2012**, 1–8 (2012)
32. M. Molati, C.M. Khalique, Symmetry classification and invariant solutions of the variable coefficient BBM equation. Appl. Math. Comput. **219**, 7917–7922 (2013)
33. P.J. Olver, *Applications of Lie Groups to Differential Equations*, 2nd edn. Graduate Texts in Mathematics, vol. 107 (Springer, Berlin, 1993)
34. I. Simbanefayi, C.M. Khalique, Travelling wave solutions and conservation laws for the Korteweg-de-Vries-Benjamin-Bona-Mahony equation. Results Phys. **8**, 57–63 (2018)
35. M. Wang, Long time behavior of a damped generalized BBM equation in low regularity spaces. Math. Methods Appl. Sci. **38**, 4852–4866 (2015)
36. L. Wang, J. Zhou, L. Ren, The exact solitary wave solutions for a family of BBM equation. Int. J. Nonlinear Sci. **1**, 58–64 (2006)
37. J. Weiss, M. Tabor, G. Carnevale, The Painlevé property for partial differential equations. J. Math. Phys. **24**(3), 522–526 (1983)
38. S. Yandong, Explicit and exact special solutions for BBM–like B (m, n) equations with fully nonlinear dispersion. Chaos, Solitons Fractals **25**, 1083–1091 (2005)

Chapter 8
Symmetry Analysis and Conservation Laws for Some Boussinesq Equations with Damping Terms

M. L. Gandarias and M. Rosa

Abstract In this work, we study some Boussinesq equations with damping term from the point of view of the Lie theory. We derive the classical Lie symmetries admitted by the equation as well as the reduced ordinary differential equations. We also present some exact solutions. Some nontrivial conservation laws for these equations are constructed by using the multiplier method.

8.1 Introduction

The Boussinesq equation

$$u_{tt} - u_{xx} + \alpha u_{xxxx} - \beta \left(u^2\right)_{xx} = 0 \tag{8.1}$$

arises in several physical applications, the first one was propagation of long waves in shallow water [1]. There have been several generalizations of the Boussinesq equation such as the modified Boussinesq equation, or the dispersive water wave. A generalization of (8.1), namely

$$u_{tt} - u_{xx} + u_{xxxx} - (f(u))_{xx} = g(x) \tag{8.2}$$

has been considered in [2]. Classical and nonclassical symmetries for Eq. (8.2) were considered in [3]. Considering the effect of viscosity in real process, Varlamov [4, 5] studied the following equation:

$$u_{tt} - 2bu_{xxt} - u_{xx} + \alpha u_{xxxx} - \beta \left(u^2\right)_{xx} = 0, \tag{8.3}$$

M. L. Gandarias (✉) · M. Rosa
Departamento de Matemáticas, Universidad de Cádiz, Cádiz, Spain
e-mail: marialuz.gandarias@uca.es; maria.rosa@uca.es

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_8

where the second term is responsible for strong internal damping [6], $\alpha > 0$ and $\beta \in R$. Replacing the dissipative term for a weaker one Grasselli et al. [7] considered

$$u_{tt} + k\,u_t + u_{xxxx} - (f(u))_{xx} = g(x) \tag{8.4}$$

with a cubic controlled growth nonlinearity, which models the rapid spinodal decomposition in non-equilibrium phase separation process.

Conservation laws have several important uses in the study of partial differential equations (PDEs), especially for determining conserved quantities and constants of motion. They are also useful in detecting integrability and linearizations, finding potentials and nonlocally related systems, as well as checking the accuracy of numerical solution methods.

In this direction, the most famous result is Noether's theorem given by Noether in 1918. And it is well known that Noether's theorem can only be applied to equations having variational structure. In [8] Ibragimov proved a theorem on conservation laws, which can be applied to equations having no variational structure and provides an elegant way to establish local conservation laws for the equations under consideration.

For variational problems, the Noether theorem [9] can be used for the derivation of conservation laws. For any PDE system of normal type, regardless of whether a Lagrangian exists, the conservation laws admitted by the system can be found by a direct method of Anco and Bluman [10, 11] which is computationally similar to Lie method for finding the symmetries [12, 13] admitted by the system. Conservation laws that are symmetry invariant have some important applications. It is well known that when a differential equation admits a Noether symmetry, a conservation law is associated with this symmetry, and furthermore that a double reduction can be achieved as a result of this association. Moreover, any symmetry-invariant conservation law will reduce to a first integral for the ODE obtained by symmetry reduction of the given PDE when symmetry-invariant solutions u(x, t) are sought. In [14–16] the relationship between symmetries and conservation laws has been used to find a double reduction of partial differential equations with two independent variables. This provides a direct reduction of order of the ODE.

In [17], we have applied Lie classical method to Eq. (8.3) in which the dissipative term has been replaced for a weaker one. We have proved that the equation is nonlinearly self-adjoint and have derived some conservation laws for this equation by using a special method that has been introduced by Ibragimov in [8]. This method, which does not require the existence of Lagrangians, is based on the concept of adjoint equations for nonlinear equations and avoids the integrals of functions. In [18] we have applied Lie classical method to the generalized equation

$$u_{tt} - u_{txx} + u_{xxxx} - (f(u))_{xx} = g(x), \tag{8.5}$$

where the second term is responsible for strong internal damping [6], in order to obtain exact solutions. We have determined the point symmetry group in terms of the functions $f(u)$ and $g(x)$ with $f(u) \neq constant$ and $g(x) \neq 0$. We have also derived conservation laws for this equation by using the conservation laws multipliers method.

8.2 Lie Classical Symmetries and Reductions

In this section, we perform Lie symmetry analysis for Eqs. (8.4) and (8.5), with $k \neq 0$, $f(u) \neq constant$, and $g(x) \neq 0$. Let us consider a one-parameter Lie group of infinitesimal transformations in (x, t, u) given by

$$\begin{aligned} x^* &= x + \varepsilon\xi(x, t, u) + \mathcal{O}(\varepsilon^2), \\ t^* &= t + \varepsilon\tau(x, t, u) + \mathcal{O}(\varepsilon^2), \\ u^* &= u + \varepsilon\phi(x, t, u) + \mathcal{O}(\varepsilon^2), \end{aligned} \tag{8.6}$$

where ε is the group parameter. Then one requires that this transformation leaves invariant the solutions of Eqs. (8.4) and (8.5). This leads to the overdetermined, linear system of twelve equations for the infinitesimals $\xi(x, t, u)$, $\tau(x, t, u)$, and $\phi(x, t, u)$. The associated Lie algebra of infinitesimal symmetries is the set of vector fields of the form

$$\mathbf{v} = \xi\frac{\partial}{\partial x} + \tau\frac{\partial}{\partial t} + \phi\frac{\partial}{\partial u}. \tag{8.7}$$

8.2.1 Lie Symmetries and Reductions for Eq. (8.4)

From the determining system for Eq. (8.4), we get that $\xi = \xi(x)$, $\tau = \tau(t)$, and $\phi = \alpha(x, t)u + \beta(x, t)$ where ξ, τ, ϕ, g, and f must satisfy the following equations:

$$2\,\alpha_x - 3\,\xi_{xx} = 0, \tag{8.8}$$

$$\tau_t - 2\,\xi_x = 0, \tag{8.9}$$

$$-\tau_{tt} + k\tau_t + 2\,\alpha_t = 0, \tag{8.10}$$

$$f_{uu}\,(\alpha\,u + \beta) + 2\,\tau_t\,f_u - 2\,\xi_x\,f_u + 4\,\xi_{xxx} - 6\,\alpha_{xx} = 0, \tag{8.11}$$

$$f_{uuu}\,(\alpha\,u + \beta) + 2\,\tau_t\,f_{uu} - 2\,\xi_x\,f_{uu} + \alpha\,f_{uu} = 0, \tag{8.12}$$

$$2\,f_{uu}\,(\alpha_x\,u + \beta_x) - \xi_{xx}\,f_u + 2\,\alpha_x\,f_u + \xi_{xxxx} - 4\,\alpha_{xxx} = 0, \tag{8.13}$$

$$\begin{aligned} &-f_u\,(\alpha_{xx}\,u + \beta_{xx}) + k\,(\alpha_t\,u + \beta_t) + \alpha_{xxxx}\,u + \alpha_{tt}\,u - \xi\,g_x \\ &\quad -2\,\tau_t\,g + \alpha\,g + \beta_{xxxx} + \beta_{tt} = 0. \end{aligned} \tag{8.14}$$

Solving Eqs. (8.8)–(8.10) we get

$$\xi = k_1 x + k_2,\ \tau = 2k_1 t + k_3,\ \alpha = -kk_1 + k_4$$

Substituting into Eq. (8.13) we obtain the following condition $f_{uu}\beta_x = 0$. Consequently we distinguish the following cases:

Case 1 If $f \neq au + b$, the following conditions must be satisfied:

$$\begin{aligned}
&-f_{u\,u}\ ((k_4 - k\,k_1\,t)\ u + \beta) - 2\,f_u\,k_1 = 0,\\
&f_{u\,u\,u}\ ((k_4 - k\,k_1\,t)\ u + \beta) + f_{u\,u}\ (k_4 - k\,k_1\,t) + 2\,f_{u\,u}\,k_1 = 0,\\
&-g_x\ (k_1\,x + k_2) + k\ (\beta_t - k\,k_1\,u) + g\ (k_4 - k\,k_1\,t) - 4\,g\,k_1 + \beta_{t\,t} = 0.
\end{aligned}$$

We obtain:

- Case 1a: For f arbitrary and g arbitrary with $k \neq 0$ we obtain that infinitesimals are

$$\xi = 0,\ \tau = k_3,\ \phi = 0.$$

- Case 1b: For f arbitrary and $g = constant$ with $k \neq 0$ we obtain that the infinitesimals are

$$\xi = k_2,\ \tau = k_3,\ \phi = 0.$$

- Case 1c: For $f = au^n + b$ and $g = c\,x^{-\frac{2(2n-1)}{n-1}}$ with $n \neq 1$ and $k = 0$ we obtain that infinitesimals are

$$\xi = k_1 x,\ \tau = 2k_1 t + k3,\ \phi = -\tfrac{2k_1}{n-1}u.$$

Case 2 If $f = au + b$, the following conditions must be satisfied:

$$\begin{aligned}
&-2\,a\,k_1 = 0,\\
&-g_x\ (k_1\,x + k_2) + k\ (\beta_t - k\,k_1\,u) + g\ (k_4 - k\,k_1\,t) - 4\,g\,k_1\\
&\quad -\beta_{x\,x}\,a + \beta_{x\,x\,x\,x} + \beta_{t\,t} = 0.
\end{aligned}$$

For $f = au + b$ and $g = g(x)$ arbitrary the infinitesimals are

$$\xi = k_2,\ \tau = k_3,\ \phi = k_4 u + \beta(x, t),$$

where for any $g = g(x)$, $\beta = \beta(x, t)$ must satisfy the following condition:

$$g\,k_4 - g_x\,k_2 + \beta_t\,k - \beta_{x\,x}\,a + \beta_{x\,x\,x\,x} + \beta_{t\,t} = 0. \tag{8.15}$$

Table 8.1 Functions and generators for Eq. (8.4)

i	f	g	$\mathbf{v}_k$	
1	*Arbitrary*	*Arbitrary*	$\mathbf{v}_2$	$(k \neq 0)$
2	*Arbitrary*	*Constant*	$\mathbf{v}_1, \mathbf{v}_2$	$(k \neq 0)$
3	$au^n + b4$	$c\,x^{-\frac{2(2n-1)}{n-1}}$	$\mathbf{v}_2, \mathbf{v}_3 = x\partial_x + 2t\partial_t - \frac{2}{n-1}u\partial_u \quad n \neq 1$	$(k = 0)$
4	$au + b$	*Arbitrary*	$\mathbf{v}_1 = \partial_x, \mathbf{v}_2 = \partial_t, \mathbf{v}_4 = (u + \beta(x,t))\partial_u$	

The functional forms of $f(u)$ and $g(x)$ as well as the corresponding generators are given in Table 8.1.
with β must satisfy Eq. (8.15). We use the method of characteristics to determine the invariants and reduced ODEs corresponding to each generator given in Table 8.1.

We obtain the following reduced ODEs for Eq. (8.4) setting $k = 1$ without loss of generality:

Reduction 1 For f and g arbitrary, by using generator $\mathbf{v}_2$ we obtain the similarity variable and similarity solution

$$z = x,\; u = h(z), \tag{8.16}$$

and the ODE_1

$$h_{zzzz} - f_h\, h_{zz} - f_{hh}\, (h_z)^2 - g(z) = 0. \tag{8.17}$$

Reduction 2 For f arbitrary and $g = c = constant$, by using generator $\lambda\mathbf{v}_1 + \mu\mathbf{v}_2$, we obtain the similarity variable and similarity solution

$$z = \mu x - \lambda t,\; u = h(z), \tag{8.18}$$

and the ODE_2

$$h_{zzzz}\,\mu^4 - f_h\, h_{zz}\,\mu^2 - f_{hh}\, (h_z)^2\, \mu^2 + \lambda^2\, h_{zz} - \lambda\, h_z - c = 0. \tag{8.19}$$

Reduction 3 For $f = au + b$ and $g = c = constant$, by using generator $\lambda\mathbf{v}_1 + \mu\mathbf{v}_2$, we obtain the similarity variable and similarity solution (8.18) and the ODE_3

$$h_{zzzz}\,\mu^4 - a\, h_{zz}\,\mu^2 + \lambda^2\, h_{zz} - \lambda\, h_z - c = 0. \tag{8.20}$$

and integrating once with respect to z we get

$$-z\,c + h_{zzz}\,\mu^4 + h_z\left(\lambda^2 - a\,\mu^2\right) - \lambda\, h + c_1 = 0$$

whose solution is

$$h(z) = \frac{\lambda c_1 - c\lambda^2 + c a\mu^2 - zc\lambda}{\lambda^2} + c_2 e^{\frac{1}{6}\frac{\left(\omega^{\frac{2}{3}} + 12\lambda^2 - 12a\mu^2\right)z}{\mu^2\sqrt[3]{\omega}}}$$

$$+ c_3 e^{\frac{-\frac{1}{12}i\left(-i\omega^{\frac{2}{3}} - 12i\lambda^2 + 12ia\mu^2 + \sqrt{3}\omega^{\frac{2}{3}} - 12\sqrt{3}\lambda^2 + 12\sqrt{3}a\mu^2\right)z}{\mu^2\sqrt[3]{\omega}}}$$

$$+ c_4 e^{\frac{\frac{1}{12}i\left(i\omega^{\frac{2}{3}} + 12i\lambda^2 - 12ia\mu^2 + \sqrt{3}\omega^{\frac{2}{3}} - 12\sqrt{3}\lambda^2 + 12\sqrt{3}a\mu^2\right)z}{\mu^2\sqrt[3]{\omega}}}$$

with

$$\omega = \left(108\lambda\mu^2 + 12\sqrt{3}\sqrt{4\lambda^6 - 12\lambda^4 a\mu^2 + 12\lambda^2 a^2\mu^4 - 4a^3\mu^6 + 27\lambda^2\mu^4}\right)$$

and $c_i,\ i = 1, 2, 3, 4$ constants.

For Eq. (8.4) with $k = 0$, we obtain the following reductions:

Reduction 4 For $f = au + b$ and $g = c = constant$, by using generator $\lambda\mathbf{v}_1 + \mu\mathbf{v}_2$, we obtain the similarity variable and similarity solution (8.18) and ODE$_4$

$$h_{zzzz}\mu^4 - a h_{zz}\mu^2 + \lambda^2 h_{zz} - c = 0. \tag{8.21}$$

Integrating once with respect to z we get

$$-zc + h_{zzz}\mu^4 + h_z\left(\lambda^2 - a\mu^2\right) + c_1 = 0,$$

whose solution is

$$h(z) = \mu^2\left(\lambda^2 - a\mu^2\right)^{\frac{-1}{2}}\left[c_3\sin\left(\frac{\sqrt{\lambda^2 - a\mu^2}z}{\mu^2}\right) - c_2\cos\left(\frac{\sqrt{\lambda^2 - a\mu^2}z}{\mu^2}\right)\right]$$

$$+ \frac{z^2c - 2zc_1}{2\left(\lambda^2 - a\mu^2\right)} + c_4$$

with $\lambda^2 - a\mu^2 > 0$ and $c_i,\ i = 1, 2, 3, 4$ arbitrary constants.

Reduction 5 For $f = au^n + b$ and $g = c\,x^{\frac{-2(2n-1)}{n-1}}$, by using generator $\mathbf{v}_2$, we obtain the similarity variable and similarity solution

$$z = x,\ u = h(z), \tag{8.22}$$

and ODE$_5$

$$h_{zzzz} + a\,n\,h^{n-1}\,h_{zz} + (n-1)\,a\,h^{n-2}\,(h_z)^2 + c\,z^{\frac{-2(2n-1)}{n-1}} = 0. \tag{8.23}$$

Reduction 6 For $f = au^n + b$ and $g = c\,x^{\frac{-2(2n-1)}{n-1}}$ by using generator $\mathbf{v}_3$, we obtain the similarity variable and similarity solution

$$z = \tfrac{x}{\sqrt{t}},\ u = \frac{h(z)}{t^{\frac{1}{n-1}}}, \tag{8.24}$$

and ODE$_6$

$$\begin{aligned}
&4\,h^2\,(n-1)^2\,z^{\frac{4n+1}{n-1}}\,h_{zzzz} + h\,(n-1)^2\left(h\,z^2 - 4\,a\,h^n\,n\right)z^{\frac{4n+1}{n-1}}\,h_{zz}\\
&\quad - 4\,a\,h^n\,(n-1)^3\,n\,z^{\frac{4n+1}{n-1}}\,(h_z)^2 + h^2\left(3\,n^2 - 2n - 1\right)z^{\frac{5n}{n-1}}\,h_z\\
&\quad + 4\,h^3\,n\,z^{\frac{4n+1}{n-1}} - 4\,h^2\,z^{\frac{3}{n-1}}\left(n^2 - 2\,n + c\right) = 0.
\end{aligned}$$

8.2.2 Lie Symmetries and Reductions for Eq. (8.5)

From the determining system for Eq. (8.5), we get that $\xi = \xi(x,t)$, $\tau = \tau(t)$, and $\phi = \alpha(x,t)u + \beta(x,t)$ where ξ, τ, ϕ, g, and f must satisfy the following equations:

$$\begin{aligned}
&-6\,\xi_{xx} + \xi_t + 4\,\alpha_x = 0,\\
&-\xi_{xx} + 2\,\xi_t + 2\,\alpha_x = 0,\\
&\tau_t - 2\,\xi_x = 0,\\
&-\tau_{tt} - \alpha_{xx} + 2\,\alpha_t = 0,\\
&f_{uu}\,(\alpha\,u + \beta) + 2\,\tau_t\,f_u - 2\,\xi_x\,f_u + 4\,\xi_{xxx} - 2\,\xi_{tx} - 6\,\alpha_{xx} + \alpha_t = 0,\\
&f_{uuu}\,(\alpha\,u + \beta) + 2\,\tau_t\,f_{uu} - 2\,\xi_x\,f_{uu} + \alpha\,f_{uu} = 0,\\
&2\,f_{uu}\,(\alpha_x\,u + \beta_x) - \xi_{xx}\,f_u + 2\,\alpha_x\,f_u + \xi_{xxxx} + \xi_{tt} - \xi_{txx}\\
&\quad -4\,\alpha_{xxx} + 2\,\alpha_{tx} = 0,\\
&u\,\alpha_{xxxx} + \beta_{xxxx} - f_u\,(u\,\alpha_{xx} + \beta_{xx}) - \xi\,g_x - \alpha_{txx}\,u + \alpha_{tt}\,u\\
&\quad -\beta_{txx} + \beta_{tt} - 2\,g\,\tau_t + \alpha\,g = 0.
\end{aligned}$$

From the determining equations we get

$$\xi = k_1\frac{x}{2} + k_3,\ \tau = k_1 t + k_2,\ \phi = k_4 u + \beta(x,t),$$

where f, g, and β must satisfy

$$\begin{aligned}
&2\,\beta_x\, f_{u\,u} = 0,\\
&f_{u\,u}\, k_4\, u + f_u\, k_1 + \beta\, f_{u\,u} = 0,\\
&f_{u\,u\,u}\, k_4\, u + f_{u\,u}\, k_4 + f_{u\,u}\, k_1 + \beta\, f_{u\,u\,u} = 0,\\
&-g_x\, k_1\, x + 2\, g\, k_4 - 2\, g_x\, k_3 - 4\, g\, k_1 - 2\, \beta_{x\,x}\, f_u + 2\, \beta_{x\,x\,x\,x}\\
&\quad +2\, \beta_{t\,t} - 2\, \beta_{t\,x\,x} = 0.
\end{aligned}$$

From the second condition we get that $\beta_x = 0$ or $f(u)$ is a linear function.

Case 1 If $\beta_x = 0$:
By substituting $\beta = \beta(t)$ into the remaining conditions we get $\beta = k_5$ where f and g must satisfy

$$f_{u\,u}\, k_4\, u + f_{u\,u}\, k_5 + f_u\, k_1 = 0, \tag{8.25}$$

$$-g_x\, k_1\, x + 2\, g\, k_4 - 2\, g_x\, k_3 - 4\, g \; = 0. \tag{8.26}$$

We distinguish

- Case 1a: For f arbitrary and g arbitrary we obtain that infinitesimals are

$$\xi = 0,\ \tau = k_2,\ \phi = 0.$$

- Case 1b: For f arbitrary and $g = constant$ with we obtain that the infinitesimals are

$$\xi = k_3,\ \tau = k_2,\ \phi = 0.$$

- Case 1c: For $f(u) = (c_1\, u + c_2)^n$ and $g(x) = c_3\, x^m$ we obtain that infinitesimals are

$$\xi = \frac{k_4(1-n)}{2}x,\ \tau = k_4(1-n)t + \frac{k_4 c_2}{c_1} + k_2,\ \phi = k_4 u + \frac{k_4 c_2}{c_1}$$

  with $m = -\frac{2(2n-1)}{n-1}$.
- Case 1d: For $f(u) = c_1\, e^{n\,u} + c_2$ and $g(x) = c_3\, x^m$ we obtain that infinitesimals are

$$\xi = -\frac{nk_6}{2}x,\ \tau = -nk_6 t + k_2,\ \phi = k_6$$

  with $m = -4$.

Table 8.2 Functions and generators for Eq. (8.5)

i	f	g	$\mathbf{v}_k$
1	*Arbitrary*	*Arbitrary*	$\mathbf{v}_2 = \partial_t$
2	*Arbitrary*	*Constant*	$\mathbf{v}_1 = \partial_x, \mathbf{v}_2$
3	$(c_1 u + c_2)^n$	$c_3 x^{-\frac{2(2n-1)}{n-1}}$	$\mathbf{v}_2$, $\mathbf{v}_3 = \frac{(1-n)}{2} x\partial_x + \left((1-n)t + \frac{c_2}{c_1}\right)\partial_t + \left(u + \frac{c_2}{c_1}\right)\partial_u$ $n \neq 1$
4	$c_1 e^{nu} + c_2$	$(c_3 x + c_4)^{-4}$	$\mathbf{v}_2, \mathbf{v}_4 = -\frac{n}{2} x\partial_x - nt\partial_t + \partial_u, \quad n \neq 0,\ c_3 \neq 0$
5	$au + b$	*Arbitrary*	$\mathbf{v}_2, \mathbf{v}_5 = k_3\partial_x + \beta(x,t)\partial_u$

Case 2 If $f = au + b$, the following conditions must be satisfied:

$$2\, a\, k_1 = 0,$$
$$\beta_{xxxx} - a\,\beta_{xx} - k_1\, x\, g_x - k_3\, g_x - \beta_{txx} + \beta_{tt} - 4\, g\, k_1 = 0.$$

For $f = au + b$ and $g = g(x)$ arbitrary the infinitesimals are

$$\xi = k_3,\ \tau = k_2,\ \phi = \beta(x,t),$$

where for any $g = g(x)$, $\beta = \beta(x,t)$ must satisfy the following condition:

$$-\, g_x\, k_3 - \beta_{x\,x}\, a + \beta_{x\,x\,x\,x} + \beta_{t\,t} - \beta_{t\,x\,x} = 0. \tag{8.27}$$

The functional forms of $f(u)$ and $g(x)$ as well as the corresponding generators are given in Table 8.2.
with β must satisfy Eq. (8.27).
We use the method of characteristics to determine the invariants and reduced ODEs corresponding to each generator given in Table 8.2. We obtain the following reduced ODEs for Eq. (8.5)

Reduction 1 For f and g arbitrary, by using generator $\mathbf{v}_2$ we obtain the similarity variable and similarity solution

$$z = x,\ u = h(z), \tag{8.28}$$

and the ODE$_1$

$$h_{z\,z\,z\,z} - f_h\, h_{z\,z} - h_{z\,z} - f_{h\,h}\ (h_z)^2 - g(z) = 0. \tag{8.29}$$

Reduction 2 For f arbitrary and $g = c = constant$, by using generator $\lambda\mathbf{v}_1 - \mu\mathbf{v}_2$, we obtain the similarity variable and similarity solution

$$z = \mu x + \lambda t,\ u = h(z), \tag{8.30}$$

and the ODE_2

$$h_{zzzz}\,\mu^4 + (\lambda^2 - f_h\,\mu^2 - \mu^2)h_{zz} - f_{hh}\,(h_z)^2\,\mu^2 - c = 0. \tag{8.31}$$

Reduction 3 Setting without loss of generality $f = u^n$ and $g = x^{\frac{2-4n}{n-1}}$ by using generator

$$\mathbf{v}_3 = \frac{x}{2}\frac{\partial}{\partial x} + t\frac{\partial}{\partial t} - \frac{u}{n-1}\frac{\partial}{\partial u},$$

we obtain the similarity variable and similarity solution

$$z = \frac{x}{\sqrt{t}},\ u = t^{\frac{1}{n-1}}h(z), \tag{8.32}$$

and the ODE_3

$$-4\,h_{zzzz} - 2\,h_{zzz}\,z + h_{zz}\left(-z^2 + 4\,h^{n-1}n - \frac{4}{n-1} - 4\right) + h_z\left(-\frac{4\,z}{n-1} - 3\,z\right)$$
$$+\,4\,h^{n-2}\,(h_z)^2\,(n-1)\,n - \frac{4\,h\,n}{(n-1)^2} + 4\,z^{\frac{2-4n}{n-1}} = 0. \tag{8.33}$$

Reduction 4 For $f = e^{nu}$ and $g = x^{-4}$, by using generator

$$\mathbf{v}_4 = \frac{x}{2}\frac{\partial}{\partial x} + t\frac{\partial}{\partial t} - \frac{1}{n}\frac{\partial}{\partial u},$$

we obtain the similarity variable and similarity solution

$$z = \frac{x}{\sqrt{t}},\ u = -\frac{1}{n}\log(t)h(z), \tag{8.34}$$

and ODE_4

$$4\,h_{zzzz}\,n\,z^4 + 2\,h_{zzz}\,n\,z^5 + h_{zz}\,n\,z^4\left(z^2 - 4\,n\,e^{h\,n} + 4\right) + 3\,h_z\,n\,z^5$$
$$-\,4\,(h_z)^2\,n^3\,e^{h\,n}\,z^4 + 4\,z^4 - 4\,n = 0. \tag{8.35}$$

Reduction 5 For $f = au + b$ we obtain the similarity variable and similarity solution

$$z = \mu x + \lambda t,\ u = h(z) + \delta(x, t) \tag{8.36}$$

and ODE_5

$$h_{zzzz}\,\mu^4 - h_{zzz}\,\mu^2\,\lambda + (\lambda^2 - a\,\mu^2)\,h_{zz} = 0,$$

where $g = g(x)$ arbitrary and $\delta = \delta(x,t) = \int \beta(x,t)\,dt$ must satisfy

$$-g(x) + \delta_{xxxx} - a\,\delta_{xx} + \delta_{tt} - \delta_{txx} = 0.$$

Integrating twice with respect to z we get the linear ODE

$$h_{zz}\,\mu^4 + h\left(\lambda^2 - a\,\mu^2\right) - \lambda\,h_z\,\mu^2 + z\,c_1 + c_2 = 0$$

whose solution is

$$h(z) = \mathrm{e}^{\frac{z\left(\lambda+\sqrt{-3\lambda^2+4a^2\mu^2}\right)}{2\mu^2}}c_3 + \mathrm{e}^{-\frac{z\left(-\lambda+\sqrt{-3\lambda^2+4a^2\mu^2}\right)}{2\mu^2}}c_4 + \frac{(-c1\,z - c2)\,\lambda^2 - \lambda\,\mu^2 c1 + a^2\mu^2\,(c2 + c1\,z)}{\left(-\lambda^2 + a^2\mu^2\right)^2}.$$

with $-3\,\lambda^2 + 4\,a^2\mu^2 > 0$ and c_i, $i = 1, \ldots, 4$ arbitrary constants.

8.3 Multiplier Conservation Laws Method

In [11] Anco and Bluman gave a general treatment of a direct conservation law method for partial differential equations expressed in the normal form. An Nth-order PDE is *normal* if it can be expressed in a solved form for some leading derivative of u such that all the other terms in the equation contain neither the leading derivative nor its differential consequences [16]. For Eqs. (8.4) and (8.5) the nontrivial conservation laws are characterized by a multiplier Λ with no dependence on u_{tt} satisfying

$$\hat{E}[u]\,(\Lambda u_{tt} - \Lambda G(x, u, u_t, u_x, u_{xx}, \ldots, u_{nx})) = 0. \tag{8.37}$$

Here

$$\hat{E}[u] := \frac{\partial}{\partial u} - D_t\frac{\partial}{\partial u_t} - D_x\frac{\partial}{\partial u_x} + D_x^2\frac{\partial}{\partial u_{xx}} + \cdots$$

is the Euler operator. The conservation law will be written

$$D_t(\Phi^t) + D_x(\Phi^x)\,|_\triangle = 0,$$

where Φ^t and Φ^x are called the conserved densities. The conserved current must satisfy

$$\Lambda = \hat{E}[u]\Phi^t \tag{8.38}$$

and the flux Φ^x is given by Euler and Euler [19]

$$\Phi^x = -D_x^{-1}(\Lambda G) - \frac{\partial \Phi^t}{\partial u_x} G + G D_x \left(\frac{\partial \Phi^t}{\partial u_{xx}} \right) + \cdots . \tag{8.39}$$

8.3.1 Multipliers and Conservation Laws for Eq. (8.4)

For Eq. (8.4) setting $k = 1$ without loss of generality, we can state:

Case 1 For $f(u)$ arbitrary and $g(x)$ arbitrary, we get the following multipliers:

$$\Lambda_1 = 1,\ \Lambda_2 = x,\ \Lambda_3 = e^t,\ \Lambda_4 = xe^t. \tag{8.40}$$

Case 2 For $f(u) = u + d$ and $g(x)$ arbitrary, applying Eq. (8.37) we get

$$\Lambda_{xxxx} = \Lambda_{xx} + \Lambda_t - \Lambda_{tt}, \tag{8.41}$$

$$\Lambda_u = 0, \tag{8.42}$$

searching for solutions of the form

$$\Lambda = X(x)T(t), \tag{8.43}$$

where $X(x)$ and $T(t)$ satisfy

$$X^{IV} = c_1 X + X'',\ T'' = -c_1 T + T',$$

then we obtain the following multipliers with $\delta_1 = \sqrt{1 - 4c_1}$, $\delta_2 = \sqrt{1 + 4c_1}$ and $c_1 \in \left[-\frac{1}{4}, \frac{1}{4}\right]$:

$$\begin{array}{lll} \Lambda_1 = 1, & \Lambda_2 = x, & \Lambda_3 = e^t, \\ \Lambda_4 = xe^t, & \Lambda_5 = \mathrm{e}^{\frac{1}{2}\left(t+t\delta_1-\sqrt{2-2\delta_2}x\right)}, & \Lambda_6 = \mathrm{e}^{\frac{1}{2}\left(t+t\delta_1+\sqrt{2-2\delta_2}x\right)}, \\ \Lambda_7 = \mathrm{e}^{\frac{1}{2}\left(t+t\delta_1-\sqrt{2+2\delta_2}x\right)}, & \Lambda_8 = \mathrm{e}^{\frac{1}{2}\left(t+t\delta_1+\sqrt{2+2\delta_2}x\right)}, & \Lambda_9 = \mathrm{e}^{\frac{1}{2}\left(t-t\delta_1-\sqrt{2-2\delta_2}x\right)}, \\ \Lambda_{10} = \mathrm{e}^{\frac{1}{2}\left(t-t\delta_1+\sqrt{2-2\delta_2}x\right)}, & \Lambda_{11} = \mathrm{e}^{\frac{1}{2}\left(t-t\delta_1-\sqrt{-2+2\delta_2}x\right)}, & \Lambda_{12} = \mathrm{e}^{\frac{1}{2}\left(t-t\delta_1+\sqrt{2+2\delta_2}x\right)}. \end{array} \tag{8.44}$$

Table 8.3 $f(u)$ and $g(x)$ are arbitrary

Multiplier	Conserved density	Flux
$\Lambda_1 = 1$	$\phi^t = u_t + u$	$\phi^x = u_{xxx} - \left(\frac{d}{du} f(u)\right) u_x - \int g(x)\, dx$
$\Lambda_2 = x$	$\phi^t = x(u_t + u)$	$\phi^x = x u_{xxx} - u_{xx} - x\left(\frac{d}{du} f(u)\right) u_x + f(u) - \int x g(x)\, dx$
$\Lambda_3 = e^t$	$\phi^t = \mathrm{e}^t (u_t - g(x))$	$\phi^x = -\mathrm{e}^t \left(-u_{xxx} + \left(\frac{d}{du} f(u)\right) u_x\right)$
$\Lambda_4 = x e^t$	$\phi^t = \mathrm{e}^t x (u_t - g(x))$	$\phi^x = \mathrm{e}^t \left(x u_{xxx} - u_{xx} - x\left(\frac{d}{du} f(u)\right) u_x + f(u)\right)$

Conservation Laws

Case 1 For $f(u)$ arbitrary and $g(x)$ arbitrary
Associated with the multipliers, from (8.38) and (8.39), we obtain the corresponding conserved densities and fluxes given for Eq. (8.4) in the following Table 8.3.

Case 2 For $f(u) = cu + d$, $g(x)$ arbitrary, $\delta_1 = \sqrt{1 - 4c_1}$ and $\delta_2 = \sqrt{1 + 4c_1}$
Associated with the multipliers, from (8.38) and (8.39), we obtain the corresponding conserved densities and fluxes:

1.

$$\begin{aligned}
\Lambda_5 &= \mathrm{e}^{\frac{1}{2}\left(t + t\delta_1 - \sqrt{2 - 2\delta_2}x\right)}, \\
\phi^t &= \frac{1}{2c_1}\left(g(-1 + \delta_1)\, \mathrm{e}^{-\frac{1}{2}\sqrt{2-2\delta_2}x + \frac{1}{2}t(1+\delta_1)} - \mathrm{e}^{\frac{1}{2}\left(t + t\delta_1 - \sqrt{2-2\delta_2}x\right)}\right. \\
&\quad \left. \times\, c_1 (-2u_t - u + u\delta_1)\right), \\
\phi^x &= \frac{\mathrm{e}^{\frac{1}{2}\left(t + t\delta_1 - \sqrt{2-2\delta_2}x\right)}}{4}\left(-4u_{xxx} - 2\sqrt{2 - 2\delta_2}u_{xx} + 2u_x + 2u_x\delta_2\right. \\
&\quad \left. + u\sqrt{2 - 2\delta_2}\,(1 + \delta_2)\right).
\end{aligned}$$

2.

$$\begin{aligned}
\Lambda_6 &= \mathrm{e}^{\frac{1}{2}\left(t + t\delta_1 + \sqrt{2 - 2\delta_2}x\right)}, \\
\phi^t &= \frac{1}{2c_1}\left(g(-1 + \delta_1)\, \mathrm{e}^{\frac{1}{2}\sqrt{2-2\delta_2}x + \frac{1}{2}t(1+\delta_1)} - \mathrm{e}^{\frac{1}{2}\left(t + t\delta_1 + \sqrt{2-2\delta_2}x\right)}\right. \\
&\quad \left. \times\, c_1 (-2u_t - u + u\delta_1)\right), \\
\phi^x &= \frac{\mathrm{e}^{\frac{1}{2}\left(t + t\delta_1 + \sqrt{2-2\delta_2}x\right)}}{4}\left(4u_{xxx} - 2\sqrt{2 - 2\delta_2}u_{xx} - 2u_x - 2u_x\delta_2\right. \\
&\quad \left. + u\sqrt{2 - 2\delta_2}\,(1 + \delta_2)\right).
\end{aligned}$$

3.

$$\Lambda_7 = e^{\frac{1}{2}\left(t+t\delta_1-\sqrt{2+2\delta_2}x\right)},$$

$$\phi^t = \frac{1}{2c_1}\Big(g\,(-1+\delta_1)\,e^{-\frac{1}{2}\sqrt{2+2\delta_2}x+\frac{1}{2}t(1+\delta_1)} - e^{\frac{1}{2}\left(t+t\delta_1-\sqrt{2+2\delta_2}x\right)} \times c_1\,(-2u_t - u + u\delta_1)\Big),$$

$$\phi^x = \frac{e^{\frac{1}{2}\left(t+t\delta_1-\sqrt{2+2\delta_2}x\right)}}{4}\Big(4u_{xxx} + 2\sqrt{2+2\delta_2}u_{xx} - 2u_x + 2u_x\delta_2 + u\sqrt{2+2\delta_2}\,(-1+\delta_2)\Big).$$

4.

$$\Lambda_8 = e^{\frac{1}{2}\left(t+t\delta_1+\sqrt{2+2\delta_2}x\right)},$$

$$\phi^t = \frac{1}{2c_1}\Big(g\,(-1+\delta_1)\,e^{\frac{1}{2}\sqrt{2+2\delta_2}x+\frac{1}{2}t(1+\delta_1)} - e^{\frac{1}{2}\left(t+t\delta_1+\sqrt{2+2\delta_2}x\right)} \times c_1\,(-2u_t - u + u\delta_1)\Big),$$

$$\phi^x = \frac{-e^{\frac{1}{2}\left(t+t\delta_1+\sqrt{2+2\delta_2}x\right)}}{4}\Big(-4u_{xxx} + 2\sqrt{2+2\delta_2}u_{xx} + 2u_x - 2u_x\delta_2 + u\sqrt{2+2\delta_2}\,(-1+\delta_2)\Big).$$

5.

$$\Lambda_9 = e^{\frac{1}{2}\left(t-t\delta_1-\sqrt{2-2\delta_2}x\right)},$$

$$\phi^t = \frac{1}{2c_1}\Big(-g\,(1+\delta_1)\,e^{-\frac{1}{2}\sqrt{2-2\,\delta_2}x-\frac{1}{2}t(-1+\delta_1)} + e^{\frac{1}{2}\left(t-t\delta_1-\sqrt{2-2\delta_2}x\right)} \times c_1\,(2u_t + u + u\delta_1)\Big),$$

$$\phi^x = \frac{-e^{\frac{1}{2}\left(t-t\delta_1-\sqrt{2-2\delta_2}x\right)}}{4}\Big(-4u_{xxx} - 2\sqrt{2-2\delta_2}u_{xx} + 2u_x + 2u_x\delta_2 + u\sqrt{2-2\delta_2}\,(1+\delta_2)\Big).$$

6.

$$\Lambda_{10} = e^{\frac{1}{2}\left(t-t\delta_1+\sqrt{2-2\delta_2}x\right)},$$

$$\phi^t = \frac{1}{2c_1}\Big(-g\,(1+\delta_1)\,e^{\frac{1}{2}\sqrt{2-2\delta_2}x-\frac{1}{2}t(-1+\delta_1)} + e^{\frac{1}{2}\left(t-t\delta_1+\sqrt{2-2\delta_2}x\right)}$$

$$\times c_1 \left(2u_t + u + u\delta_1\right)\Big),$$

$$\phi^x = \frac{e^{\frac{1}{2}\left(t - t\delta_1 + \sqrt{2-2\delta_2}x\right)}}{4}\Big(4u_{xxx} - 2\sqrt{2-2\delta_2}u_{xx} - 2u_x - 2u_x\delta_2$$
$$+ u\sqrt{2-2\delta_2}\left(1+\delta_2\right)\Big).$$

7.

$$\Lambda_{11} = e^{\frac{1}{2}\left(t - t\delta_1 - \sqrt{2+2\delta_2}x\right)},$$

$$\phi^t = \frac{1}{2c_1}\Big(-g\left(1+\delta_1\right)e^{-\frac{1}{2}\sqrt{2+2\delta_2}x - \frac{1}{2}t(-1+\delta_1)} + e^{\frac{1}{2}\left(t - t\delta_1 - \sqrt{2+2\delta_2}x\right)}$$
$$\times c_1 \left(2u_t + u + u\delta_1\right)\Big),$$

$$\phi^x = \frac{e^{\frac{1}{2}\left(t - t\delta_1 - \sqrt{2+2\delta_2}x\right)}}{4}\Big(4u_{xxx} + 2\sqrt{2+2\delta_2}u_{xx} - 2\,u_x + 2\,u_x\delta_2$$
$$+ u\sqrt{2+2\delta_2}\left(-1+\delta_2\right)\Big).$$

8.

$$\Lambda_{12} = e^{\frac{1}{2}\left(t - t\delta_1 + \sqrt{2+2\delta_2}x\right)},$$

$$\phi^t = \frac{1}{2c_1}\Big(-g\left(1+\delta_1\right)e^{\frac{1}{2}\sqrt{2+2\delta_2}x - \frac{1}{2}t(-1+\delta_1)} + e^{\frac{1}{2}\left(t - t\delta_1 + \sqrt{2+2\delta_2}x\right)}$$
$$\times c_1 \left(2\,u_t + u + u\delta_1\right)\Big),$$

$$\phi^x = \frac{-e^{\frac{1}{2}\left(t - t\delta_1 + \sqrt{2+2\delta_2}x\right)}}{4}\Big(-4u_{xxx} + 2\,\sqrt{2+2\delta}u_{xx} + 2u_x - 2u_x\delta_2$$
$$+u\sqrt{2+2\delta_2}\left(-1+\delta_2\right)\Big).$$

8.3.2 Multipliers and Conservation Laws for Eq. (8.5)

For Eq. (8.5), we can state:

Case 1 For $f(u)$ arbitrary and $g(x)$ arbitrary, we get the following multipliers:

$$\Lambda_1 = 1,\ \Lambda_2 = x,\ \Lambda_3 = t,\ \Lambda_4 = tx. \tag{8.45}$$

Table 8.4 Conserved densities and fluxes for Eq. (8.5)

Multiplier	Conserved density	Flux
$\Lambda_1 = 1$	$\phi^t = -u_{xx} + u_t$	$\phi^x = u_{xxx} - f'(u)\, u_x$ $- \int g(x)\, dx$
$\Lambda_2 = x$	$\phi^t = x\,(-u_{xx} + u_t)$	$\phi^x = -\int xg(x)\, dx + u_{xxx}x - u_{xx}$ $+ f(u) - xf'(u)u_x$
$\Lambda_3 = t$	$\phi^t = -\dfrac{g(x)\, t^2}{2}$ $+ \dfrac{(-2\,u_{xx} + 2\,u_t)\, t}{2} - u$	$\phi^x = tu_{xxx} - u_x t f'(u) + u_x$
$\Lambda_4 = tx$	$\phi^t = -\frac{1}{2} g(x)\, x\, t^2$ $- (u_{xx} - u_t)\, xt - ux$	$\phi^x = -u_x t x f'(u) + tf(u)$ $+ (xu_{xxx} - u_{xx})\, t + u_x x - u$

Case 2 **For** $f(u) = u + d$ **and** $g(x)$ **arbitrary** we obtain the following multipliers:

$$\begin{array}{lll} \Lambda_1 = 1 & \Lambda_2 = x & \Lambda_3 = t \\ \Lambda_4 = tx & \Lambda_5 = \dfrac{x^2}{2c} + \dfrac{t^2}{2} & \Lambda_6 = \dfrac{xt^2}{2} + \dfrac{x^3}{6c} \\ \Lambda_7 = \dfrac{x^2 t}{2c} + \dfrac{t^3}{6} + \dfrac{x^2}{2c^2} & \Lambda_8 = \dfrac{xt^3}{6} + \dfrac{x^3 t}{6c} + \dfrac{x^3}{6c^2}. & \end{array} \tag{8.46}$$

Conservation Laws

Case 1 **For** $f(u)$ **arbitrary and** $g(x)$ **arbitrary**
Associated with the multipliers, from (8.38) and (8.39), we obtain the corresponding conserved densities and fluxes given in the following Table 8.4.

Case 2 **For** $f(u) = cu + d$ **and** $g(x)$ **arbitrary.**
Associated with the multipliers, from (8.38) and (8.39), we obtain the corresponding conserved densities and fluxes:

1.

$$\Lambda_5 = \frac{1}{2}\left(\frac{x^2}{c} + t^2\right),$$

$$\phi^t = \frac{\left(-3x^2 t - t^3 c\right) g(x) + 3t\,((-u_{xx} + u_t)\, t - 2u)\, c + 3x^2\,(-u_{xx} + u_t)}{6c},$$

$$\phi^x = \frac{-u_x t^2 c^2 + \left(\left(2t - x^2\right) u_x + u_{xxx} t^2 + 2xu\right) c + (2u_x + x\,(u_{xxx} x - 2u_{xx}))}{2c}.$$

2.

$$
\begin{aligned}
\Lambda_6 &= \frac{1}{2}xt^2 + \frac{1}{6}\frac{x^3}{c},\\
\phi^t &= \frac{x\left(\left(t^3c + x^2t\right)g\left(x\right) - 3\,t\left(\left(-u_{xx} + u_t\right)t - 2\,u\right)c - x^2\left(-u_{xx} + u_t\right)\right)}{6c},\\
\phi^x &= \frac{t^2\left(xu_{xxx} - u_{xx}\right)c + \frac{1}{3}x^3u_{xxx} - u_{xx}x^2}{2c}.
\end{aligned}
$$

3.

$$
\begin{aligned}
\Lambda_7 &= \frac{1}{2}\frac{tx^2}{c} + \frac{1}{6}t^3 + \frac{1}{6}\frac{x^2}{c^2},\\
\phi^t &= \frac{u_{xx}x^2}{2c^2} - \frac{u_{xx}x^2t}{2c} - \frac{u_{xx}t^3}{6} + \frac{u_tx^2}{2c^2} + \frac{u_tx^2t}{2c} + \frac{u_tt^3}{6} - \frac{ux^2}{2c} - \frac{ut^2}{2}\\
&\quad - \frac{x^2g\left(x\right)t}{2c^2} - \frac{x^2g\left(x\right)t^2}{4c} - \frac{t^4g\left(x\right)}{24},\\
\phi^x &= \frac{u_{xxx}x^2}{2c^2} + \frac{u_{xxx}x^2t}{2c} + \frac{u_{xxx}t^3}{6} - \frac{xu_{xx}}{c^2} - \frac{xu_{xx}t}{c} - \frac{u_xtx^2}{2} - \frac{cu_xt^3}{6},\\
&\quad + \frac{u_x}{c^2} + \frac{u_xt}{c} + \frac{u_xt^2}{2} + txu.
\end{aligned}
$$

4.

$$
\begin{aligned}
\Lambda_8 &= \frac{x^3}{6c^2} + \frac{x^3t}{6c} + \frac{xt^3}{6},\\
\Phi^t &= -\frac{x^3u_{xx}}{6c^2} - \frac{x^3tu_{xx}}{6c} - \frac{xu_{xx}t^3}{6} + \frac{x^3u_t}{6c^2} + \frac{x^3u_tt}{6c} + \frac{xu_tt^3}{6} - \frac{x^3u}{6c},\\
&\quad - \frac{xut^2}{2} - \frac{x^3tg\left(x\right)}{6c^2} - \frac{x^3g\left(x\right)t^2}{12c} - \frac{xt^4g\left(x\right)}{24},\\
\Phi^x &= \frac{x^3u_{xxx}}{6c^2} + \frac{x^3u_{xxx}t}{6c} + \frac{xu_{xxx}t^3}{6} - \frac{u_{xx}x^2}{2c^2} - \frac{u_{xx}x^2t}{2c} - \frac{u_{xx}t^3}{6} - \frac{x^3u_xt}{6}\\
&\quad - \frac{cxu_xt^3}{6} + \frac{xu_x}{c^2} + \frac{xu_xt}{c} + \frac{xu_xt^2}{2} + \frac{tux^2}{2} + \frac{cut^3}{6} - \frac{u}{c^2} - \frac{tu}{c} - \frac{ut^2}{2}.
\end{aligned}
$$

8.4 Double Reduction Method and Exact Solutions

A powerful application of conservation laws taking into account the relationship between Lie symmetries and conservation laws is the so-called double reduction method [14, 20]. This method allows us to reduce directly Eq. (8.5) to a third order ordinary differential equation. In [14] Sjöberg introduced a method in order to get solutions of a qth partial differential equation from the solutions of an ordinary differential equation of order $q-1$ called *double reduction method*. This method can be applied when a symmetry $\mathbf{v}$ is associated with a conserved vector T [14, 16]. Considering a qth partial differential equation given by

$$F(x, t, u, u_{(1)}, \dots u_{(q)} = 0 \tag{8.47}$$

which admits a Lie symmetry associated with the conserved vector $T = (T^t, T^x)$ [14], any conservation law associated with a symmetry generator $\mathbf{v}$ can be rewritten in canonical coordinates as

$$D_r T^r + D_s T^s = 0, \tag{8.48}$$

with

$$T^s = \frac{T^t D_t(s) + T^x D_x(s)}{D_t(r) D_x(s) - D_x(r) D_t(s)}, \tag{8.49}$$

and

$$T^r = \frac{T^t D_t(r) + T^x D_x(r)}{D_t(r) D_x(s) - D_x(r) D_t(s)}. \tag{8.50}$$

The components T^x, T^t depend on $\left(x, t, u, u_{(1)}, u_{(2)}, \dots, u_{(q-1)}\right)$ which means that T^s, T^r depend on $\left(r, s, w, w_r, w_{rr}, \dots, w_{r^{q-1}}\right)$ for solutions invariant with respect to $\mathbf{v}$. Therefore Eq. (8.48) becomes

$$\frac{\partial T^s}{\partial s} + D_r T^r = 0. \tag{8.51}$$

For T associated with $\mathbf{v}$ we obtain,

$$\mathbf{v}\left(T^r\right) \equiv \frac{\partial T^r}{\partial s} = 0, \tag{8.52}$$

$$\mathbf{v}\left(T^s\right) \equiv \frac{\partial T^s}{\partial s} = 0. \tag{8.53}$$

Thus, the conservation law in canonical coordinates becomes

$$D_r T^r = 0,$$

so that

$$T^r\left(r, w, w_r, w_{rr}, \ldots, w_{r^{q-1}}\right) = k, \qquad k = cte. \tag{8.54}$$

We stress that Eq. (8.54) is an ordinary differential equation of order $q-1$, whose solutions are solutions of Eq. (8.47), by writing this solution in terms of x, t, and u.

1. Equation

$$F \equiv u_{tt} - u_{txx} + u_{xxxx} - (f(u))_{xx} = 0 \tag{8.55}$$

admits the symmetry generators $\mathbf{v_1} = \frac{\partial}{\partial x}$ and $\mathbf{v_2} = \frac{\partial}{\partial t}$ associated with the conservation law

$$\phi^t = -u_{xx} + u_t,$$
$$\phi^x = au_{xxx} - f'(u)u_x$$

Let $\mathbf{v} = \lambda\mathbf{v_1} - \mu\mathbf{v_2}$, canonical coordinates of $\mathbf{v}$ are

$$\mathbf{v}(r) = 0, \quad \mathbf{v}(s) = 1, \quad \mathbf{v}(w) = 0,$$

which leads us to

$$r = \mu x + \lambda t \quad s = \frac{1}{\mu}t, \quad w = u. \tag{8.56}$$

In this coordinates the conservation law is written as

$$D_s T^s + D_r T^r = 0 \tag{8.57}$$

with

$$T^r = \lambda^2 w_r - \lambda\mu^2 w_{rr} + \mu^4 w_{rrr} - \mu^2 f' w_r.$$

Since $T = (T^r, T^s)$ is associated with $\mathbf{v}$

$$T^r = c_1$$

that is

$$\lambda^2 w_r - \lambda\mu^2 w_{rr} + \mu^4 w_{rrr} - \mu^2 f' w_r = c_1.$$

This last equation can be integrated and we obtain

$$\lambda^2 w - \lambda\mu^2 w_r + \mu^4 w_{rr} - \mu^2 f(w) = c_1 r + c_2.$$

When f is a cubic polynomial

$$f(w) = k_1 w^3 + k_2 w^2 + k_3 w + k_4,$$

an exact solution is $w = \tanh(r)$, with $k_1 = 2\mu^2$, $k_2 = \lambda$, $k_3 = \frac{\lambda^2}{\mu^2} - 2\mu^2$ and $k_4 = -\lambda$. Consequently, for

$$f(u) = \frac{u\,\lambda^2}{\mu^2} + u^2\,\lambda - \lambda + 2\,u^3\,\mu^2 - 2\,u\,\mu^2$$
$$u = \tanh(\mu x + \lambda t)$$

is a kink solution of Eq. (8.5).

2. Equation

$$F \equiv u_{tt} - u_{txx} + u_{xxxx} - (f(u))_{xx} = g(x) \tag{8.58}$$

with $f(u) = u^3$ and $g(x) = \dfrac{1}{x^5}$ admits the symmetry generator

$$\mathbf{v} = x\frac{\partial}{\partial x} + 2t\frac{\partial}{\partial t} - u\frac{\partial}{\partial u}$$

associated with the conservation law

$$\phi^t = -\frac{t^2}{2x^5} + (u_t - u_{xx})t,$$
$$\phi^x = tu_{xxx} - 3tu^2 u_x + u_x.$$

Canonical coordinates of $\mathbf{v}$ leads us to

$$r = \frac{x}{\sqrt{t}} \quad s = \frac{1}{2}\log(t), \quad w = u\sqrt{t}. \tag{8.59}$$

In this coordinates the conservation law is written as (8.57) with

$$T^r = -4\,w_{rrr} - 2\,r\,w_{rr} + 12\,w^2\,w_r - r^2\,w_r - 4\,w_r - 3\,r\,w - \frac{1}{r^4}.$$

Since $T = (T^r, T^s)$ is associated with $\mathbf{v}$

$$T^r = c_1$$

that is

$$-4\,w_{r\,r\,r} - 2\,r\,w_{r\,r} + 12\,w^2\,w_r - r^2\,w_r - 4\,w_r - 3\,r\,w - \frac{1}{r^4} = c_1.$$

3. Equation

$$F \equiv u_{tt} - u_{txx} + u_{xxxx} - (f(u))_{xx} = g(x) \tag{8.60}$$

with $f(u) = e^{nu}$ and $g(x) = \dfrac{1}{x^4}$ admits the symmetry generator

$$\mathbf{v} = \frac{x}{2}\frac{\partial}{\partial x} + t\frac{\partial}{\partial t} - \frac{1}{n}u\frac{\partial}{\partial u}$$

associated with the conservation law

$$\phi^t = x\,(-u_{xx} + u_t)\,,$$
$$\phi^x = -\int x g\,(x)\,dx + (u_{xxx}x - u_{xx}) + f(u) - x f'(u)u_x.$$

Canonical coordinates of $\mathbf{v}$ leads us to

$$r = \frac{x}{\sqrt{t}} \quad s = log(t), \quad w = u + log(t). \tag{8.61}$$

In this coordinates the conservation law is written as (8.57) with

$$T^r = r\,w_{r\,r\,r} + \frac{r^2\,w_{r\,r}}{2} - w_{r\,r} - n\,r\,e^{n\,w}\,w_r + \frac{r^3\,w_r}{4} + e^{n\,w} + \frac{r^2}{2\,n} + \frac{k_1}{r^2}.$$

Since $T = (T^r, T^s)$ is associated with $\mathbf{v}$

$$T^r = c_1$$

that is

$$r\,w_{r\,r\,r} + \frac{r^2\,w_{r\,r}}{2} - w_{r\,r} - n\,r\,e^{n\,w}\,w_r + \frac{r^3\,w_r}{4} + e^{n\,w} + \frac{r^2}{2\,n} + \frac{k_1}{r^2} = c_1.$$

8.5 Conclusions

For a damped equation with a time-independent source term, we have derived the classical Lie symmetries admitted by the equation as well as the reduced ordinary differential equations and we have derived some exact solutions. Conservation laws for this equation are constructed for the first time by using the multiplier method. We have given a group classification for a Boussinesq equation with a strong damping term, as well as corresponding reduced ordinary differential equations. We have derived some nontrivial conservation laws by using the multipliers conservation laws method. Taking into account the relationship between symmetries and conservation laws and applying the double reduction method, we have derived a direct reduction of order of the ordinary differential equations and in particular we have found a kink solution.

Acknowledgements The authors warmly thank the reviewer of the manuscript for providing excellent suggestions to improve the quality of the paper. The support from Junta de Andalucía FQM-201 group and University of Cadiz is gratefully acknowledged.

References

1. J. Boussinesq, Théorie des ondes et des remous qui se propagent le long d' un canal rectangulaire horizontal en communiquant au liquide contenudans ce canal des vitesses sensiblement pareilles de la surface au fond. J. Math. Pures Appl. **17**, 55–108 (1872)
2. Y. Liu, Instability and blow-up of solutions to a generalized Boussinesq equation. SIAM J. Math. Anal. **26**, 1527–1546 (1995)
3. M.L. Gandarias, M.S. Bruzón, Classical and nonclassical symmetries of a generalized Boussinesq equation. J. Nonlinear Math. Phys. **5**, 8–12 (1998)
4. V. Varlamov, On spatially periodic solutions of the damped Boussinesq equation. Differ. Integral Equ. **10**, 1197–1211 (1997)
5. V. Varlamov, Long-time asymptotics of solutions of the second initial boundary value problem for the damped Boussinesq equation. Abstr. Appl. Anal. **2**, 281–289 (1998)
6. R.B. Guenther, J.W. Lee, *Partial Differential Equations of Mathematical Physics and Integral Equations* (Prentice Hall, New Jersey, 1988)
7. M. Grasselli, G. Schimperna, S. Zelik, On the 2D Cahn Hilliard equation with inertial term. Commun. Partial Differ. Equ. **34**, 137–170 (2009)
8. N.H. Ibragimov, A new conservation theorem. J. Math. Anal. Appl. **333**, 311–28 (2007)
9. E. Noether, Invariante Variations probleme, Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen 234257 (1918).
10. S. A. Anco, G. Bluman, Direct construction of conservation laws from field equations. Phys. Rev. Lett. **78**, 2869–2873 (1997)
11. S. C. Anco, G. Bluman, Direct construction method for conservation laws for partial differential equations Part II: general treatment. Eur. J. Appl. Math. **41**, 567–585 (2002)
12. P.J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, Berlin, 1986)
13. G.W. Bluman, S. Kumei, *Symmetries and Differential Equations* (Springer, Berlin, 1989)
14. A. Sjöberg, Double reduction of PDEs from the association of symmetries with conservation laws with applications. Appl. Math. Comput. **184**, 608616 (2007)

15. A.H. Kara, F.M. Mahomed, Relationship between symmetries and conservation laws. Int. J. Theor. Phys. **39**(1), 2340 (2000)
16. S.C. Anco, Symmetry properties of conservation laws. arXiv:1512.01835 (2016)
17. M.L. Gandarias, M. Rosa, Symmetries and conservation laws of a damped Boussinesq equation. Mod. Phys. Lett. B **30**(28,29), 1640012 (2016)
18. M.L. Gandarias, M. Rosa, On double reductions from symmetries and conservation laws for a damped Boussinesq equation. Chaos, Solitons Fractals **89**, 560–565 (2016)
19. N. Euler, M. Euler, On nonlocal symmetries, nonlocal conservation laws and nonlocal transformations of evolution equations: two linearisable hierarchies. J. Nonlinear Math. Phys. **6**, 489–504 (2009)
20. N. Mindu, D.P. Mason, Derivation of conservation laws for the magma equation using the multiplier method: power law and exponential law for permeability and viscosity. Abstr. Appl. Anal. **2014**, 585167 (2014). http://dx.doi.org/10.1155/2014/585167

Chapter 9
On Some Variable Exponent Problems with No-Flux Boundary Condition

Maria-Magdalena Boureanu

Abstract The variable exponent problems allow us to deal with nonhomogeneous materials for which it is not suitable to use the functional framework provided by the Lebesgue and Sobolev-type spaces with constant exponents. The no-flux boundary condition first appeared in physics and it opens the door to more real-life applications. In addition to the no-flux boundary condition, our problems involve more general operators, that is, Leray–Lions type operators. The discussion is centered on the weak solvability of such problems via the critical point theory, and it also includes the case of anisotropic exponents. For a plus of cohesion, we select only a few powerful theorems as main tools that can be applied to all these problems.

Keywords Variable exponent spaces · Anisotropic exponent · Nonlinear elliptic problems · No-flux boundary condition · Leray–Lions type operators · Weak solutions · Existence · Multiplicity

9.1 Introduction

When it comes to PDEs, some of the most common boundary conditions are the Dirichlet, Neumann, and Robin boundary conditions. But, as we all know, certain problems that arise from real-life applications may need more unusual conditions on the boundary. In that note, we make reference to the problem treated by Temam [64] in 1977,

$$\begin{cases} -\Delta u + \lambda u_- = 0 & \text{in } \Omega, \\ u = \widehat{c}\,\text{(unknown) constant} & \text{on } \partial\Omega, \\ \displaystyle\int_{\partial\Omega} \frac{\partial u}{\partial \nu}\, dS = \gamma, & \end{cases} \tag{9.1}$$

M.-M. Boureanu (✉)
Department of Applied Mathematics, University of Craiova, Craiova, Romania

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_9

where $\Omega \subseteq \mathbb{R}^2$ is an open set, ν represents the outward unit normal to the boundary $\partial\Omega$, $\gamma > 0$ is given, and, as everywhere in this chapter, to simplify the writing we use u instead of $u\big|_{\partial\Omega}$ for the trace of u on $\partial\Omega$. This problem has its origin in plasma physics since it is a simplified version of another problem previously studied by the same author in [63]. The solution of the problem discussed in [63] determines the shape at equilibrium of a confined plasma. In what concerns (9.1), the region $u < 0$ represents the region that is filled by plasma, while the region $u > 0$ corresponds to the vacuum, and once we solve problem (9.1) we can find these two regions. The existence of a free boundary to this type of problem is physically expected, since the plasma cannot touch the vacuum vessel, so the region $u = 0$ corresponds to the free boundary which separates the plasma and the vacuum.

Other contributions to problems closely related to (9.1) are due to Berestycki and Brezis [8], Kinderlehrer and Spruck [43], Puel [56], Schaeffer [62], etc., if we refer to pioneering studies from the same period of time. However, the preoccupation for such problems is carried over the years, see, for example, the recent work by Zou et al. [72, 73].

In this chapter we focus on a slightly different situation, that is, the case when $\gamma = 0$. This case gives us nonresonant surfaces, on which the wave number of the perturbations parallel to the equilibrium magnetic field is zero, see [1]. These are called no-flux surfaces, and we arrive at the following class of no-flux problems:

$$\begin{cases} -\Delta u = \lambda f & \text{in } \Omega, \\ u = \text{ constant} & \text{on } \partial\Omega, \\ \displaystyle\int_{\partial\Omega} \frac{\partial u}{\partial \nu}\, dS = 0, \end{cases} \tag{9.2}$$

where the constant boundary data is not specified, as it is the case for all the no-flux boundary conditions from what follows. Since we are concerned with generalizations of problem (9.2), we recall that Fan and Deng [34] studied the problem

$$\begin{cases} -\Delta_{p(x)}(u) + b(x)|u|^{p(x)-2}u = f(x,u) + \lambda g(x,u) & \text{in } \Omega, \\ u = \text{constant} & \text{on } \partial\Omega, \\ \displaystyle\int_{\partial\Omega} |\nabla u|^{p(x)-2}\frac{\partial u}{\partial \nu}\, dS = 0, \end{cases} \tag{9.3}$$

where $\Omega \subset \mathbb{R}^N$ $(N \geq 2)$ is a bounded open set with smooth boundary, $p : \overline{\Omega} \to (1, +\infty)$ is a variable exponent that satisfies $1 < \operatorname{ess\,inf}_{x\in\Omega} p(x) \leq \operatorname{ess\,sup}_{x\in\Omega} p(x) < \infty$, and the log-Hölder continuity condition, that is, there exists $\mathfrak{c} > 0$ such that

$$|p(x) - p(y)| \leq \frac{\mathfrak{c}}{-\log|x-y|} \quad \text{for all } x,\ y \in \Omega,\ \ 0 < |x-y| \leq \frac{1}{2} \tag{9.4}$$

and by $\Delta_{p(\cdot)}(u)$ we denote, as usual, the $p(\cdot)$-Laplace operator

$$\Delta_{p(\cdot)}(u) = \operatorname{div}(|\nabla u|^{p(x)-2}\nabla u). \tag{9.5}$$

As for the other functions involved in (9.3), $b \in L^\infty(\Omega)$ with ess $\inf_{x\in\Omega} b = b_0 > 0$ and $f, g \in C(\overline{\Omega} \times \mathbb{R})$ satisfy a subcritical growth condition. Under additional hypotheses on these functions the authors manage to obtain up to seven weak solutions to problem (9.3) and some of these results are new even for the case when the exponent p is constant.

We have brought into discussion problem (9.3) because it represents a particular case of the more general class of problems that we are interested in, and, at the same time, it makes the connection between (9.2) and our problem. As said above, the conditions on the boundary, that is, the lines 2–3 of problem (9.3), are called the no-flux boundary condition. Moreover when $N = 1$, this condition is the periodic boundary condition, see [34, 47] and [48, Remark 1.1]; therefore, the no-flux boundary conditions from our problems are in fact multidimensional generalizations of the periodic boundary condition. Since there are not many $p(\cdot)$-Laplace problems with no-flux boundary conditions, we point out that another problem of this type was studied a few years later, see [51].

The class of the no-flux problems that makes the subject of study of this chapter is represented by the variable exponent problems involving Leray–Lions type operators, see [1, 15, 16, 18, 47]. More precisely, we first discuss the following type of problem from the framework of the variable exponent spaces:

$$\begin{cases} -\operatorname{div}(a(x,\nabla u)) + |u|^{p(x)-2}u = \lambda f(x,u) & \text{for } x \in \Omega, \\ u(x) = \text{constant} & \text{for } x \in \partial\Omega, \\ \displaystyle\int_{\partial\Omega} a(x,\nabla u)\cdot \nu \, dS = 0, & \end{cases} \tag{9.6}$$

which is treated in Sect. 9.4, where all the assumptions on the functions that appear here are carefully specified. The general operators from this problem took the names of the mathematicians who introduced a first version of them in 1965 (see [49]) and have a great property: they can produce various types of operators, including Laplace type operators and the mean curvature type operators. This is the reason why multiple variants of the Leray–Lions operators appeared during the years, each variant adapted to a specific type of problem, and a variant that we are concerned with is the one from (9.6), which generalizes the $p(\cdot)$-Laplace operator (9.5). As said above, we will get into more details in Sect. 9.4, and then in Sect. 9.5, where we extend problem (9.6) to the anisotropic case. In Sects. 9.4 and 9.5 we will state the hypotheses corresponding to the problems and, in particular, to the nonhomogeneous differential operators involved in them, together with meaningful examples of operators that can be derived as particular cases of these Leray–Lions type operators. Obviously, one can refer to no-flux problems involving the Leray–Lions type operators with constant exponents too, see [48, 69]. Nonetheless, when dealing with certain materials that are highly nonhomogeneous, we need an exponent that varies. Hence the development of the theory on the variable exponent spaces and the solvability of the variable exponent problems, see the very well-written books [25, 28, 57] and the references therein. One can learn from these books that treating problems with variable exponents is quite difficult, since there are many properties that only hold when the exponent is constant, and they do not hold even when the variable exponent is very regular, that is, log-Hölder continuous or in $C^\infty(\overline{\Omega})$ with $1 < \operatorname{ess\,inf}_{x\in\Omega} p(x) \leq \operatorname{ess\,sup}_{x\in\Omega} p(x) < \infty$. To give an immediate example, since we already referred to the differential operators, we recall that, as opposed to the p-Laplace operator with constant p,

$$\Delta_p(u) = \operatorname{div}(|\nabla u|^{p-2}\nabla u),$$

the $p(\cdot)$-Laplace operator (9.5) has lost the homogeneity, which means that the $p(\cdot)$-Laplace equation is not scalable, that is, u being a solution of this equation does not imply that λu is a solution too. Other things that we can briefly mention are the fact that the space $L^{p()}$ is not rearrangement invariant, the fact that the translation operator is not bounded, or that Young's convolution inequality

$$\|f \star g\|_{L^{p()}} \leq c\|f\|_{L^1}\|g\|_{L^{p()}} \quad \text{does not hold.}$$

Also, the interpolation is not so useful because the variable exponent spaces never result as an interpolant of constant exponent spaces and many inequalities (maximal, Poincaré, Sobolev, etc.,) do not hold in a modular form.

The above described difficulties not only did not discourage the mathematicians, but they seem to have the opposite effect and many studies continue to appear, see, for example, the recent papers [4, 20, 40]. Many valuable properties were established during the last years and a small part of them is displayed in our next section. The strong interest in the variable exponent problems is fueled by multiple applications, see, for example, [19, 70] for applications concerning elastic materials, [22] for applications in image restoration, [3, 26, 50, 61] for applications due to smart fluids, [37] for applications in mathematical biology, etc.

Here we are going to focus on a collection of recently obtained results due to [15, 16, 18]. Notice that part of the results from [15] extends the results from [69], where the exponent was constant. Furthermore, continuing this line of investigation, the papers [16, 18] extend some of the results from [15], from the variable exponents that are isotropic, to the anisotropic version of them. By considering anisotropic variable exponents, that is,

$$\overrightarrow{p} : \overline{\Omega} \to (1, +\infty)^N,\ \overrightarrow{p}(\cdot) = (p_1(\cdot), p_2(\cdot) \ldots, p_N(\cdot)),$$

we can have different behavior corresponding to the distinct space directions. The importance of the anisotropic operators is emphasized by recent applications in physics [9, 29, 30, 38, 39], biology [6, 7], image processing [65], etc. The theory of the spaces with exponents that are both anisotropic and variable is quite fresh and we recall a few early papers on this research direction [5, 12, 17, 23, 33, 44]. We mention that the first problems involving anisotropic variable exponents and no-flux boundary conditions are exactly the ones from the papers [16, 18] that we are going to discuss in Sect. 9.5.

As for the organization of our work, it is as follows: This chapter is structured into six sections. After the introductory section, we present some notations, some definitions, and some basic properties of the variable exponent spaces, both isotropic and anisotropic, since they represent the abstract framework where we cast our problems. Then, taking into account the fact that our problems are weakly solved via the critical point theory, Sect. 9.3 contents a selections of results that are quite useful when adopting this strategy, and a few history notes. Sections 9.4 and 9.5 mainly present recent results obtained in [15, 16, 18] for isotropic, respectively, anisotropic, variable exponent problems with Leray–Lions type operators and no-flux boundary conditions. Finally, Sect. 9.6 provides comments and remarks, indicating possible improvements and future directions of research.

9.2 Functional Framework

Everywhere in what follows, if not otherwise stated, we will assume that $\Omega \subset \mathbb{R}^N$ ($N \geq 2$) is a bounded open set with smooth boundary and $p : \overline{\Omega} \to (1, +\infty)$, with $1 < \operatorname{ess\,inf}_{x\in\Omega} p(x) \leq \operatorname{ess\,sup}_{x\in\Omega} p(x) < \infty$, satisfies the log-Hölder continuity condition (9.4). For simplicity, for every variable exponent r, we denote

$$\operatorname*{ess\,inf}_{x\in\Omega} r(x) = r^- \quad \text{and} \quad \operatorname*{ess\,sup}_{x\in\Omega} r(x) = r^+.$$

Our goal is to insert here the basic properties from the theory of the variable exponent spaces which are relevant for the weak solvability of the problems that will be discussed later. Hence we start by recalling the definition of the Lebesgue space with variable exponent,

$$L^{p(\cdot)}(\Omega) = \{u : \Omega \to \mathbb{R} : \quad u \text{ is measurable and } \int_\Omega |u(x)|^{p(x)}\, dx < \infty\}$$

which is endowed with the Luxemburg norm,

$$\|u\|_{L^{p(\cdot)}(\Omega)} = \inf\left\{\mu > 0 : \quad \int_\Omega \left|\frac{u(x)}{\mu}\right|^{p(x)} dx \leq 1\right\},$$

and it is a separable and reflexive Banach space, see [45, Theorem 2.5, Corollary 2.7]. As one can see, it is not easy to handle this Luxemburg norm, so we rely on the application $\rho_{p(\cdot)} : L^{p(\cdot)}(\Omega) \to \mathbb{R}$,

$$\rho_{p(\cdot)}(u) = \int_\Omega |u(x)|^{p(x)}\, dx,$$

which is called the $p(\cdot)$-modular of the $L^{p(\cdot)}(\Omega)$ space. We can cite, for example, [36, Theorem 1.3, Theorem 1.4] for the following properties. If $u \in L^{p(\cdot)}(\Omega)$, then

$$\|u\|_{L^{p(\cdot)}(\Omega)} < 1 \ (= 1;\ > 1) \quad \text{if and only if} \quad \rho_{p(\cdot)}(u) < 1 \ (= 1;\ > 1);$$

$$\text{if} \quad \|u\|_{L^{p(\cdot)}(\Omega)} > 1 \quad \text{then} \quad \|u\|^{p^-}_{L^{p(\cdot)}(\Omega)} \le \rho_{p(\cdot)}(u) \le \|u\|^{p^+}_{L^{p(\cdot)}(\Omega)};$$

$$\text{if} \quad \|u\|_{L^{p(\cdot)}(\Omega)} < 1 \quad \text{then} \quad \|u\|^{p^+}_{L^{p(\cdot)}(\Omega)} \le \rho_{p(\cdot)}(u) \le \|u\|^{p^-}_{L^{p(\cdot)}(\Omega)};$$

$$\|u\|_{L^{p(\cdot)}(\Omega)} \to 0 \ (\to \infty) \quad \text{if and only if} \quad \rho_{p(\cdot)}(u) \to 0 \ (\to \infty).$$

If, in addition, $(u_n)_n \subset L^{p(\cdot)}(\Omega)$, then

$$\lim_{n\to\infty} \|u_n - u\|_{L^{p(\cdot)}(\Omega)} = 0 \quad \text{if and only if} \quad \lim_{n\to\infty} \rho_{p(\cdot)}(u_n - u) = 0.$$

When we need to make connections between functionals and norms, the above properties are very useful, as well as the embedding result below.

Theorem 9.2.1 ([45, Theorem 2.8]) *If* $0 < |\Omega| < \infty$ *and* $r_1,\ r_2 \in C(\overline{\Omega}; \mathbb{R})$*,* $1 \le r_i^- \le r_i^+ < \infty$ *(*$i = 1, 2$*), are such that* $r_1 \le r_2$ *in* Ω*, then the embedding* $L^{r_2(\cdot)}(\Omega) \hookrightarrow L^{r_1(\cdot)}(\Omega)$ *is continuous.*

Also, the following Hölder type inequality,

$$\left| \int_\Omega u(x)v(x)\, dx \right| \le 2\, \|u\|_{L^{p(\cdot)}(\Omega)} \|v\|_{L^{p'(\cdot)}(\Omega)},$$

holds for all $u \in L^{p(\cdot)}(\Omega)$ and $v \in L^{p'(\cdot)}(\Omega)$, see [45, Theorem 2.1]. Here we have denoted by $L^{p'(\cdot)}(\Omega)$ the conjugate space of $L^{p(\cdot)}(\Omega)$, obtained by conjugating the exponent pointwise, that is, $1/p(x) + 1/p'(x) = 1$, see [45, Corollary 2.7].

Furthermore, to every Carathéodory function $g : \Omega \times \mathbb{R}^m \to \mathbb{R}$, $m \in \mathbb{N}$, we can associate a Nemytsky operator N_g that maps an m-tuple of functions $(u_1, \ldots, u_m)$ into

$$N_g(u_1, \ldots, u_m)(x) = g(x, u_1(x), \ldots, u_m(x)) \quad x \in \Omega. \tag{9.7}$$

Theorem 9.2.2 ([45, Theorems 4.1–4.2]) *Let* $g : \Omega \times \mathbb{R}^m \to \mathbb{R}$*,* $m \in \mathbb{N}$*, be a Carathéodory function and* $l_i,\ l_0 \in L^\infty(\Omega)$ *with* $l_i,\ l_0 \ge 1$ *for all* $i \in \{1, 2, \ldots, m\}$*. Assume that there exist a nonnegative function* $h \in L^{l_0(\cdot)}(\Omega)$ *and a constant* $c > 0$ *such that*

$$|g(x, \xi)| \le h(x) + c \sum_{i=1}^{m} |\xi_i|^{l_i(x)/l_0(x)}$$

for all $\xi \in \mathbb{R}^N$ *and a.e.* $x \in \Omega$. *Then the Nemytsky operator* N_g *provided by formula* (9.7) *maps* $L^{l_1(\cdot)}(\Omega) \times \ldots \times L^{l_m(\cdot)}(\Omega)$ *into* $L^{l_0(\cdot)}(\Omega)$ *and it is a continuous and bounded operator.*

Let us pass now to the definition of the Sobolev space with variable exponent,

$$W^{1,p(\cdot)}(\Omega) = \left\{ u \in L^{p(\cdot)}(\Omega) : \quad |\nabla u| \in L^{p(\cdot)}(\Omega) \right\},$$

which is endowed with the norm

$$\|u\| = \|u\|_{L^{p(\cdot)}(\Omega)} + \|\nabla u\|_{L^{p(\cdot)}(\Omega)}, \tag{9.8}$$

where by $\|\nabla u\|_{L^{p(\cdot)}(\Omega)}$ we understand $\| \, |\nabla u| \, \|_{L^{p(\cdot)}(\Omega)}$. This space is a separable and reflexive Banach space, see [45, Theorem 1.3], and we have the following embedding theorem.

Theorem 9.2.3 ([32, Proposition 2.4]) *Assume* $\Omega \subset \mathbb{R}^N$ *(*$N \geq 2$*) is a bounded domain with smooth boundary and* $p \in C(\overline{\Omega})$ *with* $p^- > 1$. *In addition,* $s \in C(\overline{\Omega})$ *satisfies the condition*

$$1 \leq s(x) < p^\star(x) \qquad \forall x \in \overline{\Omega},$$

where $p^\star$ *denotes, as usual, the critical exponent given by*

$$p^\star(x) = \begin{cases} Np(x)/[N - p(x)] & \text{if } p(x) < N, \\ \infty & \text{if } p(x) \geq N. \end{cases} \tag{9.9}$$

Then the embedding $W^{1,p(\cdot)}(\Omega) \hookrightarrow L^{s(\cdot)}(\Omega)$ *is compact.*

Again, the following inequalities connect the norm and the functionals.

Proposition 9.2.4 ([35, Proposition 2.3]) *Let* $u \in W^{1,p(\cdot)}(\Omega)$. *We have*

$$\text{if} \quad \|u\| > 1 \quad \text{then} \quad \|u\|^{p^-} \leq \int_\Omega \left[|\nabla u|^{p(x)} + |u|^{p(x)} \right] dx \leq \|u\|^{p^+};$$

$$\text{if} \quad \|u\| < 1 \quad \text{then} \quad \|u\|^{p^+} \leq \int_\Omega \left[|\nabla u|^{p(x)} + |u|^{p(x)} \right] dx \leq \|u\|^{p^-}.$$

But let us not lose sight of the fact that we are preoccupied with the existence of weak solutions and, in order to be able to properly define such solutions, we need a density result.

Theorem 9.2.5 (see [27, Theorem 3.7] and [25, Section 6.5.3]) *Assume that* $\Omega \subset \mathbb{R}^N$ *(*$N \geq 2$*) is a bounded domain with Lipschitz boundary and* p *is log-Hölder continuous with* $p^- > 1$. *Then* $C^\infty(\overline{\Omega})$ *is dense in* $W^{1,p(\cdot)}(\Omega)$.

The most well-known subspace of $W^{1,p(\cdot)}(\Omega)$ is $W_0^{1,p(\cdot)}(\Omega)$, that is, the subspace of the functions that are vanishing on the boundary. To this space we associate the norm

$$\|u\|_{W_0^{1,p(\cdot)}(\Omega)} = \|\nabla u\|_{L^{p(\cdot)}(\Omega)},$$

due to the following Poincaré type inequality (see [32, Proposition 2.3]):

$$\|u\|_{L^{p(\cdot)}(\Omega)} \leq C\|\nabla u\|_{L^{p(\cdot)}(\Omega)} \quad \text{for all } u \in W_0^{1,p(\cdot)}(\Omega),$$

where C is a positive constant. Then $\left(W_0^{1,p(\cdot)}(\Omega), \|\cdot\|_{W_0^{1,p(\cdot)}(\Omega)}\right)$ is a separable and reflexive Banach space, see [32, Proposition 2.1], and this is the space where we look for weak solutions whenever we deal with problems with zero Dirichlet boundary condition.

In our case, though, the problems are with no-flux boundary condition, so we introduce another subspace of $W^{1,p(\cdot)}(\Omega)$,

$$V = \left\{u \in W^{1,p(\cdot)}(\Omega): \quad u\big|_{\partial\Omega} = \text{constant}\right\}$$

$$= \{u + c: \quad u \in W_0^{1,p(\cdot)}(\Omega),\ c \in \mathbb{R}\}.$$

With the help of the fine properties of $W_0^{1,p(\cdot)}(\Omega)$ we can establish the following.

Theorem 9.2.6 (see [15, Theorem 3]) $(V, \|\cdot\|)$ *is a separable and reflexive Banach space, where* $\|\cdot\|$ *is the norm associated with the variable exponent Sobolev space* $W^{1,p(\cdot)}(\Omega)$*, see* (9.8).

The space $(V, \|\cdot\|)$ represents the space where we are going to search for weak solutions for the isotropic variable exponent problem with no-flux boundary condition (9.6).

Let us pass now to the abstract framework corresponding to the anisotropic variable exponent problems. We recall that the anisotropic variable exponent

$$\vec{p}(\cdot) \text{ is a vectorial function, } \vec{p}: \overline{\Omega} \to (1, +\infty)^N,\ \vec{p}(\cdot) = (p_1(\cdot), p_2(\cdot) \ldots, p_N(\cdot))$$

and for all $i \in \{1, \ldots, N\}$ we will consider p_i to be log-Hölder continuous with $1 < p_i^- \leq p_i^+ < \infty$. As the reader may have noticed, the log-Hölder continuity condition that we request for the exponents is essential for the density results. We introduce more notation:

$$p_M(x) = \max\{p_1(x), \ldots, p_N(x)\}, \quad p_m(x) = \min\{p_1(x), \ldots, p_N(x)\}, \tag{9.10}$$

$$\text{and} \quad \bar{p}(x) = \frac{N}{\sum_{i=1}^{N} \big(1/p_i(x)\big)}. \tag{9.11}$$

The anisotropic variable exponent Sobolev space,

$$W^{1,\overrightarrow{p}(\cdot)}(\Omega) = \Big\{u \in L^{p_M(\cdot)}(\Omega) : \quad \partial_{x_i} u \in L^{p_i(\cdot)}(\Omega) \text{ for all } i \in \{1, \ldots, N\}\Big\},$$

is endowed with the norm

$$\|u\|_{W^{1,\overrightarrow{p}(\cdot)}(\Omega)} = \|u\|_{L^{p_M(\cdot)}(\Omega)} + \sum_{i=1}^{N} \left\|\partial_{x_i} u\right\|_{L^{p_i(\cdot)}(\Omega)}.$$

This space is a reflexive Banach space (see [33, Theorems 2.1 and 2.2]) and we recall two embedding results.

Theorem 9.2.7 ([33, Corollary 2.1]) *Let $\Omega \subset \mathbb{R}^N$ be a bounded domain. Assume that for all $i \in \{1, \ldots, N\}$, $p_i \in L^\infty(\Omega)$ and $p_i(x) \ge 1$ a.e. in Ω. Then for any $r \in L^\infty(\Omega)$ with $r(x) \ge 1$ a.e. in Ω such that* $\operatorname{ess\,inf}_{x\in\Omega}(p_M(x) - r(x)) > 0$ *we have the compact embedding $W^{1,\overrightarrow{p}(\cdot)}(\Omega) \hookrightarrow L^{r(\cdot)}(\Omega)$.*

Theorem 9.2.8 ([33, Theorem 2.5])

(i) Let $\Omega \subset \mathbb{R}^N$ be a rectangular-like domain, that is, a union of finitely many rectangular domains (or cubes) with edges parallel to the coordinate axes. Assume that $p_i \in C(\overline{\Omega})$ with $p_i^- > 1$ for all $i \in \{1, \ldots, N\}$ and $r \in C(\overline{\Omega})$ with $1 \le r(x) < \max\{\overline{p}^\star(x), p_M(x)\}$ for all $x \in \overline{\Omega}$, where $\overline{p}^\star$ and p_M are given by formulae (9.9) and (9.11). Then we have the compact embedding $W^{1,\overrightarrow{p}(\cdot)}(\Omega) \hookrightarrow L^{r(\cdot)}(\Omega)$.

(ii) The same statement as in part (i) is true for any bounded domain $\Omega \subset \mathbb{R}^N$ if we use $W_0^{1,\overrightarrow{p}(\cdot)}(\Omega)$ instead of $W^{1,\overrightarrow{p}(\cdot)}(\Omega)$, where $W_0^{1,\overrightarrow{p}(\cdot)}(\Omega)$ represents the subspace of the functions that are vanishing on the boundary, that is,

$$W_0^{1,\overrightarrow{p}(\cdot)}(\Omega) = \Big\{u \in W^{1,\overrightarrow{p}(\cdot)}(\Omega) : \quad u = 0 \text{ on } \partial\Omega\Big\}.$$

Another important theorem, which allows to formulate the definition of the weak solutions, is the following.

Theorem 9.2.9 ([33, Theorem 2.4]) *Assume that $\Omega \subset \mathbb{R}^N$ ($N \ge 2$) is a bounded domain with Lipschitz boundary and $\overrightarrow{p} \in (L^\infty(\Omega))^N$ with $p_i^- > 1$ for all $i \in \{1, \ldots, N\}$. In addition, assume that p_i, $i \in \{1, \ldots, N\}$, are log-Hölder continuous. Then*

(i) $C_0^\infty(\overline{\Omega})$ is dense in $W_0^{1,\overrightarrow{p}(\cdot)}(\Omega)$;
(ii) $C^\infty(\overline{\Omega})$ is dense in $W^{1,p(\cdot)}(\Omega)$ if Ω is a rectangular-like domain.

As in the isotropic case, we bring into discussion another subspace of $W^{1,\overrightarrow{p}(\cdot)}(\Omega)$, which is more appropriate for the weak solvability of the no-flux problems with anisotropic variable exponent,

$$\overrightarrow{V} = \left\{ u \in W^{1,\overrightarrow{p}(\cdot)}(\Omega) : \quad u\big|_{\partial\Omega} \equiv \text{constant} \right\},$$

$$= \{u + c : \quad u \in W_0^{1,\overrightarrow{p}(\cdot)}(\Omega),\ c \in \mathbb{R}\}.$$

Theorem 9.2.10 ([18, Theorem 3.3]) *The space* $\left(V, \|\cdot\|_{W^{1,\overrightarrow{p}(\cdot)}(\Omega)}\right)$ *is a reflexive Banach space.*

Now that we sketched the functional framework for our discussion, we will focus on other instruments that lead to the desired results.

9.3 Critical Point Tools

The results from the subsequent sections are based on the critical point theory, meaning that to the variable exponent problems under discussion we associate functionals such that the critical points of the functionals are weak solutions of the problems, and vice versa. This strategy allows us to search for critical points instead of weak solutions. Thus one of the most powerful theorems is the following Weierstrass type theorem.

Theorem 9.3.1 (see [24, Section 2, Theorem 1.2]) *Assume that* X *is a reflexive Banach space of norm* $\|\cdot\|_X$ *and the functional* $\Phi : X \to \mathbb{R}$ *is*

(i) coercive on X*, that is,* $\Phi(u) \to \infty$ *as* $\|u\|_X \to \infty$*;*
(ii) (sequentially) weakly lower semicontinuous on X*, that is, for any* $u \in X$ *and any subsequence* $(u_n)_n \subset X$ *such that* $u_n \rightharpoonup u$ *weakly in* X *there holds*

$$\Phi(u) \leq \liminf_{n\to\infty} \Phi(u_n).$$

Then Φ *is bounded from below on* X *and attains its infimum in* X.

Moreover, we remind two standard results that are advantageous when it comes to the weak lower semicontinuity of the functionals.

Proposition 9.3.2 (see [24, Section 2, Example B]) *If* $\Phi : X \to \mathbb{R}$ *is a convex lower semicontinuous functional on a reflexive Banach space* X*, then* Φ *is weakly lower semicontinuous on* X.

Proposition 9.3.3 (see [46, Theorem 6.2.1.]) *Let X be a reflexive Banach space, and let $\Phi : X \to \mathbb{R}$ be Gâteaux differentiable on X. Then the conditions stated below are equivalent:*

(i) *Φ is convex;*
(ii) *the first Gâteaux derivative of Φ is monotone, that is,*

$$\langle \Phi'(u) - \Phi'(v), u - v\rangle \geq 0, \quad \textit{for all } u,\ v \in X;$$

(iii) *we have*

$$\Phi(u) - \Phi(v) \geq \langle \Phi'(v), u - v\rangle \quad \textit{for all } u,\ v \in X.$$

Maybe as famous as the above Weierstrass type theorem is the mountain pass theorem. We recall that the name of Theorem 9.3.1 can be viewed in connection to the classical Dirichlet principle, which, as a funny story, first caught the attention of Green in 1833, then it was mentioned by the name of Dirichlet principle by Riemann in 1851, since Dirichlet was the one to provide a proof, but Weierstrass was the one to notice that the proof is incorrect in 1870, as he pointed out the subtle difference between the minimum and infimum. Many other great mathematicians got involved and then, using Arzelas idea, the Dirichlet principle was established for certain important cases by Hilbert in 1900. Major contributions to the critical point theory are also due to Lebesgue, Tonelli, Lagrange, Legendre, Jacobi, Hamilton, Poincaré, and the list continues. We stopped at the name of Poincaré since in 1905 he treated a variational problem whose solution corresponds neither to a minimum nor to a maximum. That was a revolutionary idea, since many years the general belief was perfectly illustrated by the words of Euler: “I am convinced that the nature acts everywhere following some principles of maximum or minimum,” see [31]. Thus the min–max theory represents a turning point in PDEs and at the beginning of its development subscribes not only Poincaré, but also Birkhoff (in 1917), Morse, Ljusternik, and Schnirelman (in the late 1920s and early 1930s), and Palais, Smale, and Rothe (in the 1960s). Finally, the contribution of Ambrosetti and Rabinowitz [2] in 1973 with their mountain pass theorem marks “the beginning of a postmodern era” in the critical point theory, see [42]. Many variants of this theorem appeared in time (see, for example, [21, 42, 55] and the references therein) and we present here two of them.

Theorem 9.3.4 (see, e.g., [55]) *Let X, endowed with the norm $\|\cdot\|_X$, be a Banach space. Assume that $\Phi \in C^1(X;\mathbb{R})$ satisfies the Palais–Smale condition, that is, any sequence $(u_n)_n \subset X$ such that $(\Phi(u_n))_n$ is bounded and $\Phi'(u_n) \to 0$ in $X^\star$ as $n \to \infty$, contains a subsequence converging to a critical point of Φ. Also, assume that Φ has a mountain pass geometry, that is,*

(i) *there exist two constants $r > 0$ and $\rho \in \mathbb{R}$ such that $\Phi(u) \geq \rho$ if $\|u\|_X = r$;*
(ii) *$\Phi(0) < \rho$ and there exists $e \in X$ such that $\|e\|_X > r$ and $\Phi(e) < \rho$.*

Then Φ *has a critical point* $u_0 \in X \setminus \{0, e\}$ *with critical value*

$$\Phi(u_0) = \inf_{\gamma \in \mathcal{P}} \sup_{u \in \gamma} \Phi(u) \geq \rho > 0,$$

where $\mathcal{P}$ *denotes the class of the paths* $\gamma \in C([0, 1]; X)$ *joining 0 to e.*

While the above theorem is used when we want to establish the existence of a solution, for the existence of infinitely many solutions we rely on the following mountain pass type theorem, that is, the fountain theorem. Let us introduce the general context first. By [68, Section 17] we know that for a separable and reflexive Banach space there exist $\{e_n\}_{n=1}^{\infty} \subset X$ and $\{f_n\}_{n=1}^{\infty} \subset X^\star$ such that

$$f_n(e_m) = \delta_{n,m} = \begin{cases} 1 \text{ if } n = m, \\ 0 \text{ if } n \neq m, \end{cases}$$

$$X = \overline{\text{span}}\{e_n : \quad n = 1, 2, \ldots\} \quad \text{and} \quad X^\star = \overline{\text{span}}\{f_n : \quad n = 1, 2, \ldots\}.$$

For $i = 1, 2, \ldots$ we denote

$$X_i = \text{span}\{e_i\}, \quad Y_i = \bigoplus_{j=1}^{i} X_j, \quad \text{and} \quad Z_i = \overline{\bigoplus_{j=i}^{\infty} X_j}. \tag{9.12}$$

Then, for a separable reflexive Banach space X and for X_i, Y_i, Z_i taken as in (9.12), we remind the fountain theorem.

Theorem 9.3.5 (see, e.g., [66]) *Assume that* $\Phi \in C^1(X, \mathbb{R})$ *is even and that for each* $i = 1, 2, \ldots,$ *there exist* $\rho_i > \gamma_i > 0$ *such that*

(H1) $\inf_{u \in Z_i,\, \|u\|_X = \gamma_i} \Phi(u) \to \infty$ *as* $i \to \infty$.
(H2) $\max_{u \in Y_i,\, \|u\|_X = \rho_i} \Phi(u) \leq 0$.
(H3) Φ *satisfies the Palais–Smale condition for every* $c > 0$.

Then Φ *has a sequence of critical values tending to* $+\infty$.

Notice that in the previous two results the functional is assumed to be even. That will reflect on the hypotheses that we impose on the problem. But there are other possibilities to obtain multiplicity results, without imposing a symmetry condition on the functional. Here we can refer to the so-called three critical points theorems. After the publication of the paper [58] from 2000 where Ricceri introduced such a theorem, this subject attracted a lot of attention and many variants and improvements of this theorem continued to appear, see, for example, [10, 54, 59, 60] and the references therein.

Theorem 9.3.6 ([10, Theorem 2.1]) *Let* X *be a separable and reflexive real Banach space, and let* $\phi, \psi : X \to \mathbb{R}$ *be two continuously Gâteaux differentiable* *functionals. Assume that there exists* $x_0 \in X$ *such that* $\phi(x_0) = \psi(x_0) = 0$ *and* $\phi(x) \geq 0$ *for every* $x \in X$ *and that there exist* $x_1 \in X$, $r_0 > 0$ *such that*

(i) $r_0 < \phi(x_1)$;
(ii) $\sup_{\phi(x)<r_0} \psi(x) < r_0 \psi(x_1)/\phi(x_1)$.

Further, put

$$\alpha = \frac{\beta r_0}{r_0 \frac{\psi(x_1)}{\phi(x_1)} - \sup_{\phi(x)<r_0} \psi(x)},$$

with $\beta > 1$, *and assume that the functional* $\phi - \lambda\psi$ *is sequentially weakly lower semicontinuous, satisfies the Palais–Smale condition, and*

(iii) $\lim_{\|x\| \to +\infty} (\phi(x) - \lambda\psi(x)) = +\infty$ *for every* $\lambda \in [0, \alpha]$.

Then, there exist an open interval $\Lambda \subseteq [0, \alpha]$ *and a positive real number* σ *such that, for every* $\lambda \in \Lambda$, *the equation*

$$\phi'(x) - \lambda\psi'(x) = 0$$

admits at least three distinct solutions in X *whose norms are less than* σ.

Before ending this section, it is worth mentioning that another possibility to obtain multiple solutions is to put together Theorems 9.3.1 and 9.3.4, as we will see in the next sections.

9.4 Problems with (Isotropic) Variable Exponent

We are concerned with problem (9.6) which is displayed again here, for the convenience of the reader:

$$\begin{cases} -\text{div}(a(x, \nabla u)) + |u|^{p(x)-2}u = \lambda f(x, u) & \text{for } x \in \Omega, \\ u(x) = \text{constant} & \text{for } x \in \partial\Omega, \\ \displaystyle\int_{\partial\Omega} a(x, \nabla u) \cdot \nu \, dS = 0, \end{cases}$$

where $\Omega \subset \mathbb{R}^N$ ($N \geq 2$) is a bounded open set with smooth boundary, $\lambda \geq 0$, and the functions involved in this problem will be given below. But first, let us say that this problem was studied in [15] and we have taken this opportunity to correct a mistake. More precisely, we refer to the fact that in [15] the authors somehow forgot to put condition $\int_{\partial\Omega} a(x, \nabla u) \cdot \nu \, dS = 0$ in their problem, although an analog condition (corresponding to the constant exponent case) appeared in the paper [69] that was extended by [15] to the variable exponent case. However, they signaled their mistake to the editors immediately after publication and we mention that the results from [15] are valid for the problem written in the correct form, that is, (9.6). Moreover, the study started in [15] was continued by the same authors and in their next papers they added this condition (in the form corresponding to their problems there), see [16, 18]. Notice that other small mistakes that appeared in [15, 16, 18] will be corrected in what follows without much emphasis.

Let us proceed now with the discussion of problem (9.6). We introduce the hypotheses on the functions involved here. As previously announced, p will always satisfy

(p) $p : \overline{\Omega} \to (1, +\infty)$ with $1 < p^- \leq p^+ < \infty$ and p fulfills the log-Hölder continuity condition (9.4).

In addition, we have the following.

(a0) $a : \overline{\Omega} \times \mathbb{R}^N \to \mathbb{R}^N$ is a Carathéodory function with the property that there exists a Carathéodory function $A : \overline{\Omega} \times \mathbb{R}^N \to \mathbb{R}$, continuously differentiable with respect to its second argument, such that

$$a(x, \xi) = \nabla_\xi A(x, \xi) \quad \text{ and } A(x, 0) = 0,$$

for all $\xi \in \mathbb{R}^N$ and all $x \in \overline{\Omega}$.

(a1) There exists $\tilde{c} > 0$ such that a satisfies the growth condition

$$|a(x, \xi)| \leq \tilde{c}(1 + |\xi|^{p(x)-1})$$

for all $x \in \Omega$ and all $\xi \in \mathbb{R}^N$, where $|\cdot|$ denotes the Euclidean norm.

(a2) The inequalities

$$|\xi|^{p(x)} \leq a(x, \xi) \cdot \xi \leq p(x) A(x, \xi)$$

hold for a.e $x \in \Omega$ and all $\xi \in \mathbb{R}^N$.

(a3) The monotonicity condition

$$0 \leq [a(x, \xi_1) - a(x, \xi_2)] \cdot (\xi_1 - \xi_2)$$

holds for all $x \in \Omega$ and all $\xi_1,\ \xi_2 \in \mathbb{R}^N$, with equality if and only if $\xi_1 = \xi_2$.

Hypotheses (a0)–(a3) will always be assumed when we discuss the weak solvability of problem (9.6). Moreover, when dealing with the multiplicity of the weak solutions, in order to be able to apply Theorem 9.3.5, we assume an additional hypothesis:

(a4) The mapping a is odd with respect to its second variable, that is,

$$a(x, -\xi) = -a(x, \xi)$$

for all $x \in \Omega$ and all $\xi \in \mathbb{R}^N$.

At a first glance, imposing assumptions (a0)–(a4) may seem a little too "heavy," but, as we said in the introductory section, the Leray–Lions type operators

$$\operatorname{div}(a(x, \nabla u))$$

are quite general, and by choosing appropriate examples of functions a that verify (a0)–(a4), we are led to different type of operators, some of them very well known. Indeed, let us consider $h \in L^\infty(\Omega)$ with the property that there exists $h_0 > 0$ such that $h(x) \geq h_0$ for all $x \in \Omega$. Then, by taking

$$a(x, \xi) = h(x)|\xi|^{p(x)-2}\xi,$$

we have

$$A(x, \xi) = \frac{h(x)}{p(x)}|\xi|^{p(x)},$$

and (a0)–(a4) are verified. This way we have arrived at the class of operators

$$\operatorname{div}(a(x, \nabla u)) = \operatorname{div}(h(x)|\nabla u|^{p(x)-2}\nabla u),$$

which, for $h \equiv 1$, produces the $p(\cdot)$-Laplace operator (9.5). Moreover, if we take

$$a(x, \xi) = h(x)(1 + |\xi|^2)^{(p(x)-2)/2}\xi,$$

we have

$$A(x, \xi) = \frac{h(x)}{p(x)}\left[(1 + |\xi|^2)^{p(x)/2} - 1\right],$$

(a0)–(a4) are verified, and we have arrived at the class of operators

$$\operatorname{div}(a(x, \nabla u)) = \operatorname{div}\left(h(x)(1 + |\nabla u|^2)^{(p(x)-2)/2}\nabla u\right),$$

which produces the generalized mean curvature operator

$$\operatorname{div}(a(x, \nabla u)) = \operatorname{div}\left((1 + |\nabla u|^2)^{(p(x)-2)/2}\nabla u\right)$$

when $h \equiv 1$. We believe that now it is clear to the reader why there is an increased interest for the problems involving Leray–Lions type operators. But we have to specify that not all the studies that analyze problems with Leray–Lions type operators preserve these hypotheses in the exact form as above; in different papers, small differences appear. For example, in [52, 53], the authors use the following assumption:

(a) There exists $k_0 > 0$ such that

$$A\left(x, \frac{\xi+\eta}{2}\right) \leq \frac{1}{2}A(x,\xi) + \frac{1}{2}A(x,\eta) - k_0|\xi-\eta|^{p(x)}$$

for all $x \in \Omega$ and $\xi, \eta \in \mathbb{R}^N$.

The $p(\cdot)$-uniform convexity condition (a) is not so convenient because it restricts the range of the exponents in the previous examples of operators, e.g., function $A(x,\xi) = \frac{h(x)}{p(x)}|\xi|^{p(x)}$ satisfies (a) only when $p \geq 2$, instead of $p > 1$, as it is in our case. Therefore imposing condition (a) would be a drawback, especially since there are studies in which it is essential for the exponents to be between 1 and 2, see, for example, [22].

Now that we have commented a little on the conditions regarding the Leray–Lions type operators, let us carry on to the assumptions on f. In this section, we always assume that the following condition holds.

(f0) $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function.

In addition to that, we will consider several hypotheses concerning f, and not all of them are going to be fulfilled at the same time. We start with

($\tilde{\text{f}}$1) there exists $k > 0$ such that f satisfies the growth condition

$$|f(x,t)| \leq k|t|^{q(x)-1} \tag{9.13}$$

for a.e. $x \in \Omega$ and all $t \in \mathbb{R}$, where $q \in C(\overline{\Omega})$ with

$$1 < q^- \leq q^+ < p^-.$$

This allows us to obtain an existence result by means of the Weierstrass type theorem. It is clear that q is subcritical, since $p < p^\star$. But restriction $q^+ < p^-$ makes us wonder what happens when there is another order between the exponents p and q. That we will see in one of the subsequent cases.

In what follows, we will distinguish between the two situations: the case when f is $p(\cdot)-1$—superlinear at infinity—and the case when f is $p(\cdot)-1$—sublinear at infinity. In each situation we will deduce not only an existence result, but also some multiplicity results. For the first case, the nonlinearity f is supposed to satisfy the following.

(f1) There exists $k > 0$ such that f satisfies the growth condition (9.13) for a.e. $x \in \Omega$ and all $t \in \mathbb{R}$, where $q \in C(\overline{\Omega})$ with

$$p^+ < q^- \leq q^+ < p^\star.$$

(f2) There exist $\theta > p^+$ and $l > 0$ such that f satisfies the Ambrosetti–Rabinowitz condition

$$0 < \theta F(x,t) \leq f(x,t)t$$

for all $|t| > l$ and a.e. $x \in \Omega$, where

$$F(x,t) = \int_0^t f(x,\tau)\,d\tau.$$

With respect to ($\tilde{\text{f}}$1), notice that (f1) considers a different order between p and q. It is easy to see from (f2) that f is $p(\cdot) - 1$—superlinear at infinity—and we can deduce the existence of at least one weak solution via the mountain pass theorem. Again, to apply Theorem 9.3.5, we assume an additional hypothesis.

(f) f is odd with respect to its second variable, that is,

$$f(x,-t) = -f(x,t)$$

for a.e. $x \in \Omega$ and all $t \in \mathbb{R}$.

More importantly, when applying Theorem 9.3.5 we are able to avoid imposing an order between the exponents p and q, other than the fact that q is subcritical. To be precise, instead of (f1) we assume

(f'1) There exists $k > 0$ such that f satisfies the growth condition (9.13) for a.e. $x \in \Omega$ and all $t \in \mathbb{R}$, where $q \in C_+(\overline{\Omega})$ with

$$1 < q^- \leq q^+ < p^\star.$$

As the reader can see, in (f'1) the values of the exponents p and q can be interleaved and this illustrates properly how much of a difference can make the utilization of the variable exponents instead of the constant ones.

In the case when f is $p(\cdot) - 1$—sublinear at infinity—we assume the following:

(f3) There exists $t_0 > 0$ such that $F(x,t_0) > 0$ for a.e. $x \in \Omega$.

(f4) $\displaystyle\lim_{|t|\to\infty} \frac{f(x,t)}{|t|^{p(x)-1}} = 0$ uniformly with respect to $x \in \Omega$.

Obviously, we deduce that f is $p(\cdot) - 1$—sublinear at infinity from (f4). Also, in order to have multiple solutions, we impose an additional hypothesis.

(f5) $\displaystyle\lim_{|t|\to 0} \frac{f(x,t)}{|t|^{p(x)-1}} = 0$ uniformly with respect to $x \in \Omega$.

Now that we have stated the assumptions, let us introduce the definition of the weak solutions, keeping in mind Theorem 9.2.5.

Definition 10 We say that $u \in V$ is a weak solution of the boundary value problem (9.6) if and only if

$$\int_\Omega a(x, \nabla u) \cdot \nabla v \, dx + \int_\Omega |u|^{p(x)-2} uv \, dx - \lambda \int_\Omega f(x, u) v \, dx = 0 \quad \text{for all } v \in V.$$

We will rely on the critical point theory for the weak solvability of the problems; therefore to (9.6), we associate a functional $I : V \to \mathbb{R}$,

$$I(u) = \int_\Omega A(x, \nabla u) \, dx + \int_\Omega \frac{1}{p(x)} |u|^{p(x)} \, dx - \lambda \int_\Omega F(x, u) \, dx. \tag{9.14}$$

Taking into consideration the above hypotheses, one can see that I is well defined and of class C^1 due to some properties from Sect. 9.2, like the inequalities involving the norm and the modular, the embedding theorems, and Theorem 9.2.2.

Note that, in the rest of the chapter we will not make a detailed reference to the results from Sect. 9.2, since they are heavily used in all the arguments and our intention is only to present and discuss the results concerning our problems, not to prove them. So the main role of Sect. 9.2 is to introduce the reader into the abstract framework corresponding to this type of problems and to let her/him see the main ingredients for a better understanding of the information behind the theorems. However, for every existence and multiplicity result presented in this chapter we will indicate the main argument, and, when it is the case, a key tool for the proof.

Being faithful to our strategy (that is, to apply the critical point theory), we reveal that

$$\langle I'(u), v \rangle = \int_\Omega a(x, \nabla u) \cdot \nabla v \, dx + \int_\Omega |u|^{p(x)-2} uv \, dx - \lambda \int_\Omega f(x, u) v \, dx,$$

for all $u, v \in V$. Hence the critical points of I are weak solutions to problem (9.6).

We will focus on obtaining critical points by means of the theorems recalled in Sect. 9.3. Thus, using Theorem 9.3.1, we arrive at our first existence theorem.

Theorem 9.4.2 ([15, Remark 2]) *Assume that hypotheses (p), (a0)–(a3), (f0), and ($\tilde{f}1$) hold. Then problem* (9.6) *has at least one nontrivial weak solution in V for every* $\lambda > 0$.

At the same time, Theorem 9.3.4 permits to achieve another existence result, with a different order between the exponents. Here we deal with the case when f has a $p(\cdot) - 1$—superlinear growth at infinity.

Theorem 9.4.3 ([15, Theorem 4]) *Assume that hypotheses (p), (a0)–(a3), and (f0)–(f2) hold. Then problem* (9.6) *has at least one nontrivial weak solution in V for every* $\lambda > 0$.

To apply Theorem 9.3.4, we show that the functional I satisfies the Palais–Smale condition. In the process, the following result is very useful.

Theorem 9.4.4 ([47, Theorem 4.1]) *Assume* $\Omega \subset \mathbb{R}^N$ *(*$N \geq 2$*) is a bounded domain with smooth boundary and* $p \in C(\overline{\Omega})$ *with* $1 < p^- \leq p^+ < \infty$*. Also, assume that the Carathéodory function* $a : \overline{\Omega} \times \mathbb{R}^N \to \mathbb{R}^N$ *fulfills hypotheses (a1), (a3) and*

(ã2) there exist $\alpha \in L^1(\Omega)$ *and* $\beta > 0$ *such that*

$$a(x,\xi)\cdot\xi \geq \beta|\xi|^{p(x)} - \alpha(x) \quad \text{for a.e } x \in \overline{\Omega} \text{ and all } \xi \in \mathbb{R}^N.$$

If $u_n \rightharpoonup u$ *(weakly) in* $W^{1,p(\cdot)}(\Omega)$ *and*

$$\limsup_{n\to\infty} \int_\Omega a(x,\nabla u_n)\cdot(\nabla u_n - \nabla u)\,dx \leq 0,$$

then $u_n \to u$ *(strongly) in* $W^{1,p(\cdot)}(\Omega)$.

Fortunately, by adding hypotheses (a4) and (f) to our set of assumptions, we can provide a multiplicity result which does not require a certain order between the exponents p and q, as we explained above. Hence, when f is $p(\cdot)-1$—superlinear at infinity—we rely on the fountain theorem to obtain infinitely many weak solutions to problem (9.6), for all $\lambda > 0$. Following the statement of Theorem 9.3.5, we take $X = V$, which is a separable and reflexive Banach space, see Theorem 9.2.6. Then, for X_i, Y_i, Z_i given by (9.12) and for I given by (9.14), we can use the fountain theorem to arrive at another main result.

Theorem 9.4.5 ([15, Theorem 10]) *Assume that hypotheses (p), (a0)–(a4), (f), (f0), (f'1), and (f2) hold. Then problem (9.6) has infinitely many weak solutions in* V *for every* $\lambda > 0$.

Besides Theorem 9.3.5, an important role in the proof of the previous theorem is represented by the next proposition.

Proposition 9.4.6 (see [14, Proposition 5]) *If for every* $i = 1, 2, \ldots$ *we denote*

$$\theta_i = \sup_{u\in Z_i,\ \|u\|\leq 1} \left| \int_\Omega \frac{1}{q(x)}|u|^{q(x)}\,dx \right|,$$

then when i goes to ∞, $\theta_i \to 0$.

We revisit the Weierstrass type theorem, that is, Theorem 9.3.1, to provide an existence result for the case when f has a $p(\cdot)-1$—sublinear growth at infinity.

Theorem 9.4.7 ([15, Theorem 11]) *Assume that hypotheses (p), (a0)–(a3), (f0), and (f3)–(f4) hold. Then there exists a constant* $\lambda_0 > 0$ *such that problem (9.6) has at least one nontrivial weak solution in* V *for every* $\lambda > \lambda_0$.

Other important ingredients for this proof are Proposition 9.3.2 and Proposition 9.3.3. Note that the case when f is $p(\cdot)-1$—sublinear at infinity—has received little attention in comparison to the case of the superlinear growth at infinity. To fill this gap in the mathematical literature, we also have two multiplicity results that are obtained by adding condition (f5) to the previous set of hypotheses.

Theorem 9.4.8 ([15, Theorem 13]) *Assume that hypotheses (p), (a0)–(a3), (f0), and (f3)–(f5) hold. Then problem* (9.6) *has at least two nontrivial weak solutions in* V *for every* $\lambda > \lambda_0 > 0$*, where* λ_0 *is the one found in Theorem 9.4.7, that is,*

$$\lambda_0 = \left[\sup_{u \in V, u \neq 0} \frac{\int_\Omega F(x,u)dx}{K(u)} \right]^{-1},$$

with K *being defined by*

$$K(u) = \int_\Omega A(x, \nabla u)\, dx + \int_\Omega \frac{1}{p(x)} |u|^{p(x)}\, dx. \tag{9.15}$$

An intriguing fact about the proof of this theorem is that it is based on a clever combination between Theorem 9.3.1 and Theorem 9.3.4. To give some details, after we find a first weak solution u_1 of problem (9.6) by means of the Weierstrass type theorem, this solution denoted by u_1 will play the role of e from the statement of Theorem 9.3.4. Then, it is shown that the functional I given by (9.14) verifies the requirements of Theorem 9.3.4 and a second nontrivial solution, distinct from u_1, is found. This is an interesting manner to use a mountain pass type theorem to obtain a multiplicity result without imposing the symmetry conditions (a4) and (f). Also, more importantly, we have applied a mountain pass theorem to a problem involving a nonlinearity which does not fulfill the Ambrosetti–Rabinowitz type condition.

Moving forward, we can use a different argument to infer a multiplicity result for problem (9.6) when we deal with the sublinear growth at infinity, that is, Theorem 9.3.6. We state the multiplicity result resulted from this theorem.

Theorem 9.4.9 ([15, Theorem 16]) *Assume that hypotheses (p), (a0)–(a3), (f0), and (f3)–(f5) hold. Then there exist an open interval* $\Lambda \subset [0, \alpha]$ *and a constant* $\sigma > 0$ *such that for all* $\lambda \in \Lambda$ *problem* (9.6) *has at least three weak solutions in* V *whose norms are less than* σ*, where*

$$\alpha = \frac{\beta r_0}{r_0 \int_\Omega F(x,v)dx / K(v) - \sup_{K(u) < r_0} \int_\Omega F(x,u)\, dx}$$

with $\beta > 1$ *and* $r_0 > 0$*,* $v \in V$ *such that* $r_0 < K(v)$*, where* K *is given by* (9.15).

Apparently under the same hypotheses as in Theorem 9.4.8 we have arrived at three solutions instead of two. But these are obtained when the values of λ are in a different set. Also, Theorem 9.4.9 does not say that the solutions are nontrivial, so we may end up with two nontrivial weak solutions after all.

9.5 Problems with Anisotropic Variable Exponent

In addition to the hypothesis that $\Omega \subset \mathbb{R}^N$ $(N \geq 2)$ is a bounded open set with smooth boundary that is assumed in the rest of the chapter, in this section we consider Ω to be a rectangular-like domain, which, as reminded in Sect. 9.2, is a union of finitely many rectangular domains (or cubes) with edges parallel to the coordinate axes. We discuss the weak solvability of the following class of anisotropic problems with no-flux boundary conditions:

$$\begin{cases} -\sum_{i=1}^{N} \partial_{x_i} a_i\left(x, \partial_{x_i} u\right) + b(x)|u|^{p_M(x)-2}u = \lambda g(x,u) & \text{for } x \in \Omega, \\ u = \text{constant} & \text{for } x \in \partial\Omega, \\ \int_{\partial\Omega} \sum_{i=1}^{N} a_i\left(x, \partial_{x_i} u\right) \nu_i \, dS = 0, \end{cases} \tag{9.16}$$

where $\lambda \geq 0$ and $\nu_i,\ i \in \{1, \ldots, N\}$, represent the components of the unit outer normal vector. For the functions involved in (9.16) we consider a set of assumptions that will be discussed in what follows.

(b) $b \in L^\infty(\Omega)$ and there exists $b_0 > 0$ such that $b(x) \geq b_0$ for all $x \in \Omega$.

For the components p_i, $i \in \{1, \ldots, N\}$, of the anisotropic variable exponent $\overrightarrow{p}(\cdot)$ we always assume a condition similar to (p), that is,

(p_i) $p_i : \overline{\Omega} \to (1, +\infty)$ with $1 < p_i^- \leq p_i^+ < \infty$ and each p_i fulfills the log-Hölder continuity condition (9.4), for all $i \in \{1, \ldots, N\}$.

Notice that in the statement of the theorems we will insert an additional hypothesis for $\overrightarrow{p}(\cdot)$ which will depend on the desired result.

We describe now the hypotheses verified by the functions a_i that provide our generalized operators.

(A0) For every $i \in \{1, \ldots, N\}$, $a_i : \overline{\Omega} \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function.
(A1) For every $i \in \{1, \ldots, N\}$, there exists a positive constant c_i such that a_i fulfills

$$|a_i(x,s)| \leq c_i \left(d_i(x) + |s|^{p_i(x)-1}\right),$$

for all $x \in \Omega$ and all $s \in \mathbb{R}$, where $d_i \in L^{p_i'(\cdot)}(\Omega)$ (with $1/p_i(x) + 1/p_i'(x) = 1$) is a nonnegative function.
(A2) For every $i \in \{1, \ldots, N\}$,

$$|s|^{p_i(x)} \leq p_i(x)\ A_i(x,s),$$

for all $x \in \Omega$ and all $s \in \mathbb{R}$, where $A_i : \overline{\Omega} \times \mathbb{R} \to \mathbb{R}$ is the antiderivative of a_i, that is,

$$A_i(x, s) = \int_0^s a_i(x, t)\, dt.$$

(A3) For every $i \in \{1, \ldots, N\}$, the monotonicity condition

$$[a_i(x, s) - a_i(x, t)](s - t) > 0$$

takes place for all $x \in \Omega$ and all $s,\ t \in \mathbb{R}$ with $s \neq t$.

($\tilde{\text{A}}$2) For every $i \in \{1, \ldots, N\}$,

$$|s|^{p_i(x)} \leq a_i(x, s)s,$$

for all $x \in \Omega$ and all $s \in \mathbb{R}$.

We mention that in most of the papers conditions (A2) and ($\tilde{\text{A}}$2) are combined into the more restrictive condition:

(A) For every $i \in \{1, \ldots, N\}$,

$$|s|^{p_i(x)} \leq a_i(x, s)s \leq p_i(x)\ A_i(x, s),$$

for all $x \in \Omega$ and all $s \in \mathbb{R}$,

see, for example, [12, 18, 44], and this is not surprising, since (A) is an extension to the anisotropic exponent of the condition (a2), from Sect. 9.2. However, a closer look, in, e.g., [18] is sufficient to impose (A2) and ($\tilde{\text{A}}$2) instead of (A), so we will slightly generalize the result from [18] in this manner in order to present it here. Moreover, sometimes we need only one of the assumptions (A2) and ($\tilde{\text{A}}$2), which is the case of some of the results that will be exposed in this section.

Furthermore, some papers add another condition to the above set of hypotheses, and we refer to the $\overrightarrow{p}(\cdot)$-uniform convexity condition: there exists $\kappa_i > 0$ such that

$$A_i\left(x, \frac{s+t}{2}\right) \leq \frac{1}{2}A_i(x, s) + \frac{1}{2}A_i(x, t) - \kappa_i|s - t|^{p_i(x)}, \tag{9.17}$$

for all $i \in \{1, \ldots, N\}$, all $x \in \Omega$, and all $s,\ t \in \mathbb{R}$, which is an extension of hypothesis (a) from Sect. 9.4. Similarly to the isotropic case, when imposing the $\overrightarrow{p}(\cdot)$-uniform convexity condition (9.17), the examples of operators that will be obtained below as particular cases of the Leray–Lions type operators

$$\sum_{i=1}^{N} \partial_{x_i} a_i \left(x, \partial_{x_i} u\right)$$

are not available for the case in which p_i are less than 2. This is not the case for the other hypotheses, that is, (A0)–(A3) and ($\tilde{\text{A}}2$), which allow us to take $p_i > 1$, $i \in \{1, \ldots, N\}$.

Let us state now some examples of operators that are produced by our Leray–Lions type operators. To this aim, we take again $h \in L^\infty(\Omega)$ with the property that there exists $h_0 > 0$ such that $h(x) \geq h_0$ for all $x \in \Omega$. Then, by choosing

$$a_i(x, s) = h(x)|s|^{p_i(x)-2}s \quad \text{for all } i \in \{1, \ldots, N\}, \tag{9.18}$$

hypotheses (A0)–(A3) and ($\tilde{\text{A}}2$) are verified for $p_i > 1$, $i \in \{1, \ldots, N\}$. This way we get the class of operators

$$\sum_{i=1}^{N} \partial_{x_i} a_i\left(x, \partial_{x_i} u\right) = \sum_{i=1}^{N} \partial_{x_i}\left(h(x)\left|\partial_{x_i} u\right|^{p_i(x)-2} \partial_{x_i} u\right)$$

and, when $h \equiv 1$, we arrive at a problem involving the $\overrightarrow{p}(\cdot)$-Laplace operator:

$$\begin{cases} -\sum_{i=1}^{N} \partial_{x_i}\left(|\partial_{x_i} u|^{p_i(x)-2}\partial_{x_i} u\right) + b(x)|u|^{p_M(x)-2}u = \lambda g(x, u) & \text{for } x \in \Omega, \\ u = \text{constant} & \text{for } x \in \partial\Omega, \\ \int_{\partial\Omega} \sum_{i=1}^{N} |\partial_{x_i} u|^{p_i(x)-2}\partial_{x_i} u \ \nu_i \, dS = 0. \end{cases}$$

On the other hand, when choosing

$$a_i(x, s) = h(x)\left(1 + |s|^2\right)^{(p_i(x)-2)/2} s \quad \text{for all } i \in \{1, \ldots, N\}, \tag{9.19}$$

hypotheses (A0)–(A3) and ($\tilde{\text{A}}2$) are verified for $p_i > 1$, $i \in \{1, \ldots, N\}$, and we are led to the following class of operators:

$$\sum_{i=1}^{N} \partial_{x_i} a_i\left(x, \partial_{x_i} u\right) = \sum_{i=1}^{N} \partial_{x_i}\left[h(x)\left(1 + \left|\partial_{x_i} u\right|^2\right)^{(p_i(x)-2)/2} \partial_{x_i} u\right].$$

By taking $h \equiv 1$ in the previous example we arrive at the following problem involving the generalized mean curvature operator:

$$\begin{cases} -\sum_{i=1}^{N} \partial_{x_i}\left(1 + |\partial_{x_i} u|^2\right)^{(p_i(x)-2)/2} \partial_{x_i} u + b(x)|u|^{p_M(x)-2}u = \lambda g(x, u) & \text{for } x \in \Omega, \\ u = \text{constant} & \text{for } x \in \partial\Omega, \\ \int_{\partial\Omega} \sum_{i=1}^{N} \left(1 + |\partial_{x_i} u|^2\right)^{(p_i(x)-2)/2} \partial_{x_i} u \ \nu_i \, dS = 0. \end{cases}$$

We have chosen to write explicitly these two particular cases of problem (9.16) because there are not many papers that treat anisotropic variable exponent problems with no-flux boundary. In fact, in addition to the papers [16, 18] which were the first to approach this subject and which are going to be discussed here, we are only aware of one other paper, that is, [1], in which g from (9.16) is replaced by $u^{q(x)-2}u$.

Let us see now the hypotheses on our nonlinearity g.

(g0) $g : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function.

(g1) There exist $k > 0$ and $q \in C(\overline{\Omega})$ with

$$p_M^+ < q^- < q^+ < \overline{p}^\star(x)$$

for all $x \in \overline{\Omega}$, such that g verifies

$$|g(x,s)| \leq k\left(1 + |s|^{q(x)-1}\right)$$

for a.e. $x \in \Omega$ and all $s \in \mathbb{R}$, where $\overline{p}^\star$ and p_M are given by formulae (9.9) and (9.11).

(g2) There exist $\gamma > p_M^+$ and $s_0 > 0$ such that the Ambrosetti–Rabinowitz condition

$$0 < \gamma G(x,s) \leq s g(x,s)$$

holds for a.e. $x \in \Omega$ and for all $s \in \mathbb{R}$ with $|s| > s_0$, where $G : \Omega \times \mathbb{R} \to \mathbb{R}$ denotes the antiderivative of g, that is,

$$G(x,s) = \int_0^s g(x,t)\,dt.$$

(g3) $\lim\limits_{|s|\to 0} \dfrac{g(x,s)}{|s|^{p_M^+-1}} = 0$ uniformly with respect to $x \in \Omega$.

At this point it is clear that problem (9.16) is a generalization of problem (9.6) that was analyzed in the previous section. Hypothesis (g2) implies, in particular, that g is $p_M^+ - 1$—superlinear at infinity—so, to create a situation that is analogous to the one from the isotropic variable exponent case, we display a hypothesis which implies that g is $p_m^- - 1$—sublinear at infinity.

(g4) $\lim\limits_{|s|\to\infty} \frac{g(x,s)}{|s|^{p_m^- -1}} = 0$ uniformly with respect to $x \in \Omega$.

Obviously, (g2) and (g4) will not be fulfilled at the same time. But under assumptions (g0) and (g4) one can obtain an existence result for problem (9.16). Still, for the nontriviality of this weak solution, we also need

(g5) There exists $s_0 > 0$ such that $G(x, s_0) > 0$ for a.e. $x \in \Omega$.

Furthermore, if we add (g3) to hypotheses (g0), (g4), and (g5), then we infer a multiplicity result. On the other hand, if instead of (g3) we add to them

(g6) $[g(x, s_1) - g(x, s_2)]\,(s_1 - s_2) < 0$ for a.e. $x \in \Omega$ and all $s_1, s_2 \in \mathbb{R}$ with $s_1 \neq s_2$,

then we infer a uniqueness result.

Taking into account Theorem 9.2.9, we introduce the notion of weak solution to problem (9.16).

Definition 11 We say that $u \in \overrightarrow{V}$ is a weak solution of the boundary value problem (9.16) if and only if

$$\int_\Omega \sum_{i=1}^N a_i(x, \partial_{x_i} u)\partial_{x_i} v \, dx + \int_\Omega b(x)|u|^{p_M(x)-2} uv \, dx - \lambda \int_\Omega g(x, u)v \, dx = 0,$$

for all $v \in \overrightarrow{V}$.

The energy functional corresponding to (9.16) is defined as $J : \overrightarrow{V} \to \mathbb{R}$,

$$J(u) = \int_\Omega \sum_{i=1}^N A_i(x, \partial_{x_i} u) \, dx + \int_\Omega \frac{b(x)}{p_M(x)} |u|^{p_M(x)} \, dx - \lambda \int_\Omega G(x, u) \, dx.$$

Based on the properties from Sect. 9.2, a standard calculus shows that functional J is well defined and of class C^1, and its derivative is described by

$$\langle J'(u), v\rangle = \int_\Omega \sum_{i=1}^N a_i(x, \partial_{x_i} u)\partial_{x_i} v \, dx + \int_\Omega b(x)|u|^{p_M(x)-2} uv \, dx - \lambda \int_\Omega g(x, u)v \, dx,$$

for all $u, v \in \overrightarrow{V}$. Since the critical points of J are weak solutions to problem (9.16), we rely on the critical point theorems from Sect. 9.3 for the weak solvability of problem (9.16). Thus, due to Theorem 9.3.4, we deduce the following result.

Theorem 9.5.2 (see [18, Theorem 3.2]) *Assume that hypotheses (b), (p_i), (A0)–(A3), ($\tilde{A}$2), and (g0)–(g3) hold, and, in addition, $p_M^+ < \overline{p}^\star(x)$ for all $x \in \Omega$. Then, problem (9.16) has at least one nontrivial weak solution in $\overrightarrow{V}$ for every $\lambda > 0$.*

As already said, in [18, Theorem 3.2] the authors imposed the more restrictive assumption (A), but the calculus works just fine when we replace (A) by conditions (A2) and ($\tilde{A}$2). Also, note that an important ingredient of the proof of Theorem 9.5.2 is represented by the next result.

Theorem 9.5.3 (see [18, Theorem 3.4]) *Let $\Omega \subset \mathbb{R}^N$, ($N \geq 2$) be a rectangular-like domain. Assume that $p_i \in C(\overline{\Omega})$, $1 < p_i^- \leq p_i^+ < \infty$ for all $i \in \{1, \ldots, N\}$*

with $p_M^+ < \overline{p}^\star(x)$*, and that (A0), (A1), (*$\tilde{A}2$*), and (A3) hold. If* $u_n \rightharpoonup u$ *(weakly) in* $W^{1,\overrightarrow{p}(\cdot)}(\Omega)$ *and*

$$\limsup_{n\to\infty} \int_\Omega \sum_{i=1}^N a_i(x, \partial_{x_i} u)(\partial_{x_i} u_n - \partial_{x_i} u)dx \le 0,$$

then $u_n \to u$ *(strongly) in* $W^{1,\overrightarrow{p}(\cdot)}(\Omega)$.

Notice that we have been a little more careful than the authors of [18, Theorem 3.4] when stating the assumptions on the Leray–Lions type operators in the above theorem. For details regarding the possibility to state a slightly improved version of Theorem 9.5.3, we send the reader to the comments from the next section.

Moving forward, to the case when g is $p_m^- - 1$—sublinear at infinity—we have another existence result, this time due to Theorem 9.3.1 combined with Propositions 9.3.2 and 9.3.3.

Theorem 9.5.4 (see [16, Theorem 3.2]) *Assume that hypotheses (b), (*p_i*), (A0)–(A3), (g0), and (g4) hold and* $p_m^- < p_M^-$*. Then problem (9.16) has at least one weak solution in* $\overrightarrow{V}$ *for every* $\lambda > 0$.

This theorem, as well as the others from the remaining part of this section, were given under conditions that are a bit more general in [16]. The details are spelled out in Sect. 9.6.

The next result concerns the nontriviality of the solution.

Corollary 9.5.5 (see [16, Corollary 1]) *Assume that hypotheses (b), (*p_i*), (A0)–(A3), (g0), (g4), and (g5) hold and* $p_m^- < p_M^-$*. Then there exists*

$$\lambda_0 = \left[\sup_{u\in \overrightarrow{V},\, u\neq 0} \frac{J_2(u)}{J_1(u)} \right]^{-1} \tag{9.20}$$

such that problem (9.16) admits a nontrivial weak solution in $\overrightarrow{V}$ *for every* $\lambda > \lambda_0$*, where*

$$J_1(u) = \int_\Omega \sum_{i=1}^N A_i(x, \partial_{x_i} u)\, dx + \int_\Omega \frac{b(x)}{p_M(x)} |u|^{p_M(x)}\, dx$$

and

$$J_2(u) = \lambda \int_\Omega F(x, u)\, dx.$$

This corollary let us know that, by adding another assumption to the set of hypotheses from Theorem 9.5.4, we can guarantee the nontriviality of the weak solution, but only for $\lambda > \lambda_0 > 0$, where λ_0 is given by (9.20). Still, one can see that under certain choices of the functions g and a_i, the weak solution is nontrivial for all $\lambda > 0$. A similar remark could be made for the uniqueness results from below.

Corollary 9.5.6 (see [16, Corollary 2]) *Assume that hypotheses (b), (p_i), (A0)–(A3), (g0), (g4), and (g6) hold and $p_m^- < p_M^-$. Then problem* (9.16) *admits a unique weak solution in $\overrightarrow{V}$ for every $\lambda > 0$. Moreover, if we add hypothesis (g5), the unique solution is nontrivial for all $\lambda > \lambda_0 > 0$, where λ_0 is given by* (9.20)*.*

Finally, we arrive at the multiplicity result, which yields when putting together Theorems 9.3.1 and 9.3.4, as in the isotropic case.

Theorem 9.5.7 (see [16, Theorem 3.3]) *Assume that hypotheses (b), (p_i), (A0)–(A3), $(\tilde{A}2)$, (g0), and (g3)–(g5) hold and $p_M^+ < \overline{p}^\star(x)$ for all $x \in \Omega$. Then, problem* (9.16) *has at least two nontrivial weak solutions in $\overrightarrow{V}$ for every $\lambda > \lambda_0 > 0$, where λ_0 is given by* (9.20)*.*

More comments and remarks on the above results can be found in the subsequent section.

9.6 Final Comments

Naturally, there is a continuous preoccupation with relaxing the hypotheses and generalizing the results. One can easily see that not all the results from Sect. 9.4 were extended to the anisotropic case. And, although several difficulties will occur on the way, we have confidence that this work can be done and it represents a possible future direction of research. But we should be careful though, because it is not an easy ride. For example, we are not aware of a proof of the fact that the space $W^{1,\overrightarrow{p}(\cdot)}(\Omega)$ is separable. Thus, even though $\overrightarrow{V}$ is a closed subspace of $\left(W^{1,\overrightarrow{p}(\cdot)}(\Omega), \|\cdot\|_{W^{1,\overrightarrow{p}(\cdot)}(\Omega)}\right)$, it cannot inherit this property. Without it, we are not able to apply the fountain theorem (see Theorem 9.3.5 and the lines above it) in the situation in which we add the symmetry conditions

($\ddot{g}$) g is odd with respect to its second variable, that is,

$$g(x, -t) = -g(x, t)$$

for a.e. $x \in \Omega$ and all $t \in \mathbb{R}$.

($\ddot{A}$) For every $i \in \{1, \ldots, N\}$, a_i is odd with respect to its second variable, that is,

$$a_i(x, -t) = -a_i(x, t)$$

for a.e. $x \in \Omega$ and all $t \in \mathbb{R}$

to the set of hypotheses. However, if we assume conditions ($\ddot{g}$) and ($\ddot{A}$), we open the door to obtaining infinitely many solutions by means of the following symmetric mountain pass theorem.

Theorem 9.6.1 ([42, Theorem 11.5]) *. Let X be a real infinite dimensional Banach space and* $\Phi \in C^1(X;\mathbb{R})$ *a functional satisfying the Palais–Smale condition. Assume that* Φ *satisfies:*

(i) $\Phi(0) = 0$ *and there are constants* $\rho, r > 0$ *such that*

$$\Phi_{|\partial B_\rho} \geq r,$$

(ii) Φ *is even, and*
(iii) for all finite dimensional subspaces $\widetilde{X} \subset X$*, there exists* $R = R(\widetilde{X}) > 0$ *such that*

$$var Phi(u) \leq 0 \, for \, u \in \widetilde{X} \setminus B_R(\widetilde{X}).$$

Then Φ *possesses an unbounded sequence of critical values characterized by a min–max argument.*

Note that functions from (9.18) and (9.19) satisfy condition ($\ddot{A}$); thus, the examples of operators presented in Sect. 9.5 would remain valid in the above described eventuality.

Also, when referring to extending Theorem 9.4.9 to the anisotropic case, one should take into account more recent and refined versions of the three critical points result from Theorem 9.3.6, see [54, 59, 60] and the references therein.

This strong connection between the results from Sect. 9.4 and the results from Sect. 9.5 works both ways. On the one hand, we can look at the results from the isotropic case and see which of them could be extended to the anisotropic case. On the other hand, by looking at the generalized results from Sect. 9.5 we can deduce which theorems from the isotropic case could of been improved. Actually, this is the case of all the results from Sect. 9.4 because we can follow the model from problem (9.16) to replace problem (9.6) by

$$\begin{cases} -\text{div}(a(x,\nabla u)) + b(x)|u|^{p(x)-2}u = \lambda f(x,u) & \text{for } x \in \Omega, \\ u(x) = \text{constant} & \text{for } x \in \partial\Omega, \\ \displaystyle\int_{\partial\Omega} a(x,\nabla u)\cdot \nu \, dS = 0, & \end{cases}$$

where the change is represented by the presence of b, which is given by (b) from Sect. 9.5. Moreover, assumption (A1) indicates that (a1) can be upgraded as follows:

($\tilde{a}1$) There exists $\tilde{c} > 0$ such that a satisfies the growth condition

$$|a(x,\xi)| \leq \tilde{c}(d(x) + |\xi|^{p(x)-1})$$

for all $x \in \Omega$ and all $\xi \in \mathbb{R}^N$, where $d \in L^{p'(\cdot)}(\Omega)$ is a nonnegative function.

In addition, (a2) can be split into

$$|\xi|^{p(x)} \leq p(x)A(x,\xi) \quad \text{for a.e } x \in \Omega \text{ and all } \xi \in \mathbb{R}^N \tag{9.21}$$

and

$$|\xi|^{p(x)} \leq a(x,\xi)\cdot\xi \quad \text{for a.e } x \in \Omega \text{ and all } \xi \in \mathbb{R}^N, \tag{9.22}$$

as in Sect. 9.5, where we have their generalizations as (A2) and (Ã2). Since all the main results from Sect. 9.5 work under (A2) and/or (Ã2) instead of (A), we expect the corresponding results from Sect. 9.4 to work under relations (9.21) and/or (9.22). Obviously, even when combined, relations (9.21) and (9.22) are less restrictive than (a2). Furthermore, since Theorem 9.4.4 is provided under the more general condition (ã2), we believe that Theorem 9.5.3 can be also improved by replacing condition (Ã2) from its statement with

(Ã'2) For every $i \in \{1,\ldots,N\}$, there exist $\alpha_i \in L^1(\Omega)$ and $\beta_i > 0$ such that

$$\beta_i|s|^{p_i(x)} - \alpha_i(x) \leq a_i(x,s)s,$$

for all $x \in \Omega$ and all $s \in \mathbb{R}$.

In fact, the authors of [16] have used a version that is more general than (Ã2) and it is a particular case of (Ã'2), obtained for $\alpha_i \equiv 0$. But we believe that, in order for Theorem 9.5.3 to hold under a more general hypothesis like (Ã'2) we need to impose an additional condition regarding the relation between the coefficients β_i from (Ã'2) and c_i from (A1), see the analogous proof from [13]. In addition, in [16] assumption (A2) is replaced by the more relaxed assumption

(A'2) For every $i \in \{1,\ldots,N\}$, there exists $k_i > 0$ such that

$$k_i|s|^{p_i(x)} \leq p_i(x)\ A_i(x,s),$$

for all $x \in \Omega$ and all $s \in \mathbb{R}$.

Without getting too technical concerning the assumptions of the problems (if it is not already too late) we have two other remarks. The first one concerns Theorem 9.5.7, in which hypothesis (g3) can be replaced by the more general

(g̃3) $\displaystyle\lim_{|s|\to 0}\frac{g(x,s)}{|s|^{p_m^- -1}} = 0$ uniformly with respect to $x \in \Omega$,

as it is written in [16, Theorem 3.3]. Here we have used (g3) just because it was needed for Theorem 9.5.2 and we tried to simplify the set of assumptions from Sect. 9.5. The second one refers to conditions (f3) and (g5), which are in fact the same condition, and can be replaced by

(f̃3) there exist $t_0 > 0$ and a ball B with $\overline{B} \subset \Omega$ such that $\int_B F(x,t_0)\,dx > 0$, where $F(x,t) = \int_0^t f(x,s)\,ds$,

see [20, 41].

Apart from the above observations on the hypotheses, we remark that the variable exponent problems with no-flux boundary conditions are gaining popularity. As an illustration, there is a fresh interest for higher order problems with no-flux boundary conditions, see [20, 71] for the first papers on this research direction. Also, we refer to the very recent study [13] for fourth-order variable exponent problems involving Leray–Lions type operators. Hence one can investigate whether similar results to those presented here could be adapted to the new context. The same question could be asked about a possible generalization of the results from this chapter to systems, and we send the reader to [67] for systems involving $p(\cdot)$ and $q(\cdot)$-Laplace operators and no-flux boundary conditions, and to [11] for anisotropic systems with variable exponents and Leray–Lions type operators.

In conclusion, there are multiple things to consider for generalization in the future; after all, in mathematics there is always some "work in progress."

References

1. G.A. Afrouzi, M. Mirzapour, V.D. Rădulescu, The variational analysis of a nonlinear anisotropic problem with no-flux boundary condition. Rev. R. Acad. Cienc. Exactas Fis. Nat. Ser. A Math. RACSAM **109**, 581–595 (2015)
2. A. Ambrosetti, P.H. Rabinowitz, Dual variational methods in critical point theory. J. Funct. Anal. **14**, 349–381 (1973)
3. S.N. Antontsev, J.F. Rodrigues, On stationary thermorheological viscous flows. Ann. Univ. Ferrara Sez. VII Sci. Mat. **52**, 19–36 (2006)
4. M. Avci, A. Pankov, Multivalued elliptic operators with nonstandard growth. Adv. Nonlinear Anal. **7**, 35–48 (2018)
5. M. Avci, R. Ayazoglu (Mashiyev), B. Cekic, Solutions of an anisotropic nonlocal problem involving variable exponent. Adv. Nonlinear Anal. **2**, 325–338 (2013)
6. M. Bendahmane, K.H. Karlsen, Renormalized solutions of an anisotropic reaction-diffusion-advection system with L^1-data. Commun. Pure Appl. Anal. **5**, 733–762 (2006)
7. M. Bendahmane, M. Langlais, M. Saad, On some anisotropic reaction-diffusion systems with L^1-data modeling the propagation of an epidemic disease. Nonlinear Anal. **54**, 617–636 (2003)
8. B. Berestycki, H. Brezis, On a free boundary problem arising in plasma physics. Nonlinear Anal. TMA **4**, 415–436 (1980)
9. M. Bojowald, H. Hernandez, H. Morales-Tecotl, A perturbative degrees of freedom in loop quantum gravity: anisotropies. Class. Quantum Grav. **23**, 3491–3516 (2006)
10. G. Bonanno, Some remarks on a three critical points theorem. Nonlinear Anal. TMA **54**, 651–665 (2003)
11. M.-M. Boureanu, A new class of nonhomogeneous differential operator and applications to anisotropic systems. Complex Var. Elliptic Equ. **61**, 712–730 (2016)
12. M.-M. Boureanu, Infinitely many solutions for a class of degenerate anisotropic elliptic problems with variable exponent. Taiwan. J. Math. **5**, 2291–2310 (2011)
13. M.-M. Boureanu, Fourth-order problems with Leray-Lions type operators in variable exponent spaces. Discrete Contin. Dyn. Syst. Ser. S **12**(2), 231–243 (2019). https://doi.org/10.3934/dcdss.2019016
14. M.-M. Boureanu, F. Preda, Infinitely many solutions for elliptic problems with variable exponent and nonlinear boundary conditions. Nonl. Diff. Eq. Appl. (NoDEA) **19**, 235–251 (2012)

15. M.M. Boureanu, D.N. Udrea, Existence and multiplicity results for elliptic problems with $p(\cdot)$ - growth conditions. Nonl. Anal. RWA **14**, 1829–1844 (2013)
16. M.M Boureanu, C. Udrea, No–flux boundary value problems with anisotropic variable exponents. Commun. Pure Appl. Anal. **14**, 881–896 (2015)
17. M.-M. Boureanu, P. Pucci, V. Rădulescu, Multiplicity of solutions for a class of anisotropic elliptic equations with variable exponent. Complex Var. Elliptic Equ. **56**, 755–767 (2011)
18. M.-M. Boureanu, C. Udrea, D.-N. Udrea, Anisotropic problems with variable exponents and constant Dirichlet condition. Electron. J. Diff. Equ. **2013**(220), 1–13 (2013)
19. M.M. Boureanu, A. Matei, M. Sofonea, Nonlinear problems with $p(\cdot)$-growth conditions and applications to antiplane contact models. Adv. Nonl. Studies **14**, 295–313 (2014)
20. M.-M. Boureanu, V. Rădulescu, D. Repovš, On a $p(\cdot)$-biharmonic problem with no-flux boundary condition. Comput. Math. Appl. **72**, 2505–2515 (2016)
21. H. Brezis, J.M. Coron, L. Nirenberg, Free vibrations for a nonlinear wave equation and a theorem of P. Rabinowitz. Commun. Pure Appl. Math. **33**, 667–684 (1980)
22. Y. Chen, S. Levine, R. Rao, Variable exponent, linear growth functionals in image restoration. SIAM J. Appl. Math. **66**, 1383–1406 (2006)
23. N.T. Chung, H.Q. Toan, On a class of anisotropic elliptic equations without Ambrosetti-Rabinowitz type conditions. Nonlinear Anal. Real World Appl. **16**, 132–145 (2014)
24. D.G. Costa, *An Invitation to Variational Methods in Differential Equations* (Birkhäuser, Boston, 2007)
25. D. Cruz-Uribe, A. Fiorenza, *Variable Lebesgue Spaces. Foundations and Harmonic Analysis* (Birkhäuser/Springer, New York, 2013)
26. L. Diening, Theoretical and numerical results for electrorheological fluids, PhD Dissertation, University of Freiburg, Germany, 2002
27. L. Diening, Maximal function on generalized Lebesgue spaces $L^{p(\cdot)}$. Math. Inequal. Appl. **7**, 245–253 (2004)
28. L. Diening, P. Harjulehto, P. Hästö, M. Růžička, *Lebesgue and Sobolev Spaces with Variable Exponents.* Lecture Notes in Mathematics, vol. 2017 (Springer-Verlag, Berlin, 2011)
29. E. Eisenriegler, Anisotropic colloidal particles in critical fluids. J. Chem. Phys. **121**, 32–99 (2004)
30. E. Eisenriegler, Anisotropic colloidal particles interacting with polymers in a good solvent. J. Chem. Phys. **124**, 144–212 (2006)
31. L. Euler, *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes sive solutio problematis isoperimetrici latissimo sensu.* (Lausanne-Genève, 1744), Opera, Lausanne-Genève, Ser.I, vol. 24 (edited by C. Carathéodory), Berne (1952)
32. X. Fan, Solutions for $p(x)$–Laplacian Dirichlet problems with singular coefficients. J. Math. Anal. Appl. **312**, 464–477 (2005)
33. X. Fan, Anisotropic variable exponent Sobolev spaces and $\overrightarrow{p}(\cdot)$–Laplacian equations. Complex Var. Elliptic Equ. **55**, 1–20 (2010)
34. X. Fan, S.-G. Deng, Remarks on Ricceri's variational principle and applications to the $p(x)$ - Laplacian equations. Nonlinear Anal. TMA **67**, 3064–3075 (2007)
35. X. Fan, X. Han, Existence and multiplicity of solutions for $p(x)$-Laplacian equations in $\mathbb{R}^N$. Nonlinear Anal. **59**, 173–188 (2004)
36. X. Fan, D. Zhao, On the spaces $L^{p(x)}(\Omega)$ and $W^{m,p(x)}(\Omega)$. J. Math. Anal. Appl. **263**, 424–446 (2001)
37. G. Fragnelli, Positive periodic solutions for a system of anisotropic parabolic equations. J. Math. Anal. Appl. **367**, 204–228 (2010)
38. J. Garnier, High-frequency asymptotics for Maxwell's equations in anisotropic media, Part I: linear geometric and diffractive optics. J. Math. Phys. **42**, 1612–1635 (2001)
39. J. Garnier, High-frequency asymptotics for Maxwell's equations in anisotropic media, Part II: nonlinear propagation and frequency conversion. J. Math. Phys. **42**, 1636–1654 (2001)
40. K. Ho, I. Sim, A-priori bounds and existence for solutions of weighted elliptic equations with a convection term. Adv. Nonlinear Anal. **6**, 427–445 (2017)

41. K. Ho, I. Sim, Existence results for degenerate $p(x)$-Laplace equations with Leray-Lions type operators. Sci. China Math. **60**, 133–146 (2017)
42. Y. Jabri, *The Mountain Pass Theorem. Variants, Generalizations and some Applications* (Cambridge University Press, Cambridge, 2003)
43. D. Kinderlehrer, J. Spruck, The shape and smoothness of stable plasma configurations. Annali della SNS **5**, 131–148 (1978)
44. B. Kone, S. Ouaro, S. Traore, Weak solutions for anisotropic nonlinear elliptic equations with variable exponents. Electr. J. Differ. Equ. **2009**(144), 1–11 (2009)
45. O. Kováčik, J. Rákosník, On spaces $L^{p(x)}$ and $W^{k,p(x)}$. Czechoslovak Math. J. **41**, 592–618 (1991)
46. A.J. Kurdila, M. Zabarankin, *Convex Functional Analysis* (Birkhäuser Verlag, Basel, 2005)
47. V.K. Le, On a sub-supersolution method for variational inequalities with Leray-Lions operators in variable exponent spaces. Nonlinear Anal. TMA **71**, 3305–3321 (2009)
48. V.K. Le, K. Schmitt, Sub-supersolution theorems for quasilinear elliptic problems: a variational approach. Electr. J. Differ. Equ. **2004**(118), 1–7 (2004)
49. J. Leray, J.-L. Lions, Quelques résultats de Višik sur les problèmes elliptiques non linéaires par les méthodes de Minty-Browder. Bull. Soc. Math. Fr. **93**, 97–107 (1965)
50. Y. Liu, R. Davidson, P. Taylor, Investigation of the touch sensitivity of ER fluid based tactile display, in *Proceedings of SPIE, Smart Structures and Materials: Smart Structures and Integrated Systems*, vol. 5764 (International Society for Optics and Photonics, Bellingham, 2005), pp. 92–99
51. Y. Liu, Y. Tian, W. He, Q. Chen, The existence of solution for p(x)-Laplacian equation with no ux boundary. Sci. Mag. **6**, 102–107 (2010)
52. R.A. Mashiyev, B. Cekic, M. Avci, Z. Yucedag, Existence and multiplicity of weak solutions for nonuniformly elliptic equations with nonstandard growth condition. Complex Var. Elliptic Equ. Int. J. **57**, 579–595 (2012)
53. M. Mihăilescu, V. Rădulescu, A multiplicity result for a nonlinear degenerate problem arising in the theory of electrorheological fluids. Proc. Roy. Soc. Lond. Ser. A **462**, 2625–2641 (2006)
54. G. Molica Bisci, D. Repovš, Multiple solutions of p-biharmonic equations with Navier boundary conditions. Complex Var. Elliptic Equ. **59**(2), 271–284 (2014)
55. P. Pucci, V. Rădulescu, The impact of the mountain pass theory in nonlinear analysis: a mathematical survey. Boll. Unione Mat. Ital. Series IX **3**, 543–584 (2010)
56. J.P. Puel, Sur un problème de valeur propre non linéaire et de frontière libre. C.R. Ac. Sc. **284**, 861–863 (1977)
57. V. Rădulescu, D. Repovš, *Partial Differential Equations with Variable Exponents: Variational Methods and Quantitative Analysis* (CRC Press, Taylor & Francis Group, Boca Raton, 2015)
58. B. Ricceri, On a three critical points theorem. Arch. Math. (Basel) **75**, 220–226 (2000)
59. B. Ricceri, A further three critical points theorem. Nonlinear Anal. **71**, 4151–4157 (2009)
60. B. Ricceri, A further refinement of a three critical points theorem. Nonlinear Anal. **74**, 7446–7454 (2011)
61. M. Růžička, *Electrorheological Fluids: Modeling and Mathematical Theory* (Springer–Verlag, Berlin, 2002)
62. D. Schaeffer, Non-uniqueness in the equilibrium shape of a confined plasma. Comm. Part. Diff, Equ. **2**, 587–600 (1977)
63. R. Temam, A non-linear eigenvalue problem: the shape at equilibrium of a confined plasma. Arch. Rat. Mech. Anal. **60**, 51–73 (1975)
64. R. Temam, Remarks on a free boundary value problem arising in plasma physics. Comm. Partial Differ. Equ. **2**, 563–585 (1977)
65. J. Weickert, *Anisotropic Diffusion in Image Processing* (Teubner-Verlag, Stuttgart, 1998)
66. M. Willem, *Minimax Theorems* (Birkhäuser, Boston, 1996)
67. Q. Zhang, Y. Guo, G. Chen, Existence and multiple solutions for a variable exponent system. Nonlinear Anal. TMA **73**, 3788–3804 (2010)
68. J.F. Zhao, *Structure Theory of Banach Spaces* (Wuhan University Press, Wuhan, 1991) [in Chinese]

69. L. Zhao, P. Zhao, X. Xie, Existence and multiplicity of solutions for divergence type elliptic equations. Electr. J. Differ. Equ. **2011**(43), 1–9 (2011)
70. V.V. Zhikov, Averaging of functionals in the calculus of variations and elasticity. Math. USSR Izv. **29**, 33–66 (1987)
71. Q.-M. Zhou, K.-Q. Wang,*Multiple solutions to a class of $p(x)$-biharmonic differential inclusion problem with no-flux boundary condition*. RACSAM (Springer, Milano, 2017)
72. W. Zou, F. Li, B. Lv, On a nonlocal problem for a confined plasma in a Tokamak. Appl. Math. **58**, 609–642 (2013)
73. W. Zou, F. Li, M. Liu, B. Lv, Existence of solutions for a nonlocal problem arising in plasma physics. J. Differ. Equ. **256**, 1653–1682 (2014)

Chapter 10
On the General Decay for a System of Viscoelastic Wave Equations

Salim A. Messaoudi and Jamilu Hashim Hassan

Abstract This work is concerned with a coupled system of nonlinear viscoelastic wave equations that models the interaction of two viscoelastic fields. This system has been extensively studied by many authors for relaxation functions decaying exponentially, polynomially, or with some general decay rate. We prove a new general decay result that improves most of the existing results in the literature related to the system of viscoelastic wave equations. Our result allows wider classes of relaxation functions.

Keywords Viscoelastic · System · Relaxation function · General decay

10.1 Introduction

In this work, we consider the following coupled system of viscoelastic wave equations:

$$\begin{cases} u_{tt} - \Delta u + \int_0^t g_1(t-s)\Delta u(\cdot,s)ds + f_1(u,v) = 0, & \text{in } \Omega \times (0,+\infty), \\ v_{tt} - \Delta v + \int_0^t g_2(t-s)\Delta v(\cdot,s)ds + f_2(u,v) = 0, & \text{in } \Omega \times (0,+\infty), \\ u = v = 0, & \text{on } \partial\Omega \times [0,+\infty), \\ u(\cdot,0) = u_0, \quad u_t(\cdot,0) = u_1, \quad v(\cdot,0) = v_0, \quad v_t(\cdot,0) = v_1, & \text{in } \Omega, \end{cases} \tag{P}$$

where Ω is a bounded domain of $\mathbb{R}^n$ with a smooth boundary $\partial\Omega$, $u_0,\ u_1,\ v_0,\ v_1$ are given initial data, $g_1,\ g_2$ are the relaxation functions, and $f_1,\ f_2$ are the

S. A. Messaoudi (✉)
Department of Mathematics, University of Sharjah, Sharjah, United Arab Emirates
e-mail: smessaoudi@sharjah.ac.ae

J. H. Hassan
Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_10

nonlinear functions to be specified in the next section. The unknowns u and v represent the displacements of waves. This system can be considered as a generalization of the well-known Klein–Gordon system that appears in the quantum field theory. For more details, see [3, 17, 29].

We start with some related results in viscoelastic wave equations in order to motivate our work. For almost a half century, viscoelastic equations had been extensively studied by many researchers since the pioneer work of Dafermos [9, 10] in which he investigated a one-dimensional viscoelastic equation and proved the well-posedness of the problem provided that the relaxation function is a positive integrable function. He also established that its solution decays asymptotically to zero if, in addition, the relaxation function is a monotone non-increasing smooth function. However, the rate of decay of the solution was not explicitly given. Hrusa [13] in 1985 considered the following one-dimensional viscoelastic problem with nonlinearity in the memory term:

$$\begin{cases} u_{tt} - cu_{xx} + \displaystyle\int_0^t g(t-s)\big(\psi(u_x(x,s))\big)_x ds = f(x,t), & \text{in} \quad (0,1)\times(0,\infty), \\ u(0,t) = u(1,t) = 0, & t \geq 0, \\ u(x,0) = u_0(x), \qquad u_t(x,0) = u_1(x), & x \in [0,1], \end{cases} \tag{10.1.1}$$

where $c > 0$ is a constant, $u_0,\ u_1$ are given initial data, and ψ is a nonlinear function. Under certain conditions on ψ, he established the global existence of a strong solution to problem (10.1.1) and showed that the solution decays exponentially to zero, if the relaxation function g decays exponentially to zero.

For multi-dimensional viscoelastic problems, we start with the work of Dassios and Zafiropoulos [11] in 1990, in which the authors studied a three-dimensional viscoelastic problem in the whole space $\mathbb{R}^3$ and proved a polynomial decay result for an exponentially decaying relaxation function. In 1994, Rivera [22] established an exponential decay result for the sum of the first and second energies of a linear viscoelastic problem in a bounded domain of $\mathbb{R}^n$ with an exponentially decaying relaxation function by imposing some extra conditions on the second derivative of the relaxation function. Rivera and Lapa [23] improved this result by proving a polynomial decay rate of the system with a relaxation function that decays polynomially. In 2002, Cavalcanti et al. [8] studied the following problem:

$$\begin{cases} u_{tt} - \Delta u + \displaystyle\int_0^t g(t-s)\Delta u(s)ds + \gamma(x)u_t = 0, & \text{in} \quad \Omega\times(0,\infty), \\ u = 0, & \text{on} \quad \partial\Omega\times[0,\infty), \\ u(\cdot,0) = u_0, \quad u_t(\cdot,0) = u_1, & \text{in} \quad \Omega, \end{cases} \tag{10.1.2}$$

where Ω is a bounded domain of $\mathbb{R}^n$ $(n \geq 1)$ with a smooth boundary $\partial\Omega$ and $\gamma : \Omega \longrightarrow \mathbb{R}_+$ is bounded and satisfies

$$\gamma(x) \geq \gamma_0 \quad a.e. \quad \text{on} \quad \omega \subset \Omega.$$

They imposed the following assumptions on the relaxation function, g:

$$g(0) > 0, \qquad \int_0^\infty g(s)ds < 1,$$

and there exist two positive constants $\xi_1,\ \xi_2$ such that

$$-\xi_1 g(t) \leq g'(t) \leq -\xi_2 g(t), \qquad \forall\, t \geq 0. \tag{10.1.3}$$

They proved an exponential rate of decay for the solution of (10.1.2) under some geometric conditions on ω. Berrimi and Messaoudi [5] showed that one can drop the geometric condition imposed on ω in [8] and still maintain the exponential decay of the solution of (10.1.2). They established their result under weaker conditions on g. Furthermore, the same authors in [6] extended and improved their result to the case where a source term is competing with a viscoelastic damping.

Up to the year 2008, most of the studies of viscoelastic problems were concerned with relaxation functions satisfying

$$g'(t) \leq -\xi g^p(t), \qquad \forall\, t \geq 0, \tag{10.1.4}$$

where $\xi > 0$ and $1 \leq p < \frac{3}{2}$ which, in turn, yielded either uniform or polynomial decay. In 2008, Messaoudi [18, 19] proved a general decay rate from which the exponential and polynomial decay rates are only special cases. Precisely, he studied the following problem:

$$\begin{cases} u_{tt} - \Delta u + \displaystyle\int_0^t g(t-s)\Delta u(s)ds = \gamma |u|^{m-2}u, & \text{in } \ \Omega \times (0,\infty), \\ u = 0, & \text{on } \ \partial\Omega \times [0,\infty), \\ u(\cdot,0) = u_0, \quad u_t(\cdot,0) = u_1, & \text{in } \ \Omega, \end{cases} \tag{10.1.5}$$

with $\gamma = 0$ or $\gamma = 1$ and g satisfying

$$g'(t) \leq -\xi(t)g(t), \qquad \forall\, t \geq 0, \tag{10.1.6}$$

where $\xi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is a non-increasing differentiable function and showed that the energy of the solution of (10.1.5) decays with the same rate as g. Motivated by these results of Messaoudi, many general decay results using (10.1.6) have been established, see Cao [7], Han and Wang [12], Liu [14, 15], and references therein.

In 2009, Alabau-Boussouira and Cannarsa [2] announced, without a proof, a general decay result for the solution of problem (10.1.5) with $\gamma = 0$ for a class of relaxation functions satisfying

$$g'(t) \leq -H(g(t)), \qquad \forall\, t \geq 0,$$

where $H : [0, \infty) \longrightarrow [0, \infty)$ is a strictly increasing, strictly convex, and $\boldsymbol{C}^1$ function on $[0, k_0]$ with $H(0) = H'(0) = 0$ and satisfies the following extra conditions:

$$\int_0^{k_0} \frac{1}{H(s)} ds = \infty, \qquad \int_0^{k_0} \frac{s}{H(s)} ds < 1 \qquad \text{and} \qquad \liminf_{s \to 0^+} \frac{H(s)}{sH(s)} > \frac{1}{2}.$$

Moreover, if H satisfies

$$g'(t) = -H(g(t)), \qquad \forall t \geq 0 \qquad \text{and} \qquad \limsup_{s \to 0^+} \frac{H(s)}{sH'(s)} < 1,$$

then an explicit optimal decay rate is claimed. They also asked the following question:

Q. What about a more general class of relaxation functions satisfying

$$g'(t) \leq -\xi(t) H(g(t)), \qquad \forall t \geq 0?$$

It is worth noting that the result of Messaoudi in [18] answered **Q** when $H = Id$ and ξ is a positive non-increasing differentiable function. In 2012, Mustafa and Messaoudi [26] relaxed most of the unnecessary conditions imposed on H in [2] and answered **Q** with $\xi \equiv 1$. In 2016, Messaoudi and Al-Khulafi [20] proved a general and optimal decay rate of the solution of (10.1.5) with $\gamma = 0$ for a class of relaxation functions, satisfying

$$g'(t) \leq -\xi(t) g^p(t), \qquad \forall t \geq 0, \qquad 1 \leq p < \frac{3}{2}. \tag{10.1.7}$$

This result answered **Q** with ξ being a non-increasing differentiable function and $H(s) = s^p$, for $1 \leq p < \frac{3}{2}$. Very recently, Mustafa [25] gave a complete answer to **Q** by assuming that H is either linear or strictly increasing and strictly convex $\boldsymbol{C}^2$ function on $(0, r]$, for $r \leq g(0)$ and ξ is a positive non-increasing differentiable function. His result generalizes and improves all the existing results in the literature related to the decay of the solution of viscoelastic equations.

For the general decay results of a solution of the system of viscoelastic wave equations, Messaoudi and Tatar [21] studied the following problem:

$$\begin{cases} u_{tt} - \Delta u + \displaystyle\int_0^t g(t-s)\Delta u(s) ds + f(u, v) = 0, & \text{in } \Omega \times (0, \infty), \\ v_{tt} - \Delta v + \displaystyle\int_0^t h(t-s)\Delta v(s) ds + k(u, v) = 0, & \text{in } \Omega \times (0, \infty), \\ u = v = 0, & \text{on } \partial\Omega \times [0, \infty), \\ u(\cdot, 0) = u_0, \quad u_t(\cdot, 0) = u_1, \quad v(\cdot, 0) = v_0, \quad v_t(\cdot, 0) = v_1, & \text{in } \Omega, \end{cases} \tag{10.1.8}$$

where Ω is a bounded domain of $\mathbb{R}^n$ $(n \geq 1)$ with a smooth boundary $\partial\Omega$ and u_0, v_0, u_1, v_1 are given initial data. The functions f and k satisfy, for all $(u, v) \in \mathbb{R}^2$, the following assumption:

$$\begin{cases} |f(u,v)| \leq d(|u|^{\beta_1} + |v|^{\beta_2}) \\ \\ |k(u,v)| \leq d(|u|^{\beta_3} + |v|^{\beta_4}) \end{cases}$$

for some constant $d > 0$ and

$$\beta_i \geq 1 \qquad (n-2)\beta_i < n, \qquad i = 1, 2, 3, 4.$$

Under the following hypothesis: there exist two positive constants $\xi_1, \ \xi_2$ such that

$$g'(t) \leq -\xi_1 g^p(t), \quad t \geq 0, \ 1 \leq p < \tfrac{3}{2}$$

$$h'(t) \leq -\xi_2 h^q(t), \quad t \geq 0, \ 1 \leq q < \tfrac{3}{2},$$

they proved an exponential decay result if $(p, q) = (1, 1)$ and a polynomial decay otherwise. This result improves that of Santos [28] in which some extra conditions on g'' and h'' were required. Mustafa [24] discussed (10.1.8) and gave sufficient conditions to guarantee the well-posedness of the system. In addition, under the following assumptions on the relaxation functions:

$$\begin{aligned} g'(t) &\leq -\xi_1(t) g(t), \quad t \geq 0 \\ h'(t) &\leq -\xi_2(t) h(t), \quad t \geq 0, \end{aligned} \tag{10.1.9}$$

where $\xi_1, \ \xi_2 : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ are non-increasing functions, he proved the existence and uniqueness result and established a generalized stability result from which exponential and polynomial decay rates are only special cases. Said-Houari et al. [27] considered a system of viscoelastic wave equations with nonlinear damping terms acting on both equations. Their work was mainly concerned with the following problem:

$$\begin{cases} u_{tt} - \Delta u + \int_0^t g(t-s)\Delta u(s)ds + |u_t|^{m-1}u_t = f_1(u,v), & \text{in } \Omega \times (0,\infty), \\ \\ v_{tt} - \Delta v + \int_0^t h(t-s)\Delta v(s)ds + |v_t|^{r-1}v_t = f_2(u,v), & \text{in } \Omega \times (0,\infty), \\ u = v = 0, & \text{on } \partial\Omega \times [0,\infty), \\ u(\cdot,0) = u_0, \quad u_t(\cdot,0) = u_1, \quad v(\cdot,0) = v_0, \quad v_t(\cdot,0) = v_1, & \text{in } \Omega, \end{cases} \tag{10.1.10}$$

with

$$f_1(u,v) = a|u+v|^{2(\rho+1)}(u+v) + b|u|^{\rho}u|v|^{\rho+2}$$

$$f_2(u,v) = a|u+v|^{2(\rho+1)}(u+v) + b|u|^{\rho+2}|v|^{\rho}v.$$

Under some conditions on the initial data, $\partial\Omega$, ρ, m, r, g, h, they proved the existence (local and global) and uniqueness result. By imposing (10.1.9) on g and h, they established a generalized decay rate of the solution of (10.1.10). Their result improves the ones in Messaoudi and Tatar [21] and Liu [16]. Very recently, Al-Gharabli and Kafini [1] established a general decay result for (10.1.8) with the relaxation functions $g_i's$ satisfying

$$g_i'(t) \le -H_i(g_i(t)), \qquad \forall\, t \ge 0, \qquad i = 1, 2 \tag{10.1.11}$$

with $H_i : [0,\infty) \longrightarrow [0,\infty)$ with $H_i(0) = 0$ and each H_i is linear or strictly increasing and strictly convex $\boldsymbol{C}^2$ function on $(0, r]$ for some $r > 0$. This later result allowed larger class of relaxation functions and generalizes, in some cases, those in [21, 24, 28].

The aim of this work is to investigate problem (10.1.1) with the general class of relaxation functions g and h and use the idea developed by Mustafa in [25], taking into consideration the nature of the system (P), to prove a new general decay result. Our result generalizes and improves all the existing results related to system of viscoelastic equations. This paper is organized as follows: In Sect. 10.2, we state some preliminary results. In Sect. 10.3, we state and prove some technical lemmas needed for the entire work. We state and prove our main result in Sect. 10.4, followed by some examples to demonstrate our result.

10.2 Preliminaries

In this section, we give our assumptions, state the existence theorem, and present some useful lemmas. We use $c > 1$ to denote a positive generic constant.

Assumptions We assume that the relaxation functions satisfy the following hypotheses:

(A.1) $g_i : [0,+\infty) \longrightarrow (0,+\infty)$ (for $i = 1, 2$) are non-increasing differentiable functions such that

$$g_i(0) > 0, \qquad 1 - \int_0^{+\infty} g_i(s)ds =: l_i > 0.$$

(**A.2**) There exist non-increasing differentiable functions $\xi_i : [0,+\infty) \longrightarrow (0,+\infty)$ and $\boldsymbol{C}^1$ functions $H_i : [0,+\infty) \longrightarrow [0,+\infty)$ which are linear or strictly increasing and strictly convex $\boldsymbol{C}^2$ functions on $(0,r]$, $r < g_i(0)$, with $H_i(0) = H_i'(0) = 0$ such that

$$g_i'(t) \leq -\xi_i(t)H_i(g_i(t)), \qquad \forall t \geq 0 \quad \text{and for } i = 1,2.$$

(**A.3**) $f_i : \mathbb{R}^2 \longrightarrow \mathbb{R}$ (for $i = 1,2$) are $\boldsymbol{C}^1$ functions with $f_i(0,0) = 0$ and there exists a function F such that

$$f_1(x,y) = \frac{\partial F}{\partial x}(x,y), \qquad f_2(x,y) = \frac{\partial F}{\partial y}(x,y),$$

$$F \geq 0, \qquad xf_1(x,y) + yf_2(x,y) - F(x,y) \geq 0,$$

and

$$\left|\frac{\partial f_i}{\partial x}(x,y)\right| + \left|\frac{\partial f_i}{\partial y}(x,y)\right| \leq d\left(1 + |x|^{\beta_i-1} + |y|^{\beta_i-1}\right), \quad \forall (x,y) \in \mathbb{R}^2, \tag{10.2.1}$$

for some constants $d > 0$ and

$$\beta_i \geq 1, \qquad \text{if } n = 1,2; \qquad 1 \leq \beta_i \leq \frac{n}{n-2}, \qquad \text{if } n \geq 3.$$

Remark 10.2.1

(1) It follows from assumption (**A.1**) that

$$\lim_{t\to+\infty} g_i(t) = 0 \quad \text{and} \quad g_i(t) \leq \frac{1-l_i}{t}, \qquad \forall t > 0 \quad \text{and} \quad \text{for} \quad i = 1,2.$$

Also, assumption (**A.2**) entails that there exists $t_i > 0$ (for $i = 1,2$) such that

$$g_i(t_i) = r \qquad \text{and} \qquad g_i(t) \leq r, \qquad \forall t \geq t_0 := \max\{t_1, t_2\}.$$

The non-increasing property of g_i gives

$$0 < g_i(t_i) \leq g_i(t) \leq g_i(0), \qquad \forall t \in [0, t_0].$$

A combination of this with the continuity of H_i yields (for $i = 1,2$)

$$a_i \leq H_i(g_i(t)) \leq b_i, \qquad \forall t \in [0, t_0],$$

for some constants $a_i, b_i > 0$, $i = 1,2$. Consequently, for any $t \in [0, t_0]$ and for $i = 1,2$, we have

$$g_i'(t) \leq -\xi_i(t)H_i(g_i(t)) \leq -a_i\xi_i(t) = -\frac{a_i}{g_i(0)}\xi_i(t)g_i(0) \leq -\frac{a_i}{g_i(0)}\xi_i(t)g_i(t).$$

This implies that

$$\xi_i(t)g_i(t) \le -\frac{g_i(0)}{a_i}g_i'(t), \qquad \forall t \in [0, t_0] \qquad \text{and} \qquad \text{for} \qquad i = 1, 2. \tag{10.2.2}$$

(2) If H is a strictly increasing and strictly convex $\boldsymbol{C}^2$ function on $(0, r]$, with $H(0) = H'(0) = 0$, then it has an extension $\bar{H}$ which is a strictly increasing and strictly convex $\boldsymbol{C}^2$-function on $(0, +\infty)$. For instance, we can define $\bar{H}$, for any $t > r$, by

$$\bar{H}(t) := \frac{H''(r)}{2}t^2 + \big(H'(r) - H''(r)r\big)t + \left(H(r) + \frac{H''(r)}{2}r^2 - H'(r)r\right).$$

(3) Inequality (10.2.1) yields, for some positive constant k, that

$$|f_i(x, y)| \le k\left(|x| + |y| + |x|^{\beta_i} + |y|^{\beta_i}\right) \tag{10.2.3}$$

for all $(x, y) \in \mathbb{R}^2$ and $i = 1, 2$.

For completeness, we state, without proof, the global existence and regularity result whose proof can be found in [24].

Theorem 10.2.1 *Let $(u_0, u_1), (v_0, v_1) \in H_0^1(\Omega) \times L^2(\Omega)$ be given. Assume that hypotheses (**A.1**) and (**A.3**) are satisfied. Then, problem (P) has a unique weak solution*

$$(u,\ v) \in \boldsymbol{C}\big([0, \infty);\ H_0^1(\Omega)\big) \cap \boldsymbol{C}^1\big([0, \infty);\ L^2(\Omega)\big) \cap \boldsymbol{C}^2\big([0, \infty);\ H^{-1}(\Omega)\big).$$

Moreover, if $(u_0, u_1), (v_0, v_1) \in \big(H^2(\Omega) \cap H_0^1(\Omega)\big) \times H_0^1(\Omega)$, then problem ($P$) has a unique strong solution

$$(u,\ v) \in L^\infty\big([0, \infty);\ H^2(\Omega) \cap H_0^1(\Omega)\big) \cap W^{1,\infty}\big([0, \infty);\ H_0^1(\Omega)\big) \cap W^{2,\infty}\big([0, \infty);\ L^2(\Omega)\big).$$

Now, we introduce the energy functional

$$\begin{aligned} E(t) := &\frac{1}{2}\left[\|u_t\|_2^2 + \left(1 - \int_0^t g_1(s)ds\right)\|\nabla u\|_2^2 + (g_1 \circ \nabla u)(t)\right] \\ &+ \frac{1}{2}\left[\|v_t\|_2^2 + \left(1 - \int_0^t g_2(s)ds\right)\|\nabla v\|_2^2 + (g_2 \circ \nabla v)(t)\right] \\ &+ \int_\Omega F(u, v)dx, \end{aligned} \tag{10.2.4}$$

where, for any $w \in L^2_{loc}\big([0, +\infty);\ L^2(\Omega)\big)$ and $i = 1, 2$,

$$(g_i \circ w)(t) := \int_0^t g_i(t - s)\|w(t) - w(s)\|_2^2 ds.$$

Lemma 10.2.1 *Let (u, v) be the solution of (P). Then,*

$$E'(t) = -\frac{1}{2}g_1(t)\|\nabla u\|_2^2 + \frac{1}{2}(g_1' \circ \nabla u)(t) - \frac{1}{2}g_2(t)\|\nabla v\|_2^2 + \frac{1}{2}(g_2' \circ \nabla v)(t) \leq 0, \ \forall t \geq 0. \quad (10.2.5)$$

As in [25], we set, for any $0 < \alpha < 1$ and $i = 1, 2$,

$$C_{\alpha,i} := \int_0^\infty \frac{g_i^2(s)}{\alpha g_i(s) - g_i'(s)} ds \qquad \text{and} \qquad h_i(t) := \alpha g_i(t) - g_i'(t).$$

Lemma 10.2.2 ([25]) *Assume that conditions (**A.1**) hold. Then for any $w \in L^2_{loc}\big([0, +\infty); L^2(\Omega)\big)$, we have*

$$\int_\Omega \left(\int_0^t g_i(t-s)(w(t)-w(s))ds\right)^2 dx \leq C_{\alpha,i}(h_i \circ w)(t), \quad \forall t \geq 0, \quad \text{for } i = 1, 2. \quad (10.2.6)$$

Lemma 10.2.3 (Jensen's Inequality) *Let $G : [a, b] \longrightarrow \mathbb{R}$ be a convex function. Assume that the functions $f : \Omega \longrightarrow [a, b]$ and $h : \Omega \longrightarrow \mathbb{R}$ are integrable such that $h(x) \geq 0$, for any $x \in \Omega$ and $\int_\Omega h(x)dx = k > 0$. Then,*

$$G\left(\frac{1}{k}\int_\Omega f(x)h(x)dx\right) \leq \frac{1}{k}\int_\Omega G(f(x))h(x)dx.$$

We will also need the following embedding $H_0^1(\Omega) \hookrightarrow L^q(\Omega)$, for $q \geq 2$ if $n = 1, 2$ or $2 \leq q \leq \frac{2n}{n-2}$ if $n \geq 3$, that is,

$$\|w\|_q \leq c\|\nabla w\|_2, \qquad \forall w \in H_0^1(\Omega). \quad (10.2.7)$$

10.3 Technical Lemmas

In this section, we state and prove some lemmas needed to establish our main result.

Lemma 10.3.1 *Assume that (**A.1**)–(**A.3**) hold. Then, the functional I defined by*

$$I(t) := \int_\Omega u u_t dx + \int_\Omega v v_t dx$$

satisfies, along the solution of (P)*, the estimates*

$$I'(t) \le \|u_t\|_2^2 - \frac{l_1}{2}\|\nabla u\|_2^2 + cC_{\alpha,1}(h_1 \circ \nabla u)(t) + \|v_t\|_2^2$$
$$-\frac{l_2}{2}\|\nabla v\|_2^2 + cC_{\alpha,2}(h_2 \circ \nabla v)(t) - \int_\Omega F(u,v)dx. \qquad (10.3.1)$$

Proof Differentiating I and using equations in (P), integrating by parts, and using Young's inequality, (**A.3**), and Lemma 10.2.2, we get

$$\begin{aligned}
I'(t) = &\|u_t\|_2^2 - \left(1 - \int_0^t g_1(s)ds\right)\|\nabla u\|_2^2 + \int_\Omega \nabla u(t)\cdot\int_0^t g_1(t-s)(\nabla u(s)\\
&-\nabla u(t))dsdx + \|v_t\|_2^2 - \left(1 - \int_0^t g_2(s)ds\right)\|\nabla v\|_2^2 + \int_\Omega \nabla v(t)\\
&\cdot\int_0^t g_2(t-s)(\nabla v(s) - \nabla v(t))dsdx - \int_\Omega [uf_1(u,v) + vf_2(u,v)]dx\\
\le &\|u_t\|_2^2 - l_1\|\nabla u\|_2^2 + \frac{l_1}{2}\|\nabla u\|_2^2\\
&+\frac{1}{2l_1}\int_\Omega\left(\int_0^t g_1(t-s)|\nabla u(s) - \nabla u(t)|ds\right)^2 dx\\
&+\|v_t\|_2^2 - l_2\|\nabla v\|_2^2 + \frac{l_2}{2}\|\nabla u\|_2^2\\
&+\frac{1}{2l_2}\int_\Omega\left(\int_0^t g_2(t-s)|\nabla v(s) - \nabla v(t)|ds\right)^2 dx - \int_\Omega F(u,v)dx\\
\le &\|u_t\|_2^2 - \frac{l_1}{2}\|\nabla u\|_2^2 + cC_{\alpha,1}(h_1\circ\nabla u)(t)\\
&+\|v_t\|_2^2 - \frac{l_2}{2}\|\nabla v\|_2^2 + cC_{\alpha,2}(h_2\circ\nabla v)(t) - \int_\Omega F(u,v)dx.
\end{aligned}$$

□

Lemma 10.3.2 *Assume that (**A.1**)–(**A.3**) hold. Then, the functional K defined by*

$$K(t) := K_1(t) + K_2(t)$$

with

$$K_1(t) := -\int_\Omega u_t \int_0^t g_1(t-s)(u(t) - u(s))dsdx$$

and

$$K_2(t) := -\int_\Omega v_t \int_0^t g_2(t-s)(v(t) - v(s))dsdx$$

satisfies, along the solution of (P) *and for any* $0 < \delta < 1$, *the estimate*

$$K'(t) \leq -\left(\int_0^t g_1(s)ds - \delta\right) \|u_t\|_2^2 + c\delta\|\nabla u\|_2^2 + \frac{c}{\delta}(C_{\alpha,1}+1)(h_1 \circ \nabla u)(t)$$
$$-\left(\int_0^t g_2(s)ds - \delta\right) \|v_t\|_2^2 + c\delta\|\nabla v\|_2^2 + \frac{c}{\delta}(C_{\alpha,2}+1)(h_2 \circ \nabla v)(t). \tag{10.3.2}$$

Proof By exploiting equations in (P) and integrating by parts, we have

$$K_1'(t) = \left(1 - \int_0^t g_1(s)ds\right)\int_\Omega \nabla u(t)\cdot\int_0^t g_1(t-s)(\nabla u(t) - \nabla u(s))dsdx$$
$$+\int_\Omega\left(\int_0^t g_1(t-s)|\nabla u(t) - \nabla u(s)|ds\right)^2 dx$$
$$+\int_\Omega f_1(u,v)\int_0^t g_1(t-s)(u(t)-u(s))dsdx$$
$$-\int_\Omega u_t\int_0^t g_1'(t-s)(u(t)-u(s))dsdx - \left(\int_0^t g_1(s)ds\right)\|u_t\|_2^2.$$

Now, we estimate the terms in the right-hand side of the above equality.

Applying Young's inequality and Lemma 10.2.2, we obtain, for any $0 < \delta < 1$,

$$\left(1 - \int_0^t g_1(s)ds\right)\int_\Omega \nabla u(t)\cdot\int_0^t g_1(t-s)(\nabla u(t) - \nabla u(s))dsdx$$
$$+\int_\Omega\left(\int_0^t g_1(t-s)|\nabla u(t) - \nabla u(s)|ds\right)^2 dx$$
$$\leq \delta\|\nabla u\|_2^2 + \frac{c}{\delta}\int_\Omega\left(\int_0^t g_1(t-s)|\nabla u(t) - \nabla u(s)|ds\right)^2 dx$$
$$\leq \delta\|\nabla u\|_2^2 + \frac{c}{\delta}C_{\alpha,1}(h_1 \circ \nabla u)(t).$$

Using $\|\nabla u\|_2^2 + \|\nabla v\|_2^2 \leq cE(t) \leq cE(0)$ and inequalities (10.2.3) and (10.2.7), we have

$$\int_\Omega f_1(u,v)\int_0^t g_1(t-s)(u(t)-u(s))dsdx$$
$$\leq c\delta\int_\Omega\left(|u|^2 + |v|^2 + |u|^{2\beta_1} + |v|^{2\beta_2}\right)dx$$

$$
\begin{aligned}
&+\frac{c}{\delta}\int_{\Omega}\left(\int_0^t g_1(t-s)(u(t)-u(s))ds\right)^2 dx\\
&\leq c\delta\left(\|\nabla u\|_2^2+\|\nabla v\|_2^2+\|\nabla u\|_2^{2\beta_1}+\|\nabla v\|_2^{2\beta_2}\right)\\
&+\frac{c}{\delta}C_{\alpha,1}(h_1\circ\nabla u)(t)\\
&=c\delta\left(\|\nabla u\|_2^2+\|\nabla v\|_2^2+\|\nabla u\|_2^{2(\beta_1-1)}\|\nabla u\|_2^2+\|\nabla v\|_2^{2(\beta_2-1)}\|\nabla v\|_2^2\right)\\
&+\frac{c}{\delta}C_{\alpha,1}(h_1\circ\nabla u)(t)\\
&\leq c\delta\|\nabla u\|_2^2+c\delta\|\nabla v\|_2^2+\frac{c}{\delta}C_{\alpha,1}(h_1\circ\nabla u)(t).
\end{aligned}
$$

Exploiting Young's inequality and Lemma 10.2.2 again, we obtain, for any $0<\delta<1$,

$$
\begin{aligned}
&-\int_{\Omega}u_t\int_0^t g_1'(t-s)(u(t)-u(s))dsdx\\
&=\int_{\Omega}u_t\int_0^t h_1(t-s)(u(t)-u(s))dsdx\\
&-\int_{\Omega}u_t\int_0^t \alpha g_1(t-s)(u(t)-u(s))dsdx\\
&\leq\frac{\delta}{2}\|u_t\|_2^2+\frac{1}{2\delta}\int_{\Omega}\left(\int_0^t\sqrt{h_1(t-s)}\sqrt{h_1(t-s)}(u(t)-u(s))ds\right)^2dx\\
&+\frac{\delta}{2}\|u_t\|_2^2+\frac{1}{2\delta}\alpha^2\int_{\Omega}\left(\int_0^t g_1(t-s)(u(t)-u(s))ds\right)^2dx\\
&\leq\delta\|u_t\|_2^2+\frac{1}{2\delta}\left(\int_0^t h_1(s)ds\right)(h_1\circ\psi)(t)+\frac{c}{\delta}C_{\alpha,1}(h_1\circ u)(t)\\
&\leq\delta\|u_t\|_2^2+\frac{c}{\delta}(C_{\alpha,1}+1)(h_1\circ\nabla u)(t).
\end{aligned}
$$

A combination of all the above estimates gives

$$
K_1'(t)\leq-\left(\int_0^t g_1(s)ds-\delta\right)\|u_t\|_2^2+c\delta\|\nabla u\|_2^2+\frac{c}{\delta}(C_{\alpha,1}+1)(h_1\circ\nabla u)(t)+c\delta\|\nabla v\|_2^2.
$$

Similarly,

$$
K_2'(t)\leq-\left(\int_0^t g_2(s)ds-\delta\right)\|v_t\|_2^2+c\delta\|\nabla v\|_2^2+\frac{c}{\delta}(C_{\alpha,2}+1)(h_2\circ\nabla v)(t)+c\delta\|\nabla u\|_2^2.
$$

The last two estimates lead to the desired result. □

Lemma 10.3.3 ([25]) *Assume that (***A.1***)–(***A.3***) hold. Then, the functionals* J_1 *and* J_2 *defined by*

$$J_1(t) := \int_\Omega \int_0^t G_1(t-s)|\nabla u(s)|^2 ds dx$$

and

$$J_2(t) := \int_\Omega \int_0^t G_2(t-s)|\nabla v(s)|^2 ds dx$$

with $G_i(t) := \int_t^\infty g_i(s)ds$ *(for* $i = 1, 2$*) satisfy, along the solution of (P), the estimates*

$$J_1'(t) \leq 3(1-l)\|\nabla u\|_2^2 - \frac{1}{2}(g_1 \circ \nabla u)(t) \tag{10.3.3}$$

and

$$J_2'(t) \leq 3(1-l)\|\nabla v\|_2^2 - \frac{1}{2}(g_2 \circ \nabla v)(t), \tag{10.3.4}$$

where $l = \min\{l_1, l_2\}$.

Lemma 10.3.4 *The functional* $\mathcal{L}$ *defined by*

$$\mathcal{L}(t) := NE(t) + N_1 I(t) + N_2 K(t)$$

satisfies, for a suitable choice of $N,\ N_1,\ N_2 \geq 1$,

$$\mathcal{L}(t) \sim E(t) \tag{10.3.5}$$

and the estimate

$$\begin{aligned} \mathcal{L}'(t) \leq & -4(1-l)\big(\|\nabla u\|_2^2 + \|\nabla v\|_2^2\big) - \big(\|u_t\|_2^2 + \|v_t\|_2^2\big) \\ & - c\int_\Omega F(u,v)dx + \frac{1}{4}\big[(g_1 \circ \nabla u)(t) + (g_2 \circ \nabla v)(t)\big], \quad \forall\, t \geq t_0, \end{aligned} \tag{10.3.6}$$

where $l = \min\{l_1, l_2\}$ *and* t_0 *has been introduced in Remark 10.2.1.*

Proof It is not difficult to establish that $\mathcal{L}(t) \sim E(t)$. To prove (10.3.6), set

$$g_0 = \min\left\{\int_0^{t_0} g_1(s)ds, \int_0^{t_0} g_2(s)ds\right\} > 0, \quad \delta = \frac{l}{4cN_2}, \text{ and } C_\alpha = \max\{C_{\alpha,1},\ C_{\alpha,2}\}.$$

Exploiting (10.3.1), (10.3.2) and recalling that $g_i' = \alpha g_i - h_i$, we obtain, for any $t \geq t_0$,

$$\begin{aligned}\mathcal{L}'(t) \leq & -\frac{l}{4}(2N_1 - 1)\big(\|\nabla u\|_2^2 + \|\nabla v\|_2^2\big) - \left(g_0 N_2 - \frac{l}{4c} - N_1\right)\left(\|u_t\|_2^2 + \|v_t\|_2^2\right) \\ & -N_1 \int_\Omega F(u,v)dx + \frac{\alpha}{2}N\big[(g_1 \circ \nabla u)(t) + (g_2 \circ \nabla v)(t)\big] \\ & -\left[\frac{1}{2}N - \frac{4c^2}{l}N_2^2 - C_\alpha\left(\frac{4c^2}{l}N_2^2 + cN_1\right)\right]\big[(h_1 \circ \nabla u)(t) + (h_2 \circ \nabla v)(t)\big].\end{aligned}$$

We start by choosing N_1 large enough so that

$$\frac{l}{4}(2N_1 - 1) > 4(1 - l),$$

then we select N_2 so large that

$$g_0 N_2 - \frac{l}{4c} - N_1 > 1.$$

As $\dfrac{\alpha g_i^2(s)}{\alpha g_i(s) - g_i'(s)} < g_i(s)$ for $i = 1, 2$, it follows from the Lebesgue dominated convergence theorem that

$$\lim_{\alpha \to 0^+} \alpha C_{\alpha,i} = \lim_{\alpha \to 0^+} \int_0^\infty \frac{\alpha g_i^2(s)}{\alpha g_i(s) - g_i'(s)} ds = 0 \qquad \text{for} \qquad i = 1, 2.$$

This gives

$$\lim_{\alpha \to 0^+} \alpha C_\alpha = 0.$$

Consequently, there exists $\alpha_0 \in (0, 1)$ such that if $\alpha < \alpha_0$, then

$$\alpha C_\alpha < \frac{1}{8\left[\frac{4c^2}{l}N_2^2 + cN_1\right]}.$$

Now, we choose N large enough so that

$$N > \max\left\{\frac{16c^2}{l}N_2^2,\ \frac{1}{2\alpha_0}\right\}$$

and set

$$\alpha = \frac{1}{2N}.$$

Then

$$\frac{1}{4}N - \frac{4c^2}{l}N_2^2 > 0 \qquad \text{and} \qquad \alpha = \frac{1}{2N} < \alpha_0.$$

These imply

$$\frac{1}{2}N - \frac{4c^2}{l}N_2^2 - C_\alpha\left[\frac{4c^2}{l}N_2^2 + cN_1\right] > \frac{1}{2}N - \frac{4c^2}{l}N_2^2 - \frac{1}{8\alpha}$$
$$= \frac{1}{4}N - \frac{4c^2}{l}N_2^2 > 0.$$

Hence, we arrive at the required estimate. □

10.4 General Decay Result

In this section, we state and prove our main result.

Theorem 10.4.1 *Let* $(u_0, u_1), (v_0, v_1) \in \left(H^2(\Omega) \cap H_0^1(\Omega)\right) \times H_0^1(\Omega)$ *be given. Suppose that assumptions (**A.1**)–(**A.3**) hold. Then there exist two positive constants* k_1 *and* k_2 *such that the solution to problem* (P) *satisfies the estimate*

$$E(t) \le k_2 G_*^{-1}\left(k_1 \int_{t_0}^{t} \xi(s)ds\right), \qquad \forall\, t > t_0, \tag{10.4.1}$$

where $t_0 = \min\{t_1, t_2\}$ *is introduced in Remark 10.2.1,* $\xi(t) = \min\{\xi_1(t),\ \xi_2(t)\}$, *and* G_* *is given by*

$$G_*(t) = \int_t^r \frac{1}{sG(s)}ds \qquad \text{with} \qquad G(t) = \min\{H_1'(t),\ H_2'(t)\}.$$

Proof We start by using estimates (10.2.2) and (10.2.5) to deduce, for any $t \ge t_0$,

$$\int_0^{t_0} g_1(s)\|\nabla u(t) - \nabla u(t-s)\|_2^2 ds + \int_0^{t_0} g_2(s)\|\nabla v(t) - \nabla v(t-s)\|_2^2 ds$$
$$\le \frac{1}{\xi_1(t_0)} \int_0^{t_0} \xi_1(s) g_1(s)\|\nabla u(t) - \nabla u(t-s)\|_2^2 ds$$

$$
\begin{aligned}
&+\frac{1}{\xi_2(t_0)}\int_0^{t_0}\xi_2(s)g_2(s)\|\nabla v(t)-\nabla v(t-s)\|_2^2 ds\\
&\leq -\frac{g_1(0)}{a_1\xi_1(t_0)}\int_0^{t_0} g_1'(s)\|\nabla u(t)-\nabla u(t-s)\|_2^2 ds\\
&-\frac{g_2(0)}{a_2\xi_2(t_0)}\int_0^{t_0} g_2'(s)\|\nabla v(t)-\nabla v(t-s)\|_2^2 ds\\
&\leq -cE'(t).
\end{aligned}
$$

Exploiting this estimate, inequality (10.3.6) becomes, for some $m > 0$ and for any $t \geq t_0$,

$$
\begin{aligned}
\mathcal{L}'(t) &\leq -mE(t) + c\big[(g_1\circ\nabla u)(t) + (g_2\circ\nabla v)(t)\big]\\
&\leq -mE(t) - cE'(t) + c\int_{t_0}^{t} g_1(s)\|\nabla u(t)-\nabla u(t-s)\|_2^2 ds\\
&\quad + c\int_{t_0}^{t} g_2(s)\|\nabla v(t)-\nabla v(t-s)\|_2^2 ds.
\end{aligned}
$$

By setting $\mathcal{F} := \mathcal{L} + cE \sim E$, we obtain

$$
\begin{aligned}
\mathcal{F}'(t) &\leq -mE(t) + c\int_{t_0}^{t} g_1(s)\|\nabla u(t)-\nabla u(t-s)\|_2^2 ds\\
&\quad + c\int_{t_0}^{t} g_2(s)\|\nabla v(t)-\nabla v(t-s)\|_2^2 ds, \qquad \forall\, t \geq t_0. \qquad (10.4.2)
\end{aligned}
$$

Case I H_1 and H_2 are linear: Set $\xi(t) = \min\{\xi_1(t), \xi_2(t)\} > 0$, for any $t \geq 0$, then ξ is differentiable almost everywhere and non-increasing on $[0, +\infty)$. Multiply both sides of (10.4.2) by $\xi(t)$ and exploit (**A.2**) and (10.2.5) to get

$$
\begin{aligned}
\xi(t)\mathcal{F}'(t) &\leq -m\xi(t)E(t) + c\xi(t)\int_{t_0}^{t} g_1(s)\|\nabla u(t)-\nabla u(t-s)\|_2^2 ds\\
&\quad + c\xi(t)\int_{t_0}^{t} g_2(s)\|\nabla v(t)-\nabla v(t-s)\|_2^2 ds\\
&\leq -m\xi(t)E(t) + c\int_{t_0}^{t} \xi_1(s)g_1(s)\|\nabla u(t)-\nabla u(t-s)\|_2^2 ds\\
&\quad + c\int_{t_0}^{t} \xi_2(s)g_2(s)\|\nabla v(t)-\nabla v(t-s)\|_2^2 ds\\
&\leq -m\xi(t)E(t) - c\int_{t_0}^{t} g_1'(s)\|\nabla u(t)-\nabla u(t-s)\|_2^2 ds
\end{aligned}
$$

$$
\begin{aligned}
&-c\int_{t_0}^{t} g_2'(s)\|\nabla v(t)-\nabla v(t-s)\|_2^2 ds\\
&\le -m\xi(t)E(t)-cE'(t), \qquad \forall t\ge t_0.
\end{aligned}
$$

Using the non-increasing property of ξ we have $\xi\mathcal{F}+cE\sim E$ and

$$
(\xi\mathcal{F}+cE)'(t)\le -m\xi(t)E(t), \qquad \forall t\ge t_0.
$$

A simple integration over (t_0,t) yields, for two positive constants k_1 and k_2,

$$
E(t)\le k_2\exp\left(-k_1\int_{t_0}^{t}\xi(s)ds\right), \qquad \forall t>t_0.
$$

Continuity of E (see [25]) gives

$$
E(t)\le k_2\exp\left(-k_1\int_{0}^{t}\xi(s)ds\right), \qquad \forall t>0.
$$

Case II H_1 or H_2 is nonlinear: First, we use Lemmas 10.3.3 and 10.3.4 to conclude that

$$
\mathscr{L}(t):=\mathcal{L}(t)+J_1(t)+J_2(t)
$$

is nonnegative and satisfies, for any $t\ge t_0$,

$$
\begin{aligned}
\mathscr{L}'(t)&\le -(1-l)\big(\|\nabla u\|_2^2+\|\nabla v\|_2^2\big)-\big(\|u_t\|_2^2+\|v_t\|_2^2\big)\\
&\quad -c\int_{\Omega}F(u,v)dx-\frac{1}{4}\big[(g_1\circ\nabla u)(t)+(g_2\circ\nabla v)(t)\big]\\
&\le -\beta E(t),
\end{aligned}
$$

for some $\beta>0$. Consequently, we arrive at

$$
\int_0^{\infty}E(s)ds<+\infty. \tag{10.4.3}
$$

Now, we define functionals η_i (for $i=1,2$) by

$$
\eta_1(t):=\gamma\int_{t_0}^{t}\|\nabla u(t)-\nabla u(t-s)\|_2^2\, ds
$$

and

$$
\eta_2(t):=\gamma\int_{t_0}^{t}\|\nabla v(t)-\nabla v(t-s)\|_2^2\, ds,
$$

where (10.4.3) allows us to choose $0 < \gamma < 1$ so that

$$\eta_i(t) < 1, \qquad \forall\, t \geq t_0 \qquad \text{and} \qquad i = 1, 2. \tag{10.4.4}$$

We further assume that $\eta_i(t) > 0$, for any $t > t_0$. Also, we define another functional θ_i (for $i = 1, 2$) by

$$\theta_1(t) := -\int_{t_0}^{t} g_1'(s)\|\nabla u(t) - \nabla u(t-s)\|_2^2 ds,$$

$$\theta_2(t) := -\int_{t_0}^{t} g_2'(s)\|\nabla v(t) - \nabla v(t-s)\|_2^2 ds$$

and observe that

$$\theta_1(t) + \theta_2(t) \leq -cE'(t), \qquad \forall\, t \geq t_0. \tag{10.4.5}$$

In addition, it follows from the strict convexity of H_i and the fact that $H_i(0) = 0$ that

$$H_i(s\tau) \leq sH_i(\tau), \qquad \text{for} \qquad 0 \leq s \leq 1, \qquad \tau \in (0, r] \qquad \text{and} \qquad i = 1, 2.$$

These facts, hypothesis (**A.2**), estimates (10.4.4), and Jensen's inequality lead to

$$\begin{aligned}
\theta_1(t) &= -\frac{1}{\gamma\eta_1(t)} \int_{t_0}^{t} \gamma\eta_1(t) g_1'(s)\|\nabla u(t) - \nabla u(t-s)\|_2^2 ds \\
&\geq \frac{1}{\gamma\eta_1(t)} \int_{t_0}^{t} \gamma\eta_1(t)\xi_1(s) H_1(g_1(s))\|\nabla u(t) - \nabla u(t-s)\|_2^2 ds \\
&\geq \frac{\xi_1(t)}{\gamma\eta_1(t)} \int_{t_0}^{t} \gamma H_1(\eta_1(t) g_1(s))\|\nabla u(t) - \nabla u(t-s)\|_2^2 ds \\
&\geq \frac{\xi_1(t)}{\gamma} H_1\left(\frac{1}{\eta_1(t)} \int_{t_0}^{t} \gamma\eta_1(t) g_1(s)\|\nabla u(t) - \nabla u(t-s)\|_2^2 ds\right) \\
&= \frac{\xi_1(t)}{\gamma} H_1\left(\gamma \int_{t_0}^{t} g_1(s)\|\nabla u(t) - \nabla u(t-s)\|_2^2 ds\right) \\
&= \frac{\xi_1(t)}{\gamma} \bar{H}_1\left(\gamma \int_{t_0}^{t} g_1(s)\|\nabla u(t) - \nabla u(t-s)\|_2^2 ds\right), \qquad \forall\, t \geq t_0,
\end{aligned}$$

where $\bar{H}_1$ is a $\boldsymbol{C}^2$ extension of H_1 that is strictly increasing and strictly convex on $(0, \infty)$. This implies that

$$\int_{t_0}^{t} g_1(s)\|\nabla u(t) - \nabla u(t-s)\|_2^2 ds \leq \frac{1}{\gamma} \bar{H}_1^{-1}\left(\frac{\gamma\theta_1(t)}{\xi_1(t)}\right), \qquad \forall\, t \geq t_0.$$

Similarly, we have

$$\int_{t_0}^{t} g_2(s)\|\nabla v(t) - \nabla v(t-s)\|_2^2 ds \le \frac{1}{\gamma}\bar{H}_2^{-1}\left(\frac{\gamma\theta_2(t)}{\xi_2(t)}\right), \qquad \forall\, t \ge t_0.$$

Thus, (10.4.2) becomes

$$\mathcal{F}'(t) \le -mE(t) + c\bar{H}_1^{-1}\left(\frac{\gamma\theta_1(t)}{\xi_1(t)}\right) + c\bar{H}_2^{-1}\left(\frac{\gamma\theta_2(t)}{\xi_2(t)}\right), \qquad \forall\, t \ge t_0. \tag{10.4.6}$$

Set $G = \min\{\bar{H}_1', \bar{H}_2'\}$ and, for a fixed $0 < \varepsilon < r$, define a functional $\mathcal{F}_1$ by

$$\mathcal{F}_1(t) := G\left(\varepsilon\frac{E(t)}{E(0)}\right)\mathcal{F}(t) + E(t), \qquad \forall\, t \ge 0.$$

Then, using the fact that $E' \le 0$, $\bar{H}_i' > 0$, and $\bar{H}_i'' > 0$, we deduce that $\mathcal{F}_1 \sim E$ and, we, further, have

$$\mathcal{F}_1'(t) = \varepsilon\frac{E'(t)}{E(0)}G'\left(\varepsilon\frac{E(t)}{E(0)}\right)\mathcal{F}(t) + G\left(\varepsilon\frac{E(t)}{E(0)}\right)\mathcal{F}'(t) + E'(t), \quad \text{for } a.e\ \ t \ge t_0.$$

By dropping the first and last terms of the above identity, since they are non-positive, and using estimate (10.4.6), we get

$$\begin{aligned}\mathcal{F}_1'(t) \le -mE(t)G\left(\varepsilon\frac{E(t)}{E(0)}\right) + cG\left(\varepsilon\frac{E(t)}{E(0)}\right)\bar{H}_1^{-1}\left(\frac{\gamma\theta_1(t)}{\xi_1(t)}\right)\\ +cG\left(\varepsilon\frac{E(t)}{E(0)}\right)\bar{H}_2^{-1}\left(\frac{\gamma\theta_2(t)}{\xi_2(t)}\right), \qquad \text{for} \quad a.e\ \ t \ge t_0.\end{aligned} \tag{10.4.7}$$

Let $\bar{H}_i^*$ be the convex conjugate of $\bar{H}_i$ in the sense of Young (see [4, pp. 61–64]), which has the form

$$\bar{H}_i^*(s) = s\big(\bar{H}_i'\big)^{-1}(s) - \bar{H}_i\left[\big(\bar{H}_i'\big)^{-1}(s)\right], \qquad \text{for} \qquad i = 1, 2, \tag{10.4.8}$$

and satisfies the following generalized Young inequality:

$$AB_i \le \bar{H}_i^*(A) + \bar{H}_i(B_i), \qquad \text{for} \qquad i = 1, 2. \tag{10.4.9}$$

By taking $A = G\left(\varepsilon\dfrac{E(t)}{E(0)}\right)$, $B_i = \bar{H}_i^{-1}\left(\dfrac{\gamma\theta_i(t)}{\xi_i(t)}\right)$, for $i = 1, 2$, and combining (10.4.7)–(10.4.9), we obtain, for almost every $t \geq t_0$,

$$\begin{aligned}
\mathcal{F}_1'(t) &\leq -mE(t)G\left(\varepsilon\frac{E(t)}{E(0)}\right) + c\bar{H}_1^*\left[G\left(\varepsilon\frac{E(t)}{E(0)}\right)\right] + c\frac{\gamma\theta_1(t)}{\xi_1(t)} \\
&\quad + c\bar{H}_2^*\left[G\left(\varepsilon\frac{E(t)}{E(0)}\right)\right] + c\frac{\gamma\theta_2(t)}{\xi_2(t)} \\
&\leq -mE(t)G\left(\varepsilon\frac{E(t)}{E(0)}\right) + cG\left(\varepsilon\frac{E(t)}{E(0)}\right)\left(\bar{H}_1'\right)^{-1}\left[G\left(\varepsilon\frac{E(t)}{E(0)}\right)\right] + c\frac{\gamma\theta_1(t)}{\xi_1(t)} \\
&\quad + cG\left(\varepsilon\frac{E(t)}{E(0)}\right)\left(\bar{H}_2'\right)^{-1}\left[G\left(\varepsilon\frac{E(t)}{E(0)}\right)\right] + c\frac{\gamma\theta_2(t)}{\xi_2(t)} \\
&\leq -mE(t)G\left(\varepsilon\frac{E(t)}{E(0)}\right) + cG\left(\varepsilon\frac{E(t)}{E(0)}\right)\left(\bar{H}_1'\right)^{-1}\left[\bar{H}_1'\left(\varepsilon\frac{E(t)}{E(0)}\right)\right] + c\frac{\gamma\theta_1(t)}{\xi_1(t)} \\
&\quad + cG\left(\varepsilon\frac{E(t)}{E(0)}\right)\left(\bar{H}_2'\right)^{-1}\left[\bar{H}_2'\left(\varepsilon\frac{E(t)}{E(0)}\right)\right] + c\frac{\gamma\theta_2(t)}{\xi_2(t)} \\
&\leq -\Big(mE(0) - c\varepsilon\Big)\frac{E(t)}{E(0)}G\left(\varepsilon\frac{E(t)}{E(0)}\right) + c\left(\frac{\gamma\theta_1(t)}{\xi_1(t)} + \frac{\gamma\theta_2(t)}{\xi_2(t)}\right).
\end{aligned}$$

Multiplying this estimate by $\xi(t) = \min\{\xi_1(t), \xi_2(t)\} > 0$ and using inequality (10.4.5), we obtain

$$\begin{aligned}
\xi(t)\mathcal{F}_1'(t) &\leq -(mE(0) - c\varepsilon)\xi(t)\frac{E(t)}{E(0)}G\left(\varepsilon\frac{E(t)}{E(0)}\right) + c\gamma\big(\theta_1(t) + \theta_2(t)\big) \\
&\leq -(mE(0) - c\varepsilon)\xi(t)\frac{E(t)}{E(0)}G\left(\varepsilon\frac{E(t)}{E(0)}\right) - cE'(t), \quad \text{for} \quad a.e \ \ t \geq t_0.
\end{aligned}$$

Take ε smaller, if needed, to get, for some $k_0 > 0$,

$$\xi(t)\mathcal{F}_1'(t) \leq -k_0\xi(t)\frac{E(t)}{E(0)}G\left(\varepsilon\frac{E(t)}{E(0)}\right) - cE'(t), \qquad \text{for} \quad a.e \ \ t \geq t_0.$$

Consequently, by setting $\mathcal{F}_2 = \xi\mathcal{F}_1 + cE$, we obtain, for some $\alpha_1, \alpha_2 > 0$

$$\alpha_1\mathcal{F}_2(t) \leq E(t) \leq \alpha_2\mathcal{F}_2(t), \qquad \forall\, t \geq t_0 \tag{10.4.10}$$

and

$$\mathcal{F}_2'(t) \leq -k_0\xi(t)\frac{E(t)}{E(0)}G\left(\varepsilon\frac{E(t)}{E(0)}\right), \qquad \text{for} \quad a.e \ \ t \geq t_0. \tag{10.4.11}$$

It follows from $0 \leq \varepsilon \dfrac{E(t)}{E(0)} < r$ that

$$G\left(\varepsilon\frac{E(t)}{E(0)}\right) = \min\left\{\bar{H}_1'\left(\varepsilon\frac{E(t)}{E(0)}\right), \bar{H}_2'\left(\varepsilon\frac{E(t)}{E(0)}\right)\right\}$$
$$= \min\left\{H_1'\left(\varepsilon\frac{E(t)}{E(0)}\right), H_2'\left(\varepsilon\frac{E(t)}{E(0)}\right)\right\}, \qquad \forall t \geq 0.$$

Now, set

$$G_0(\tau) = \tau G(\varepsilon\tau), \qquad \forall \tau \in [0, 1],$$

we deduce from $H_i' > 0$ and $H_i'' > 0$ on $(0, r]$ (for $i = 1, 2$) that $G_0, G_0' > 0$ $a.e.$ on $(0, 1]$. Define a functional R by

$$R(t) := \frac{\alpha_1 \mathcal{F}_2(t)}{E(0)}$$

and exploit (10.4.10) and (10.4.11) to notice that $R \sim E$ and, for some $k_1 > 0$,

$$R'(t) \leq -k_1 \xi(t) G_0(R(t)), \qquad \text{for} \quad a.e \;\; t \geq t_0.$$

An integration over (t_0, t) gives

$$-\int_{t_0}^{t} \frac{R'(s)}{G_0(R(s))} ds \geq k_1 \int_{t_0}^{t} \xi(s) ds$$

or equivalently,

$$\int_{\varepsilon R(t)}^{\varepsilon R(t_0)} \frac{1}{sG(s)} ds \geq k_1 \int_{t_0}^{t} \xi(s) ds,$$

which implies that

$$R(t) \leq \frac{1}{\varepsilon} G_*^{-1}\left(k_1 \int_{t_0}^{t} \xi(s) ds\right) \qquad \forall t > t_0,$$

where $G_*(t) := \displaystyle\int_t^r \frac{1}{sG(s)} ds$. A combination of this estimate with the fact that $R \sim E$ gives

$$E(t) \leq k_2 G_*^{-1}\left(k_1 \int_{t_0}^{t} \xi(s) ds\right) \qquad \forall t > t_0.$$

- If $\eta_i(t) = 0$, for $t \geq t_0$ and $i = 1, 2$, then we get an exponential decay from (10.4.2).
- If $\eta_1(t) > 0$ and $\eta_2(t) = 0$, for any $t > t_0$, then we set $\theta_2(t) = 0$ and (10.4.6) becomes

$$\mathcal{F}'(t) \leq -mE(t) + c\bar{H}_1^{-1}\left(\frac{\gamma\theta_1(t)}{\xi_1(t)}\right), \qquad \forall t \geq t_0.$$

Repeating the above steps, we arrive at

$$E(t) \leq k_2 G_*^{-1}\left(k_1 \int_{t_0}^{t} \xi_1(s)ds\right) \qquad \forall t > t_0,$$

where

$$G_*(t) := \int_t^r \frac{1}{sH_1'(s)}ds.$$

- Similarly, if $\eta_1(t) = 0$ and $\eta_2(t) > 0$, for any $t > t_0$, we have

$$E(t) \leq k_2 G_*^{-1}\left(k_1 \int_{t_0}^{t} \xi_2(s)ds\right) \qquad \forall t > t_0,$$

where

$$G_*(t) := \int_t^r \frac{1}{sH_2'(s)}ds.$$

This completes the proof.

□

Example 10.4.1

(1) Consider the relaxation functions $g_1(t) = ae^{-\alpha t}$ and $g_2(t) = \dfrac{b}{(1+t)^\mu}$, $\mu > 1$, a and b are chosen so that hypothesis (**A.1**) holds. Then there exists $C > 0$ such that

$$E(t) \leq \frac{C}{(1+t)^\mu}, \qquad \forall t > t_0.$$

(2) Let $g_1(t) = \dfrac{a}{(1+t)^\mu}$ and $g_2(t) = \dfrac{b}{(1+t)^\nu}$ with $\mu,\ \nu > 1$, a and b are chosen so that hypothesis (**A.1**) holds. Then, there exists $C > 0$ such that, for any $t > t_0$,

$$E(t) \leq \frac{C}{(1+t)^\gamma}, \quad \text{with} \quad \gamma = \min\{\mu, \nu\}.$$

(3) If $g_1(t) = ae^{-\alpha t}$ and $g_2(t) = be^{-(1+t)^\nu}$ with $0 < \nu < 1$, a and b are chosen so that hypothesis (**A.1**) holds. Then, there exist positive constants C and k_1 such that

$$E(t) \leq Ce^{-k_1(1+t)^\nu}, \qquad \text{for} \quad t \quad \text{large.}$$

(4) If $g_1(t) = ae^{-(1+t)^\nu}$ with $0 < \nu < 1$ and $g_2(t) = \dfrac{b}{(1+t)^\mu}$ with $\mu > 1$, a and b are chosen so that hypothesis (**A.1**) holds. Then, there exists $C > 0$ such that

$$E(t) \leq \frac{C}{(1+t)^\mu}, \qquad \text{for} \quad t \quad \text{large.}$$

Acknowledgements The authors would like to express their profound gratitude to King Fahd University of Petroleum and Minerals (KFUPM) and University of Sharjah for their support.

References

1. M.M. Al-Gharabli, M.M. Kafini, A general decay result of a coupled system of nonlinear wave equations. Rend. Circ. Mat. Palermo, II. Ser **67**, 145–157 (2018)
2. F. Alabau-Boussouira, P. Cannarsa, A general method for proving sharp energy decay rates for memory-dissipative evolution equations. C.R. Math. **347**(15–16), 867–872 (2009)
3. D. Andrade, A. Mognon, global solutions for a system of Klein–Gordon equations with memory. Bol. Soc. Paran. Mat **21**(1/2), 127–138 (2003)
4. V.I. Arnold, *Mathematical Methods of Classical Mechanics*. Graduate Texts in Mathematics, vol. 60 (Springer, New York, 1989)
5. S. Berrimi, S.A. Messaoudi, Exponential decay of solutions to a viscoelastic. Electron. J. Differ. Equ. **88**, 1–10 (2004)
6. S. Berrimi, S.A. Messaoudi, Existence and decay of solutions of a viscoelastic equation with a nonlinear source. Nonlinear Anal. Theory Methods Appl. **64**(10), 2314–2331 (2006)
7. X. Cao, Energy decay of solutions for a variable-coefficient viscoelastic wave equation with a weak nonlinear dissipation. J. Math. Phys. **57**(2), 021509 (2016)
8. M.M. Cavalcanti, V.N. Cavalcanti Domingos, J.A. Soriano, Exponential decay for the solution of semilinear viscoelastic wave equations with localized damping. Electron. J. Differ. Equ. **44**, 1–14 (2002)
9. C.M. Dafermos, An abstract Volterra equation with applications to linear viscoelasticity. J. Differ. Equ. **7**(3), 554–569 (1970)
10. C.M. Dafermos, Asymptotic stability in viscoelasticity. Arch. Ration. Mech. Anal. **37**(4), 297–308 (1970)
11. G. Dassios, F. Zafiropoulos, Equipartition of energy in linearized 3-d viscoelasticity. Q. Appl. Math. **484**, 43–89 (1990)
12. X. Han, M. Wang, General decay of energy for a viscoelastic equation with nonlinear damping. Math. Methods Appl. Sci. **32**(3), 346–358 (2009)
13. W.J. Hrusa, Global existence and asymptotic stability for a semilinear hyperbolic Volterra equation with large initial data. SIAM J. Math. Anal. **16**(1), 110–134 (1985)
14. W. Liu, General decay of solutions to a viscoelastic wave equation with nonlinear localized damping. Ann. Acad. Sci. Fenn. Math. **34**, 291–302 (2009)

15. W. Liu, General decay rate estimate for a viscoelastic equation with weakly nonlinear time-dependent dissipation and source terms. J. Math. Phys. **50**(11), 113506 (2009)
16. W. Liu, Uniform decay of solutions for a quasilinear system of viscoelastic equations. Nonlinear Anal. Theory Methods Appl. **71**(5), 2257–2267 (2009)
17. L.A. Medeiros, M.M. Miranda, Weak solutions for a system of nonlinear Klein–Gordon equations. Ann. di Mat. Pura ed Appl. **146**(1), 173–183 (1986)
18. S.A. Messaoudi, General decay of solutions of a viscoelastic equation. J. Math. Anal. Appl. **341**(2), 1457–1467 (2008)
19. S.A. Messaoudi, General decay of the solution energy in a viscoelastic equation with a nonlinear source. Nonlinear Anal. Theory Methods Appl. **69**(8), 2589–2598 (2008)
20. S.A. Messaoudi, W. Al-Khulaifi, General and optimal decay for a quasilinear viscoelastic equation. Appl. Math. Lett. **66**, 16–22 (2017)
21. S.A. Messaoudi, N.-E. Tatar, Uniform stabilization of solutions of a nonlinear system of viscoelastic equations. Appl. Anal. **87**(3), 247–263 (2008)
22. J.E. Muñoz Rivera, Asymptotic behaviour in linear viscoelasticity. Q. Appl. Math. **52**(4), 629–648 (1994)
23. J.E. Muñoz Rivera, E.C. Lapa, Decay rates of solutions of an anisotropic inhomogeneous n-dimensional viscoelastic equation with polynomially decaying kernels. Commun. Math. Phys. **177**(3), 583–602 (1996)
24. M.I. Mustafa, Well posedness and asymptotic behavior of a coupled system of nonlinear viscoelastic equations. Nonlinear Anal. Real World Appl. **13**(1), 452–463 (2012)
25. M.I. Mustafa, General decay result for nonlinear viscoelastic equations. J. Math. Anal. Appl. **457**(1), 134–152 (2018)
26. M.I. Mustafa, S.A. Messaoudi, General stability result for viscoelastic wave equations. J. Math. Phys. **53**(5), 053702 (2012)
27. B. Said-Houari, S.A. Messaoudi, A. Guesmia, General decay of solutions of a nonlinear system of viscoelastic wave equations. Nonlinear Differ. Equ. Appl. **18**(6), 659–684 (2011)
28. M.d.L. Santos, Decay rates for solutions of a system of wave equations with memory. Electron. J. Differ. Equ. **2002**(38), 1–17 (2002)
29. I.E. Segal, The global Cauchy problem for a relativistic scalar field with power interaction. Bull. Soc. Math. France **91**(2), 129–135 (1963)

Chapter 11
Mathematical Theory of Incompressible Flows: Local Existence, Uniqueness, and Blow-Up of Solutions in Sobolev–Gevrey Spaces

Wilberclay G. Melo, Natã Firmino Rocha, and Ezequiel Barbosa

Abstract This work establishes the local existence and uniqueness as well as the blow-up criteria for solutions of the Navier–Stokes equations in Sobolev–Gevrey spaces. More precisely, if the maximal time of existence of solutions for these equations is finite, we demonstrate the explosion, near this instant, of some limits superior and integrals involving a specific usual Lebesgue spaces and, as a consequence, we prove the lower bounds related to Sobolev–Gevrey spaces.

Keywords Navier–Stokes equations · Local existence and uniqueness of solutions · Blow-up criteria · Sobolev–Gevrey spaces

This chapter presents a study that determines the local existence, uniqueness, and blow-up criteria of solutions for the following Navier–Stokes equations:

$$\begin{cases} u_t + u \cdot \nabla u + \nabla p = \mu \Delta u, \quad x \in \mathbb{R}^3, t \in [0, T^*), \\ \operatorname{div} u = 0, \quad x \in \mathbb{R}^3, t \in [0, T^*), \\ u(x, 0) = u_0(x), \quad x \in \mathbb{R}^3, \end{cases} \tag{11.1}$$

where $T^* > 0$ gives the solution's existence time, $u(x,t) = (u_1(x,t), u_2(x,t), u_3(x,t)) \in \mathbb{R}^3$ denotes the incompressible velocity field, and $p(x,t) \in \mathbb{R}$ the hydrostatic pressure. The positive constant μ is the kinematic viscosity and the initial data for the velocity field, given by u_0 in (11.1), is assumed to be divergence free, i.e., $\operatorname{div} u_0 = 0$.

W. G. Melo (✉)
Departamento de Matemática, Universidade Federal de Sergipe, São Cristóvão, SE, Brazil

N. F. Rocha · E. Barbosa
Departamento de Matemática, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
e-mail: nata@ufmg.br; ezequiel@mat.ufmg.br

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_11

Let us inform that we shall study the above system using Sobolev–Gevrey spaces $H^s_{a,\sigma}(\mathbb{R}^3)$, with $a > 0$, $\sigma \geq 1$, and $s \in \left(\frac{1}{2}, \frac{3}{2}\right)$.

It is important to emphasize that there are two main goals to be accomplished in this chapter: proving the local existence and uniqueness of a solution $u(x, t)$ for the Navier–Stokes equations (11.1) and establishing blow-up criteria obeyed for $u(x, t)$. It is important to point out that the authors were mainly inspired by Benameur and Jlali [4].

Assuming that the initial data u_0 belongs to $H^s_{a,\sigma}(\mathbb{R}^3)$, with $s \in \left(\frac{1}{2}, \frac{3}{2}\right)$, $a > 0$, and $\sigma \geq 1$, this chapter assures that there exist a positive time T and a unique solution $u \in C([0, T]; H^s_{a,\sigma}(\mathbb{R}^3))$ of the Navier–Stokes equations (11.1) (let us recall that it is not known if $T = \infty$ always holds for these famous equations). Besides, the local existence and uniqueness result obtained in [4] is a particular case of ours; in fact, it is enough to take $s = 1$.

Under the same assumptions adopted above for s and a, if it is considered that $\sigma > 1$ and the maximal time interval of existence, $0 \leq t < T^*$, is finite, then the blow-up inequality

$$\|u(t)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \geq \frac{C_2 \exp\{C_3(T^* - t)^{-\frac{1}{3\sigma}}\}}{(T^* - t)^{\frac{2(s\sigma+\sigma_0)+1}{6\sigma}}}, \quad \forall\, t \in [0, T^*), \tag{11.2}$$

is valid, where C_2 and C_3 are positive constants that rely only on a, μ, s, σ, and u_0, and $2\sigma_0$ is the integer part of 2σ. As a direct result, it is easy to check that (11.2) implies

$$\|u(t)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \geq \frac{C_2}{(T^* - t)^{\frac{2(s\sigma+\sigma_0)+1}{6\sigma}}}, \quad \forall\, t \in [0, T^*).$$

In order to give more details on what it is going to be done in this chapter, we shall also demonstrate the following blow-up criteria related to the space $L^1(\mathbb{R}^3)$:

$$\int_t^{T^*} \left\| e^{\frac{a}{\sigma(\sqrt{\sigma})^{(n-1)}}|\cdot|^{\frac{1}{\sigma}}} \widehat{u}(\tau) \right\|^2_{L^1(\mathbb{R}^3)} d\tau = \infty, \tag{11.3}$$

and

$$\left\| e^{\frac{a}{\sigma(\sqrt{\sigma})^{(n-1)}}|\cdot|^{\frac{1}{\sigma}}} \widehat{u}(t) \right\|_{L^1(\mathbb{R}^3)} \geq \frac{8\pi^3\sqrt{\mu}}{\sqrt{T^* - t}}, \tag{11.4}$$

for all $t \in [0, T^*)$, $n \in \mathbb{N} \cup \{0\}$. Let us inform that the criterion (11.3) follows from the limit superior

$$\limsup_{t \nearrow T^*} \|u(t)\|_{H^s_{\frac{a}{(\sqrt{\sigma})^{(n-1)}},\sigma}(\mathbb{R}^3)} = \infty, \quad \forall\, n \in \mathbb{N} \cup \{0\}. \tag{11.5}$$

Notice that (11.4) is not trivial, provided that $\left\| e^{\frac{a}{\sigma(\sqrt{\sigma})^{(n-1)}}|\cdot|^{\frac{1}{\sigma}}} \widehat{u}(t) \right\|_{L^1(\mathbb{R}^3)}$ is finite for all $t \in [0, T^*)$, $n \in \mathbb{N} \cup \{0\}$. It can be concluded that due to the estimate (11.17) and the standard continuous embedding $H^s_{a,\sigma}(\mathbb{R}^3) \hookrightarrow H^s_{\frac{a}{(\sqrt{\sigma})^{(n-1)}},\sigma}(\mathbb{R}^3)$. Furthermore, by applying dominated convergence theorem in (11.4), one obtains

$$\|\widehat{u}(t)\|_{L^1(\mathbb{R}^3)} = \lim_{n\to\infty} \left\| e^{\frac{a}{\sigma(\sqrt{\sigma})^{(n-1)}}|\cdot|^{\frac{1}{\sigma}}} \widehat{u}(t) \right\|_{L^1(\mathbb{R}^3)} \geq \frac{8\pi^3\sqrt{\mu}}{\sqrt{T^*-t}}, \quad \forall\, t \in [0, T^*). \tag{11.6}$$

Besides, inequality (11.6) is not trivial as well. In fact, it follows from Lemmas 11.2.7 and 11.2.8, and (11.64).

It is also important to clarify that the lower bound given in (11.2) is not the only one that is obtained assuming the $H^s_{a,\sigma}(\mathbb{R}^3)$-norm. More specifically, we shall assure that

$$\|u(t)\|_{H^s_{\frac{a}{(\sqrt{\sigma})^n},\sigma}(\mathbb{R}^3)} \geq \frac{8\pi^3\sqrt{\mu}}{C_1\sqrt{T^*-t}}, \quad \forall\, t \in [0, T^*), n \in \mathbb{N} \cup \{0\}, \tag{11.7}$$

provided that C_1 depends only on a, σ, s, and n.

Observe that all the blow-up criteria obtained in [4] are particular cases of ours, it suffices to assume $s = 1$.

Finally, it is important to point out that the constants that appear in this chapter may change line by line, but they are denoted the same way. Moreover, $C_{p,l}$ means the constants that rely on p and l, and we sometimes drop the dependence of t or x in the domain of the vector fields.

As follows, we shall be interested in proving all the propositions stated above.

11.1 Local Existence and Uniqueness of Solutions

In this section, we demonstrate the existence of an instant $T > 0$ and a unique solution $u \in C([0, T]; H^s_{a,\sigma}(\mathbb{R}^3))$,[1] with $s \in \left(\frac{1}{2}, \frac{3}{2}\right)$, $a > 0$, and $\sigma \geq 1$, for the Navier–Stokes system (11.1).

First of all, we start presenting some basic lemmas that will be useful in the demonstration of the statement above.

[1] $H^s_{a,\sigma}(\mathbb{R}^3) := \left\{ f \in S'(\mathbb{R}^3) : \|f\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} := \int_{\mathbb{R}^3} (1+|\xi|^2)^s e^{2a|\xi|^{\frac{1}{\sigma}}} |\widehat{f}(\xi)|^2\, d\xi < \infty \right\}$ is Gevrey–Sobolev space endowed with the inner product $\langle f, g\rangle_{H^s_{a,\sigma}(\mathbb{R}^3)} := \int_{\mathbb{R}^3} (1+|\xi|^2)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \widehat{f}(\xi) \cdot \widehat{g}(\xi)\, d\xi$, where $\cdot$ and $|\cdot|$ denote Euclidean inner product and norm, respectively ($S'(\mathbb{R}^3)$ is the set of all the distributions). Here $\mathcal{F}(f)(\xi) = \widehat{f}(\xi) := \int_{\mathbb{R}^3} e^{-i\xi\cdot x} f(x)\, dx$ and $\mathcal{F}^{-1}(f)(\xi) := (2\pi)^{-3} \int_{\mathbb{R}^3} e^{i\xi\cdot x} f(x)\, dx, \forall\, \xi \in \mathbb{R}^3$.

Lemma 11.1.1 (see [8]) *Let* $(X, \|\cdot\|)$ *be a Banach space and* $B : X \times X \to X$ *be a continuous bilinear operator, i.e., there exists a positive constant* C_1 *such that*

$$\|B(x, y)\| \leq C_1 \|x\| \|y\|, \quad \forall\, x, y \in X.$$

Then, for each $x_0 \in X$ *that satisfies* $4C_1\|x_0\| < 1$*, one has that the equation*

$$a = x_0 + B(a, a) \tag{11.8}$$

admits a solution $x = a \in X$*. Moreover,* x *obeys the inequality* $\|x\| \leq 2\|x_0\|$ *and it is the only one such that* $\|x\| \leq \frac{1}{2C_1}$*.*

The following result has been proved by [3] and it is useful in order to obtain some important inequalities related to the elementary exponential function.

Lemma 11.1.2 *The following inequality holds:*

$$(a + b)^r \leq ra^r + b^r, \quad \forall\, 0 \leq a \leq b, r \in (0, 1]. \tag{11.9}$$

Proof First of all, notice that if $b = 0$, then $a = 0$ and, consequently, (11.9) follows. Thus, assume that $b > 0$ and let $c = a/b \in [0, 1]$. Now, apply Taylor's theorem to the function $t \mapsto (1 + t)^r$, with $t \in [0, c]$, in order to obtain $\gamma \in [0, c]$ such that

$$(1 + c)^r = 1 + rc + \frac{r(r - 1)(1 + \gamma)^{r-2}}{2} c^2.$$

By using the fact that $r, \gamma \in [0, 1]$, one has $(1+c)^r \leq 1+rc$. Moreover, $c, r \in [0, 1]$ implies that $c \leq c^r$. As a result, $(1 + c)^r \leq 1 + rc^r$. Replace $c = a/b$ in this last inequality to prove (11.9). □

Now, let us introduce two consequences of Lemma 11.1.2.

Lemma 11.1.3 *The inequality below is valid:*

$$e^{a|\xi|^{\frac{1}{\sigma}}} \leq e^{a \max\{|\xi-\eta|,|\eta|\}^{\frac{1}{\sigma}}} e^{\frac{a}{\sigma} \min\{|\xi-\eta|,|\eta|\}^{\frac{1}{\sigma}}}, \quad \forall\, \xi, \eta \in \mathbb{R}^3, a > 0, \sigma \geq 1.$$

Proof Lemma 11.1.2 assures that

$$\begin{aligned} a|\xi|^{\frac{1}{\sigma}} &= a|\xi - \eta + \eta|^{\frac{1}{\sigma}} \leq a(|\xi - \eta| + |\eta|)^{\frac{1}{\sigma}} \\ &\leq a(\max\{|\xi - \eta|, |\eta|\} + \min\{|\xi - \eta|, |\eta|\})^{\frac{1}{\sigma}} \\ &\leq a \max\{|\xi - \eta|, |\eta|\}^{\frac{1}{\sigma}} + \frac{a}{\sigma} \min\{|\xi - \eta|, |\eta|\}^{\frac{1}{\sigma}}. \end{aligned}$$

Hence, one has

$$e^{a|\xi|^{\frac{1}{\sigma}}} \leq e^{a \max\{|\xi-\eta|,|\eta|\}^{\frac{1}{\sigma}} + \frac{a}{\sigma} \min\{|\xi-\eta|,|\eta|\}^{\frac{1}{\sigma}}} = e^{a \max\{|\xi-\eta|,|\eta|\}^{\frac{1}{\sigma}}} e^{\frac{a}{\sigma} \min\{|\xi-\eta|,|\eta|\}^{\frac{1}{\sigma}}}.$$

It finishes the proof of Lemma 11.1.3. □

Lemma 11.1.4 *Let* $\xi, \eta \in \mathbb{R}^3$, $a > 0$, *and* $\sigma \geq 1$. *Then, it holds*

$$e^{a|\xi|^{\frac{1}{\sigma}}} \leq e^{a|\xi-\eta|^{\frac{1}{\sigma}}} e^{a|\eta|^{\frac{1}{\sigma}}}. \tag{11.10}$$

Proof It is an implication that comes from Lemma 11.1.3 and the fact that $\sigma \geq 1$. □

The below lemma presents an interpolation property involving the space $\dot{H}^s(\mathbb{R}^3)$,[2] and it has been proved by Chemin [9].

Lemma 11.1.5 (see [9]) *Let* $(s_1, s_2) \in \mathbb{R}^2$, *such that* $s_1 < \frac{3}{2}$ *and* $s_1 + s_2 > 0$. *Then, there exists a positive constant* C_{s_1,s_2} *such that, for all* $f, g \in \dot{H}^{s_1}(\mathbb{R}^3) \cap \dot{H}^{s_2}(\mathbb{R}^3)$, *we have*

$$\|fg\|_{\dot{H}^{s_1+s_2-\frac{3}{2}}(\mathbb{R}^3)} \leq C_{s_1,s_2}\big[\|f\|_{\dot{H}^{s_1}(\mathbb{R}^3)}\|g\|_{\dot{H}^{s_2}(\mathbb{R}^3)} + \|f\|_{\dot{H}^{s_2}(\mathbb{R}^3)}\|g\|_{\dot{H}^{s_1}(\mathbb{R}^3)}\big].$$

If $s_1 < \frac{3}{2}$, $s_2 < \frac{3}{2}$, *and* $s_1 + s_2 > 0$, *then there is a positive constant* C_{s_1,s_2} *such that*

$$\|fg\|_{\dot{H}^{s_1+s_2-\frac{3}{2}}(\mathbb{R}^3)} \leq C_{s_1,s_2}\|f\|_{\dot{H}^{s_1}(\mathbb{R}^3)}\|g\|_{\dot{H}^{s_2}(\mathbb{R}^3)}.$$

Benameur and Jlali [4] have proved a version of Chemin's lemma (see [9]) by considering Sobolev–Gevrey space $\dot{H}^s_{a,\sigma}(\mathbb{R}^3)$.[3] Let us introduce this result exactly as it has been enunciated and shown in [4].

Lemma 11.1.6 (see [4]) *Let* $a > 0$, $\sigma \geq 1$, *and* $(s_1, s_2) \in \mathbb{R}^2$, *such that* $s_1 < \frac{3}{2}$ *and* $s_1 + s_2 > 0$. *Then, there exists a positive constant* C_{s_1,s_2} *such that, for all* $f, g \in \dot{H}^{s_1}_{a,\sigma}(\mathbb{R}^3) \cap \dot{H}^{s_2}_{a,\sigma}(\mathbb{R}^3)$, *we have*

$$\|fg\|_{\dot{H}^{s_1+s_2-\frac{3}{2}}_{a,\sigma}(\mathbb{R}^3)} \leq C_{s_1,s_2}\big[\|f\|_{\dot{H}^{s_1}_{a,\sigma}(\mathbb{R}^3)}\|g\|_{\dot{H}^{s_2}_{a,\sigma}(\mathbb{R}^3)} + \|f\|_{\dot{H}^{s_2}_{a,\sigma}(\mathbb{R}^3)}\|g\|_{\dot{H}^{s_1}_{a,\sigma}(\mathbb{R}^3)}\big]. \tag{11.11}$$

If $s_1 < \frac{3}{2}$, $s_2 < \frac{3}{2}$, *and* $s_1 + s_2 > 0$, *then there is a positive constant* C_{s_1,s_2} *such that*

$$\|fg\|_{\dot{H}^{s_1+s_2-\frac{3}{2}}_{a,\sigma}(\mathbb{R}^3)} \leq C_{s_1,s_2}\|f\|_{\dot{H}^{s_1}_{a,\sigma}(\mathbb{R}^3)}\|g\|_{\dot{H}^{s_2}_{a,\sigma}(\mathbb{R}^3)}. \tag{11.12}$$

---

[2] $\dot{H}^s(\mathbb{R}^3) = \big\{f \in S'(\mathbb{R}^3) : \|f\|^2_{\dot{H}^s(\mathbb{R}^3)} := \int_{\mathbb{R}^3} |\xi|^{2s}|\widehat{f}(\xi)|^2\, d\xi < \infty\big\}$ is the homogenous Sobolev space.

[3] The Sobolev–Gevrey space $\dot{H}^s_{a,\sigma}(\mathbb{R}^3) := \big\{f \in S'(\mathbb{R}^3) : \|f\|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} := \int_{\mathbb{R}^3} |\xi|^{2s} e^{2a|\xi|^{\frac{1}{\sigma}}} |\widehat{f}(\xi)|^2\, d\xi < \infty\big\}$ is endowed with the inner product $\langle f, g\rangle_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} := \int_{\mathbb{R}^3} |\xi|^{2s} e^{2a|\xi|^{\frac{1}{\sigma}}}\, \widehat{f}(\xi) \cdot \widehat{g}(\xi)\, d\xi$.

Proof We aim to apply Lemma 11.1.5. Thereby, to accomplish this goal, it is necessary to use the elementary equality

$$\widehat{fg}(\xi) = (2\pi)^{-3}(\widehat{f} * \widehat{g})(\xi), \quad \forall \xi \in \mathbb{R}^3.$$

Thus, we estimate the expression on the left-hand side of inequalities (11.11) and (11.12) as follows[4]:

$$\begin{aligned}
\|fg\|^2_{\dot{H}^{s_1+s_2-\frac{3}{2}}_{a,\sigma}(\mathbb{R}^3)} &= \int_{\mathbb{R}^3} |\xi|^{2s_1+2s_2-3} e^{2a|\xi|^{\frac{1}{\sigma}}} |\widehat{fg}(\xi)|^2 \, d\xi \\
&= (2\pi)^{-6} \int_{\mathbb{R}^3} |\xi|^{2s_1+2s_2-3} e^{2a|\xi|^{\frac{1}{\sigma}}} |\widehat{f} * \widehat{g}(\xi)|^2 \, d\xi \\
&\leq (2\pi)^{-6} \int_{\mathbb{R}^3} |\xi|^{2s_1+2s_2-3} e^{2a|\xi|^{\frac{1}{\sigma}}} \left(\int_{\mathbb{R}^3} |\widehat{f}(\xi - \eta)||\widehat{g}(\eta)| \, d\eta \right)^2 d\xi \\
&= (2\pi)^{-6} \int_{\mathbb{R}^3} |\xi|^{2s_1+2s_2-3} \left(\int_{\mathbb{R}^3} e^{a|\xi|^{\frac{1}{\sigma}}} |\widehat{f}(\xi - \eta)||\widehat{g}(\eta)| \, d\eta \right)^2 d\xi.
\end{aligned}$$

Moreover, inequality (11.10) implies the following results:

$$\begin{aligned}
\|fg\|^2_{\dot{H}^{s_1+s_2-\frac{3}{2}}_{a,\sigma}(\mathbb{R}^3)} &\leq (2\pi)^{-6} \int_{\mathbb{R}^3} |\xi|^{2s_1+2s_2-3} \\
&\quad \times \left(\int_{\mathbb{R}^3} e^{a|\xi-\eta|^{\frac{1}{\sigma}}} \big|\widehat{f}(\xi - \eta)\big| e^{a|\eta|^{\frac{1}{\sigma}}} |\widehat{g}(\eta)| \, d\eta \right)^2 d\xi \\
&= (2\pi)^{-6} \int_{\mathbb{R}^3} |\xi|^{2s_1+2s_2-3} \Big\{ \Big[\big(e^{a|\cdot|^{\frac{1}{\sigma}}} \big|\widehat{f}\big| \big) * \big(e^{a|\cdot|^{\frac{1}{\sigma}}} |\widehat{g}| \big) \Big](\xi) \Big\}^2 d\xi \\
&= \int_{\mathbb{R}^3} |\xi|^{2s_1+2s_2-3} \\
&\quad \times \Big\{ \mathcal{F}\Big[\mathcal{F}^{-1}\big(e^{a|\xi|^{\frac{1}{\sigma}}} \big|\widehat{f}(\xi)\big| \big) \mathcal{F}^{-1}\big(e^{a|\xi|^{\frac{1}{\sigma}}} |\widehat{g}(\xi)| \big) \Big] \Big\}^2 d\xi \\
&= \big\| \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} \big|\widehat{f}\big| \big) \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} |\widehat{g}| \big) \big\|^2_{\dot{H}^{s_1+s_2-\frac{3}{2}}(\mathbb{R}^3)}.
\end{aligned}$$

Now, we are ready to apply Lemma 11.1.5 and, consequently, deduce (11.11). In fact, one has

$$\begin{aligned}
\|fg\|_{\dot{H}^{s_1+s_2-\frac{3}{2}}_{a,\sigma}(\mathbb{R}^3)} &\leq \big\| \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} \big|\widehat{f}\big| \big) \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} |\widehat{g}| \big) \big\|_{\dot{H}^{s_1+s_2-\frac{3}{2}}(\mathbb{R}^3)} \\
&\leq C_{s_1,s_2} \big[\big\| \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} \big|\widehat{f}\big| \big) \big\|_{\dot{H}^{s_1}(\mathbb{R}^3)} \big\| \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} |\widehat{g}| \big) \big\|_{\dot{H}^{s_2}(\mathbb{R}^3)}
\end{aligned}$$

[4]The usual convolution is given by $\varphi * \psi(x) = \int_{\mathbb{R}^3} \varphi(x-y)\psi(y)\,dy$, where $\varphi, \psi : \mathbb{R}^3 \to \mathbb{R}$.

$$+ \left\| \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} |\widehat{f}|\big) \right\|_{\dot{H}^{s_2}(\mathbb{R}^3)} \left\| \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} |\widehat{g}|\big) \right\|_{\dot{H}^{s_1}(\mathbb{R}^3)} \Big]$$
$$= C_{s_1,s_2}\Big[\|f\|_{\dot{H}^{s_1}_{a,\sigma}(\mathbb{R}^3)} \|g\|_{\dot{H}^{s_2}_{a,\sigma}(\mathbb{R}^3)} + \|f\|_{\dot{H}^{s_2}_{a,\sigma}(\mathbb{R}^3)} \|g\|_{\dot{H}^{s_1}_{a,\sigma}(\mathbb{R}^3)} \Big].$$

On the other hand, if $s_1, s_2 < \frac{3}{2}$ and $s_1 + s_2 > 0$, use Lemma 11.1.5 again in order to obtain

$$\begin{aligned}
\|fg\|_{\dot{H}^{s_1+s_2-\frac{3}{2}}_{a,\sigma}(\mathbb{R}^3)} &\le \left\| \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} |\widehat{f}|\big) \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} |\widehat{g}|\big) \right\|^2_{\dot{H}^{s_1+s_2-\frac{3}{2}}(\mathbb{R}^3)} \\
&\le C_{s_1,s_2} \left\| \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} |\widehat{f}|\big) \right\|_{\dot{H}^{s_1}(\mathbb{R}^3)} \left\| \mathcal{F}^{-1}\big(e^{a|\cdot|^{\frac{1}{\sigma}}} |\widehat{g}|\big) \right\|_{\dot{H}^{s_2}(\mathbb{R}^3)} \\
&= C_{s_1,s_2} \|f\|_{\dot{H}^{s_1}_{a,\sigma}(\mathbb{R}^3)} \|g\|_{\dot{H}^{s_2}_{a,\sigma}(\mathbb{R}^3)}.
\end{aligned}$$

It completes the proof of (11.12). □

The next result presents our extension of Lemma 2.5 given in [4]; since, this last lemma is the same as Lemma 11.1.7, if it is assumed that $s = 1$.

Lemma 11.1.7 *Let $a > 0$, $\sigma > 1$, and $s \in \big[0, \frac{3}{2}\big)$. For every $f, g \in H^s_{a,\sigma}(\mathbb{R}^3)$, we have $fg \in H^s_{a,\sigma}(\mathbb{R}^3)$. More precisely, one obtains*

i) $\|fg\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \le 2^{\frac{2s-5}{2}} \pi^{-3} \Big[\big\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{f} \big\|_{L^1(\mathbb{R}^3)} \|g\|_{H^s_{a,\sigma}(\mathbb{R}^3)} + \big\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{g} \big\|_{L^1(\mathbb{R}^3)} \|f\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \Big]$;
ii) $\|fg\|_{H^s_{a,\sigma}(\mathbb{R}^3))} \le 2^{s-2} \pi^{-3} C_{a,\sigma,s} \|f\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \|g\|_{H^s_{a,\sigma}(\mathbb{R}^3)}$,

where $C_{a,\sigma,s} := \sqrt{\frac{4\pi\sigma\Gamma(\sigma(3-2s))}{\left[2(a-\frac{a}{\sigma})\right]^{\sigma(3-2s)}}} < \infty$, and Γ is the elementary gamma function.[5]

Proof This result is a consequence of Lemma 11.1.3. First of all, let us estimate the left-hand side of the inequalities given in **i)** and **ii)**. Thus,

$$\begin{aligned}
\|fg\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} &= \int_{\mathbb{R}^3} \big(1+|\xi|^2\big)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \big|\widehat{fg}(\xi)\big|^2 \, d\xi \\
&= (2\pi)^{-6} \int_{\mathbb{R}^3} \big(1+|\xi|^2\big)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \big|\widehat{f} * \widehat{g}(\xi)\big|^2 \, d\xi \\
&\le (2\pi)^{-6} \int_{\mathbb{R}^3} \left(\int_{\mathbb{R}^3} \big(1+|\xi|^2\big)^{\frac{s}{2}} e^{a|\xi|^{\frac{1}{\sigma}}} \big|\widehat{f}(\xi-\eta)\big| |\widehat{g}(\eta)| \, d\eta \right)^2 d\xi \\
&\le (2\pi)^{-6} \int_{\mathbb{R}^3} \Big(\int_{|\eta|\le|\xi-\eta|} \big(1+|\xi|^2\big)^{\frac{s}{2}} e^{a|\xi|^{\frac{1}{\sigma}}} \big|\widehat{f}(\xi-\eta)\big| |\widehat{g}(\eta)| \, d\eta \\
&\quad + \int_{|\eta|>|\xi-\eta|} \big(1+|\xi|^2\big)^{\frac{s}{2}} e^{a|\xi|^{\frac{1}{\sigma}}} \big|\widehat{f}(\xi-\eta)\big| |\widehat{g}(\eta)| \, d\eta \Big)^2 d\xi.
\end{aligned}$$

[5] $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} \, dx$, for all $z = x + iy \in \mathbb{C}$, with $x > 0$.

By using basic arguments, it is easy to check that

$$\Big(1+|\xi|^2\Big)^{\frac{s}{2}} \leq \Big[1+(|\xi-\eta|+|\eta|)^2\Big]^{\frac{s}{2}} \leq \Big[1+(2\max\{|\xi-\eta|,|\eta|\})^2\Big]^{\frac{s}{2}}$$
$$\leq 2^s\Big[1+\max\{|\xi-\eta|,|\eta|\}^2\Big]^{\frac{s}{2}}. \qquad (11.13)$$

Now, we are interested in applying Lemma 11.1.3 in order to obtain

$$\|fg\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \leq (2\pi)^{-6}2^{2s}\int_{\mathbb{R}^3}\bigg(\int_{|\eta|\leq|\xi-\eta|}\Big(1+|\xi-\eta|^2\Big)^{\frac{s}{2}}e^{a|\xi-\eta|^{\frac{1}{\sigma}}}$$
$$\times\big|\widehat{f}(\xi-\eta)\big|\,e^{\frac{a}{\sigma}|\eta|^{\frac{1}{\sigma}}}|\widehat{g}(\eta)|\,d\eta$$
$$+\,int_{|\eta|>|\xi-\eta|}e^{\frac{a}{\sigma}|\xi-\eta|^{\frac{1}{\sigma}}}\,\big|\widehat{f}(\xi-\eta)\big|\Big(1+|\eta|^2\Big)^{\frac{s}{2}}e^{a|\eta|^{\frac{1}{\sigma}}}|\widehat{g}(\eta)|\,d\eta\bigg)^2d\xi$$
$$\leq (2\pi)^{-6}2^{2s+1}\bigg[\int_{\mathbb{R}^3}\bigg(\int_{\mathbb{R}^3}(1+|\xi-\eta|^2)^{\frac{s}{2}}e^{a|\xi-\eta|^{\frac{1}{\sigma}}}$$
$$\times\big|\widehat{f}(\xi-\eta)\big|e^{\frac{a}{\sigma}|\eta|^{\frac{1}{\sigma}}}|\widehat{g}(\eta)|\,d\eta\bigg)^2d\xi$$
$$+\int_{\mathbb{R}^3}\bigg(\int_{\mathbb{R}^3}e^{\frac{a}{\sigma}|\xi-\eta|^{\frac{1}{\sigma}}}\big|\widehat{f}(\xi-\eta)\big|(1+|\eta|^2)^{\frac{s}{2}}e^{a|\eta|^{\frac{1}{\sigma}}}|\widehat{g}(\eta)|\,d\eta\bigg)^2d\xi\bigg].$$

Rewriting the last inequality above reached, it holds

$$\|fg\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \leq (2\pi)^{-6}2^{2s+1}\big\|\big[(1+|\cdot|^2)^{\frac{s}{2}}e^{a|\cdot|^{\frac{1}{\sigma}}}\big|\widehat{f}\big|\big]*\big[e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}|\widehat{g}|\big]\big\|^2_{L^2(\mathbb{R}^3)}$$
$$+\,(2\pi)^{-6}2^{2s+1}\big\|\big[e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\big|\widehat{f}\big|\big]*\big[(1+|\cdot|^2)^{\frac{s}{2}}e^{a|\cdot|^{\frac{1}{\sigma}}}|\widehat{g}|\big]\big\|^2_{L^2(\mathbb{R}^3)}.$$

Consequently, it follows from Young's inequality for convolutions[6] that

$$\|fg\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \leq 2^{2s-5}\pi^{-6}\big[\big\|\big(1+|\cdot|^2\big)^{\frac{s}{2}}e^{a|\cdot|^{\frac{1}{\sigma}}}\widehat{f}\big\|^2_{L^2(\mathbb{R}^3)}\big\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{g}\big\|^2_{L^1(\mathbb{R}^3)}$$
$$+\,\big\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{f}\big\|^2_{L^1(\mathbb{R}^3)}\big\|\big(1+|\cdot|^2\big)^{\frac{s}{2}}e^{a|\cdot|^{\frac{1}{\sigma}}}\widehat{g}\big\|^2_{L^2(\mathbb{R}^3)}\big]. \qquad (11.14)$$

[6]Let $1\leq p,q,r\leq\infty$ such that $1+\frac{1}{r}=\frac{1}{p}+\frac{1}{q}$. Assume that $f\in L^p(\mathbb{R}^3)$ and $g\in L^q(\mathbb{R}^3)$; then, $\|f*g\|_{L^r(\mathbb{R}^3)}\leq\|f\|_{L^p(\mathbb{R}^3)}\|g\|_{L^q(\mathbb{R}^3)}$.

Notice that the $L^2(\mathbb{R}^3)$-norm of $\left(1+|\xi|^2\right)^{\frac{s}{2}} e^{a|\xi|^{\frac{1}{\sigma}}} \widehat{f}(\xi)$ presented above can be replaced by the $H^s_{a,\sigma}(\mathbb{R}^3)$-norm of f. More precisely, we have

$$\left\|\left(1+|\cdot|^2\right)^{\frac{s}{2}} e^{a|\cdot|^{\frac{1}{\sigma}}} \widehat{f}\right\|^2_{L^2(\mathbb{R}^3)} = \int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left|\hat{f}(\xi)\right|^2 d\xi = \|f\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}. \tag{11.15}$$

This same process can be applied to the equivalent term related to g in (11.14). Thereby, it is true that

$$\left\|\left(1+|\cdot|^2\right)^{\frac{s}{2}} e^{a|\cdot|^{\frac{1}{\sigma}}} \widehat{g}\right\|^2_{L^2(\mathbb{R}^3)} = \|g\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}. \tag{11.16}$$

As a consequence, replace (11.15) and (11.16) in (11.14) in order to get

$$\begin{aligned}\|fg\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \leq 2^{\frac{2s-5}{2}} \pi^{-3} \Big[&\left\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{f} \right\|_{L^1(\mathbb{R}^3)} \|g\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \\ &+ \left\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{g} \right\|_{L^1(\mathbb{R}^3)} \|f\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \Big].\end{aligned}$$

It completes the proof of **i)**.

Now, let us prove **ii)** by using the results established above. Thus, applying Cauchy–Schwarz's inequality,[7] one infers

$$\begin{aligned}&\left\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{g} \right\|_{L^1(\mathbb{R}^3)} \\ &= \int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}} |\widehat{g}(\xi)|\, d\xi \\ &\leq \left(\int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^{-s} e^{2\left(\frac{a}{\sigma}-a\right)|\xi|^{\frac{1}{\sigma}}}\, d\xi \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} |\widehat{g}(\xi)|^2\, d\xi \right)^{\frac{1}{2}} \\ &\leq \left(\int_{\mathbb{R}^3} |\xi|^{-2s} e^{2\left(\frac{a}{\sigma}-a\right)|\xi|^{\frac{1}{\sigma}}}\, d\xi \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} |\widehat{g}(\xi)|^2\, d\xi \right)^{\frac{1}{2}} \\ &=: C_{a,\sigma,s} \|g\|_{H^s_{a,\sigma}(\mathbb{R}^3)},\end{aligned} \tag{11.17}$$

where

$$C^2_{a,\sigma,s} = \frac{4\pi\sigma\Gamma(\sigma(3-2s))}{\left[2\left(a-\frac{a}{\sigma}\right)\right]^{\sigma(3-2s)}},$$

[7] Let $p, q \in [1,\infty]$ such that $\frac{1}{p}+\frac{1}{q}=1$. Consider that $f \in L^p(\mathbb{R}^3)$ and $g \in L^q(\mathbb{R}^3)$; then, $\|fg\|_{L^1(\mathbb{R}^3)} \leq \|f\|_{L^p(\mathbb{R}^3)} \|g\|_{L^q(\mathbb{R}^3)}$. Here $L^p(\mathbb{R}^3)$ denotes the usual Lebesgue space, where $\|f\|_{L^p(\mathbb{R}^3)} := \left(\int_{\mathbb{R}^3} |f(x)|^p\, dx\right)^{\frac{1}{p}}$ and $\|f\|_{L^\infty(\mathbb{R}^3)} := \operatorname{esssup}_{x\in\mathbb{R}^3}\{|f(x)|\}$.

since $\sigma > 1$ and $0 \leq s < 3/2$. Analogously, we obtain

$$\left\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{f} \right\|_{L^1(\mathbb{R}^3)} \leq C_{a,\sigma,s} \| f \|_{H^s_{a,\sigma}(\mathbb{R}^3)}. \tag{11.18}$$

Hence, by combining (11.14)–(11.18), we have

$$\| fg \|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \leq 2^{2s-4} \pi^{-6} C^2_{a,\sigma,s} \| f \|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \| g \|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}.$$

It proves **ii)**. □

The next result is our version of Lemma 2.8 in [4], provided that this last lemma is the same as Lemma 11.1.8, whether it is considered $s = 1$.

Lemma 11.1.8 *Let $s \geq 0, a > 0, \sigma \geq 1$, and $f \in H^s_{a,\sigma}(\mathbb{R}^3)$. Then, the following inequalities hold:*

$$\begin{aligned} \| f \|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} &\leq 2^s \left[e^{2a} (2\pi)^3 \| f \|^2_{L^2(\mathbb{R}^3)} + \| f \|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} \right] \\ &\leq 2^s \left[e^{2a} + 1 \right] \| f \|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}. \end{aligned} \tag{11.19}$$

Proof This result follows directly from the definition of the spaces $H^s_{a,\sigma}(\mathbb{R}^3)$, $\dot{H}^s_{a,\sigma}(\mathbb{R}^3)$, and $L^2(\mathbb{R}^3)$. In fact, note that, by using Parseval's identity, i.e.,

$$\| f \|^2_{L^2(\mathbb{R}^3)} = (2\pi)^{-3} \| \hat{f} \|^2_{L^2(\mathbb{R}^3)},$$

one obtains

$$\begin{aligned} \| f \|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} &= \int_{\mathbb{R}^3} \left(1 + |\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left| \hat{f}(\xi) \right|^2 d\xi \\ &= \int_{|\xi| \leq 1} \left(1 + |\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left| \hat{f}(\xi) \right|^2 d\xi \\ &\quad + \int_{|\xi| > 1} \left(1 + |\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left| \hat{f}(\xi) \right|^2 d\xi \\ &\leq 2^s e^{2a} \int_{\mathbb{R}^3} \left| \hat{f}(\xi) \right|^2 d\xi + 2^s \int_{\mathbb{R}^3} |\xi|^{2s} e^{2a|\xi|^{\frac{1}{\sigma}}} \left| \hat{f}(\xi) \right|^2 d\xi \\ &= 2^s e^{2a} (2\pi)^3 \| f \|^2_{L^2(\mathbb{R}^3)} + 2^s \| f \|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)}. \end{aligned}$$

It demonstrates the first inequality in (11.19).

By applying the last equality above, one infers

$$\begin{aligned} 2^s e^{2a} (2\pi)^3 \| f \|^2_{L^2(\mathbb{R}^3)} + 2^s \| f \|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} = 2^s e^{2a} \int_{\mathbb{R}^3} \left| \hat{f}(\xi) \right|^2 d\xi + 2^s \\ \times \int_{\mathbb{R}^3} |\xi|^{2s} e^{2a|\xi|^{\frac{1}{\sigma}}} \left| \hat{f}(\xi) \right|^2 d\xi \end{aligned}$$

$$
\begin{aligned}
&\le 2^s\big[e^{2a}+1\big]\int_{\mathbb{R}^3}\big(1+|\xi|^2\big)^s \\
&\quad\times e^{2a|\xi|^{\frac{1}{\sigma}}}\big|\hat{f}(\xi)\big|^2\,d\xi \\
&= 2^s\big[e^{2a}+1\big]\|f\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}.
\end{aligned}
$$

Therefore, the proof of the second inequality in (11.19) is given. □

Remark 11.1.9 It is worth to observe that the demonstration of the lemma above establishes, for instance, the standard embeddings $H^s_{a,\sigma}(\mathbb{R}^3) \hookrightarrow L^2(\mathbb{R}^3)$ and $H^s_{a,\sigma}(\mathbb{R}^3) \hookrightarrow \dot{H}^s_{a,\sigma}(\mathbb{R}^3)$ ($s \ge 0$). In fact, note that in the proof of Lemma 11.1.8, we have proved

$$
\begin{aligned}
\|f\|^2_{L^2(\mathbb{R}^3)} &= (2\pi)^{-3}\int_{\mathbb{R}^3}\big|\hat{f}(\xi)\big|^2\,d\xi \le (2\pi)^{-3}\int_{\mathbb{R}^3}\big(1+|\xi|^2\big)^s e^{2a|\xi|^{\frac{1}{\sigma}}}\big|\hat{f}(\xi)\big|^2\,d\xi \\
&= (2\pi)^{-3}\|f\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}.
\end{aligned}
$$

Consequently, the continuous embedding $H^s_{a,\sigma}(\mathbb{R}^3) \hookrightarrow L^2(\mathbb{R}^3)$ ($s \ge 0$) is given by the inequality

$$
\|f\|_{L^2(\mathbb{R}^3)} \le (2\pi)^{-\frac{3}{2}}\|f\|_{H^s_{a,\sigma}(\mathbb{R}^3)}.
$$

The other embedding related above follows directly from the following results:

$$
\begin{aligned}
\|f\|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} &= \int_{\mathbb{R}^3}|\xi|^{2s}e^{2a|\xi|^{\frac{1}{\sigma}}}\big|\hat{f}(\xi)\big|^2\,d\xi \le \int_{\mathbb{R}^3}\big(1+|\xi|^2\big)^s e^{2a|\xi|^{\frac{1}{\sigma}}}\big|\hat{f}(\xi)\big|^2\,d\xi \\
&= \|f\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}.
\end{aligned}
$$

Finally, we enunciate the next elementary result, which follows from the basic tools obtained in calculus.

Lemma 11.1.10 (See [5]) *Let $a, b > 0$. Then, $\lambda^a e^{-b\lambda} \le a^a (eb)^{-a}$ for all $\lambda > 0$.*

Proof Consider the real function f defined by $f(\lambda) = \lambda^a e^{-b\lambda}$, for all $\lambda > 0$. It is easy to verify that f attains its maximum at a/b since

$$
f'(\lambda) = \lambda^a e^{-b\lambda}\left[\frac{a}{\lambda} - b\right] \text{ and } f''(\lambda) = \lambda^a e^{-b\lambda}\left[\left(\frac{a}{\lambda} - b\right)^2 - \frac{a}{\lambda^2}\right], \quad \forall\,\lambda > 0.
$$

Therefore, the demonstration of this lemma is complete. □

Now, let us precisely enunciate our result that assures the local existence and uniqueness of solutions for the Navier–Stokes equations (11.1).

Theorem 11.1.11 *Assume that* $a > 0$, $\sigma \geq 1$, *and* $s \in \left(\frac{1}{2}, \frac{3}{2}\right)$. *Let* $u_0 \in H^s_{a,\sigma}(\mathbb{R}^3)$ *such that* $\operatorname{div} u_0 = 0$. *Then, there exist an instant* $T = T_{s,a,\mu,u_0} > 0$ *and a unique solution* $u \in C\big([0, T]; H^s_{a,\sigma}(\mathbb{R}^3)\big)$ *for the Navier–Stokes equations* (11.1).

Proof Our aim in this demonstration is to assure that all the assumptions presented in Lemma 11.1.1 are satisfied if (11.25) and (11.26) are considered; thus, first of all, let us rewrite the Navier–Stokes equations (11.1) as in (11.8).

Use the heat semigroup $e^{\mu\Delta(t-\tau)}$, with $\tau \in [0, t]$, in the first equation[8] given in (11.1), and then, integrate the obtained result over the interval $[0, t]$ to reach

$$\int_0^t e^{\mu\Delta(t-\tau)} u_\tau \, d\tau + \int_0^t e^{\mu\Delta(t-\tau)} [u \cdot \nabla u + \nabla p] \, d\tau = \mu \int_0^t e^{\mu\Delta(t-\tau)} \Delta u \, d\tau.$$

By applying integration by parts to the first integral above and using the properties of the heat semigroup, one deduces

$$u(t) = e^{\mu\Delta t} u_0 - \int_0^t e^{\mu\Delta(t-\tau)} [u \cdot \nabla u + \nabla p] \, d\tau. \tag{11.20}$$

Let us recall that Helmontz's projector P_H (see [10] and the references therein) is well defined and it is a linear operator such that

$$P_H(u \cdot \nabla u) = u \cdot \nabla u + \nabla p, \tag{11.21}$$

and also

$$\mathcal{F}[P_H(f)](\xi) = \widehat{f}(\xi) - \frac{\widehat{f}(\xi) \cdot \xi}{|\xi|^2} \xi. \tag{11.22}$$

Notice that equality (11.22) implies that

$$|\mathcal{F}[P_H(f)](\xi)|^2 = \left| \widehat{f}(\xi) - \frac{\widehat{f}(\xi) \cdot \xi}{|\xi|^2} \xi \right|^2 = |\widehat{f}(\xi)|^2 - \frac{|\widehat{f}(\xi) \cdot \xi|^2}{|\xi|^2} \leq |\widehat{f}(\xi)|^2. \tag{11.23}$$

On the other hand, by replacing (11.21) in (11.20), it follows that

$$u(t) = e^{\mu\Delta t} u_0 - \int_0^t e^{\mu\Delta(t-\tau)} P_H[u \cdot \nabla u] \, d\tau.$$

---

[8] In the Navier–Stokes equations (11.1), the usual Laplacian for $f = (f_1, f_2, \ldots, f_n)$ is established by $\Delta f = (\Delta f_1, \Delta f_2, \ldots, \Delta f_n)$, where $\Delta f_j = \sum_{i=1}^3 D_i^2 f_j$. The gradient field is defined by $\nabla f = (\nabla f_1, \nabla f_2, \ldots, \nabla f_n)$, where $\nabla f_j = (D_1 f_j, D_2 f_j, D_3 f_j)$ $(j = 1, 2, \ldots, n)$. Besides, $f \cdot \nabla g = \sum_{i=1}^3 f_i D_i g$, where $f = (f_1, f_2, f_3)$ and $g = (g_1, g_2, g_3)$. At last, the standard divergent is given by $\operatorname{div} f = \sum_{i=1}^3 D_i f_i$ provided that $f = (f_1, f_2, f_3)$ and $D_i = \partial/\partial x_i$ $(i = 1, 2, 3)$.

It is well known that $u \cdot \nabla u = \sum_{j=1}^{3} u_j D_j u$. Thereby, one has

$$\begin{aligned} u(t) &= e^{\mu\Delta t}u_0 - \int_0^t e^{\mu\Delta(t-\tau)} P_H(u \cdot \nabla u)\, d\tau \\ &= e^{\mu\Delta t}u_0 - \int_0^t e^{\mu\Delta(t-\tau)} P_H\left[\sum_{j=1}^{3}(u_j D_j u)\right] d\tau \\ &= e^{\mu\Delta t}u_0 - \int_0^t e^{\mu\Delta(t-\tau)} P_H\left[\sum_{j=1}^{3} D_j(u_j u)\right] d\tau, \end{aligned}$$

provided that $\operatorname{div} u = 0$. Rewriting this last equality above, we get

$$u(t) = e^{\mu\Delta t}u_0 - \int_0^t e^{\mu\Delta(t-\tau)} P_H\left[\sum_{j=1}^{3} D_j(u_j u)\right] d\tau, \tag{11.24}$$

or equivalently,

$$u(t) = e^{\mu\Delta t}u_0 + B(u,u)(t), \tag{11.25}$$

where

$$B(w,v)(t) = -\int_0^t e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big] d\tau. \tag{11.26}$$

In order to apply Lemma 11.1.1, let X be the Banach space $C\big([0,T]; H^s_{a,\sigma}(\mathbb{R}^3)\big)$ ($T > 0$ will be revealed as follows). It is important to notice that (11.25) is the same equation as (11.8) if it is considered that $a = u$ and $x_0 = e^{\mu\Delta t}u_0$. Moreover, it is easy to check that B is a bilinear operator. Therefore, we shall prove that B is continuous by choosing T small enough.

At first, let us estimate $B(w,v)(t)$ in $\dot{H}^s_{a,\sigma}(\mathbb{R}^3)$. Thereby, it follows from the definition of the space $\dot{H}^s_{a,\sigma}(\mathbb{R}^3)$ that

$$\begin{aligned} &\Big\| e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} \\ &\quad = \int_{\mathbb{R}^3} |\xi|^{2s} e^{2a|\xi|^{\frac{1}{\sigma}}} \Big|\mathcal{F}\Big\{e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\}(\xi)\Big|^2 d\xi. \end{aligned}$$

It is also well known that

$$\mathcal{F}\{e^{\Delta t} f\}(\xi) = e^{-t|\xi|^2} \hat{f}(\xi), \quad \forall \xi \in \mathbb{R}^3, t \geq 0.$$

As a consequence, we have

$$\begin{aligned} &\Big\| e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} \\ &\quad = \int_{\mathbb{R}^3} e^{-2\mu(t-\tau)|\xi|^2}|\xi|^{2s} e^{2a|\xi|^{\frac{1}{\sigma}}} \Big|\mathcal{F}\Big\{P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\}(\xi)\Big|^2 d\xi. \end{aligned}$$

By applying (11.23), one can write[9]

$$\begin{aligned} \Big\| e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} &\leq \int_{\mathbb{R}^3} e^{-2\mu(t-\tau)|\xi|^2}|\xi|^{2s} \\ &\qquad \times e^{2a|\xi|^{\frac{1}{\sigma}}} \Big|\sum_{j=1}^{3} \mathcal{F}[D_j(v_j w)](\xi)\Big|^2 d\xi \\ &\leq \int_{\mathbb{R}^3} e^{-2\mu(t-\tau)|\xi|^2}|\xi|^{2s} \\ &\qquad \times e^{2a|\xi|^{\frac{1}{\sigma}}} |\mathcal{F}(w \otimes v)(\xi) \cdot \xi|^2 d\xi \\ &\leq \int_{\mathbb{R}^3} e^{-2\mu(t-\tau)|\xi|^2}|\xi|^{2s+2} \\ &\qquad \times e^{2a|\xi|^{\frac{1}{\sigma}}} |\mathcal{F}(w \otimes v)(\xi)|^2 d\xi. \end{aligned}$$

Rewriting the last integral above with the goal of applying Lemma 11.1.10, we have

$$\begin{aligned} &\Big\| e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} \\ &\quad \leq \int_{\mathbb{R}^3} |\xi|^{5-2s} e^{-2\mu(t-\tau)|\xi|^2}|\xi|^{4s-3} e^{2a|\xi|^{\frac{1}{\sigma}}} |\mathcal{F}(w \otimes v)(\xi)|^2 d\xi. \end{aligned}$$

[9]The tensor product is given by $f \otimes g := (g_1 f, g_2 f, g_3 f)$, where $f, g : \mathbb{R}^3 \to \mathbb{R}^3$.

As a result, by using Lemma 11.1.10, it follows

$$\Big\|e^{\mu\Delta(t-\tau)}P_H\Big[\sum_{j=1}^{3}D_j(v_jw)\Big]\Big\|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} \leq \frac{\left(\frac{5-2s}{4e\mu}\right)^{\frac{5-2s}{2}}}{(t-\tau)^{\frac{5-2s}{2}}}\int_{\mathbb{R}^3}|\xi|^{4s-3} \times e^{2a|\xi|^{\frac{1}{\sigma}}}|\mathcal{F}(w\otimes v)(\xi)|^2\,d\xi =: \frac{C_{s,\mu}}{(t-\tau)^{\frac{5-2s}{2}}}\|w\otimes v\|^2_{\dot{H}^{2s-\frac{3}{2}}_{a,\sigma}(\mathbb{R}^3)}, \tag{11.27}$$

where $C_{s,\mu}=\left(\frac{5-2s}{4e\mu}\right)^{\frac{5-2s}{2}}$ ($s<3/2$). On the other hand, let us estimate the term $\|w\otimes v\|_{\dot{H}^{2s-\frac{3}{2}}_{a,\sigma}(\mathbb{R}^3)}$ presented in the last equality above. Lemma 11.1.6 is the tool that provides a suitable result related to our goal in this proof. Thus, by using this lemma, one infers

$$\begin{aligned}\|w\otimes v\|^2_{\dot{H}^{2s-\frac{3}{2}}_{a,\sigma}(\mathbb{R}^3)} &= \int_{\mathbb{R}^3}|\xi|^{4s-3}e^{2a|\xi|^{\frac{1}{\sigma}}}|\widehat{w\otimes v}(\xi)|^2\,d\xi\\ &= \sum_{j,k=1}^{3}\int_{\mathbb{R}^3}|\xi|^{4s-3}e^{2a|\xi|^{\frac{1}{\sigma}}}|\widehat{v_jw_k}(\xi)|^2\,d\xi\\ &= \sum_{j,k=1}^{3}\|v_jw_k\|^2_{\dot{H}^{2s-\frac{3}{2}}_{a,\sigma}(\mathbb{R}^3)}\\ &\leq C_s\|w\|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)}\|v\|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)},\end{aligned} \tag{11.28}$$

provided that $0<s<3/2$. Therefore, by replacing (11.28) in (11.27), one deduces

$$\Big\|e^{\mu\Delta(t-\tau)}P_H\Big[\sum_{j=1}^{3}D_j(v_jw)\Big]\Big\|_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} \leq \frac{C_{s,\mu}}{(t-\tau)^{\frac{5-2s}{4}}}\|w\|_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)}\|v\|_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)}.$$

By integrating over $[0,t]$, the above estimate, we conclude[10]

$$\begin{aligned}\int_0^t\Big\|e^{\mu\Delta(t-\tau)}P_H\Big[\sum_{j=1}^{3}D_j(v_jw)\Big]\Big\|_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)}d\tau &\leq C_{s,\mu}\int_0^t\frac{\|w\|_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)}\|v\|_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)}}{(t-\tau)^{\frac{5-2s}{4}}}\,d\tau\\ &\leq C_{s,\mu}T^{\frac{2s-1}{4}}\|w\|_{L^\infty([0,T];\dot{H}^s_{a,\sigma}(\mathbb{R}^3))}\|v\|_{L^\infty([0,T];\dot{H}^s_{a,\sigma}(\mathbb{R}^3))},\end{aligned} \tag{11.29}$$

for all $t\in[0,T]$ (recall that $s>1/2$).

[10] Assuming that $(X,\|\cdot\|)$ is a Banach space and $T>0$, the space $L^\infty([0,T];X)$ contains all measurable functions $f:[0,T]\to X$ such that $\|f\|\in L^\infty([0,T])$. Here $\|f\|_{L^\infty([0,T];X)}:=\operatorname{esssup}_{t\in[0,T]}\{\|f(t)\|\}$.

By (11.26), we can assure that (11.29) implies

$$\|B(w,v)(t)\|_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} \leq C_{s,\mu} T^{\frac{2s-1}{4}} \|w\|_{L^\infty([0,T];\dot{H}^s_{a,\sigma}(\mathbb{R}^3))} \|v\|_{L^\infty([0,T];\dot{H}^s_{a,\sigma}(\mathbb{R}^3))}, \tag{11.30}$$

for all $t \in [0, T]$. It is important to observe that (11.30) presents our estimate to the operator B related to the space $\dot{H}^s_{a,\sigma}(\mathbb{R}^3)$.

Now, let us estimate $B(w,v)(t)$ in $H^s_{a,\sigma}(\mathbb{R}^3)$. By Lemma 11.1.8 and (11.30), it is enough to get a lower bound to $B(w,v)(t)$ in $L^2(\mathbb{R}^3)$. Following a similar process to the one presented above, we have

$$\Big\|e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\|^2_{L^2(\mathbb{R}^3)} = \int_{\mathbb{R}^3} \Big|e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big](\xi)\Big|^2\, d\xi.$$

Parseval's identity implies the following equality:

$$\Big\|e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\|^2_{L^2(\mathbb{R}^3)} = (2\pi)^{-3} \int_{\mathbb{R}^3} \Big|\mathcal{F}\Big\{e^{\mu\Delta(t-\tau)} P_H \times \Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\}(\xi)\Big|^2\, d\xi.$$

As a result, we obtain the next result:

$$\Big\|e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\|^2_{L^2(\mathbb{R}^3)} = (2\pi)^{-3} \int_{\mathbb{R}^3} e^{-2\mu(t-\tau)|\xi|^2} \Big|\mathcal{F}\Big\{P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\}(\xi)\Big|^2\, d\xi.$$

By using (11.23), it is true that

$$\Big\|e^{\mu\Delta(t-\tau)} P_H\Big[\sum_{j=1}^{3} D_j(v_j w)\Big]\Big\|^2_{L^2(\mathbb{R}^3)} \leq (2\pi)^{-3} \int_{\mathbb{R}^3} |\xi|^2 \times e^{-2\mu(t-\tau)|\xi|^2} \Big|\mathcal{F}(w \otimes v)(\xi)\Big|^2\, d\xi.$$

Rewriting the last integral in order to apply Lemma 11.1.10, one has

$$\Big\|e^{\mu\Delta(t-\tau)}P_H\Big[\sum_{j=1}^{3}D_j(v_jw)\Big]\Big\|^2_{L^2(\mathbb{R}^3)}$$
$$\leq (2\pi)^{-3}\int_{\mathbb{R}^3}|\xi|^{5-2s}e^{-2\mu(t-\tau)|\xi|^2}|\xi|^{2s-3}|\mathcal{F}(w\otimes v)(\xi)|^2\,d\xi.$$

As a result, by using Lemma 11.1.10, it follows

$$\Big\|e^{\mu\Delta(t-\tau)}P_H\Big[\sum_{j=1}^{3}D_j(v_jw)\Big]\Big\|^2_{L^2(\mathbb{R}^3)} \leq \frac{C_{s,\mu}}{(t-\tau)^{\frac{5-2s}{2}}}\|w\otimes v\|^2_{\dot{H}^{s-\frac{3}{2}}(\mathbb{R}^3)},$$

since $s < 3/2$. Now we are interested in estimating the term $\|w \otimes v\|_{\dot{H}^{s-\frac{3}{2}}(\mathbb{R}^3)}$ obtained above. Lemma 11.1.5 is the tool that lets us obtain this specific bound. More precisely, by utilizing this lemma, one has

$$\|w\otimes v\|^2_{\dot{H}^{s-\frac{3}{2}}(\mathbb{R}^3)} = \int_{\mathbb{R}^3}|\xi|^{2s-3}|\widehat{w\otimes v}(\xi)|^2\,d\xi = \sum_{j,k=1}^{3}\int_{\mathbb{R}^3}|\xi|^{2s-3}|\widehat{v_jw_k}(\xi)|^2\,d\xi$$
$$= \sum_{j,k=1}^{3}\|v_jw_k\|^2_{\dot{H}^{s-\frac{3}{2}}(\mathbb{R}^3)} \leq C_s\|w\|^2_{\dot{H}^s(\mathbb{R}^3)}\|v\|^2_{L^2(\mathbb{R}^3)}.$$

It is easy to check out that the continuous embedding $H^s_{a,\sigma}(\mathbb{R}^3) \hookrightarrow \dot{H}^s(\mathbb{R}^3)$ $(s \geq 0)$ holds and by applying Lemma 11.1.8, we deduce

$$\Big\|e^{\mu\Delta(t-\tau)}P_H\Big[\sum_{j=1}^{3}D_j(v_jw)\Big]\Big\|_{L^2(\mathbb{R}^3)} \leq \frac{C_{s,\mu}}{(t-\tau)^{\frac{5-2s}{4}}}\|w\|_{H^s_{a,\sigma}(\mathbb{R}^3)}\|v\|_{H^s_{a,\sigma}(\mathbb{R}^3)}.$$

By integrating over $[0, t]$, the above estimate, we conclude

$$\int_0^t \Big\|e^{\mu\Delta(t-\tau)}P_H\Big[\sum_{j=1}^{3}D_j(v_jw)\Big]\Big\|_{L^2(\mathbb{R}^3)}\,d\tau$$
$$\leq C_{s,\mu}T^{\frac{2s-1}{4}}\|w\|_{L^\infty([0,T];H^s_{a,\sigma}(\mathbb{R}^3))}\|v\|_{L^\infty([0,T];H^s_{a,\sigma}(\mathbb{R}^3))}, \tag{11.31}$$

for all $t \in [0, T]$ (since that $s > 1/2$).

By using definition (11.26) and applying (11.31), one concludes

$$\|B(w,v)(t)\|_{L^2(\mathbb{R}^3)} \le C_{s,\mu} T^{\frac{2s-1}{4}} \|w\|_{L^\infty([0,T];H^s_{a,\sigma}(\mathbb{R}^3))} \|v\|_{L^\infty([0,T];H^s_{a,\sigma}(\mathbb{R}^3))}, \tag{11.32}$$

for all $t \in [0,T]$.

Finally, by using Lemma 11.1.8, (11.30), (11.32) and the fact that $H^s_{a,\sigma}(\mathbb{R}^3) \hookrightarrow \dot{H}^s_{a,\sigma}(\mathbb{R}^3)$ ($s \ge 0$), it results

$$\|B(w,v)(t)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \le C_{s,a,\mu} T^{\frac{2s-1}{4}} \|w\|_{L^\infty([0,T];H^s_{a,\sigma}(\mathbb{R}^3))} \|v\|_{L^\infty([0,T];H^s_{a,\sigma}(\mathbb{R}^3))}, \tag{11.33}$$

for all $t \in [0,T]$.

To use Lemma 11.1.1, it is enough to guarantee that

$$4C_{s,a,\mu} T^{\frac{2s-1}{4}} \|e^{\mu\Delta t} u_0\|_{L^\infty([0,T];H^s_{a,\sigma}(\mathbb{R}^3))} < 1.$$

Thus, first of all, as we did before, one concludes

$$\begin{aligned}
\left\|e^{\mu\Delta t}u_0\right\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} &= \int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left|\mathcal{F}\left\{e^{\mu\Delta t}u_0\right\}(\xi)\right|^2 d\xi \\
&= \int_{\mathbb{R}^3} e^{-2\mu t|\xi|^2} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left|\widehat{u}_0(\xi)\right|^2 d\xi \\
&\le \int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} |\widehat{u}_0(\xi)|^2 d\xi \\
&= \|u_0\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}.
\end{aligned}$$

As a result, we write

$$\|e^{\mu\Delta t}u_0\|_{L^\infty([0,T];H^s_{a,\sigma}(\mathbb{R}^3))} \le \|u_0\|_{H^s_{a,\sigma}(\mathbb{R}^3)}.$$

At last, choose

$$T < \frac{1}{\left[4C_{s,a,\mu}\|u_0\|_{H^s_{a,\sigma}(\mathbb{R}^3)}\right]^{\frac{4}{2s-1}}},$$

where $C_{s,a,\mu}$ is given in (11.33), and apply Lemma 11.1.1 in order to obtain a unique solution $u \in C([0,T]; H^s_{a,\sigma}(\mathbb{R}^3))$ for Eq. (11.25).

The arguments given above also establish the local existence of a unique solution for the Navier–Stokes equations (11.1). □

11.2 Blow-Up Criteria for the Solution

In this section, we establish the veracity of the blow-up criteria for the solution of the Navier–Stokes equations (11.1), previously presented, by proving appropriate theorems. Let us inform that we argue similarly to references [1–4, 6, 7, 11].

11.2.1 Limit Superior Related to $H^s_{a,\sigma}(\mathbb{R}^3)$

The first blow-up criterion is related to the limit superior given in (11.5) (case $n = 1$).

Theorem 11.2.1 *Assume that* $a > 0$, $\sigma > 1$, *and* $s \in \left(\frac{1}{2}, \frac{3}{2}\right)$. *Let* $u_0 \in H^s_{a,\sigma}(\mathbb{R}^3)$ *such that* $\operatorname{div} u_0 = 0$. *Consider that* $u \in C([0, T^*); H^s_{a,\sigma}(\mathbb{R}^3))$ *is the maximal solution for the Navier–Stokes equations* (11.1) *obtained in Theorem* 11.1.11. *If* $T^* < \infty$, *then*

$$\limsup_{t \nearrow T^*} \|u(t)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = \infty. \tag{11.34}$$

Proof Consider by absurd that (11.34) is not valid, i.e., assume that

$$\limsup_{t \nearrow T^*} \|u(t)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} < \infty. \tag{11.35}$$

As a result, we shall prove that the solution $u(\cdot, t)$ can be extended beyond $t = T^*$ (it is the absurd that we shall obtain). Let us prove this statement.

Assuming (11.35) to be true, and using Theorem 11.1.11, there is a non-negative constant C such that

$$\|u(t)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \leq C, \quad \forall\, t \in [0, T^*). \tag{11.36}$$

Integrating over $[0, t]$ inequality (11.52), and applying (11.36) and (11.17), one concludes

$$\|u(t)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} + \mu \int_0^t \|\nabla u(\tau)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}\, d\tau \leq \|u_0\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} + C_{s,a,\sigma,\mu} C^4 T^*,$$

for all $t \in [0, T^*)$. As we are interested in using the fact that the integral above is bounded, we can write

$$\int_0^t \|\nabla u(\tau)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}\, d\tau \leq \frac{1}{\mu} \|u_0\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} + C_{s,a,\sigma,\mu} C^4 T^* =: C_{s,a,\sigma,\mu,u_0,T^*}, \tag{11.37}$$

for all $t \in [0, T^*)$.

Now, let $(\kappa_n)_{n\in\mathbb{N}}$ be a sequence such that $\kappa_n \nearrow T^*$, where $\kappa_n \in (0, T^*)$, for all $n \in \mathbb{N}$ (choose $\kappa_n = T^* - 1/n$, for n large enough, for instance). We guarantee that $(u(\kappa_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in the space $H^s_{a,\sigma}(\mathbb{R}^3)$, that is,

$$\lim_{n,m\to\infty} \|u(\kappa_n) - u(\kappa_m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0. \tag{11.38}$$

Let us inform that the limit (11.38) does not depend on the sequence $(\kappa_n)_{n\in\mathbb{N}}$. This fact will be shown later. First of all, we begin with the demonstration of (11.38). Thereby, one can apply (11.25) and (11.26) in order to obtain

$$u(\kappa_n) - u(\kappa_m) = I_1(n, m) + I_2(n, m) + I_3(n, m), \tag{11.39}$$

where

$$I_1(n, m) = \big[e^{\mu\Delta\kappa_n} - e^{\mu\Delta\kappa_m}\big]u_0, \tag{11.40}$$

$$I_2(n, m) = \int_0^{\kappa_m} \big[e^{\mu\Delta(\kappa_m-\tau)} - e^{\mu\Delta(\kappa_n-\tau)}\big] P_H[u \cdot \nabla u]\, d\tau, \tag{11.41}$$

and also

$$I_3(n, m) = -\int_{\kappa_m}^{\kappa_n} e^{\mu\Delta(\kappa_n-\tau)} P_H[u \cdot \nabla u]\, d\tau. \tag{11.42}$$

Let us prove that $I_j(n, m) \to 0$ in $H^s_{a,\sigma}(\mathbb{R}^3)$, as $n, m \to \infty$, for $j = 1, 2, 3$.

In order to prove the veracity of the limit related to $I_1(n, m)$, notice that

$$\begin{aligned}\|I_1(n, m)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} &= \big\|\big[e^{\mu\Delta\kappa_n} - e^{\mu\Delta\kappa_m}\big]u_0\big\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}\\ &= \int_{\mathbb{R}^3} \big[e^{-\mu\kappa_n|\xi|^2} - e^{-\mu\kappa_m|\xi|^2}\big]^2 \big(1 + |\xi|^2\big)^s e^{2a|\xi|^{\frac{1}{\sigma}}} |\widehat{u}_0(\xi)|^2\, d\xi\\ &\le \int_{\mathbb{R}^3} \big[e^{-\mu\kappa_n|\xi|^2} - e^{-\mu T^*|\xi|^2}\big]^2 \big(1 + |\xi|^2\big)^s e^{2a|\xi|^{\frac{1}{\sigma}}} |\widehat{u}_0(\xi)|^2\, d\xi.\end{aligned}$$

By utilizing the fact that $u_0 \in H^s_{a,\sigma}(\mathbb{R}^3)$ and that $e^{-\mu\kappa_n|\xi|^2} - e^{-\mu T^*|\xi|^2} \le 1$, for all $n \in \mathbb{N}$, it results from dominated convergence theorem that

$$\lim_{n,m\to\infty} \|I_1(n, m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0.$$

Now, our imminent goal is to establish the limit $\lim_{n,m\to\infty} \|I_2(n,m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0$. Thus, we have

$$\begin{aligned}\|I_2(n,m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} &\le \int_0^{\kappa_m} \left\| \left[e^{\mu\Delta(\kappa_m-\tau)} - e^{\mu\Delta(\kappa_n-\tau)} \right] P_H(u\cdot\nabla u) \right\|_{H^s_{a,\sigma}(\mathbb{R}^3)} d\tau \\ &= \int_0^{\kappa_m} \Big(\int_{\mathbb{R}^3} \left[e^{-\mu(\kappa_m-\tau)|\xi|^2} - e^{-\mu(\kappa_n-\tau)|\xi|^2} \right]^2 \\ &\quad \times \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left|\mathcal{F}\left[P_H(u\cdot\nabla u)\right](\xi)\right|^2 d\xi \Big)^{\frac{1}{2}} d\tau.\end{aligned}$$

By (11.23), we can write $|\mathcal{F}[P_H(f)](\xi)| \le \left|\hat{f}(\xi)\right|$ and, consequently,

$$\begin{aligned}&\|I_2(n,m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \\ &\le \int_0^{T^*} \left(\int_{\mathbb{R}^3} \left[1 - e^{-\mu(T^*-\kappa_m)|\xi|^2}\right]^2 \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left|\mathcal{F}[u\cdot\nabla u](\xi)\right|^2 d\xi \right)^{\frac{1}{2}} d\tau.\end{aligned}$$

Use Cauchy–Schwarz's inequality in order to obtain

$$\begin{aligned}&\|I_2(n,m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \\ &\le \sqrt{T^*} \left(\int_0^{T^*} \int_{\mathbb{R}^3} \left[1 - e^{-\mu(T^*-\kappa_m)|\xi|^2}\right]^2 \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left|\mathcal{F}[u\cdot\nabla u](\xi)\right|^2 d\xi\, d\tau \right)^{\frac{1}{2}}.\end{aligned}$$

On the other hand, observe that by Lemma 11.1.7 **ii)**, (11.36) and (11.37), one infers

$$\begin{aligned}\int_0^{T^*} \|u\cdot\nabla u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} d\tau &\le C^2_{s,a,\sigma} \int_0^{T^*} \|u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \|\nabla u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} d\tau \\ &\le C^2_{s,a,\sigma} C^2 \int_0^{T^*} \|\nabla u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} d\tau < \infty. \qquad (11.43)\end{aligned}$$

As $1 - e^{-\mu(T^*-\kappa_m)|\xi|^2} \le 1$, for all $m \in \mathbb{N}$; then, by dominated convergence theorem, we deduce

$$\lim_{n,m\to\infty} \|I_2(n,m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0.$$

Finally, we are interested in demonstrating that $\lim_{n,m\to\infty} \|I_3(n,m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0$. Hence, one obtains

$$\begin{aligned}\|I_3(n,m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} &\le \int_{\kappa_m}^{\kappa_n} \left\| e^{\mu\Delta(\kappa_n-\tau)} P_H(u\cdot\nabla u) \right\|_{H^s_{a,\sigma}(\mathbb{R}^3)} d\tau \\ &= \int_{\kappa_m}^{\kappa_n} \left(\int_{\mathbb{R}^3} e^{-2\mu(\kappa_n-\tau)|\xi|^2} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left|\mathcal{F}[P_H(u\cdot\nabla u)](\xi)\right|^2 d\xi \right)^{\frac{1}{2}} d\tau.\end{aligned}$$

By (11.23), we can write $|\mathcal{F}[P_H(f)](\xi)| \leq \left|\hat{f}(\xi)\right|$ and, consequently,

$$\|I_3(n,m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \leq \int_{\kappa_m}^{\kappa_n} \left(\int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}} \left|\mathcal{F}[u\cdot\nabla u](\xi)\right|^2 d\xi \right)^{\frac{1}{2}} d\tau$$
$$\leq \int_{\kappa_m}^{T^*} \|u\cdot\nabla u\|_{H^s_{a,\sigma}(\mathbb{R}^3)}\, d\tau.$$

By Cauchy–Schwarz's inequality, (11.43) and (11.37), one infers

$$\|I_3(n,m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \leq \sqrt{T^*-\kappa_m} \left(\int_{\kappa_m}^{T^*} \|u\cdot\nabla u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}\, d\tau \right)^{\frac{1}{2}}$$
$$\leq CC_{s,a,\sigma}\sqrt{T^*-\kappa_m} \left(\int_{\kappa_m}^{T^*} \|\nabla u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}\, d\tau \right)^{\frac{1}{2}}$$
$$\leq C_{s,a,\sigma,\mu,u_0,T^*}\sqrt{T^*-\kappa_m}.$$

As a result, we infer that $\lim_{n,m\to\infty} \|I_3(n,m)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0$. Thus, (11.39) implies (11.38). In addition, (11.38) means that $(u(\kappa_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in the Banach space $H^s_{a,\sigma}(\mathbb{R}^3)$. Hence, there exists $u_1 \in H^s_{a,\sigma}(\mathbb{R}^3)$ such that

$$\lim_{n\to\infty} \|u(\kappa_n) - u_1\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0.$$

Now, we shall prove that the above limit does not depend on the sequence $(\kappa_n)_{n\in\mathbb{N}}$. Thus, choose an arbitrary sequence $(\rho_n)_{n\in\mathbb{N}} \subseteq (0,T^*)$ such that $\rho_n \nearrow T^*$ and

$$\lim_{n\to\infty} \|u(\rho_n) - u_2\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0,$$

for some $u_2 \in H^s_{a,\sigma}(\mathbb{R}^3)$. Let us verify that $u_2 = u_1$. In fact, define $(\varsigma_n)_{n\in\mathbb{N}} \subseteq (0,T^*)$ by $\varsigma_{2n} = \kappa_n$ and $\varsigma_{2n-1} = \rho_n$, for all $n \in \mathbb{N}$. It is easy to check that $\varsigma_n \nearrow T^*$. By rewriting the process above, we guarantee that there is $u_3 \in H^s_{a,\sigma}(\mathbb{R}^3)$ such that

$$\lim_{n\to\infty} \|u(\varsigma_n) - u_3\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0.$$

As a consequence, one has

$$\lim_{n\to\infty} \|u(\kappa_n) - u_3\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = \lim_{n\to\infty} \|u(\varsigma_{2n}) - u_3\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0$$

and also

$$\lim_{n\to\infty} \|u(\rho_n) - u_3\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = \lim_{n\to\infty} \|u(\varsigma_{2n-1}) - u_3\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0.$$

By uniqueness of the limit, one infers $u_1 = u_3 = u_2$. Therefore, the limit (11.38) does not rely on the sequence $(\kappa_n)_{n\in\mathbb{N}}$.

It means that $\lim_{t\nearrow T^*} \|u(t) - u_1\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = 0$. Thereby, assuming (11.1) with the initial data u_1, instead of u_0, we assure, by Theorem 11.1.11, the local existence and uniqueness of $\bar{u} \in C([0, \bar{T}]; H^s_{a,\sigma}(\mathbb{R}^3))$ ($\bar{T} > 0$) for system (11.1). Hence, $\tilde{u} \in C([0, \bar{T} + T^*]; H^s_{a,\sigma}(\mathbb{R}^3))$ defined by

$$\tilde{u}(t) = \begin{cases} u(t), & t \in [0, T^*); \\ \bar{u}(t - T^*), & t \in [T^*, \bar{T} + T^*] \end{cases}$$

solves (11.1) in $[0, \bar{T} + T^*]$. Thus, the solution of (11.1) can be extended beyond $t = T^*$. It is a contradiction. Consequently, one must have

$$\limsup_{t\nearrow T^*} \|u(t)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} = \infty.$$

□

11.2.2 Blow-Up of the Integral Related to $L^1(\mathbb{R}^3)$

Now, we present the proof of inequality (11.3) in the case $n = 1$. It is important to let the reader know that the next theorem might be written as a corollary of Theorem 11.2.1 since the first one follows from this last result.

Theorem 11.2.2 *Assume that $a > 0$, $\sigma > 1$, and $s \in \left(\frac{1}{2}, \frac{3}{2}\right)$. Let $u_0 \in H^s_{a,\sigma}(\mathbb{R}^3)$ such that* $\operatorname{div} u_0 = 0$. *Consider that $u \in C([0, T^*); H^s_{a,\sigma}(\mathbb{R}^3))$ is the maximal solution for the Navier–Stokes equations* (11.1) *obtained in Theorem* 11.1.11. *If $T^* < \infty$, then*

$$\int_t^{T^*} \left\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{u}(\tau) \right\|^2_{L^1(\mathbb{R}^3)} d\tau = \infty.$$

Proof This result follows from the limit superior presented in Theorem 11.2.1. Thus, let us start taking the $H^s_{a,\sigma}(\mathbb{R}^3)$-inner product, with $u(t)$, in the first equation of (11.1) to get

$$\langle u, u_t \rangle_{H^s_{a,\sigma}(\mathbb{R}^3)} = \langle u, -u \cdot \nabla u - \nabla p + \mu \Delta u \rangle_{H^s_{a,\sigma}(\mathbb{R}^3)}. \tag{11.44}$$

In order to study some terms on the right-hand side of the equality above, use the fact that

$$\mathcal{F}(D_j f)(\xi) = i\xi_j \hat{f}(\xi), \quad \forall\, \xi = (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3,$$

to get

$$\mathcal{F}(u)\cdot\mathcal{F}[\nabla p](\xi) = -i\sum_{j=1}^{3}\mathcal{F}(u_j)(\xi)\xi_j\overline{\hat{p}(\xi)} = -\sum_{j=1}^{3}\mathcal{F}(D_ju_j)(\xi)\overline{\hat{p}(\xi)}$$
$$= -\mathcal{F}(\operatorname{div} u)(\xi)\overline{\hat{p}(\xi)} = 0, \tag{11.45}$$

because u is divergence free (see (11.1)). Thereby, the term related to the pressure in (11.44) is null. In fact, we have

$$\langle u, \nabla p\rangle_{H^s_{a,\sigma}(\mathbb{R}^3)} = \int_{\mathbb{R}^3}\left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}}\mathcal{F}(u)\cdot\mathcal{F}[\nabla p](\xi)\,d\xi = 0. \tag{11.46}$$

On the other hand, following a similar argument, one infers

$$\widehat{u}\cdot\widehat{\Delta u}(\xi) = \sum_{j=1}^{3}\widehat{u}\cdot\widehat{D_j^2 u}(\xi) = -i\sum_{j=1}^{3}\widehat{u}\cdot\left[\xi_j\widehat{D_j u}(\xi)\right]$$
$$= -\sum_{j=1}^{3}\widehat{D_j u}\cdot\widehat{D_j u}(\xi) = -\left|\widehat{\nabla u}(\xi)\right|^2. \tag{11.47}$$

Therefore, the term related to Δu in (11.44) satisfies

$$\langle u, \Delta u\rangle_{H^s_{a,\sigma}(\mathbb{R}^3)} = \int_{\mathbb{R}^3}\left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}}\widehat{u}\cdot\widehat{\Delta u}(\xi)\,d\xi$$
$$= -\int_{\mathbb{R}^3}\left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}}\left|\widehat{\nabla u}(\xi)\right|^2 d\xi$$
$$= -\|\nabla u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}. \tag{11.48}$$

By replacing (11.46) and (11.48) in (11.44), we have

$$\frac{1}{2}\frac{d}{dt}\|u(t)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} + \mu\|\nabla u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \leq |\langle u, u\cdot\nabla u\rangle_{H^s_{a,\sigma}(\mathbb{R}^3)}|. \tag{11.49}$$

Now, let us study the inner product above. Thus, as $\operatorname{div} u = 0$, one obtains

$$\mathcal{F}(\nabla u)\cdot\mathcal{F}(u\otimes u)(\xi) = \sum_{j=1}^{3}\mathcal{F}(\nabla u_j)\cdot\mathcal{F}(u_ju)(\xi)$$
$$= \sum_{j,k=1}^{3}\mathcal{F}(D_ku_j)(\xi)\overline{\mathcal{F}(u_ju_k)(\xi)}$$

$$
\begin{aligned}
&= i \sum_{j,k=1}^{3} \xi_k \mathcal{F}(u_j)(\xi)\overline{\mathcal{F}(u_j u_k)(\xi)} \\
&= -\sum_{j,k=1}^{3} \mathcal{F}(u_j)(\xi)\overline{\mathcal{F}(D_k(u_j u_k))(\xi)} \\
&= -\sum_{j,k=1}^{3} \mathcal{F}(u_j)(\xi)\overline{\mathcal{F}(u_k D_k u_j)(\xi)} \\
&= -\sum_{j=1}^{3} \mathcal{F}(u_j)(\xi)\overline{\mathcal{F}(u\cdot\nabla u_j)(\xi)} \\
&= -\mathcal{F}(u)\cdot\mathcal{F}(u\cdot\nabla u)(\xi).
\end{aligned}
$$

As a result, by using the tensor product, it follows that

$$
\begin{aligned}
\langle u, u\cdot\nabla u\rangle_{H^s_{a,\sigma}(\mathbb{R}^3)} &= \int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}}\mathcal{F}(u)\cdot\mathcal{F}(u\cdot\nabla u)(\xi)\,d\xi \\
&= -\int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}}\mathcal{F}(\nabla u)\cdot\mathcal{F}(u\otimes u)(\xi)\,d\xi \\
&= -\langle \nabla u, u\otimes u\rangle_{H^s_{a,\sigma}(\mathbb{R}^3)}.
\end{aligned} \tag{11.50}
$$

Hence, applying Cauchy–Schwarz's inequality,[11] (11.49) and (11.50) imply

$$
\frac{1}{2}\frac{d}{dt}\|u(t)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} + \mu\|\nabla u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \le \|\nabla u\|_{H^s_{a,\sigma}(\mathbb{R}^3)}\|u\otimes u\|_{H^s_{a,\sigma}(\mathbb{R}^3)}. \tag{11.51}
$$

Now, our interest is to find an estimate for the term $\|u\otimes u\|_{H^s_{a,\sigma}(\mathbb{R}^3)}$ obtained above. Thus, by applying Lemma 11.1.7 **i)**, one has

$$
\begin{aligned}
\|u\otimes u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} &= \int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}}\left|\mathcal{F}(u\otimes u)(\xi)\right|^2 d\xi \\
&= \sum_{j,k=1}^{3}\int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{2a|\xi|^{\frac{1}{\sigma}}}\left|\mathcal{F}(u_j u_k)(\xi)\right|^2 d\xi
\end{aligned}
$$

[11] $|\langle f, g\rangle_{H^s_{a,\sigma}(\mathbb{R}^3)}| \le \|f\|_{H^s_{a,\sigma}(\mathbb{R}^3)}\|g\|_{H^s_{a,\sigma}(\mathbb{R}^3)}$, for all $f, g \in H^s_{a,\sigma}(\mathbb{R}^3)$.

$$
\begin{aligned}
&= \sum_{j,k=1}^{3} \|u_j u_k\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \\
&\le C_s \sum_{j,k=1}^{3} \Big[\Big\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{u_j} \Big\|_{L^1(\mathbb{R}^3)} \|u_k\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \\
&\quad + \Big\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{u_k} \Big\|_{L^1(\mathbb{R}^3)} \|u_j\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \Big]^2 \\
&\le C_s \sum_{j,k=1}^{3} \Big[\Big\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{u_j} \Big\|^2_{L^1(\mathbb{R}^3)} \|u_k\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \\
&\quad + \Big\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{u_k} \Big\|^2_{L^1(\mathbb{R}^3)} \|u_j\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \Big] \\
&\le C_s \| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{u} \|^2_{L^1(\mathbb{R}^3)} \|u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)},
\end{aligned}
$$

or equivalently,

$$
\|u \otimes u\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \le C_s \Big\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{u} \Big\|_{L^1(\mathbb{R}^3)} \|u\|_{H^s_{a,\sigma}(\mathbb{R}^3)}.
$$

By replacing this inequality in (11.51), we deduce

$$
\frac{1}{2}\frac{d}{dt}\|u(t)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} + \mu\|\nabla u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \le C_s \Big\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{u} \Big\|_{L^1(\mathbb{R}^3)} \|u\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \|\nabla u\|_{H^s_{a,\sigma}(\mathbb{R}^3)}.
$$

By Young's inequality,[12] it results that

$$
\frac{1}{2}\frac{d}{dt}\|u(t)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} + \frac{\mu}{2}\|\nabla u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \le C_{s,\mu} \| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{u} \|^2_{L^1(\mathbb{R}^3)} \|u\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}. \tag{11.52}
$$

Consider $0 \le t \le T < T^*$ in order to obtain, by Gronwall's inequality,[13] the following estimate:

$$
\|u(T)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \le \|u(t)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)} \exp\left\{ C_{s,\mu} \int_t^T \Big\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}} \widehat{u}(\tau) \Big\|^2_{L^1(\mathbb{R}^3)} d\tau \right\}.
$$

[12]Let p and q be positive real numbers such that $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. Then, $ab \le \frac{a^p}{p} + \frac{b^q}{q}$, for all $a, b > 0$.

[13]Let $f, g : [t, T] \to \mathbb{R}$ be differential functions in (t, T) such that $f'(s) \le g(s)f(s)$, for all $s \in [t, T]$. Then, $f(s) \le f(t)\exp\{\int_t^s g(\tau)\, d\tau\}$, for all $s \in [t, T]$.

Passing to the limit superior, as $T \nearrow T^*$, Theorem 11.2.1 implies

$$\int_t^{T^*} \|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(\tau)\|^2_{L^1(\mathbb{R}^3)}\, d\tau = \infty, \qquad \forall\, t \in [0, T^*).$$

The proof of Theorem 11.2.2 is given. □

11.2.3 Blow-Up Inequality Involving $L^1(\mathbb{R}^3)$

Below, it is presented the proof of blow-up inequality (11.4) in the case $n = 1$. Let us inform that the below theorem could be enunciated as a corollary of Theorem 11.2.2.

Theorem 11.2.3 *Assume that $a > 0$, $\sigma > 1$, and $s \in (\frac{1}{2}, \frac{3}{2})$. Let $u_0 \in H^s_{a,\sigma}(\mathbb{R}^3)$ such that $\operatorname{div} u_0 = 0$. Consider that $u \in C([0, T^*); H^s_{a,\sigma}(\mathbb{R}^3))$ is the maximal solution for the Navier–Stokes equations (11.1) obtained in Theorem 11.1.11. If $T^* < \infty$, then*

$$\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(t)\|_{L^1(\mathbb{R}^3)} \geq \frac{8\pi^3\sqrt{\mu}}{\sqrt{T^* - t}}, \qquad \forall\, t \in [0, T^*).$$

Proof Allow us to inform that this result is a consequence of Theorem 11.2.2. Thereby, apply the Fourier Transform and take the scalar product in $\mathbb{C}^3$ of the first equation of (11.1), with $\widehat{u}(t)$, in order to obtain

$$\widehat{u} \cdot \widehat{u}_t = -\mu\big|\widehat{\nabla u}\big|^2 - \widehat{u} \cdot \widehat{u \cdot \nabla u},$$

see (11.45) and (11.47). Consequently, one infers

$$\frac{1}{2}\partial_t|\widehat{u}(t)|^2 + \mu\big|\widehat{\nabla u}\big|^2 \leq \big|\widehat{u} \cdot \widehat{u \cdot \nabla u}\big|. \tag{11.53}$$

For $\delta > 0$ arbitrary, by applying Cauchy–Schwarz's inequality,[14] it is easy to check that

$$\partial_t\sqrt{|\widehat{u}(t)|^2 + \delta} + \mu\frac{\big|\widehat{\nabla u}\big|^2}{\sqrt{\big|\widehat{u}\big|^2 + \delta}} \leq \frac{|\hat{u}|}{\sqrt{|\widehat{u}|^2 + \delta}}\big|\widehat{u \cdot \nabla u}\big| \leq \big|\widehat{u \cdot \nabla u}\big|.$$

By integrating from t to T, with $0 \leq t \leq T < T^*$, one has

$$\sqrt{|\widehat{u}(T)|^2 + \delta} + \mu|\xi|^2\int_t^T \frac{|\widehat{u}(\tau)|^2}{\sqrt{|\widehat{u}(\tau)|^2 + \delta}}\, d\tau \leq \sqrt{|\widehat{u}(t)|^2 + \delta} + \int_t^T \big|(\widehat{u \cdot \nabla u})(\tau)\big|\, d\tau,$$

[14] $|v \cdot w| \leq |v||w|$.

since $\left|\widehat{\nabla u}\right| = |\xi||\widehat{u}|$. Passing to the limit, as $\delta \to 0$, multiplying by $e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}$, and integrating over $\xi \in \mathbb{R}^3$, we obtain

$$\begin{aligned}\left\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(T)\right\|_{L^1(\mathbb{R}^3)} + \mu\int_t^T \left\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{\Delta u}(\tau)\right\|_{L^1(\mathbb{R}^3)} d\tau \leq \left\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(t)\right\|_{L^1(\mathbb{R}^3)} \\ + \int_t^T\int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}\left|\widehat{(u\cdot\nabla u)}(\tau)\right| d\xi d\tau,\end{aligned} \tag{11.54}$$

because $\left|\widehat{\Delta u}\right| = |\xi|^2|\widehat{u}|$. Studying the last module above, we can assure

$$\begin{aligned}\left|\widehat{(u\cdot\nabla u)}(\xi)\right| &= \left|\sum_{j=1}^3 \widehat{u_j D_j u}(\xi)\right| = (2\pi)^{-3}\left|\sum_{j=1}^3 \widehat{u_j} * \widehat{D_j u}(\xi)\right| \\ &= (2\pi)^{-3}\left|\sum_{j=1}^3\int_{\mathbb{R}^3}\widehat{u_j}(\eta)\widehat{D_j u}(\xi-\eta)\,d\eta\right| \\ &\leq (2\pi)^{-3}\left|\int_{\mathbb{R}^3}\widehat{u}(\eta)\cdot\widehat{\nabla u}(\xi-\eta)\,d\eta\right| \\ &\leq (2\pi)^{-3}\int_{\mathbb{R}^3}|\widehat{u}(\eta)|\left|\widehat{\nabla u}(\xi-\eta)\right| d\eta.\end{aligned}$$

Therefore, by (11.10), the last integral in (11.54) can be estimated as follows:

$$\begin{aligned}\int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}\left|\widehat{(u\cdot\nabla u)}(\xi)\right| d\xi &\leq (2\pi)^{-3}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}|\widehat{u}(\eta)|\left|\widehat{\nabla u}(\xi-\eta)\right| d\eta d\xi \\ &\leq (2\pi)^{-3}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\eta|^{\frac{1}{\sigma}}}|\widehat{u}(\eta)| \\ &\quad\times e^{\frac{a}{\sigma}|\xi-\eta|^{\frac{1}{\sigma}}}\left|\widehat{\nabla u}(\xi-\eta)\right| d\eta d\xi \\ &= (2\pi)^{-3}\int_{\mathbb{R}^3}\left[e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}|\widehat{u}(\xi)|\right] * \left[e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}\left|\widehat{\nabla u}(\xi)\right|\right] d\xi \\ &= (2\pi)^{-3}\left\|\left[e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}|\widehat{u}|\right] * \left[e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\left|\widehat{\nabla u}\right|\right]\right\|_{L^1(\mathbb{R}^3)}.\end{aligned}$$

Apply Young's inequality for convolutions in order to obtain the following inequality:

$$\int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}\left|\widehat{(u\cdot\nabla u)}(\xi)\right| d\xi \leq (2\pi)^{-3}\left\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}\right\|_{L^1(\mathbb{R}^3)}\left\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{\nabla u}\right\|_{L^1(\mathbb{R}^3)}. \tag{11.55}$$

Let us obtain an estimate for the term $\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{\nabla u}\|_{L^1(\mathbb{R}^3)}$ above. Thus, Cauchy–Schwarz's inequality implies

$$\begin{aligned}\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{\nabla u}\|_{L^1(\mathbb{R}^3)} &= \int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}\left|\widehat{\nabla u}(\xi)\right| d\xi = \int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}|\xi||\widehat{u}(\xi)|\,d\xi \\ &\le \left(\int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}|\xi|^2|\widehat{u}(\xi)|\,d\xi\right)^{\frac{1}{2}}\left(\int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}|\widehat{u}(\xi)|\,d\xi\right)^{\frac{1}{2}} \\ &= \|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{\Delta u}\|_{L^1(\mathbb{R}^3)}^{\frac{1}{2}}\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}\|_{L^1(\mathbb{R}^3)}^{\frac{1}{2}},\end{aligned} \tag{11.56}$$

since $|\xi|^2|\widehat{u}| = \left|\widehat{\Delta u}\right|$ and $\left|\widehat{\nabla u}\right| = |\xi||\widehat{u}|$. Thereby, by replacing (11.56) in (11.55), one deduces

$$\int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}\left|\widehat{(u\cdot\nabla u)}(\xi)\right| d\xi \le (2\pi)^{-3}\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}\|_{L^1(\mathbb{R}^3)}^{\frac{3}{2}}\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{\Delta u}\|_{L^1(\mathbb{R}^3)}^{\frac{1}{2}}.$$

By using Cauchy–Schwarz's inequality once again, we conclude

$$\begin{aligned}(2\pi)^{-3}\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}\|_{L^1(\mathbb{R}^3)}^{\frac{3}{2}}\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{\Delta u}\|_{L^1(\mathbb{R}^3)}^{\frac{1}{2}} &\le \frac{1}{128\pi^6\mu}\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}\|_{L^1(\mathbb{R}^3)}^{3} \\ &+ \frac{\mu}{2}\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{\Delta u}\|_{L^1(\mathbb{R}^3)}.\end{aligned}$$

Consequently, (11.54) can be rewritten as follows:

$$\begin{aligned}\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(T)\|_{L^1(\mathbb{R}^3)} + \frac{\mu}{2}\int_t^T \|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{\Delta u}(\tau)\|_{L^1(\mathbb{R}^3)}\,d\tau &\le \|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(t)\|_{L^1(\mathbb{R}^3)} \\ &+ \frac{1}{128\pi^6\mu}\int_t^T \|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(\tau)\|_{L^1(\mathbb{R}^3)}^3 d\tau.\end{aligned}$$

By Gronwall's inequality,[15] one gets

$$\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(T)\|_{L^1(\mathbb{R}^3)}^2 \le \|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(t)\|_{L^1(\mathbb{R}^3)}^2 \exp\left\{\frac{1}{64\pi^6\mu}\int_t^T \|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(\tau)\|_{L^1(\mathbb{R}^3)}^2 d\tau\right\},$$

for all $0 \le t \le T < T^*$, or equivalently,

$$\left(-64\pi^6\mu\right)\frac{d}{dT}\left[\exp\left\{-\frac{1}{64\pi^6\mu}\int_t^T \|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(\tau)\|_{L^1(\mathbb{R}^3)}^2 d\tau\right\}\right] \le \|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(t)\|_{L^1(\mathbb{R}^3)}^2.$$

[15] Let $f, g : [a, b] \to \mathbb{R}$ be continuous functions in $[a, b]$ such that $f(s) \le f(a) + \int_a^s g(\tau)f(\tau)\,d\tau$, for all $s \in [a, b]$. Then, $f(s) \le f(a)\exp\left\{\int_a^s g(\tau)\,d\tau\right\}$, for all $s \in [a, b]$.

Integrate from t to t_0, with $0 \leq t \leq t_0 < T^*$, in order to get

$$\left(-64\pi^6\mu\right)\exp\left\{-\frac{1}{64\pi^6\mu}\int_t^{t_0}\left\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(\tau)\right\|^2_{L^1(\mathbb{R}^3)}d\tau\right\}+64\pi^6\mu$$

$$\leq \left\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(t)\right\|^2_{L^1(\mathbb{R}^3)}(t_0-t).$$

By passing to the limit, as $t_0 \nearrow T^*$, and using Theorem 11.2.2, we have

$$64\pi^6\mu \leq \left\|e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(t)\right\|^2_{L^1(\mathbb{R}^3)}(T^*-t), \quad \forall\, t \in [0, T^*).$$

It proves Theorem 11.2.3. □

11.2.4 Blow-Up Inequality Involving $H^s_{a,\sigma}(\mathbb{R}^3)$

The lower bound (11.7), in the case $n = 1$, can be rewritten as below. Let us clarify that from now on $T^*_\omega < \infty$ denotes the first blow-up time for the solution $u \in C([0, T^*_\omega); H^s_{\omega,\sigma}(\mathbb{R}^3))$, where $\omega > 0$.

Theorem 11.2.4 *Assume that* $a > 0$, $\sigma > 1$, *and* $s \in \left(\frac{1}{2}, \frac{3}{2}\right)$. *Let* $u_0 \in H^s_{a,\sigma}(\mathbb{R}^3)$ *such that* $\operatorname{div} u_0 = 0$. *Consider that* $u \in C([0, T^*_a); H^s_{a,\sigma}\left(\mathbb{R}^3\right))$ *is the maximal solution for the Navier–Stokes equations* (11.1) *obtained in Theorem* 11.1.11. *If* $T^*_a < \infty$, *then*

$$\|u(t)\|_{H^s_{\frac{a}{\sqrt{\sigma}},\sigma}(\mathbb{R}^3)} \geq \frac{8\pi^3\sqrt{\mu}}{C_1\sqrt{T^*_a - t}}, \quad \forall\, t \in [0, T^*_a),$$

where $C_1 := \left\{4\pi\sigma\left[2a\left(\frac{1}{\sqrt{\sigma}}-\frac{1}{\sigma}\right)\right]^{-\sigma(3-2s)}\Gamma(\sigma(3-2s))\right\}^{\frac{1}{2}}$.

Proof This theorem is a direct implication of Theorem 11.2.3. First of all, notice that $\frac{a}{\sqrt{\sigma}} \in (0, a)$. As a result, it holds the following continuous embedding $H^s_{a,\sigma}\left(\mathbb{R}^3\right) \hookrightarrow H^s_{\frac{a}{\sqrt{\sigma}},\sigma}\left(\mathbb{R}^3\right)$ that comes from the inequality:

$$\|u\|_{H^s_{\frac{a}{\sqrt{\sigma}},\sigma}(\mathbb{R}^3)} \leq \|u\|_{H^s_{a,\sigma}(\mathbb{R}^3)}.$$

Thereby, we can guarantee, by Theorem 11.1.11 and the inequality above, that $u \in C([0, T^*_a), H^s_{\frac{a}{\sqrt{\sigma}},\sigma}\left(\mathbb{R}^3\right))$ (since $u \in C([0, T^*_a), H^s_{a,\sigma}\left(\mathbb{R}^3\right)))$ and also that

$$T^*_{\frac{a}{\sqrt{\sigma}}} \geq T^*_a. \tag{11.57}$$

Moreover, by applying Theorem 11.2.3 and Cauchy–Schwarz's inequality, it follows that

$$\begin{aligned}
\frac{8\pi^3\sqrt{\mu}}{\sqrt{T_a^*-t}} &\le \left\| e^{\frac{a}{\sigma}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(t)\right\|_{L^1(\mathbb{R}^3)} = \int_{\mathbb{R}^3} e^{\frac{a}{\sigma}|\xi|^{\frac{1}{\sigma}}}|\widehat{u}(\xi)|\,d\xi \\
&\le \left(\int_{\mathbb{R}^3}\left(1+|\xi|^2\right)^{-s} e^{2\left(\frac{a}{\sigma}-\frac{a}{\sqrt{\sigma}}\right)|\xi|^{\frac{1}{\sigma}}}\,d\xi\right)^{\frac{1}{2}} \\
&\quad \times \left(\int_{\mathbb{R}^3}\left(1+|\xi|^2\right)^{s} e^{2\frac{a}{\sqrt{\sigma}}|\xi|^{\frac{1}{\sigma}}}|\widehat{u}(\xi)|^2\,d\xi\right)^{\frac{1}{2}} \\
&\le \left(\int_{\mathbb{R}^3}|\xi|^{-2s} e^{2\left(\frac{a}{\sigma}-\frac{a}{\sqrt{\sigma}}\right)|\xi|^{\frac{1}{\sigma}}}\,d\xi\right)^{\frac{1}{2}} \\
&\quad \times \left(\int_{\mathbb{R}^3}\left(1+|\xi|^2\right)^{s} e^{2\frac{a}{\sqrt{\sigma}}|\xi|^{\frac{1}{\sigma}}}|\widehat{u}(\xi)|^2\,d\xi\right)^{\frac{1}{2}} \\
&\le C_{a,\sigma,s}\|u(t)\|_{H^s_{\frac{a}{\sqrt{\sigma}},\sigma}(\mathbb{R}^3)},
\end{aligned} \tag{11.58}$$

for all $t \in [0, T_a^*)$, where

$$\begin{aligned}
C^2_{a,\sigma,s} &:= \int_{\mathbb{R}^3}\frac{1}{|\xi|^{2s}}e^{-2a\left(\frac{1}{\sqrt{\sigma}}-\frac{1}{\sigma}\right)|\xi|^{\frac{1}{\sigma}}}\,d\xi \\
&= 4\pi\sigma\left[2a\left(\frac{1}{\sqrt{\sigma}}-\frac{1}{\sigma}\right)\right]^{-\sigma(3-2s)}\Gamma(\sigma(3-2s)).
\end{aligned}$$

(Recall that $s < 3/2$ and $\sigma > 1$). It demonstrates Theorem 11.2.4. □

11.2.5 Generalization of the Blow-Up Criteria

We are ready to prove the blow-up criteria given in (11.3)–(11.5) and (11.7) with $n > 1$. Actually, it is enough to obtain a demonstration for the case $n = 2$; since, the proof of this last statement follows by applying a simple argument of induction.

Theorem 11.2.5 *Assume that $a > 0$, $\sigma > 1$, and $s \in \left(\frac{1}{2}, \frac{3}{2}\right)$. Let $u_0 \in H^s_{a,\sigma}\left(\mathbb{R}^3\right)$ such that $\operatorname{div} u_0 = 0$. Consider that $u \in C([0, T_a^*); H^s_{a,\sigma}\left(\mathbb{R}^3\right))$ is the maximal solution for the Navier–Stokes equations (11.1) obtained in Theorem 11.1.11. If $T_a^* < \infty$, then*

i) $\limsup_{t\nearrow T_a^*}\|u(t)\|_{H^s_{\frac{a}{\sqrt{\sigma}},\sigma}(\mathbb{R}^3)} = \infty$;

ii) $\displaystyle\int_t^{T_a^*}\left\|e^{\frac{a}{\sigma\sqrt{\sigma}}|\cdot|^{\frac{1}{\sigma}}}\widehat{u}(\tau)\right\|^2_{L^1(\mathbb{R}^3)}d\tau = \infty$;

iii) $\left\| e^{\frac{a}{\sigma\sqrt{\sigma}}|\cdot|^{\frac{1}{\sigma}}} \widehat{u}(t) \right\|_{L^1(\mathbb{R}^3)} \geq \dfrac{8\pi^3\sqrt{\mu}}{\sqrt{T_a^* - t}};$

iv) $\|u(t)\|_{H^s_{\frac{a}{\sigma},\sigma}(\mathbb{R}^3)} \geq \dfrac{8\pi^3\sqrt{\mu}}{C_1\sqrt{T_a^* - t}},$

for all $t \in [0, T_a^*)$, *and*

$$C_1 = C_{a,\sigma,s} := \left\{ 4\pi\sigma \left[2\frac{a}{\sqrt{\sigma}} \left(\frac{1}{\sqrt{\sigma}} - \frac{1}{\sigma} \right) \right]^{-\sigma(3-2s)} \Gamma(\sigma(3-2s)) \right\}^{\frac{1}{2}}.$$

Proof First of all, let us inform that this result is, in its most part, an adaptation of the proofs of the theorems established before. Understanding this, notice that (11.58) implies

$$\limsup_{t \nearrow T_a^*} \|u(t)\|_{H^s_{\frac{a}{\sqrt{\sigma}},\sigma}(\mathbb{R}^3)} = \infty. \tag{11.59}$$

It demonstrates **i)**.

By applying **i)**, as in the proof of Theorem 11.2.2, one can infer that

$$\int_t^{T_a^*} \left\| e^{\frac{a}{\sigma\sqrt{\sigma}}|\cdot|^{\frac{1}{\sigma}}} \widehat{u}(\tau) \right\|^2_{L^1(\mathbb{R}^3)} d\tau = \infty, \quad \forall t \in [0, T_a^*).$$

It proves **ii)**.

Consequently, **iii)** follows from **ii)** and the proof of Theorem 11.2.3.

Moreover, as an immediate consequence of (11.59), one obtains

$$T_a^* \geq T^*_{\frac{a}{\sqrt{\sigma}}}. \tag{11.60}$$

Thus, using inequalities (11.57) and (11.60), we reach

$$T_a^* = T^*_{\frac{a}{\sqrt{\sigma}}}. \tag{11.61}$$

Thereby, as in (11.58), by Cauchy–Schwarz's inequality, we obtain

$$\frac{8\pi^3\sqrt{\mu}}{\sqrt{T^*_{\frac{a}{\sqrt{\sigma}}} - t}} \leq \left\| e^{\frac{a}{\sigma\sqrt{\sigma}}|\cdot|^{\frac{1}{\sigma}}} \widehat{u}(t) \right\|_{L^1(\mathbb{R}^3)} = \int_{\mathbb{R}^3} e^{\frac{a}{\sigma\sqrt{\sigma}}|\xi|^{\frac{1}{\sigma}}} |\widehat{u}(\xi)|\, d\xi$$

$$\leq \left(\int_{\mathbb{R}^3} \left(1 + |\xi|^2\right)^{-s} e^{-2\left(\frac{a}{\sigma} - \frac{a}{\sigma\sqrt{\sigma}}\right)|\xi|^{\frac{1}{\sigma}}} d\xi \right)^{\frac{1}{2}}$$

$$\times \left(\int_{\mathbb{R}^3} \left(1+|\xi|^2\right)^s e^{\frac{2a}{\sigma}|\xi|^{\frac{1}{\sigma}}} |\widehat{u}(\xi)|^2 \, d\xi \right)^{\frac{1}{2}}$$
$$\leq C_{a,\sigma,s} \|u(t)\|_{H^s_{\frac{a}{\sigma},\sigma}(\mathbb{R}^3)}, \tag{11.62}$$

for all $t \in \left[0, T^*_{\frac{a}{\sqrt{\sigma}}}\right)$, where

$$C^2_{a,\sigma,s} = \int_{\mathbb{R}^3} \frac{1}{|\xi|^{2s}} e^{-2a\left(\frac{1}{\sigma}-\frac{1}{\sigma\sqrt{\sigma}}\right)|\xi|^{\frac{1}{\sigma}}} \, d\xi$$
$$= 4\pi\sigma \left[2\frac{a}{\sqrt{\sigma}}\left(\frac{1}{\sqrt{\sigma}}-\frac{1}{\sigma}\right)\right]^{-\sigma(3-2s)} \Gamma(\sigma(3-2s)).$$

By (11.61) and (11.62), one has

$$\|u(t)\|_{H^s_{\frac{a}{\sigma},\sigma}(\mathbb{R}^3)} \geq \frac{8\pi^3\sqrt{\mu}}{C_{a,\sigma,s}\sqrt{T^*_a - t}}, \quad \forall\, t \in [0, T^*_a). \tag{11.63}$$

It completes the proof of **iv)**. □

Remark 11.2.6 Passing to the limit superior, as $t \nearrow T^*_a$, in (11.63), we deduce

$$\limsup_{t \nearrow T^*_a} \|u(t)\|_{H^s_{\frac{a}{\sigma},\sigma}(\mathbb{R}^3)} = \infty.$$

Consequently, inequality (11.5), with $n = 3$, holds and the process above established can be rewritten in order to guarantee the veracity of (11.3)–(11.5) and (11.7) with $n = 3$. Therefore, inductively, one concludes that our blow-up criteria are valid by assuming $n > 1$.

11.2.6 Main Blow-Up Criterion Involving $H^s_{a,\sigma}(\mathbb{R}^3)$

To guarantee the veracity of the blow-up criterion stated in (11.2), it will be necessary to present two basic tolls. These will play an important role in the proof of (11.2). The first one was obtained by Benameur [3] and we shall prove it for convenience.

Lemma 11.2.7 *Let $\delta > 3/2$, and $f \in \dot{H}^\delta(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$. Then, the following inequality is valid:*

$$\|\widehat{f}\|_{L^1(\mathbb{R}^3)} \leq C_\delta \|f\|^{1-\frac{3}{2\delta}}_{L^2(\mathbb{R}^3)} \|f\|^{\frac{3}{2\delta}}_{\dot{H}^\delta(\mathbb{R}^3)},$$

where

$$C_\delta = 2(2\pi)^{\frac{3}{2}\left(1-\frac{3}{2\delta}\right)}\sqrt{\frac{\pi}{3}}\left[\left(\frac{2\delta}{3}-1\right)^{\frac{3}{4\delta}} + \left(\frac{2\delta}{3}-1\right)^{-1+\frac{3}{4\delta}}\right].$$

Moreover, for each $\delta_0 > 3/2$ *there exists a positive constant* C_{δ_0} *such that* $C_\delta \leq C_{\delta_0}$, *for all* $\delta \geq \delta_0$.

Proof Consider $\epsilon > 0$ arbitrary. Thereby, by using Cauchy–Schwarz's inequality, it results

$$\begin{aligned}\|\widehat{f}\|_{L^1(\mathbb{R}^3)} &= \int_{|\xi|\leq\epsilon} \left|\widehat{f}(\xi)\right| d\xi + \int_{|\xi|>\epsilon} \left|\widehat{f}(\xi)\right| d\xi \\ &\leq \left(\int_{|\xi|\leq\epsilon} d\xi\right)^{\frac{1}{2}} \left(\int_{|\xi|\leq\epsilon} \left|\widehat{f}(\xi)\right|^2 d\xi\right)^{\frac{1}{2}} \\ &\quad + \left(\int_{|\xi|>\epsilon} \frac{1}{|\xi|^{2\delta}} d\xi\right)^{\frac{1}{2}} \left(\int_{|\xi|>\epsilon} |\xi|^{2\delta} \left|\widehat{f}(\xi)\right|^2 d\xi\right)^{\frac{1}{2}}.\end{aligned}$$

Now, apply Parseval's identity and the fact that $\delta > 3/2$ to reach

$$\begin{aligned}\|\widehat{f}\|_{L^1(\mathbb{R}^3)} &\leq 2\sqrt{\frac{\pi}{3}}\epsilon^{\frac{3}{2}}(2\pi)^{\frac{3}{2}}\|f\|_{L^2(\mathbb{R}^3)} + 2\sqrt{\frac{\pi}{2\delta-3}}\epsilon^{\frac{3}{2}-\delta}\|f\|_{\dot{H}^\delta(\mathbb{R}^3)} \\ &= 2\sqrt{\frac{\pi}{3}}\left[\epsilon^{\frac{3}{2}}(2\pi)^{\frac{3}{2}}\|f\|_{L^2(\mathbb{R}^3)} + \frac{\epsilon^{\frac{3}{2}-\delta}}{\sqrt{\frac{2\delta}{3}-1}}\|f\|_{\dot{H}^\delta(\mathbb{R}^3)}\right].\end{aligned}$$

Thus, we can guarantee that the function given by

$$\epsilon \mapsto \epsilon^{\frac{3}{2}}(2\pi)^{\frac{3}{2}}\|f\|_{L^2(\mathbb{R}^3)} + \frac{\epsilon^{\frac{3}{2}-\delta}}{\sqrt{\frac{2\delta}{3}-1}}\|f\|_{\dot{H}^\delta(\mathbb{R}^3)}$$

attains its maximum at

$$\left[\frac{\sqrt{\frac{2\delta}{3}-1}\|f\|_{\dot{H}^\delta(\mathbb{R}^3)}}{(2\pi)^{\frac{3}{2}}\|f\|_{L^2(\mathbb{R}^3)}}\right]^{\frac{1}{\delta}}.$$

Consequently, we have

$$\begin{aligned}\|\widehat{f}\|_{L^1(\mathbb{R}^3)} &\leq 2(2\pi)^{\frac{3}{2}\left(1-\frac{3}{2\delta}\right)}\sqrt{\frac{\pi}{3}}\left[\left(\frac{2\delta}{3}-1\right)^{\frac{3}{4\delta}} + \left(\frac{2\delta}{3}-1\right)^{-1+\frac{3}{4\delta}}\right] \\ &\quad \times \|f\|_{L^2(\mathbb{R}^3)}^{1-\frac{3}{2\delta}}\|f\|_{\dot{H}^\delta(\mathbb{R}^3)}^{\frac{3}{2\delta}}.\end{aligned}$$

It is easy to check that

$$\lim_{\delta\to\infty} 2(2\pi)^{\frac{3}{2}\left(1-\frac{3}{2\delta}\right)}\sqrt{\frac{\pi}{3}}\left[\left(\frac{2\delta}{3}-1\right)^{\frac{3}{4\delta}}+\left(\frac{2\delta}{3}-1\right)^{-1+\frac{3}{4\delta}}\right]=2(2\pi)^{\frac{3}{2}}\sqrt{\frac{\pi}{3}}.$$

As a consequence, for each $\delta_0 > 3/2$, one deduces that

$$2(2\pi)^{\frac{3}{2}\left(1-\frac{3}{2\delta}\right)}\sqrt{\frac{\pi}{3}}\left[\left(\frac{2\delta}{3}-1\right)^{\frac{3}{4\delta}}+\left(\frac{2\delta}{3}-1\right)^{-1+\frac{3}{4\delta}}\right]$$

is bounded in the interval $[\delta_0, \infty)$.

It finishes the demonstration of Lemma 11.2.7. □

It is important to point out that the affirmation below can also be used in order to assure that (11.6) above is not trivial.

Lemma 11.2.8 *Let $a > 0$, $\sigma \geq 1$, $s \in \left[0, \frac{3}{2}\right)$, and $\delta \geq \frac{3}{2}$. For every $f \in H^s_{a,\sigma}\left(\mathbb{R}^3\right)$, we have that $f \in \dot{H}^\delta\left(\mathbb{R}^3\right)$. More precisely, one concludes that there is a positive constant $C_{a,s,\delta,\sigma}$ such that*

$$\|f\|_{\dot{H}^\delta(\mathbb{R}^3)} \leq C_{a,s,\delta,\sigma}\|f\|_{H^s_{a,\sigma}(\mathbb{R}^3)}.$$

Proof It is well known that $\mathbb{R}_+ \subseteq \cup_{n\in\mathbb{N}\cup\{0\}}[n, n+1)$. Notice that $2\sigma(\delta - s) \in \mathbb{R}_+$. As a result, there is $n_0 \in \mathbb{N}\cup\{0\}$ that depends on σ, δ, and s such that $\frac{n_0}{\sigma} \leq 2\delta - 2s < \frac{n_0+1}{\sigma}$. Consequently, one obtains $t \in [0, 1]$ such that, by Young's inequality, we infer

$$\begin{aligned}|\xi|^{2\delta-2s} &= |\xi|^{t\cdot\frac{n_0}{\sigma}+(1-t)\cdot\frac{n_0+1}{\sigma}} = |\xi|^{t\cdot\frac{n_0}{\sigma}}|\xi|^{(1-t)\cdot\frac{n_0+1}{\sigma}}\\ &\leq t|\xi|^{\frac{n_0}{\sigma}}+(1-t)|\xi|^{\frac{n_0+1}{\sigma}} \leq |\xi|^{\frac{n_0}{\sigma}}+|\xi|^{\frac{n_0+1}{\sigma}}.\end{aligned}$$

Therefore, one has

$$\begin{aligned}\|f\|^2_{\dot{H}^\delta(\mathbb{R}^3)} &= \int_{\mathbb{R}^3}|\xi|^{2\delta}\left|\hat{f}(\xi)\right|^2 d\xi \leq \int_{\mathbb{R}^3}\left[|\xi|^{\frac{n_0}{\sigma}}+|\xi|^{\frac{n_0+1}{\sigma}}\right]|\xi|^{2s}\left|\hat{f}(\xi)\right|^2 d\xi\\ &\leq \int_{\mathbb{R}^3}\Big[\frac{(2a+1)(2a)^{n_0}(n_0+1)!}{(2a)^{n_0+1}n_0!}|\xi|^{\frac{n_0}{\sigma}}+\frac{(2a+1)(2a)^{n_0+1}(n_0+1)!}{(2a)^{n_0+1}(n_0+1)!}\\ &\quad\times|\xi|^{\frac{n_0+1}{\sigma}}\Big]|\xi|^{2s}\left|\hat{f}(\xi)\right|^2 d\xi.\end{aligned}$$

As a result, we get

$$\|f\|^2_{\dot H^\delta(\mathbb{R}^3)} \le \frac{(n_0+1)!(2a+1)}{(2a)^{n_0+1}} \int_{\mathbb{R}^3} \Big[\frac{(2a|\xi|^{\frac{1}{\sigma}})^{n_0}}{n_0!} + \frac{(2a|\xi|^{\frac{1}{\sigma}})^{n_0+1}}{(n_0+1)!}\Big] |\xi|^{2s} \left|\hat f(\xi)\right|^2 d\xi.$$

Hence, we deduce

$$\|f\|^2_{\dot H^\delta(\mathbb{R}^3)} \le \frac{(n_0+1)!(2a+1)}{(2a)^{n_0+1}} \int_{\mathbb{R}^3} |\xi|^{2s} e^{2a|\xi|^{\frac{1}{\sigma}}} \times |\hat f(\xi)|^2 d\xi \le \frac{(n_0+1)!(2a+1)}{(2a)^{n_0+1}} \|f\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)}.$$

It completes the proof of Lemma 11.2.8. □

At last, let us prove the lower bound given in (11.2). This inequality is our main blow-up criterion of the solution obtained in Theorem 11.1.11.

Theorem 11.2.9 *Assume that $a > 0$, $\sigma > 1$, and $s \in \left(\frac{1}{2}, \frac{3}{2}\right)$. Let $u_0 \in H^s_{a,\sigma}\left(\mathbb{R}^3\right)$ such that $\operatorname{div} u_0 = 0$. Consider that $u \in C\left([0,T^*); H^s_{a,\sigma}\left(\mathbb{R}^3\right)\right)$ is the maximal solution for the Navier–Stokes equations (11.1) obtained in Theorem 11.1.11. If $T^* < \infty$, then*

$$\|u(t)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \ge \frac{a^{\sigma_0+\frac{1}{2}} C_2 \exp\{aC_3(T^*-t)^{-\frac{1}{3\sigma}}\}}{(T^*-t)^{\frac{2(s\sigma+\sigma_0)+1}{6\sigma}}}, \quad \forall t \in [0,T^*),$$

where $C_2 = C_{\mu,s,\sigma,u_0}$, $C_3 = C_{\mu,s,\sigma,u_0}$, and $2\sigma_0$ is the integer part of 2σ.

Proof This result follows from Lemma 11.2.7. In fact, choose $\delta = s + \frac{k}{2\sigma}$, with $k \in \mathbb{N} \cup \{0\}$ and $k \ge 2\sigma$, and $\delta_0 = s+1$. By using Lemmas 11.2.7 and 11.2.8, and (11.6), we obtain

$$\frac{8\pi^3\sqrt{\mu}}{\sqrt{T^*-t}} \le \|\widehat{u}(t)\|_{L^1(\mathbb{R}^3)} \le C_s \|u(t)\|^{1-\frac{3}{2(s+\frac{k}{2\sigma})}}_{L^2(\mathbb{R}^3)} \|u(t)\|^{\frac{3}{2(s+\frac{k}{2\sigma})}}_{\dot H^{s+\frac{k}{2\sigma}}(\mathbb{R}^3)}.$$

By using the energy estimate

$$\|u(t)\|_{L^2(\mathbb{R}^3)} \le \|u(t_0)\|_{L^2(\mathbb{R}^3)}, \quad \forall\, 0 \le t_0 \le t < T^*, \tag{11.64}$$

see (2) in [7], one has

$$\frac{C_{\mu,s,u_0}}{(T^*-t)^{\frac{2s}{3}}} \left(\frac{D_{\sigma,s,\mu,u_0}}{(T^*-t)^{\frac{1}{3\sigma}}}\right)^k \le \|u(t)\|^2_{\dot H^{s+\frac{k}{2\sigma}}(\mathbb{R}^3)}, \tag{11.65}$$

where $C_{\mu,s,u_0} = \left(C_s^{-1}8\pi^3\sqrt{\mu}\right)^{\frac{4s}{3}} \|u_0\|_{L^2(\mathbb{R}^3)}^{\frac{6-4s}{3}}$ and $D_{\sigma,s,\mu,u_0} = \left(C_s^{-1}8\pi^3 \sqrt{\mu}\|u_0\|_{L^2(\mathbb{R}^3)}^{-1}\right)^{\frac{2}{3\sigma}}$. Multiplying (11.65) by $\frac{(2a)^k}{k!}$, one obtains

$$\frac{C_{\mu,s,u_0}}{(T^*-t)^{\frac{2s}{3}}} \frac{\left(\frac{2aD_{\sigma,s,\mu,u_0}}{(T^*-t)^{\frac{1}{3\sigma}}}\right)^k}{k!} \leq \int_{\mathbb{R}^3} \frac{(2a)^k}{k!}|\xi|^{2\left(s+\frac{k}{2\sigma}\right)} |\widehat{u}(t)|^2 \, d\xi$$
$$= \int_{\mathbb{R}^3} \frac{\left(2a|\xi|^{\frac{1}{\sigma}}\right)^k}{k!}|\xi|^{2s} |\widehat{u}(t)|^2 \, d\xi.$$

By summing over the set $\{k \in \mathbb{N}; k \geq 2\sigma\}$ and applying monotone convergence theorem, it results

$$\frac{C_{\mu,s,u_0}}{(T^*-t)^{\frac{2s}{3}}} \left[\exp\left\{\frac{2aD_{\sigma,s,\mu,u_0}}{(T^*-t)^{\frac{1}{3\sigma}}}\right\} - \sum_{0\leq k<2\sigma} \frac{\left(\frac{2aD_{\sigma,s,\mu,u_0}}{(T^*-t)^{\frac{1}{3\sigma}}}\right)^k}{k!}\right]$$
$$\leq \int_{\mathbb{R}^3} \left[e^{2a|\xi|^{\frac{1}{\sigma}}} - \sum_{0\leq k<2\sigma} \frac{\left(2a|\xi|^{\frac{1}{\sigma}}\right)^k}{k!}\right] |\xi|^{2s} |\widehat{u}(t)|^2 \, d\xi$$
$$\leq \int_{\mathbb{R}^3} |\xi|^{2s} e^{2a|\xi|^{\frac{1}{\sigma}}} |\widehat{u}(t)|^2 \, d\xi$$
$$\leq \|u(t)\|^2_{H^s_{a,\sigma}(\mathbb{R}^3)},$$

for all $t \in [0, T^*)$. Finally, if we define

$$f(x) = \left[e^x - \sum_{k=0}^{2\sigma_0} \frac{x^k}{k!}\right] \left[x^{-(2\sigma_0+1)} e^{-\frac{x}{2}}\right], \quad \forall x \in (0, \infty),$$

where $2\sigma_0$ is the integer part of 2σ; then, f is continuous on $(0, \infty)$, $f > 0$, $\lim_{x\to\infty} f(x) = \infty$ (it means that f is bounded below as $x \to \infty$) and $\lim_{x\nearrow 0} f(x) = \frac{1}{(2\sigma_0+1)!}$ (it implies that f is bounded below as $x \nearrow 0$). Hence, there is a positive constant C_{σ_0} such that $f(x) \geq C_{\sigma_0}$, for all $x > 0$. Thereby, we can write

$$\|u(t)\|^2_{\dot{H}^s_{a,\sigma}(\mathbb{R}^3)} \geq \frac{C_{\mu,s,\sigma_0,u_0}}{(T^*-t)^{\frac{2s}{3}}} \left(\frac{2aD_{\sigma,s,\mu,u_0}}{(T^*-t)^{\frac{1}{3\sigma}}} \right)^{2\sigma_0+1} \exp\left\{ \frac{aD_{\sigma,s,\mu,u_0}}{(T^*-t)^{\frac{1}{3\sigma}}} \right\}$$
$$= \frac{a^{2\sigma_0+1} C_{\mu,s,\sigma,\sigma_0,u_0}}{(T^*-t)^{\frac{2(s\sigma+\sigma_0)+1}{3\sigma}}} \exp\left\{ \frac{aD_{\sigma,s,\mu,u_0}}{(T^*-t)^{\frac{1}{3\sigma}}} \right\},$$

for all $t \in [0, T^*)$. Therefore, the demonstration of Theorem 11.2.9 is given.

□

In order to finish this chapter, let us combine the results obtained above with the ones presented in the paper [3]. Thus, if it is considered that $a > 0$, $\sigma \geq 1$, $s > \frac{1}{2}$ with $s \neq \frac{3}{2}$, and $u_0 \in H^s_{a,\sigma}(\mathbb{R}^3)$ satisfies $\operatorname{div} u_0 = 0$; then, there are an instant $T > 0$ and a unique solution $u \in C([0,T]; H^s_{a,\sigma}(\mathbb{R}^3))$ for the Navier–Stokes equations (11.1). Moreover, the blow-up criteria (11.3)–(11.5) and (11.7) established above are valid as well (it is necessary to replace the hypothesis $\sigma \geq 1$ by $\sigma > 1$ in the case $s \in \left(\frac{1}{2}, \frac{3}{2}\right)$). On the other hand, the lower bound given in (11.2) must be replaced by

$$\|(u,b)(t)\|_{H^s_{a,\sigma}(\mathbb{R}^3)} \geq \frac{C_1 \|(u,b)(t)\|^{1-\frac{2s}{3}}_{L^2(\mathbb{R})} \exp\{aC_2 \|(u,b)(t)\|^{-\frac{2}{3\sigma}}_{L^2(\mathbb{R})} (T^*-t)^{-\frac{1}{3\sigma}}\}}{(T^*-t)^{\frac{s}{3}}},$$

for all $t \in [0, T^*)$, in the case $s > \frac{3}{2}$. In fact, it follows from the proof established in [3] without applying the estimate (11.64).

It is also important to point out that, considering the critical cases $s = \frac{1}{2}$ and $s = \frac{3}{2}$, the local existence, uniqueness, and blow-up of solution for (11.1) are not discussed here and are still open problems in the mathematical theory of incompressible flows.

Acknowledgements The author Natã Firmino Rocha was partially supported by CAPES grant 1579575. The author Ezequiel Barbosa was partially supported by CNPq.

References

1. J. Benameur, On the blow-up criterion of 3D Navier-Stokes equations. J. Math. Anal. Appl. **371**, 719–727 (2010)
2. J. Benameur, On the blow-up criterion of the periodic incompressible fluids. Math. Methods Appl. Sci. **36**, 143–153 (2013)
3. J. Benameur, On the exponential type explosion of Navier-Stokes equations. Nonlinear Anal. **103**, 87–97 (2014)

4. J. Benameur, L. Jlali, On the blow-up criterion of 3D-NSE in Sobolev–Gevrey spaces. J. Math. Fluid Mech. **18**, 805–822 (2016)
5. A. Biswas, Local existence and Gevrey regularity of 3-D Navier-Stokes equations with initial data. J. Differ. Equ. **215**, 429–447 (2005)
6. P. Braz e Silva, W.G. Melo, P.R. Zingano, Some remarks on the paper "On the blow up criterion of 3D Navier-Stokes equations" by J. Benameur. C. R. Acad. Sci. Paris Ser. I **352**, 913–915 (2014)
7. P. Braz e Silva, W.G. Melo, P.R. Zingano, Lower bounds on blow-up of solutions for magneto-micropolar fluid systems in homogeneous Sobolev spaces. Acta Appl. Math. **147**, 1–17 (2017)
8. M. Cannone, Harmonic analysis tools for solving the incompressible Navier-Stokes equations, in *Handbook of Mathematical Fluid Dynamics*, vol. 3 (Elsevier, North Holland, 2004), pp. 161–244
9. J.-Y. Chemin, About Navier-Stokes Equations. Publication du Laboratoire Jacques-Louis Lions, Universite de Paris VI, R96023 (1996)
10. J. Lorenz, P.R. Zingano, Properties at potential blow-up times for the incompressible Navier-Stokes equations. Bol. Soc. Paran. Mat. **35**, 127–158 (2017)
11. W.G. Melo, The magneto-micropolar equations with periodic boundary conditions: solution properties at potential blow-up times. J. Math. Anal. Appl. **435**, 1194–1209 (2016)

Chapter 12
Mathematical Research for Models Which is Related to Chemotaxis System

Jiashan Zheng

Abstract This paper proposes a survey and critical analysis focused on a variety of chemotaxis models in biology, namely the Chemotaxis system and its subsequent modifications, which, in several cases, have been developed to obtain models that prevent the non-physical blow-up of solutions. The presentation is organized in six parts. The first part focuses on background of the models which is related to Chemotaxis system and its development. The second–five part are devoted to the qualitative analysis of the (quasilinear) Keller–Segel model, the (quasilinear) chemotaxis–haptotaxis model, the (quasilinear) chemotaxis system with consumption of chemoattractant, and the (quasilinear) Keller–Segel–Navier–Stokes system. Finally, an overview of the entire contents leads to suggestions for future research activities.

Keywords Boundedness · Navier–Stokes system · Keller–Segel model · Chemotaxis models · Chemotaxis-haptotaxis model · Global existence · Nonlinear diffusion

12.1 Introduction

Mathematical analysis of biological phenomena has become more and more important in understanding these complex processes [62]. Thus, the number of mathematicians studying biological and medical phenomena and problems is continuously increasing in recent years. Mathematical models of chemotaxis, the cell movement induced by chemical substances, were introduced for the first time by Patlak [67], and further developed by Keller and Segel [43–45]. Their model, consisting of two coupled parabolic equations, has been studied in great detail in the literature (see, e.g., Burger et al. [5], Winkler et al. [3, 39, 105, 108], Osaki and

J. Zheng (✉)
School of Mathematics and Statistics Science, Ludong University, Yantai, P.R. China

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_12

Yagi [64], Horstmann [37]). To describe chemotaxis of cell populations, the signal is produced by the cells, an important variant of the quasilinear chemotaxis model

$$\begin{cases} u_t = \nabla \cdot (\phi(u)\nabla u) - \chi\nabla \cdot (\psi(u)\nabla v), \\ v_t = \Delta v - v + u \end{cases} \tag{12.1.1}$$

was initially proposed by Painter and Hillen ([66], see also Winkler et al. [3, 80, 103]), where u denotes the cell density, while v describes the concentration of the chemical signal. The function ψ measures the chemotactic sensitivity, which may depend on u, $\phi(u)$ is the diffusion function. Various variants of the chemotaxis model (12.1.1) have been extensively studied both theoretically and experimentally. Throughout the main issue of the studies was whether the chemotaxis model allows for a global-in-time solution or a chemotactic collapse (a solution blows up in finite).

If chemotaxis occurs in incompressible fluid, then the original chemotaxis system needs to be coupled with another equation which characterizes the motion of the fluid, and the resulting system can read as

$$\begin{cases} u_t + w \cdot \nabla u = \Delta u^m - \nabla \cdot (u\nabla v), \quad x \in \Omega, t > 0, \\ v_t + w \cdot \nabla v = \Delta v - v + u, \quad x \in \Omega, t > 0, \\ w_t + \kappa(w \cdot \nabla)w + \nabla P = \Delta w + u\nabla\phi, \quad x \in \Omega, t > 0, \\ \nabla \cdot w = 0, \quad x \in \Omega, t > 0, \end{cases} \tag{12.1.2}$$

where u and v are denoted as before, and $m > 0, \phi$ is a given gravitational potential. Here $w = w(x,t)$, $P = P(x,t)$, and $\kappa \in \mathbb{R}$ denote the velocity of incompressible fluid, the associated pressure, the strength of nonlinear fluid convection, respectively. System (12.1.2) is called as Keller–Segel–(Navier)–Stokes system with nonlinear diffusion (see Zheng [128], Wang et al. [91, 94, 95]), which arises in the modeling of bacterial populations in which individuals, besides moving randomly, partially adjust their movement according to the concentration gradients of a chemical which they produce themselves.

Before going into our mathematical analysis, we recall some important progresses on system (12.1.2) and its variants. The following chemotaxis-Navier-Stokes system involving tensor-valued sensitivity with saturation

$$\begin{cases} u_t + w \cdot \nabla u = \Delta u^m - \nabla \cdot (uS(x,u,v)\nabla v), \quad x \in \Omega, t > 0, \\ v_t + w \cdot \nabla v = \Delta v - uf(v), \quad x \in \Omega, t > 0, \\ w_t + \kappa(w \cdot \nabla)w + \nabla P = \Delta w + u\nabla\phi, \quad x \in \Omega, t > 0, \\ \nabla \cdot w = 0, \quad x \in \Omega, t > 0. \end{cases} \tag{12.1.3}$$

was proposed by Tuval et al. in [87]. Systems of this type arise in the modeling of populations of aerobic bacteria when suspended into sessile drops of water ([22, 87]). If $S(x,u,v) := S(v)$ is a scalar function, by making use of energy-type functionals, some local and global solvability of corresponding initial value problem for (12.1.3) in either bounded or unbounded domains have been obtained in the past years (see Lorz et al. [24, 54], Winkler et al. [3, 79, 110, 113], Chae et al. [10] and references therein). If the chemotactic sensitivity $S(x, n, c)$ fulfills

$$|S(x, n, c)| \leq C_S(1+n)^{-\alpha} \quad \text{for some } C_S > 0 \text{ and } \alpha > 0, \tag{12.1.4}$$

which is regarded as a tensor rather than a scalar one ([116]), then system (12.1.2) (with rotational flux) has only been studied very rudimentarily so far due to loss of some natural gradient-like structure (see Ishida [40], Winkler [112]).

The aim of this section is to develop a survey and critical analysis of the aforementioned analytic problems, namely a qualitative analysis of the solutions and derivation of models at the tissue scale from the underlying description at the cellular dynamics level. Hopefully, the critical analysis proposed in this section will provide and answer the aforementioned questions and generate new trends in research activity in this field.

Before giving the main results, we will give some preliminary lemmas, which play a crucial role in the following proofs. As for the proofs of these lemmas, here we will not repeat them again.

Lemma 12.1.1 ([86]) *Let $y(t) \geq 0$ be a solution of problem*

$$\begin{cases} y'(t) + Ay^p \leq B & t > 0, \\ y(0) = y_0 \end{cases} \tag{12.1.5}$$

with $A > 0$, $p > 0$, and $B \geq 0$. Then we have

$$y(t) \leq \max\left\{y_0, \left(\frac{B}{A}\right)^{\frac{1}{p}}\right\}, \quad t > 0.$$

Lemma 12.1.2 ([96]) *Let $\theta \in (0, p)$. There exists a positive constant C_{GN} such that for all $u \in W^{1,2}(\Omega) \cap L^{\theta}(\Omega)$,*

$$\|u\|_{L^p(\Omega)} \leq C_{GN}(\|\nabla u\|^a_{L^2(\Omega)}\|u\|^{1-a}_{L^\theta(\Omega)} + \|u\|_{L^\theta(\Omega)}),$$

is valid with $a = \dfrac{\frac{N}{\theta} - \frac{N}{p}}{1 - \frac{N}{2} + \frac{N}{\theta}} \in (0, 1)$.

Lemma 12.1.3 ([28, 41]) *Let $s \geq 1$ and $q \geq 1$. Assume that $p > 0$ and $a \in (0, 1)$ satisfy*

$$\frac{1}{2} - \frac{p}{N} = (1-a)\frac{q}{s} + a\left(\frac{1}{2} - \frac{1}{N}\right) \text{ and } p \leq a.$$

Then there exist $c_0, c_0' > 0$ such that for all $u \in W^{1,2}(\Omega) \cap L^{\frac{s}{q}}(\Omega)$,

$$\|u\|_{W^{p,2}(\Omega)} \leq c_0\|\nabla u\|^a_{L^2(\Omega)}\|u\|^{1-a}_{L^{\frac{s}{q}}(\Omega)} + c_0'\|u\|_{L^{\frac{s}{q}}(\Omega)}.$$

Lemma 12.1.4 ([60]) *Let* Ω *be a bounded domain in* $\mathbb{R}^N$ *with smooth boundary. If* $w \in C^2(\bar{\Omega})$ *satisfies* $\frac{\partial w}{\partial \nu} = 0$, *then*

$$\frac{\partial |\nabla w|^2}{\partial \nu} \leq C_\Omega |\nabla w|,$$

where $C_\Omega > 0$ *is a constant depending only on the curvatures of* Ω.

Lemma 12.1.5 ([7, 33]) *Suppose* $\gamma \in (1, +\infty)$ *and* $g \in L^\gamma((0, T); L^\gamma(\Omega))$. *On the other hand, assuming v is a solution of the following initial boundary value*

$$\begin{cases} v_t - \Delta v + v = g, \\ \frac{\partial v}{\partial \nu} = 0, \\ v(x, 0) = v_0(x). \end{cases} \tag{12.1.6}$$

Then there exists a positive constant C_γ *such that if* $s_0 \in [0, T)$, $v(\cdot, s_0) \in W^{2,\gamma}(\Omega)$ *with* $\frac{\partial v(\cdot, s_0)}{\partial \nu} = 0$, *then*

$$\begin{aligned} &\int_{s_0}^T e^{\gamma s} \|\Delta v(\cdot, t)\|^\gamma_{L^\gamma(\Omega)} ds \\ &\leq C_\gamma \left(\int_{s_0}^T e^{\gamma s} \|g(\cdot, s)\|^\gamma_{L^\gamma(\Omega)} ds + e^{\gamma s_0} (\|v_0(\cdot, s_0)\|^\gamma_{L^\gamma(\Omega)} + \|\Delta v_0(\cdot, s_0)\|^\gamma_{L^\gamma(\Omega)}) \right). \end{aligned} \tag{12.1.7}$$

Lemma 12.1.6 ([134]) *Let*

$$A_1 = \frac{1}{\delta + 1} \left[\frac{\delta + 1}{\delta} \right]^{-\delta} \left(\frac{\delta - 1}{\delta} \right)^{\delta+1} \tag{12.1.8}$$

and $H(y) = y + A_1 y^{-\delta} \chi^{\delta+1} C_{\delta+1}$ *for* $y > 0$. *For any fixed* $\delta \geq 1$, $\chi, C_{\delta+1} > 0$, *then*

$$\min_{y>0} H(y) = \frac{(\delta - 1)}{\delta} C_{\delta+1}^{\frac{1}{\delta+1}} \chi.$$

Proof It is easy to verify that

$$H'(y) = 1 - A_1 \delta C_{\delta+1} \left(\frac{\chi}{y} \right)^{\delta+1}.$$

Let $H'(y) = 0$, we have

$$y = (A_1 C_{\delta+1} \delta)^{\frac{1}{\delta+1}} \chi.$$

On the other hand, by $\lim_{y\to 0^+} H(y) = +\infty$ and $\lim_{y\to+\infty} H(y) = +\infty$, we have

$$\min_{y>0} H(y) = H\left[(A_1C_{\delta+1}\delta)^{\frac{1}{\delta+1}}\chi\right] = (A_1C_{\delta+1})^{\frac{1}{\delta+1}}\left(\delta^{\frac{1}{\delta+1}} + \delta^{-\frac{\delta}{\delta+1}}\right)\chi$$
$$= \frac{(\delta-1)}{\delta}C_{\delta+1}^{\frac{1}{\delta+1}}\chi.$$

□

12.2 The (Quasilinear) Keller–Segel Model

The structure of the original formulation of the (quasilinear) Keller–Segel model [43, 44] with logistic term is as follows:

$$\begin{cases} u_t = \nabla\cdot(\phi(u)\nabla u) - \chi\nabla\cdot(\psi(u)\nabla v) + f(u), & x\in\Omega,\ t>0,\\ \tau v_t = \Delta v - v + u, & x\in\Omega,\ t>0. \end{cases} \tag{12.2.1}$$

During the past four decades, the Keller–Segel models (12.2.1) have been studied extensively by many authors, where the main issue of the investigation was whether the solutions of the models are bounded or blow-up (see, e.g., Cieślak and Winkler [18], Calvez and Carrillo [6], Kiselev and Ryzhik [46], Osaki [64, 65], Corrias et al. [19, 20], Painter and Hillen [34, 66], Perthame [68], Rascle and Ziti [69], Winkler et al. [2, 102, 104, 106], Zheng et al. [123–125, 133, 134], Nagai et al. [63], Mizukami et al. [59, 61], Viglialoro [88]), where $\tau = 0$ or 1.

When $f \equiv 0$ in (12.2.1), the classical parabolic–elliptic (or fully parabolic type) chemotaxis model has been extensively studied over the past decades. For instance, for the case $\phi(u) \equiv 1$, either $\tau = 0$ or $\tau = 1$ has been thoroughly investigated to study the question whether solutions blow up in finite time or exist globally (see, for instance, Gajewski and Zacharias [25], Herrero and Velázquez [31, 32], Horstmann et al. [37–39]). In particular, in [38], Horstmann and Wang showed that the solutions are global and bounded provided that $\psi(u) \le c(u+1)^{\frac{2}{N}-\varepsilon}$ for all $u \ge 0$ with some $\varepsilon > 0$ and $c > 0$; on the other hand, if $\psi(u) \ge c(u+1)^{\frac{2}{N}+\varepsilon}$ for all $u \ge 0$ with $\varepsilon > 0$ and $c > 0$, $\Omega \subset \mathbb{R}^N (N \ge 2)$ is a ball, and some further technical conditions are satisfied, then the solutions become unbounded in finite or infinite time. Moreover, for more general $\phi(u)$, there exist many results concerning the question whether the solutions are bounded or blow-up (see, e.g., Tao and Winkler [78], Ishida et al. [41], Winkler [108], Cieślak and Stinner [15, 16]). In [78], Tao and Winkler proved that if $\dfrac{\psi(u)}{\phi(u)} \le c(u+1)^{\frac{2}{N}+\varepsilon}$ for all $u \ge 0$ with some $\varepsilon > 0$ and $c > 0$, then the corresponding solutions are global and bounded provided that $\phi(u)$ satisfies some another technical conditions. Recently, Ishida et al. [41] improve the results of [78] in the case of degenerate diffusion on a bounded non-convex domain. The breakthrough of the proof for the finite-time blow-up to the fully parabolic–parabolic Keller–Segel model has been made recently in [15] (see also [14, 18, 108]) by using a new method which strongly depends on the existence of a Lyapunov functional. In the case of a nonlinear diffusion system they even prove an optimal (with respect to possible nonlinear diffusions generating explosion in finite time of solutions) finite time blow-up result.

As we all know that logistic-type growth restrictions have been detected to prevent blow-up in (12.2.1). Particularly, for the case $f(u) = au - bu^2$ (the logistic source), the problem has been studied extensively by many authors (see, e.g., Tello and Winkler [76], Herrero and Velázquez [30], Winkler [102, 109, 111], Lankeit [49], Cao et al. [7, 9], Li and Xiang [51]). In particular, if $\tau = 0, \phi(u) \equiv 1, \psi(u) = u$, Tello and Winkler [76] discussed the existence of global bounded classical solutions to problem (12.2.1) under the assumption that either $N \leq 2$, or that the logistic damping effect $b > \frac{N-2}{N}\chi$. In [104], Winkler proved the parabolic–parabolic models are global and bounded provided that b is sufficiently large. When $\tau = 0, \phi(u) \equiv 1$, and logistic damping effect b possibly being smaller than two, it is proved in [102] that the global very weak solutions of system (12.2.1) were constructed for rather arbitrary initial data under the assumption that $b > 2 - \frac{1}{N}, N \geq 2$. Moreover, for more general $f(u)$, there have been many papers which dealt with the question whether the solutions are global bounded or blow-up (see Wang et al. [98, 100], Winkler [109]). For example, the logistic source $f(u)$ satisfies (12.2.3), $\tau = 0, \psi(u) = u, \phi \in C^2([0,\infty)), \phi(u) \geq C_\phi u^{m-1}($ for all $u > 0), \phi(u) > 0($ for all $u \geq 0)$, if one of the following cases holds: (i) $r \geq 2$ and $b > b^*$, where

$$b^* := \begin{cases} \dfrac{N[2-m]-2}{(2-m)N}\chi \ \text{ if } \ 2 > m + \dfrac{2}{N}, \\ 0 \ \text{ if } \ 2 \leq m + \dfrac{2}{N}, \end{cases} \tag{12.2.2}$$

(ii) $r \in (1,2)$ and $m \geq 2 - \dfrac{2}{N}$, the authors of [100] used the standard Moser's technique to prove that model (12.2.1) has a unique nonnegative classical solution (u, v) which is global bounded. It should be pointed out that Wang et al. [98] obtained the unique global uniformly bounded classical solution (u, v) of problem (12.2.1) when the nonlinearities $\phi(u) = (u+1)^{-\alpha}$ and $\psi(u) = u(u+1)^{\beta-1}$ with $0 < \alpha + \beta < \frac{2}{N}$, and f satisfies (12.2.3).

Going beyond these boundedness statements, a number of results are available which show that the interplay of chemotactic cross-diffusion and cell kinetics of logistic-type may lead to quite a colorful dynamics. For instance, if $N = 1, \tau = 0, \phi(u) \equiv 1, \psi(u) = u, f(u) = au - bu^2$ with $b \geq 1$, Winkler ([109]) obtained that the solutions of (12.2.1) may become large at intermediate time scales, thus exceeding the system's carrying capacity to an arbitrary extent (though not blowing up). On the other hand, the result in [106] indicates that chemotaxis models may admit finite-time blow-up solutions even in the presence of certain logistic-type growth inhibitions, provided the latter are suitably weak.

In this part, we assume that $\psi(u)$ describes the chemotactic sensitivity of cell population and the function $f : [0, \infty) \mapsto \mathbb{R}$ is smooth and satisfies $f(0) = 0$ as well as

$$f(u) \leq a - bu^r \text{ for all } u \geq 0 \tag{12.2.3}$$

with some $a \geq 0, b > 0$ and $r > 1$. Moreover, we assume that the functions $\phi(u)$ and $\psi(u)$ satisfy

$$\phi, \psi \in C^2([0, \infty)), \quad \phi(u) > 0 \text{ and } \psi(u) \geq 0 \text{ for all } u \geq 0. \tag{12.2.4}$$

Moreover, in order to prove our results, we need to impose the conditions that there exist some constants $q > 0, m \geq 0, C_\phi$ and $C_\psi > 0$ such that

$$\phi(u) \geq C_\phi u^{m-1} \text{ for all } u \geq 1 \tag{12.2.5}$$

and

$$\psi(u) \leq C_\psi u^q \text{ for all } u \geq 1. \tag{12.2.6}$$

The following local existence result is rather standard; since a similar reasoning in [7, 18, 78, 97, 100, 114], see for example. However, we could not find a precise reference in the literature that exactly matches to our situation, we include a short proof for the sake of completeness.

Lemma 12.2.1 ([120, 121]) *Assume that the nonnegative function* $u_0(x) \in W^{1,\infty}(\Omega)$, ϕ *and* ψ *satisfy* (12.2.4), $f \in W^{1,\infty}_{loc}([0, \infty))$ *with* $f(0) \geq 0$. *Then there exist a maximal existence time* $T_{max} \in (0, \infty]$ *and a pair nonnegative functions* $(u, v) \in C^0(\Omega \times [0, T_{max})) \cap C^{2,1}(\Omega \times [0, T_{max}))$ *classically solving* (12.2.1) *in* $\Omega \times [0, T_{max})$. *Moreover, if* $T_{max} < +\infty$, *then*

$$\|u(\cdot, t)\|_{L^\infty(\Omega)} \to \infty \text{ as } t \to T_{max} \tag{12.2.7}$$

is fulfilled.

Proof Let $T \in (0, 1)$ be specified below and we consider the closed bounded convex subset

$$S := \{\bar{u} \in X | \|\bar{u}(\cdot, t)\|_{L^\infty(\Omega)} \leq R \text{ for all } t \in [0, T]\}$$

of the space

$$X := C^0(\bar{\Omega} \times [0, T]),$$

where $R = \|u_0\|_{L^\infty(\Omega)} + 1$ and $\delta > 0$. For $\bar{u} \in X$, we introduce a mapping $\Phi : S \mapsto S$ such that $\Phi(\bar{u}) = u$, where u is the solution of

$$\begin{cases} u_t - \operatorname{div}\{\phi(\bar{u})\nabla u\} = -\chi\nabla\cdot(\psi(u)\nabla v) + f(u), \ x \in \Omega, t \in (0,T), \\ \dfrac{\partial u}{\partial \nu} = 0, \ x \in \partial\Omega, t \in (0,T), \\ u(x,0) = u_0(x), \quad x \in \Omega \end{cases} \tag{12.2.8}$$

with v being the solution of

$$\begin{cases} \tau v_t = \Delta v - v + \bar{u}, \quad x \in \Omega, t \in (0,T), \\ \dfrac{\partial v}{\partial \nu} = 0, \quad x \in \partial\Omega, t \in (0,T). \end{cases} \tag{12.2.9}$$

In the following we will use the Schauder fixed point theorem to show that for T small enough Φ has a fixed point. To this end, firstly, by the regularity theory of the partial differential equation (Theorem 8.34 of [27], see also [29]) there is a unique solution $v(\cdot,t) \in C^{1+\delta}(\Omega)$ to (12.2.9) for some $\delta \in (0,1)$. Moreover, due to the Sobolev embedding theorem, we have

$$\|\nabla v\|_{L^\infty(\Omega\times(0,T))} \le C_1\|\Delta v\|_{L^\infty(0,T);W^{2,p}(\Omega))} \le C_2\|\bar{u}\|_{L^\infty((0,T);L^p(\Omega))} \tag{12.2.10}$$

with $p > n$ and some positive constants C_1 and C_2. On the other hand, by (12.2.4), we obtain that there exists constant $C > 0$ such that

$$\|\phi(\bar{u})\|_{L^\infty(\Omega\times(0,T))} \le C.$$

Hence, applying the classical parabolic regularity theory (Theorem V 6.1 of [48]) to conclude that there exists $\theta \in (0,1)$ and $C > 0$ such that $u \in C^{\theta,\frac{\theta}{2}}(\Omega\times(0,T))$ and

$$\|u\|_{C^{\theta,\frac{\theta}{2}}(\Omega\times(0,T))} \le C, \tag{12.2.11}$$

where C depending on $\min\limits_{0\le s\le R}\phi(\bar{u})$ and $\|\nabla v\|_{L^\infty((0,T);C^\theta(\bar{\Omega}))}$. Thus, we have

$$\max_{0\le t\le T}\|u(\cdot,t)\|_{L^\infty(\Omega)} \le \|u_0\|_{L^\infty(\Omega)} + Ct^{\frac{\theta}{2}}. \tag{12.2.12}$$

Now, choosing $T < C^{-\frac{2}{\theta}}$ in (12.2.12) we conclude that

$$\max_{0\le t\le T}\|u(\cdot,t)\|_{L^\infty(\Omega)} \le \|u_0\|_{L^\infty(\Omega)} + 1. \tag{12.2.13}$$

Hence, it then follows from Φ is a compact mapping that Φ maps S into itself for such T. From (12.2.11) we also infer that Φ is a compact mapping. Therefore, the Schauder fixed point theorem ensures the existence of a fixed point $u \in S$ of Φ. Employing the regularity theory for elliptic equations we have $v(\cdot, t) \in C^{2+\theta}(\Omega)$. Then by (12.2.11) we have $v \in C^{2+\theta,1+\frac{\theta}{2}}(\Omega)(\Omega \times [\tau, T])$ for all $\tau \in (0, T)$. The regularity theory for parabolic equations (Theorem V 6.1 of [48]) thus entails $u \in C^{2+\theta,1+\frac{\theta}{2}}(\Omega)(\Omega \times [\tau, T])$. The solution may be prolonged in the interval $[0, T_{\max})$, and either if $T_{\max} = \infty$ or $T_{\max} < \infty$, the latter case entails that (12.2.7) holds.

Since $f(0) \geq 0$ the parabolic comparison principle ensures u is nonnegative, and hence the elliptic comparison principle applied to the second equation in (12.2.1) implies v is nonnegative. □

12.2.1 The Quasilinear Parabolic–Elliptic Keller–Segel System ($\tau = 0$)

$$\begin{cases} u_t = \nabla \cdot (\phi(u)\nabla u) - \chi \nabla \cdot (\psi(u)\nabla v) + f(u), \quad x \in \Omega,\ t > 0, \\ 0 = \Delta v - v + u, \quad x \in \Omega,\ t > 0. \\ \dfrac{\partial u}{\partial \nu} = \dfrac{\partial v}{\partial \nu} = 0, \quad x \in \partial\Omega,\ t > 0, \\ u(x, 0) = u_0(x), \quad v(x, 0) = v_0(x), x \in \Omega \end{cases} \tag{12.2.14}$$

Theorem 12.2.1 ([120]) *Assume that* $\tau = 0$ *and the initial data* $u_0(x)$ *is nonnegative function with* $u_0(x) \in W^{1,\infty}(\Omega)$, f *satisfies* (12.2.3) *with some* $a \geq 0, b > 0$ *and* $r > 1$, ϕ *and* ψ *satisfy* (12.2.4)–(12.2.6)*. If one of the following cases holds:*

(i) $q + 1 < \max\{r, m + \frac{2}{N}\}$;

(ii)

$$b > b^* := \frac{N[r - m] - 2}{(r - m)N + 2(r - 2)} \chi C_\psi \ \ if\ q + 1 = r;$$

there exists a pair $(u, v) \in C^0(\bar{\Omega} \times [0, \infty)) \cap C^{2,1}((\bar{\Omega} \times (0, \infty))$ *which solves* (12.2.14) *in the classical sense. Moreover, both* u *and* v *are bounded in* $\Omega \times (0, \infty)$.

Lemma 12.2.2 *Assume that* f *satisfies* (12.2.3) *and* (u, v) *is the solution of* (12.2.14)*. Then for any* $T \in (s, T_{max})$*, there exists* $C > 0$ *such that*

$$\int_\Omega u^\alpha(x, t)dx \leq C \ \ for\ all\ \ t \in (s, T)$$

and

$$\int_s^T \int_\Omega u^r(x,t)dxdt \le C(T+1),$$

where $0 < \alpha \le 1$.

Proof Integrating $(12.2.14)_1$ (the first equation of (12.2.14)) over Ω and using (12.2.3), we obtain

$$\frac{d}{dt}\int_\Omega u(x,t)dx = \int_\Omega f(u(x,t))dx \le a|\Omega| - b\int_\Omega u^r(x,t)dx. \quad (12.2.15)$$

Due to $r > 1$ and the Hölder inequality, we conclude that

$$\frac{d}{dt}\int_\Omega u(x,t)dx + b|\Omega|^{1-r}\left(\int_\Omega u(x,t)dx\right)^r \le a|\Omega|.$$

Hence, using Lemma 12.1.1 we can get

$$\int_\Omega u(x,t)dx \le \max\{K|\Omega|, (\frac{a}{b})^{\frac{1}{r}}|\Omega|\} \text{ for all } t \in (s,T). \quad (12.2.16)$$

If $0 < \alpha < 1$, by (12.2.16) and the Hölder inequality we have

$$\int_\Omega u^\alpha(x,t)dx \le \left(\int_\Omega u(x,t)dx\right)^\alpha |\Omega|^{1-\alpha} \le (\max\{K, (\frac{a}{b})^{\frac{1}{r}}\})^\alpha|\Omega|. \quad (12.2.17)$$

On the other hand, integrating (12.2.15) over (s,T) with respect to t and using (12.2.19), we have

$$\begin{aligned}\int_s^T \int_\Omega u^r(x,t)dxdt &\le \frac{1}{b}\left(a|\Omega|T + \int_\Omega u(x,s)dx\right)\\ &\le \frac{|\Omega|}{b}(aT+K)\\ &\le \frac{|\Omega|}{b}(a+K)(T+1).\end{aligned} \quad (12.2.18)$$

Finally, choosing $C = [(\max\{K, (\frac{a}{b})^{\frac{1}{r}}\})^\alpha + \frac{a+K}{b}]|\Omega|$ and using (12.2.17) and (12.2.18), we can get the results. □

In order to discuss the boundedness and classical solution of (12.2.14) (or (12.2.63), see Sect. 12.2.2), in light of Lemma 12.2.1, firstly, let us pick any $s_0 \in (0, T_{\max})$ and $s_0 \le 1$, there exists $K > 0$ such that

$$\begin{aligned}\|u(\tau)\|_{L^\infty(\Omega)} \le K, \quad \|v(\tau)\|_{L^\infty(\Omega)} \le K \text{ and } \|\Delta v(\tau)\|_{L^\infty(\Omega)}\\ \le K \text{ for all } \tau \in [0, s_0].\end{aligned} \quad (12.2.19)$$

Lemma 12.2.3 *Assume that f satisfies (12.2.3) with $q+1 < r$, ϕ and ψ satisfy (12.2.4)–(12.2.6). Let (u, v) be a solution to (12.2.14) on $(0, T_{max})$ and*

$$\kappa_r = \begin{cases} r-1 \ \ if \ 1 < r \le 2, \\ 1 \ \ if \ r \ge 2. \end{cases} \tag{12.2.20}$$

Then there exist $M_0 > 0$ and $M > 0$, depending on a, b, q, r K and $|\Omega|$ only, such that

$$\int_\Omega u^{\mu_k+\kappa_r}(x,t)dx \le C_2 M^{\mu_k+\kappa_r}(T+1)\ \forall\, t \in (s, T),$$

where $\mu_k = (r-q)^k \kappa_r + r - 1 - q$ and $k \ge 1$.

Proof For any $\beta \ge -1$, multiplying $(12.2.14)_1$ by $u^{\mu_k+\beta}$, integrating over Ω and using (12.2.5), we get

$$\begin{aligned} &\frac{1}{\mu_k+\beta+1}\frac{d}{dt}\|u\|^{\mu_k+\beta+1}_{L^{\mu_k+\beta+1}(\Omega)} + C_\phi(\mu_k+\beta)\int_\Omega u^{m+\mu_k+\beta-2}|\nabla u|^2 dx \\ &\le -\chi \int_\Omega \nabla\cdot(\psi(u)\nabla v)u^{\mu_k+\beta}dx + \int_\Omega u^{\mu_k+\beta} f(u)dx. \end{aligned} \tag{12.2.21}$$

Integrating by parts to the first term on the right-hand side of (12.2.21) and using $q+1 < r$ and the Young inequality, we obtain from the second equation in (12.2.14)

$$\begin{aligned} &-\chi\int_\Omega \nabla\cdot(\psi(u)\nabla v)u^{\mu_k+\beta}dx \\ &= (\mu_k+\beta)\chi\int_\Omega \psi(u)u^{\mu_k+\beta-1}\nabla u\cdot\nabla v dx \\ &= (\mu_k+\beta)\chi\int_\Omega \nabla\Psi(u)\cdot\nabla v dx \\ &= -(\mu_k+\beta)\chi\int_\Omega \Psi(u)\Delta v dx \\ &\le (\mu_k+\beta)\chi\int_\Omega \Psi(u)u dx \\ &= (\mu_k+\beta)\chi\int_\Omega\int_1^u \psi(\tau)\tau^{\mu_k+\beta-1}d\tau u dx \\ &\le \frac{\mu_k+\beta}{\mu_k+\beta+q}\chi C_\psi\int_\Omega u^{\mu_k+\beta+q+1}dx \\ &\le \chi C_\psi \int_\Omega u^{\mu_k+\beta+q+1}dx \\ &\le \frac{b}{2}\int_\Omega u^{\mu_k+\beta+r}dx + C_1, \end{aligned} \tag{12.2.22}$$

where

$$\Psi(u) = \int_1^u \psi(\tau)\tau^{\mu_k+\beta-1}d\tau, \tag{12.2.23}$$

$$\begin{aligned}
C_1 &:= \frac{r-q-1}{\mu_k+\beta+r}\left(\frac{b}{2}\times\frac{\mu_k+\beta+r}{\mu_k+\beta+q+1}\right)^{-\frac{\mu_k+\beta+q+1}{r-q-1}}(\chi C_\psi)^{\frac{\mu_k+\beta+r}{r-q-1}}|\Omega| \\
&= \frac{r-q-1}{\mu_k+\beta+r}\left(\frac{b}{2}\right)^{-\frac{q}{r-q-1}}(\chi C_\psi)^{\frac{r-1}{r-q-1}}\left(1+\frac{r-q-1}{\mu_k+\beta+q+1}\right)^{-\frac{\mu_k+\beta+q+1}{r-q-1}} \\
&\quad\times\left[\left(\frac{2\chi C_\psi}{b}\right)^{\frac{1}{r-q-1}}\right]^{\mu_k+\beta+1}|\Omega| \\
&\le (r-q-1)\left(\frac{b}{2}\right)^{-\frac{q}{r-q-1}}(\chi C_\psi)^{\frac{r-1}{r-q-1}}\left(1+\frac{r-q-1}{\mu_k+\beta+q+1}\right)^{-\frac{\mu_k+\beta+q+1}{r-q-1}}|\Omega| \\
&\quad\times\frac{\left[\left(\frac{2\chi C_\psi}{b}\right)^{\frac{1}{r-q-1}}\right]^{\mu_k+\beta+1}}{\mu_k+\beta+1}.
\end{aligned} \tag{12.2.24}$$

Here we have used the fact that $r > 1$, u and v are nonnegative functions. On the other hand, due to (12.2.3), we have

$$\int_\Omega u^{\mu_k+\beta}f(u)dx \le \int_\Omega u^{\mu_k+\beta}(a-bu^r)dx. \tag{12.2.25}$$

Inserting (12.2.25) and (12.2.22) into (12.2.21), we have

$$\begin{aligned}
&\frac{1}{\mu_k+\beta+1}\frac{d}{dt}\|u\|^{\mu_k+\beta+1}_{L^{\mu_k+\beta+1}(\Omega)} + C_\phi(\mu_k+\beta)\int_\Omega u^{m+\mu_k+\beta-2}|\nabla u|^2dx \\
&\quad\le -\frac{b}{2}\int_\Omega u^{\mu_k+\beta+r}dx + a\int_\Omega u^{\mu_k+\beta}dx + C_1.
\end{aligned} \tag{12.2.26}$$

Since $r > 1$, and with the help of the Young inequality, we see that

$$\begin{aligned}
&\frac{1}{\mu_k+\beta+1}\frac{d}{dt}\|u\|^{\mu_k+\beta+1}_{L^{\mu_k+\beta+1}(\Omega)} + C_\phi(\mu_k+\beta)\int_\Omega u^{m+\mu_k+\beta-2}|\nabla u|^2dx \\
&\quad\le -\frac{1}{4}b\int_\Omega u^{\mu_k+\beta+r}dx + C_1 + C_2,
\end{aligned} \tag{12.2.27}$$

where

$$\begin{aligned} C_2 &:= \frac{r}{\mu_k+\beta+r}\left(\frac{b}{4}\times\frac{\mu_k+\beta+r}{\mu_k+\beta}\right)^{-\frac{\mu_k+\beta}{r}}(a)^{\frac{\mu_k+\beta+r}{r}}|\Omega| \\ &= \frac{r}{\mu_k+\beta+r}\left(\frac{b}{4}\right)^{\frac{1}{r}}(a)^{\frac{r-1}{r}}\left(1+\frac{r}{\mu_k+\beta}\right)^{-\frac{\mu_k+\beta}{r}}\left[\left(\frac{4a}{b}\right)^{\frac{1}{r}}\right]^{\mu_k+\beta+1}|\Omega| \\ &\le r\left(\frac{b}{4}\right)^{\frac{1}{r}}(a)^{\frac{r-1}{r}}\left(1+\frac{r}{\mu_k+\beta}\right)^{-\frac{\mu_k+\beta}{r}}|\Omega|\frac{\left[\left(\frac{4a}{b}\right)^{\frac{1}{r}}\right]^{\mu_k+\beta+1}}{\mu_k+\beta+1}. \end{aligned} \tag{12.2.28}$$

Now, choosing

$$M_1=\max\left\{(r-q-1)\left(\frac{b}{2}\right)^{-\frac{q}{r-q-1}}(\chi C_\psi)^{\frac{r-1}{r-q-1}},\ r\left(\frac{b}{4}\right)^{\frac{1}{r}}(a)^{\frac{r-1}{r}}\right\}|\Omega|$$

and

$$M_2=\max\left\{1+\left(\frac{4a}{b}\right)^{\frac{1}{r}},\ 1+\left(\frac{2\chi C_\psi}{b}\right)^{\frac{1}{r-q-1}}\right\},$$

then by (12.2.24), (12.2.27) and (12.2.28), we have

$$\frac{1}{\mu_k+\beta+1}\frac{d}{dt}\|u\|^{\mu_k+\beta+1}_{L^{\mu_k+\beta+1}(\Omega)}+\frac{1}{4}b\int_\Omega u^{\mu_k+\beta+r}dx\le M_1\frac{M_2^{\mu_k+\beta+1}}{\mu_k+\beta+1}. \tag{12.2.29}$$

After integrating (12.2.29) over (s,T), then we have

$$\int_\Omega u^{\mu_k+\beta+1}(x,t)dx\le\int_\Omega u^{\mu_k+\beta+1}(x,s)dx+M_1M_2^{\mu_k+\beta+1}T\ \text{ for any }\ t\in(s,T), \tag{12.2.30}$$

which together with (12.2.19) implies that

$$\int_\Omega u^{\mu_k+\beta+1}(x,t)dx\le K^{\mu_k+\beta+1}|\Omega|+M_1M_2^{\mu_k+\beta+1}T\le M_0M^{\mu_k+\beta+1}(T+1) \tag{12.2.31}$$

with $M_0=M_1+|\Omega|$, $M=M_2+K$. In particular, choosing $\beta=\kappa_r-1$ in (12.2.31) and using $\kappa_r>0$, then we have

$$\int_\Omega u^{\mu_k+\kappa_r}(x,t)dx\le M_0M^{\mu_k+\kappa_r}(T+1). \tag{12.2.32}$$

This completes the proof of Lemma 12.2.3. □

Lemma 12.2.4 *Assume that f satisfies (12.2.3) with $q+1 = r$, ϕ and ψ satisfy (12.2.4)–(12.2.6). Let (u, v) be a solution to (12.2.14) on $(0, T_{max})$. Then for any $T \in (s, T_{max})$ and $k \in (1, \frac{\chi C_\psi + bq - b}{(\chi C_\psi - b)^+})$, there exists a positive constant C such that*

$$\|u(\cdot, t)\|_{L^k(\Omega)} \le C \text{ for all } t \in (s, T) \tag{12.2.33}$$

holds.

Proof Multiplying $(12.2.14)_1$ by u^{k-1} and integrating over Ω, we get

$$\begin{aligned}&\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} + (k-1)\int_\Omega u^{k-2}\phi(u)|\nabla u|^2 dx \\ &\quad \le -\chi \int_\Omega \nabla \cdot (\psi(u)\nabla v)u^{k-1}dx + \int_\Omega u^{k-1} f(u)dx.\end{aligned} \tag{12.2.34}$$

Integrating by parts to the first term on the right-hand side of (12.2.34), we obtain from the second equation in (12.2.14)

$$\begin{aligned}&-\chi \int_\Omega \nabla \cdot (\psi(u)\nabla v)u^{k-1}dx \\ &= (k-1)\chi \int_\Omega \psi(u)u^{k-2}\nabla u \cdot \nabla v dx \\ &= (k-1)\chi \int_\Omega \nabla\tilde{\Psi}(u) \cdot \nabla v dx \\ &\le (k-1)\chi \int_\Omega \tilde{\Psi}(u)u dx \\ &\le \frac{k-1}{k+q-1}\chi C_\psi \int_\Omega u^{q+k}dx,\end{aligned} \tag{12.2.35}$$

where

$$\tilde{\Psi}(u) = \int_1^u \psi(\tau)\tau^{k-2}d\tau. \tag{12.2.36}$$

Inserting (12.2.35) into (12.2.34) and using (12.2.5), we have

$$\begin{aligned}&\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} + (k-1)C_\phi \int_\Omega u^{m+k-3}\phi(u)|\nabla u|^2 dx \\ &\quad \le \frac{k-1}{k+q-1}\chi C_\psi \int_\Omega u^{q+k}dx + \int_\Omega u^{k-1} f(u)dx,\end{aligned} \tag{12.2.37}$$

which combined with (12.2.3) and $q+1 = r$ implies that

$$\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} \le -\left(b - \frac{k-1}{k+q-1}\chi C_\psi\right)\int_\Omega u^{q+k}dx + a\int_\Omega u^{k-1}dx. \tag{12.2.38}$$

Due to $k \in (1, \frac{\chi C_\psi + bq - b}{(\chi C_\psi - b)^+})$, $b - \frac{k-1}{k+q-1}\chi C_\psi > 0$, it then follows from the Young inequality that

$$\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} \le -\frac{1}{2}\left(b - \frac{k-1}{k+q-1}\chi C_\psi\right)\int_\Omega u^{q+k}dx + C_1 \tag{12.2.39}$$

with $C_1 > 0$ depends only on χ, C_ψ, b, a, and q. Thus, by the Höder inequality, we have

$$\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} + \frac{1}{2}\left(b - \frac{k-1}{k+q-1}\chi\right)|\Omega|^{-\frac{q}{k}}\left(\int_\Omega u^k dx\right)^{\frac{q+k}{k}} \le C_1. \tag{12.2.40}$$

By using Lemma 12.1.1, we have the boundedness of $\|u(\cdot,t)\|_{L^k(\Omega)}$ for all $t \in (s, T)$. The proof Lemma 12.2.4 is complete. □

Lemma 12.2.5 *Assume that f satisfies (12.2.3) with $q + 1 = r$ and $b > b^*$. Suppose ϕ and ψ satisfy (12.2.4)–(12.2.6). Let (u, v) be a solution to (12.2.14) on $(0, T_{max})$. Then for any $T \in (s, T_{max})$ and for all $k > 1$, there exists a positive constant C such that*

$$\|u(\cdot,t)\|_{L^k(\Omega)} \le C \quad \textit{for all} \quad t \in (s, T) \tag{12.2.41}$$

holds.

Proof Multiplying $(12.2.14)_1$ by u^{k-1} and integrating over Ω, we get

$$\begin{aligned}\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} &+ (k-1)\int_\Omega u^{k-2}\phi(u)|\nabla u|^2dx\\ &\le -\chi\int_\Omega \nabla\cdot(\psi(u)\nabla v)u^{k-1}dx + \int_\Omega u^{k-1}f(u)dx.\end{aligned} \tag{12.2.42}$$

Integrating by parts to the first term on the right-hand side of (12.2.42), we obtain from the second equation in (12.2.14)

$$\begin{aligned}&-\chi\int_\Omega \nabla\cdot(\psi(u)\nabla v)u^{k-1}dx\\ =\ &(k-1)\chi\int_\Omega \psi(u)u^{k-2}\nabla u\cdot\nabla v dx\\ &= (k-1)\chi\int_\Omega \nabla\tilde\Psi(u)\cdot\nabla v dx\\ &\le (k-1)\chi\int_\Omega \tilde\Psi(u)u dx\\ &\le \frac{k-1}{k+q-1}\chi C_\psi\int_\Omega u^{q+k}dx,\end{aligned} \tag{12.2.43}$$

where $\tilde\Psi(u)$ is given by (12.2.36).

Inserting (12.2.43) into (12.2.42) and using (12.4.8), we have

$$\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} + (k-1)C_\phi \int_\Omega u^{m+k-3}\phi(u)|\nabla u|^2 dx \\ \leq \frac{k-1}{k+q-1}\chi C_\psi \int_\Omega u^{q+k}dx + \int_\Omega u^{k-1} f(u)dx,$$

which together with $q+1=r$ implies that

$$\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} + \frac{4C_\phi(k-1)}{(m+k-1)^2}|\nabla u^{\frac{m+k-1}{2}}|_2^2 \\ \leq \left(\frac{k-1}{k+q-1}\chi C_\psi - b\right)\int_\Omega u^{q+k}dx + a\int_\Omega u^{k-1}dx. \tag{12.2.44}$$

It then follows from the Young inequality that

$$\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} + \frac{4C_\phi(k-1)}{(m+k-1)^2}|\nabla u^{\frac{m+k-1}{2}}|_2^2 \\ \leq \left(\frac{k-1}{k+q-1}\chi C_\psi + a - b\right)\int_\Omega u^{q+k}dx + a|\Omega|. \tag{12.2.45}$$

On the other hand, due to $b > b^*$, we have

$$\frac{\chi C_\psi + bq - b}{(\chi C_\psi - b)^+} > \frac{N(r-m)}{2}.$$

Hence, we can choose some $k' \in (\frac{N[r-m]}{2}, \frac{\chi C_\psi + bq - b}{(\chi C_\psi - b)^+})$. Therefore, by Lemma 12.2.4 we obtain that there exists a positive constant $C := C(\|u_0\|_{L^{k'}(\Omega)})$ such that

$$\|u(\cdot,t)\|_{L^{k'}(\Omega)} \leq C \quad \text{for all} \quad t \in (s, T_{max}). \tag{12.2.46}$$

Thus, choosing $k > k'$ and using the Gagliarde–Nirenberg inequality, we obtain

$$\int_\Omega u^{q+k}dx \\ = \|u^{\frac{m+k-1}{2}}\|^{\frac{2(q+k)}{m+k-1}}_{\frac{2(q+k)}{m+k-1}} \\ \leq C_1(\|\nabla u^{\frac{m+k-1}{2}}\|_2^\lambda \|u^{\frac{m+k-1}{2}}\|^{1-\lambda}_{\frac{2k'}{m+k-1}} + \|u^{\frac{m+k-1}{2}}\|_{\frac{2k'}{m+k-1}})^{\frac{2(q+k)}{m+k-1}} \\ \leq C_2(\|\nabla u^{\frac{m+k-1}{2}}\|_2^{\frac{2\lambda(q+k)}{m+k-1}} + 1) \tag{12.2.47}$$

with some positive constants C_1, C_2 and

$$\lambda = \frac{\frac{N[m+k-1]}{2k'} - \frac{N[m+k-1]}{2(q+k)}}{1 - \frac{N}{2} + \frac{N[m+k-1]}{2k'}} = [m+k-1]\frac{\frac{N}{2k'} - \frac{N}{2(q+k)}}{1 - \frac{N}{2} + \frac{N[m+k-1]}{2k'}} \in (0,1),$$

where we have used that $k > \frac{[r-m]N}{2} > \frac{[r-m]N-2}{2}$. By $k' > \frac{[r-m]N}{2}$, we have

$$\begin{aligned} &\frac{2\lambda(q+k)}{m+k-1} \\ &= 2(q+k)\frac{\frac{N}{2k'} - \frac{N}{2(q+k)}}{1 - \frac{N}{2} + \frac{N[m+k-1]}{2k'}} \\ &= \frac{\frac{N(q+k)}{k'} - N}{1 - \frac{N}{2} + \frac{N[m+k-1]}{2k'}} \\ &< 2. \end{aligned} \tag{12.2.48}$$

Thus, combining (12.2.47) with (12.2.48) and using the Young inequality there exists a positive constant C_3 such that

$$\left(\frac{k-1}{k+q-1}\chi C_\psi + a\right)\int_\Omega u^{q+k}dx \le \frac{4C_\phi(k-1)}{(m+k-1)^2}\|\nabla u^{\frac{m+k-1}{2}}|_2^2 + C_3, \tag{12.2.49}$$

which together with (12.2.45) implies that

$$\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} + b\int_\Omega u^{q+k}dx \le C_4, \tag{12.2.50}$$

where C_4 is a positive constant. Employing the Höder inequality to the second term on the left-hand side of (12.2.50) and using Lemma 12.1.1, we obtain the desired result. □

Lemma 12.2.6 *Assume that f satisfies* (12.2.3) *with $q+1 < m + \frac{2}{N}$. Suppose ϕ and ψ satisfy* (12.2.4)–(12.2.6)*. Let (u, v) be a solution to* (12.2.14) *on $(0, T_{max})$. Then for any $T \in (s, T_{max})$ and $k \ge 1$, there exists a positive constant C such that*

$$\|u(\cdot,t)\|_{L^k(\Omega)} \le C \ \text{for all}\ t \in (s,T) \tag{12.2.51}$$

holds.

Proof Multiplying $(12.2.14)_1$ by u^{k-1}, integrating over Ω, employing the same arguments as in the proof of (12.2.37) and using (12.2.3), we have

$$\begin{aligned} &\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} + C_\phi(k-1)\int_\Omega u^{m+k-3}|\nabla u|^2dx \\ &\le \frac{k-1}{k+q-1}\chi C_\psi\int_\Omega u^{q+k}dx + a\int_\Omega u^{k-1}dx - b\int_\Omega u^{r+k-1}dx. \end{aligned} \tag{12.2.52}$$

Since $r > 1$, using the Hölder inequality yields

$$b|\Omega|^{-\frac{r-1}{k}} \left(\int_\Omega u^k dx \right)^{\frac{r+k-1}{k}} \leq b \int_\Omega u^{r+k-1} dx, \tag{12.2.53}$$

which together with (12.2.52) and the Young inequality implies that

$$\begin{aligned} &\frac{1}{k} \frac{d}{dt} \|u\|^k_{L^k(\Omega)} + \frac{4C_\phi(k-1)}{[m+k-1]^2} \left\| \nabla u^{\frac{m+k-1}{2}} \right\|^2_{L^2(\Omega)} + b|\Omega|^{-\frac{r-1}{k}} \left(\int_\Omega u^k dx \right)^{\frac{r+k-1}{k}} \\ &\leq C_1 \int_\Omega u^{q+k} dx + C_1, \end{aligned} \tag{12.2.54}$$

where C_1 is a positive constant. On the other hand, by Lemma 12.2.2, we have $\|u(\cdot, t)\|_{L^1(\Omega)}$ is bounded for all $t \in (s, T)$ and thanks to the Gagliarde–Nirenberg inequality, there exist positive constants C_2 and C_3 such that

$$\begin{aligned} &C_1 \int_\Omega u^{q+k} dx \\ &= C_1 \left\| u^{\frac{m+k-1}{2}} \right\|^{\frac{2(q+k)}{m+k-1}}_{L^{\frac{2(q+k)}{m+k-1}}(\Omega)} \\ &\leq C_2 \left(\left\| \nabla u^{\frac{m+k-1}{2}} \right\|^{\lambda'}_{L^2(\Omega)} \left\| u^{\frac{m+k-1}{2}} \right\|^{1-\lambda'}_{L^{\frac{2}{m+k-1}}(\Omega)} + \left\| u^{\frac{m+k-1}{2}} \right\|_{\frac{2}{m+k-1}} \right)^{L^{\frac{2(q+k)}{m+k-1}}(\Omega)} \\ &\leq C_3 \left(\left\| \nabla u^{\frac{m+k-1}{2}} \right\|^{\frac{\lambda' 2(q+k)}{m+k-1}}_{L^2(\Omega)} + 1 \right) \end{aligned} \tag{12.2.55}$$

where

$$\lambda' = \frac{\frac{N[m+k-1]}{2} - \frac{N[m+k-1]}{2(q+k)}}{1 - \frac{N}{2} + \frac{N[m+k-1]}{2}} = [m+k-1] \frac{\frac{N}{2} - \frac{N}{2(q+k)}}{1 - \frac{N}{2} + \frac{N[m+k-1]}{2}} \in (0, 1).$$

Next, due to $q + 1 < m + \dfrac{2}{N}$, we have

$$\begin{aligned} &\frac{\lambda' 2(q+k)}{m+k-1} \\ &= 2(q+k) \frac{\frac{N}{2} - \frac{1}{2(q+k)}}{1 - \frac{N}{2} + \frac{N[m+k-1]}{2}} \\ &= \frac{N(q+k-1)}{1 - \frac{N}{2} + \frac{N[m+k-1]}{2}} \\ &< 2. \end{aligned} \tag{12.2.56}$$

Hence, by (12.2.55), (12.2.56) and using the Young inequality we find a constant $C_4 > 0$ such that

$$C_1 \int_\Omega u^{q+k} dx \le \frac{4C_\phi(k-1)}{[m+k-1]^2} \left| \nabla u^{\frac{m+k-1}{2}} \right|_2^2 + C_4. \tag{12.2.57}$$

Inserting (12.2.57) into (12.2.54), we derive

$$\frac{1}{k}\frac{d}{dt}\|u\|^k_{L^k(\Omega)} + b|\Omega|^{-\frac{r-1}{k}} \left(\int_\Omega u^k dx \right)^{\frac{r+k-1}{k}} \le C_5. \tag{12.2.58}$$

By using Lemma 12.1.1, we have the boundedness of $\|u(\cdot,t)\|_{L^k(\Omega)}$ for all $t \in (s,T)$. □

In the following, we will set up the iteration procedure to derive the main result. To this end, we we first give the $\|\cdot\|_{L^\infty(\Omega)}$ estimate of $u(t)$ for all $t \in (0,T)$, where $T \in (0, T_{\max})$.

Lemma 12.2.7 *Let $q+1 < r$. Then there exists a constant $C > 0$ independent of T such that for any $T \in (0, T_{\max})$, $\|u(\cdot,t)\|_{L^\infty(\Omega)} \le C$ for all $t \in (0,T)$.*

Proof Let $\mu_k = (r-q)^k \kappa_r + r - 1 - q$, where κ_r is given by (12.2.20). Hence, Lemma 12.2.3 gives us

$$\int_\Omega u^{\mu_k+\kappa_r}(x,t)dx \le M_0 M^{\mu_k+\kappa_r}(T+1) \ \text{ for all } \ t \in (s,T) \text{ and } \ k \ge 1, \tag{12.2.59}$$

that is,

$$\|u(\cdot,t)\|_{L^{\mu_k+\kappa_r}} \le M_0^{\frac{1}{\mu_k+\kappa_r}} M(T+1)^{\frac{1}{\mu_k+\kappa_r}} \ \text{ for all } \ t \in (s,T) \text{ and } \ k \ge 1. \tag{12.2.60}$$

Because $q+1 < r$ implies that $\mu_k \to \infty$ as $k \to \infty$. Thus letting $k \to \infty$ on both sides of (12.2.60), we have

$$\|u(\cdot,t)\|_{L^\infty} \le M \ \text{ for all } \ t \in (s,T). \tag{12.2.61}$$

Here we have used the fact

$$\lim_{k\to\infty} \frac{1}{\mu_k + \kappa_r} = 0.$$

On the other hand, it follows from Lemma 12.2.1 that

$$\|u(\cdot,t)\|_{L^\infty(\Omega)} \le K, \quad \text{for all } \ t \in (0,s). \tag{12.2.62}$$

Now, choosing $C := \max\{K, M\}$, we complete the proof. □

The Proof of Theorem 12.2.1 In view of Lemmas 12.2.2–12.2.6 and 12.2.7 and using Lemma A.1 in [78] (see also [1]), we obtain that u is uniformly bounded in $\Omega \times (0, T_{\max})$. Theorem 12.2.1 will be proved if we can show $T_{\max} = \infty$. Suppose on contrary that $T_{\max} < \infty$. Then $\|u(\cdot, t)\|_{L^\infty(\Omega)} \le C$ for all $t \in (0, T)$, where C is independent of $T_{\max}$. This contradicts Lemma 12.2.1. Thanks to Lemmas 12.2.1, 12.2.2–12.2.6 and 12.2.7, the solution (u, v) is global in time and bounded.

12.2.2 The Quasilinear Parabolic–Parabolic Keller–Segel System ($\tau = 1$)

$$\begin{cases} u_t = \nabla \cdot (\phi(u)\nabla u) - \chi \nabla \cdot (\psi(u)\nabla v) + f(u), \quad x \in \Omega,\ t > 0, \\ v_t = \Delta v - v + u, \quad x \in \Omega,\ t > 0. \\ \dfrac{\partial u}{\partial \nu} = \dfrac{\partial v}{\partial \nu} = 0, \quad x \in \partial\Omega,\ t > 0, \\ u(x, 0) = u_0(x), \quad v(x, 0) = v_0(x), x \in \Omega \end{cases} \tag{12.2.63}$$

Theorem 12.2.2 *Assume that $u_0 \in C^0(\bar{\Omega})$ and $v_0 \in W^{1,\theta}(\bar{\Omega})$ (with some $\theta > n$) both are nonnegative, f satisfies (12.2.3), ϕ and ψ satisfies (12.2.4)–(12.2.6). If $q + 1 < \max\{r, m + \frac{2}{N}\}$,*
or

$$b \text{ is big enough}, \quad \text{if } q = r - 1,$$

then there exists a pair $(u, v) \in (C^0(\bar{\Omega} \times [0, \infty)) \cap C^{2,1}(\Omega \times (0, \infty))^2$ which solves (12.2.63) in the classical sense. Moreover, both u and v are bounded in $\Omega \times (0, \infty)$. If $\phi(u) \equiv \psi(u) \equiv 1$, $f(u) = au - bu^2$ and $b > \frac{(N-2)_+}{N}\chi C_{\frac{N}{2}+1}^{\frac{1}{\frac{N}{2}+1}}$, then (12.2.63) possesses a unique classical solution (u, v) which is globally bounded in $\Omega \times (0, \infty)$.

Lemma 12.2.8 *Let $\Omega \subset \mathbb{R}^N (N \ge 1)$ be a smooth bounded domain. Assume that $b > \frac{(N-2)_+}{N}\chi C_{\frac{N}{2}+1}^{\frac{1}{\frac{N}{2}+1}}$, where $C_{\frac{N}{2}+1}$ is given by Lemma 12.1.5 (with $\gamma = \frac{N}{2} + 1$ in Lemma 12.1.5). Let (u, v) be a solution to (12.2.63) on $(0, T_{max})$. Then for all $p > 1$, there exists a positive constant $C := C(p, |\Omega|, b, \chi, K)$ such that*

$$\int_\Omega u^p(x, t) \le C \quad \text{for all } t \in (0, T_{max}). \tag{12.2.64}$$

Proof Multiplying the first equation of (12.2.63) by u^{r-1} and integrating over Ω, we get

$$\begin{aligned}&\frac{1}{r}\frac{d}{dt}\|u\|^r_{L^r(\Omega)}+(r-1)\int_\Omega u^{r-2}|\nabla u|^2\\&=-\chi\int_\Omega \nabla\cdot(u\nabla v)u^{r-1}+\int_\Omega u^{r-1}(au-bu^2)\ \text{ for all }\ t\in(0,T_{\max}),\end{aligned}\tag{12.2.65}$$

that is,

$$\begin{aligned}&\frac{1}{r}\frac{d}{dt}\|u\|^r_{L^r(\Omega)}\\&\le -\frac{r+1}{r}\int_\Omega u^r-\chi\int_\Omega \nabla\cdot(u\nabla v)u^{r-1}\\&+\int_\Omega\left(\frac{r+1}{r}u^r+u^{r-1}(au-bu^2)\right)\ \text{ for all }\ t\in(0,T_{\max}).\end{aligned}\tag{12.2.66}$$

Hence, by the Young inequality, it reads that

$$\begin{aligned}&\int_\Omega\left(\frac{r+1}{r}u^r+u^{r-1}\left(au-bu^2\right)\right)\\&\le\frac{r+1}{r}\int_\Omega u^r+a\int_\Omega u^r-b\int_\Omega u^{r+1}\\&\le(\varepsilon_1-b)\int_\Omega u^{r+1}+C_1(\varepsilon_1,r),\end{aligned}\tag{12.2.67}$$

where

$$C_1(\varepsilon_1,r)=\frac{1}{r+1}\left(\varepsilon_1\frac{r+1}{r}\right)^{-r}\left(\frac{r+1}{r}+a\right)^{r+1}|\Omega|.$$

Next, integrating by parts to the first term on the right-hand side of (12.2.65), using the Young inequality we obtain

$$\begin{aligned}&-\chi\int_\Omega\nabla\cdot(u\nabla v)u^{r-1}\\&=(r-1)\chi\int_\Omega u^{r-1}\nabla u\cdot\nabla v\\&=-\frac{r-1}{r}\chi\int_\Omega u^r\Delta v\\&\le\frac{r-1}{r}\chi\int_\Omega u^r|\Delta v|\ \text{ for all }\ t\in(0,T_{\max}).\end{aligned}\tag{12.2.68}$$

Now, let

$$\lambda_0:=(A_1C_{r+1}r)^{\frac{1}{r+1}}\chi,\tag{12.2.69}$$

where A_1 is given by (12.1.8). While from (12.2.68) and the Young inequality, we have

$$
\begin{aligned}
&-\chi \int_\Omega \nabla \cdot (u\nabla v)u^{r-1} \\
&\leq \lambda_0 \int_\Omega u^{r+1} + \frac{1}{r+1}\left[\lambda_0 \frac{r+1}{r}\right]^{-r} \left(\frac{r-1}{r}\chi\right)^{r+1} \int_\Omega |\Delta v|^{r+1} \\
&= \lambda_0 \int_\Omega u^{r+1} + A_1\lambda_0^{-r}\chi^{r+1} \int_\Omega |\Delta v|^{r+1} \ \text{ for all } \ t \in (0, T_{\max}).
\end{aligned}
\tag{12.2.70}
$$

Thus, inserting (12.2.67) and (12.2.70) into (12.2.66), we get

$$
\begin{aligned}
\frac{1}{r}\frac{d}{dt}\|u\|^r_{L^r(\Omega)} \leq (\varepsilon_1 + \lambda_0 - b) \int_\Omega u^{r+1} - \frac{r+1}{r}\int_\Omega u^r \\
+A_1\lambda_0^{-r}\chi^{r+1} \int_\Omega |\Delta v|^{r+1} + C_1(\varepsilon_1, r) \ \text{ for all } \ t \in (0, T_{\max}).
\end{aligned}
$$

For any $t \in (s_0, T_{\max})$, employing the variation-of-constants formula to the above inequality, we obtain

$$
\begin{aligned}
&\frac{1}{r}\|u(t)\|^r_{L^r(\Omega)} \\
&\leq \frac{1}{r}e^{-(r+1)(t-s_0)}\|u(s_0)\|^r_{L^r(\Omega)} + (\varepsilon_1 + \lambda_0 - b) \int_{s_0}^t e^{-(r+1)(t-s)} \int_\Omega u^{r+1} \\
&\quad +A_1\lambda_0^{-r}\chi^{r+1} \int_{s_0}^t e^{-(r+1)(t-s)} \int_\Omega |\Delta v|^{r+1} + C_1(\varepsilon_1, r) \int_{s_0}^t e^{-(r+1)(t-s)} \\
&\leq (\varepsilon_1 + \lambda_0 - b) \int_{s_0}^t e^{-(r+1)(t-s)} \int_\Omega u^{r+1} \\
&\quad +A_1\lambda_0^{-r}\chi^{r+1} \int_{s_0}^t e^{-(r+1)(t-s)} \int_\Omega |\Delta v|^{r+1} + C_2(r, \varepsilon_1),
\end{aligned}
\tag{12.2.71}
$$

where

$$
C_2 := C_2(r, \varepsilon_1) = \frac{1}{r}\|u(s_0)\|^r_{L^r(\Omega)} + C_1(\varepsilon_1, r) \int_{s_0}^t e^{-(r+1)(t-s)} ds
$$

and s_0 is the same as (12.2.19).

Now, by Lemma 12.1.5, we have

$$
\begin{aligned}
&A_1\lambda_0^{-r}\chi^{r+1} \int_{s_0}^t e^{-(r+1)(t-s)} \int_\Omega |\Delta v|^{r+1} \\
&= A_1\lambda_0^{-r}\chi^{r+1} e^{-(r+1)t} \int_{s_0}^t e^{(r+1)s} \int_\Omega |\Delta v|^{r+1} \\
&\leq A_1\lambda_0^{-r}\chi^{r+1} e^{-(r+1)t} C_{r+1}[\int_{s_0}^t \int_\Omega e^{(r+1)s} u^{r+1} \\
&\quad +e^{(r+1)s_0}\left(\|v(\cdot, s_0)\|^{r+1}_{L^{r+1}(\Omega)} + \|\Delta v(\cdot, s_0)\|^{r+1}_{L^{r+1}(\Omega)}\right)]
\end{aligned}
\tag{12.2.72}
$$

for all $t \in (s_0, T_{\max})$. By substituting (12.2.72) into (12.2.71), using (12.2.69) and Lemma 12.1.6, we get

$$\begin{aligned}
&\frac{1}{r}\|u(t)\|^r_{L^r(\Omega)} \\
&\le \left(\varepsilon_1 + \lambda_0 + A_1\lambda_0^{-r}\chi^{r+1}C_{r+1} - b\right)\int_{s_0}^t e^{-(r+1)(t-s)}\int_\Omega u^{r+1} \\
&\quad + A_1\lambda_0^{-r}\chi^{r+1}e^{-(r+1)(t-s_0)}C_{r+1}\left(\|v(\cdot,s_0)\|^{r+1}_{L^{r+1}(\Omega)} + \|\Delta v(\cdot,s_0)\|^{r+1}_{L^{r+1}(\Omega)}\right) \\
&\quad + C_2(r,\varepsilon_1) = \left(\varepsilon_1 + \frac{(r-1)}{r}C_{r+1}^{\frac{1}{r+1}}\chi - b\right)\int_{s_0}^t e^{-(r+1)(t-s)}\int_\Omega u^{r+1} \\
&\quad + A_1\lambda_0^{-r}\chi^{r+1}e^{-(r+1)(t-s_0)}C_{r+1}\left(\|v(\cdot,s_0)\|^{r+1}_{L^{r+1}(\Omega)} + \|\Delta v(\cdot,s_0)\|^{r+1}_{L^{r+1}(\Omega)}\right) \\
&\quad + C_2(r,\varepsilon_1).
\end{aligned} \tag{12.2.73}$$

Since, $b > \frac{(N-2)_+}{N}\chi C_{\frac{N}{2}+1}^{\frac{1}{\frac{N}{2}+1}}$, we may choose $r := q_0 > \frac{N}{2}$ in (12.2.73) such that

$$b > \frac{q_0 - 1}{q_0}\chi C_{q_0+1}^{\frac{1}{q_0+1}},$$

thus, pick ε_1 appropriating small such that

$$0 < \varepsilon_1 < b - \frac{q_0 - 1}{q_0}\chi C_{q_0+1}^{\frac{1}{q_0+1}},$$

then in light of (12.2.73), we derive that there exists a positive constant C_3 such that

$$\int_\Omega u^{q_0}(x,t)dx \le C_3 \quad \text{for all } \ t \in (s_0, T_{\max}). \tag{12.2.74}$$

Next, we fix $q < \frac{Nq_0}{(N-q_0)^+}$ and choose some $\alpha > \frac{1}{2}$ such that

$$q < \frac{1}{\frac{1}{q_0} - \frac{1}{N} + \frac{2}{N}(\alpha - \frac{1}{2})} \le \frac{Nq_0}{(N-q_0)^+}. \tag{12.2.75}$$

Now, involving the variation-of-constants formula for v, we have

$$v(t) = e^{-\tau(A+1)}v(s_0) + \int_{s_0}^t e^{-(t-s)(A+1)}u(s)ds, \quad t \in (s_0, T_{\max}), \tag{12.2.76}$$

where $A := A_p$ denote the sectorial operator defined by

$$A_p u := -\Delta u \text{ for all } u \in D(A_p) := \{\varphi \in W^{2,p}(\Omega) | \frac{\partial \varphi}{\partial \nu}|_{\partial\Omega} = 0\}.$$

Hence, it follows from (12.2.19) and (12.2.76) that

$$\begin{aligned} &\|(A+1)^{\alpha} v(t)\|_{L^q(\Omega)} \\ &\leq C_4 \int_{s_0}^{t} (t-s)^{-\alpha-\frac{N}{2}(\frac{1}{q_0}-\frac{1}{q})} e^{-b(t-s)} \|u(s)\|_{L^{q_0}(\Omega)} ds \\ &\quad + C_4 s_0^{-\alpha-\frac{N}{2}(1-\frac{1}{q})} \|v(s_0,t)\|_{L^1(\Omega)} \\ &\leq C_4 \int_0^{+\infty} \sigma^{-\alpha-\frac{N}{2}(\frac{1}{q_0}-\frac{1}{q})} e^{-b\sigma} d\sigma + C_4 s_0^{-\alpha-\frac{N}{2}(1-\frac{1}{q})} K, \end{aligned} \tag{12.2.77}$$

where s_0 is the same as (12.2.19). Hence, due to (12.2.75) and (12.2.77), we have

$$\int_{\Omega} |\nabla v(t)|^q \leq C_5 \text{ for all } t \in (s_0, T_{\max}) \tag{12.2.78}$$

and $q \in [1, \frac{Nq_0}{(N-q_0)^+})$. Finally, in view of (12.2.19) and (12.2.78), we can get

$$\int_{\Omega} |\nabla v(t)|^q \leq C_6 \text{ for all } t \in (0, T_{\max}) \text{ and } q \in \left[1, \frac{Nq_0}{(N-q_0)^+}\right). \tag{12.2.79}$$

with some positive constant C_6.

Multiplying both sides of the first equation in (12.2.63) by u^{p-1}, integrating over Ω and integrating by parts, we arrive at

$$\begin{aligned} &\frac{1}{p}\frac{d}{dt}\|u\|^p_{L^p(\Omega)} + (p-1)\int_{\Omega} u^{p-2}|\nabla u|^2 \\ &= -\chi \int_{\Omega} \nabla \cdot (u\nabla v) u^{p-1} + \int_{\Omega} u^{p-1}\left(au - bu^2\right) \\ &= \chi(p-1)\int_{\Omega} u^{p-1}\nabla u \cdot \nabla v + \int_{\Omega} u^{p-1}\left(au - bu^2\right), \end{aligned} \tag{12.2.80}$$

which together with the Young inequality implies that

$$\begin{aligned} &\frac{1}{p}\frac{d}{dt}\|u\|^p_{L^p(\Omega)} + (p-1)\int_{\Omega} u^{p-2}|\nabla u|^2 \\ &\leq \frac{p-1}{2}\int_{\Omega} u^{p-2}|\nabla u|^2 + \frac{\chi^2(p-1)}{2}\int_{\Omega} u^p|\nabla v|^2 - \frac{b}{2}\int_{\Omega} u^{p+1} + C_7 \end{aligned} \tag{12.2.81}$$

for some positive constant C_7. Since, $q_0 > \frac{N}{2}$ yields $q_0 < \frac{Nq_0}{2(N-q_0)^+}$, in light of the Hölder inequality and (12.2.79), we derive at

$$\begin{aligned} \frac{\chi^2(p-1)}{2}\int_\Omega u^p|\nabla v|^2 &\le \frac{\chi^2(p-1)}{2}\left(\int_\Omega u^{\frac{q_0}{q_0-1}p}\right)^{\frac{q_0-1}{q_0}}\left(\int_\Omega |\nabla v|^{2q_0}\right)^{\frac{1}{q_0}} \\ &\le C_8\|u^{\frac{p}{2}}\|^2_{L^{2\frac{q_0}{q_0-1}}(\Omega)}, \end{aligned} \tag{12.2.82}$$

where C_8 is a positive constant. Since $q_0 > \frac{N}{2}$ and $p > q_0 - 1$, we have

$$\frac{q_0}{p} \le \frac{q_0}{q_0-1} \le \frac{N}{N-2},$$

which together with the Gagliardo–Nirenberg inequality implies that

$$\begin{aligned} C_8\left\|u^{\frac{p}{2}}\right\|^2_{L^{2\frac{q_0}{q_0-1}}(\Omega)} &\le C_9\left(\left\|\nabla u^{\frac{p}{2}}\right\|^{b_1}_{L^2(\Omega)}\left\|u^{\frac{p}{2}}\right\|^{1-b_1}_{L^{\frac{2q_0}{p}}(\Omega)} + \left\|u^{\frac{p}{2}}\right\|_{L^{\frac{2q_0}{p}}(\Omega)}\right)^2 \\ &\le C_{10}\left(\left\|\nabla u^{\frac{p}{2}}\right\|^{2b_1}_{L^2(\Omega)} + 1\right) \\ &= C_{10}\left(\left\|\nabla u^{\frac{p}{2}}\right\|^{\frac{2N(p-q_0+1)}{Np+2q_0-Nq_0}}_{L^2(\Omega)} + 1\right) \end{aligned} \tag{12.2.83}$$

with some positive constants C_9, C_{10} and

$$b_1 = \frac{\frac{Np}{2q_0} - \frac{Np}{2\frac{q_0}{q_0-1}p}}{1 - \frac{N}{2} + \frac{Np}{2q_0}} = p\frac{\frac{N}{2q_0} - \frac{N}{2\frac{q_0}{q_0-1}p}}{1 - \frac{N}{2} + \frac{Np}{2q_0}} \in (0, 1).$$

Now, in view of the Young inequality, we derive that

$$\frac{\chi^2(p-1)}{2}\int_\Omega u^p|\nabla v|^2 \le \frac{p-1}{4}\int_\Omega u^{p-2}|\nabla u|^2 + C_{11}. \tag{12.2.84}$$

Inserting (12.2.84) into (12.2.85), we conclude that

$$\frac{1}{p}\frac{d}{dt}\|u\|^p_{L^p(\Omega)} + \frac{p-1}{4}\int_\Omega u^{p-2}|\nabla u|^2 + \frac{b}{2}\int_\Omega u^{p+1} \le C_{12}. \tag{12.2.85}$$

Therefore, letting $y := \int_\Omega u^p$ in (12.2.85) yields to

$$\frac{d}{dt}y(t) + C_{13}y^h(t) \le C_{14} \text{ for all } t \in (0, T_{\max})$$

with some positive constant h. Thus a standard ODE comparison argument implies

$$\|u(\cdot,t)\|_{L^p(\Omega)} \le C_{15} \quad \text{for all} \quad p \ge 1 \quad \text{and} \quad t \in (0, T_{\max}) \tag{12.2.86}$$

for some positive constant C_{15}. The proof Lemma 12.2.8 is complete. □

Our main result on global existence and boundedness thereby becomes a straightforward consequence of Lemma 12.2.1 and Lemma 12.2.8.

The Proof of Theorem 12.2.2 Theorem 12.2.2 will be proved if we can show $T_{\max} = \infty$. Suppose on contrary that $T_{\max} < \infty$. Due to $\|u(\cdot,t)\|_{L^p(\Omega)}$ is bounded for any large p, we infer from the fundamental estimates for Neumann semigroup (see Lemma 4.1 of [39]) or the standard regularity theory of parabolic equation (see, e.g., Ladyzenskaja et al. [48]) that

$$\|\nabla v(\cdot,t)\|_{L^\infty(\Omega)} \le C_1 \quad \text{for all} \quad t \in (0, T_{\max}) \tag{12.2.87}$$

and some positive constant C_1.

Upon an application of the well-known Moser–Alikakos iteration procedure (see Lemma A.1 in [78]), we see that

$$\|u(\cdot,t)\|_{L^\infty(\Omega)} \le C_2 \quad \text{for all} \quad t \in (0, T_{\max}) \tag{12.2.88}$$

and a positive constant C_2.

In view of (12.2.87) and (12.2.88), we apply Lemma 12.2.1 to reach a contradiction. Hence the classical solution (u, v) of (12.2.63) is global in time and bounded. Finally, employing the same arguments as in the proof of Lemma 1.1 in [104], and taking advantage of (12.2.88), we conclude the uniqueness of solution to (12.2.63).

12.3 The (Quasilinear) Chemotaxis System with Consumption of Chemoattractant

In this section, we consider the following parabolic–parabolic Keller–Segel chemotaxis system with nonlinear diffusion and consumption of chemoattractant under homogeneous Neumann boundary conditions

$$\begin{cases} u_t = \nabla \cdot (\phi(u)\nabla u) - \nabla \cdot (u\nabla v), & x \in \Omega, t > 0, \\ v_t = \Delta v - uv, & x \in \Omega, t > 0, \\ \dfrac{\partial u}{\partial \nu} = \dfrac{\partial v}{\partial \nu} = 0, & x \in \partial\Omega, t > 0, \\ u(x,0) = u_0(x), v(x,0) = v_0(x), & x \in \Omega, \end{cases} \tag{12.3.1}$$

where $\Omega \subset \mathbb{R}^N (N \geq 2)$ is a bounded convex domain with smooth boundary $\partial\Omega$, $\frac{\partial}{\partial\nu}$ denotes the outward normal derivative on $\partial\Omega$, the initial data u_0 and v_0 are assumed to be nonnegative functions. $u = u(x, t)$ and $v = v(x, t)$ denote the density of the cell population and the oxygen concentration, respectively.

In this section, we assume that

$$\phi \in C^2([0, \infty)) \tag{12.3.2}$$

as well as

$$\phi(u) \geq C_\phi u^{m-1} \text{ for all } u > 0 \tag{12.3.3}$$

with some $m > 1$ and $C_\phi > 0$. In addition to (12.3.2) and (12.3.3), $D(u)$ satisfies

$$\phi(u) > 0 \text{ for all } u \geq 0. \tag{12.3.4}$$

Definition 12.3.1 Let $T \in (0, \infty)$. A pair of nonnegative functions (u, v) defined in $\Omega \times (0, T)$ is called a weak solution of model (12.3.1), if

1.

$$u \in L^2(0, T; L^2(\Omega)), v \in L^2(0, T; W^{1,2}(\Omega)), D(u)\nabla u \in L^2(0, T; L^2(\Omega)),$$

2.

$$u\nabla v \in L^2(0, T; L^2(\Omega)) \text{ and } uv \in L^2(0, T; L^2(\Omega));$$

the integral equalities

$$-\int_0^T \int_\Omega u\varphi_t - \int_\Omega u_0\varphi(\cdot, 0) = -\int_0^T \int_\Omega D(u)\nabla u \cdot \nabla\varphi + \int_0^T \int_\Omega u\nabla v \cdot \nabla\varphi \tag{12.3.5}$$

and

$$-\int_0^T \int_\Omega v\varphi_t - \int_\Omega v_0\varphi(\cdot, 0) = -\int_0^T \int_\Omega \nabla v \cdot \nabla\varphi + \int_0^T \int_\Omega uv\varphi \tag{12.3.6}$$

hold for all $\varphi \in C_0^\infty(\bar{\Omega} \times [0, T))$. If (u, v) is a weak solution of model (12.3.1) in $\Omega \times (0, T)$ for all $T \in (0, \infty)$, then we call (u, v) a global weak solution.

Theorem 12.3.1 ([131]) *Let $\Omega \subset \mathbb{R}^N (N \geq 2)$ be a smooth bounded convex domain and the diffusion function D satisfy* (12.3.2)–(12.3.4) *with some*

$$m > \frac{3N}{2N + 2}. \tag{12.3.7}$$

Assume that the initial data $u_0(x)$ *and* $v_0(x)$ *satisfy that*

$$\begin{cases} u_0 \in W^{1,l}(\Omega) \ \textit{for all} \ l > N \ \textit{with} \ u_0 > 0 \ \textit{in} \ \bar{\Omega}, \\ v_0 \in W^{1,l}(\Omega) \ \textit{for all} \ l > N \ \textit{with} \ v_0 > 0 \ \textit{in} \ \bar{\Omega}. \end{cases} \tag{12.3.8}$$

Then problem (12.3.1) *possesses a unique global classical solution* (u, v).

Now, we display an important auxiliary interpolation lemma by using the idea which comes from the references [50, 93, 112].

Lemma 12.3.1 *Suppose that* $q > \max\{1, \frac{N-2}{2}\}$ *and* $\Omega \subset \mathbb{R}^N (N \geq 2)$ *is a bounded domain with smooth boundary. Moreover, assume that*

$$\lambda \in [2q + 2, L_{q,N}], \tag{12.3.9}$$

where

$$L_{q,N} \begin{cases} = \frac{N(2q+1)-2(q+1)}{N-2} \ \textit{if} \ N \geq 3, \\ < +\infty \ \textit{if} \ N = 2. \end{cases}$$

Then there exists $C > 0$ *such that for all* $\varphi \in C^2(\bar{\Omega})$ *fulfilling* $\varphi \cdot \frac{\partial \varphi}{\partial \nu} = 0$ *on* $\partial\Omega$ *we have*

$$\|\nabla\varphi\|_{L^\lambda(\Omega)} \leq C \||\nabla\varphi|^{q-1} D^2\varphi\|_{L^2(\Omega)}^{\frac{2(\lambda-N)}{(2q-N+2)\lambda}} \|\varphi\|_{L^\infty(\Omega)}^{\frac{2Nq-(N-2)\lambda}{(2q-N+2)\lambda}} + C\|\varphi\|_{L^\infty(\Omega)}. \tag{12.3.10}$$

Proof An integration by parts together with $\varphi \cdot \frac{\partial\varphi}{\partial\nu} = 0$ on $\partial\Omega$ yields that

$$\|\nabla\varphi\|_{L^\lambda(\Omega)}^\lambda = -\int_\Omega \varphi|\nabla\varphi|^{\lambda-2}\Delta\varphi - (\lambda-2)\int_\Omega \varphi|\nabla\varphi|^{\lambda-4}\nabla\varphi \cdot (D^2\varphi \cdot \nabla\varphi), \tag{12.3.11}$$

which combined with the Cauchy–Schwarz inequality we see that

$$\begin{aligned} &\left|-(\lambda-2)\int_\Omega \varphi|\nabla\varphi|^{\lambda-4}\nabla\varphi \cdot (D^2\varphi \cdot \nabla\varphi)\right| \\ &\leq (\lambda-2)\|\varphi\|_{L^\infty(\Omega)} \cdot I \cdot \left(\int_\Omega |\nabla\varphi|^{2(\lambda-q-1)}\right)^{\frac{1}{2}} \end{aligned} \tag{12.3.12}$$

with $I := \||\nabla\varphi|^{q-1}D^2\varphi\|_{L^2(\Omega)}$. Likewise, using that $|\Delta\varphi| \le \sqrt{N}|D^2\varphi|$ we can estimate

$$\begin{aligned} |-\int_\Omega \varphi|\nabla\varphi|^{\lambda-2}\Delta\varphi| &\le \sqrt{N}\|\varphi\|_{L^\infty(\Omega)}\int_\Omega |\nabla\varphi|^{\lambda-2}|D^2\varphi| \\ &\le \sqrt{N}\|\varphi\|_{L^\infty(\Omega)}\cdot I\cdot\left(\int_\Omega |\nabla\varphi|^{2(\lambda-q-1)}\right)^{\frac{1}{2}}. \end{aligned} \tag{12.3.13}$$

Observing that $\lambda \le \frac{N(2q+1)-2(q+1)}{N-2}$ implies that $\frac{2(\lambda-q-1)}{q} \le \frac{2N}{N-2}$, by $\lambda \ge 2q+2$ and the Gagliardo–Nirenberg inequality, there exist $C_1 > 0$ and $C_2 > 0$ such that

$$\begin{aligned} \left(\int_\Omega |\nabla\varphi|^{2(\lambda-q-1)}\right)^{\frac{1}{2}} &= \||\nabla\varphi|^q\|_{L^{\frac{2(\lambda-q-1)}{q}}(\Omega)}^{\frac{\lambda-q-1}{q}} \\ &\le C_1\|\nabla|\nabla\varphi|^q\|_{L^2(\Omega)}^{\frac{(\lambda-q-1)a}{q}}\||\nabla\varphi|^q\|_{L^{\frac{\lambda}{q}}(\Omega)}^{\frac{(\lambda-q-1)(1-a)}{q}} \\ &\quad +C_1\||\nabla\varphi|^q\|_{L^{\frac{\lambda}{q}}(\Omega)}^{\frac{\lambda-q-1}{q}} \\ &\le C_2\cdot I^{\frac{(\lambda-q-1)a}{q}}\cdot\|\nabla\varphi\|_{L^\lambda(\Omega)}^{(\lambda-q-1)(1-a)} + C_1\|\nabla\varphi\|_{L^\lambda(\Omega)}^{\lambda-q-1} \end{aligned} \tag{12.3.14}$$

with

$$a = \frac{\frac{Nq}{\lambda} - \frac{Nq}{2(\lambda-q-1)}}{1-\frac{N}{2}+\frac{Nq}{\lambda}} \in [0,1].$$

Hence combining (12.3.12)–(12.3.14) yields $C_3 > 0$ fulfilling

$$\begin{aligned} &\|\nabla\varphi\|_{L^\lambda(\Omega)}^\lambda \\ &\le C_3\|\varphi\|_{L^\infty(\Omega)}\cdot I^{1+\frac{(\lambda-q-1)a}{q}}\cdot\|\nabla\varphi\|_{L^\lambda(\Omega)}^{(\lambda-q-1)(1-a)} + C_3\|\varphi\|_{L^\infty(\Omega)}\cdot I\cdot\|\nabla\varphi\|_{L^\lambda(\Omega)}^{\lambda-q-1} \\ &= C_3\|\varphi\|_{L^\infty(\Omega)}\cdot I^{\frac{2(\lambda-N)}{2Nq-(N-2)\lambda}}\cdot\|\nabla\varphi\|_{L^\lambda(\Omega)}^{\frac{\lambda[(2N-2)q-(N-2)(\lambda-1)]}{2Nq-(N-2)\lambda}} \\ &+C_3\|\varphi\|_{L^\infty(\Omega)}I\cdot\|\nabla\varphi\|_{L^\lambda(\Omega)}^{\lambda-q-1}. \end{aligned} \tag{12.3.15}$$

Now, invoking the Young inequality, we find $C_4 > 0$ and $C_5 > 0$ such that

$$\begin{aligned} C_3\|\varphi\|_{L^\infty(\Omega)}\cdot I^{\frac{2(\lambda-N)}{2Nq-(N-2)\lambda}}\cdot\|\nabla\varphi\|_{L^\lambda(\Omega)}^{\frac{\lambda[(2N-2)q-(N-2)(\lambda-1)]}{2Nq-(N-2)\lambda}} &\le \frac{1}{4}\|\nabla\varphi\|_{L^\lambda(\Omega)}^\lambda \\ &+C_4\|\varphi\|_{L^\infty(\Omega)}^{\frac{2Nq-(N-2)\lambda}{2q-N+2}}\cdot I^{\frac{2(\lambda-N)}{2q-N+2}} \end{aligned} \tag{12.3.16}$$

and

$$
\begin{aligned}
&C_3\|\varphi\|_{L^\infty(\Omega)} I \cdot \|\nabla\varphi\|^{\lambda-q-1}_{L^\lambda(\Omega)}\\
&\le \frac{1}{4}\|\nabla\varphi\|^{\lambda}_{L^\lambda(\Omega)} + C_5\|\varphi\|^{\frac{\lambda}{q+1}}_{L^\infty(\Omega)} \cdot I^{\frac{\lambda}{q+1}}\\
&= \frac{1}{4}\|\nabla\varphi\|^{\lambda}_{L^\lambda(\Omega)} + C_5\left(\|\varphi\|^{\frac{2Nq-(N-2)\lambda}{2q-N+2}}_{L^\infty(\Omega)} \cdot I^{\frac{2(\lambda-N)}{2q-N+2}}\right)^{\frac{\lambda(2q-N+2)}{2(\lambda-N)(q+1)}} \|\varphi\|^{\frac{N\lambda(\lambda-2q-2)}{2(q+1)(\lambda-N)}}_{L^\infty(\Omega)}\\
&\le \frac{1}{4}\|\nabla\varphi\|^{\lambda}_{L^\lambda(\Omega)} + C_5\|\varphi\|^{\frac{2Nq-(N-2)\lambda}{2q-N+2}}_{L^\infty(\Omega)} \cdot I^{\frac{2(\lambda-N)}{2q-N+2}} + C_5\|\varphi\|^{\lambda}_{L^\infty(\Omega)}.
\end{aligned}
\tag{12.3.17}
$$

Finally, (12.3.15)–(12.3.17) prove (12.3.10). □

In light of the well-established fixed point arguments (see Wang et al. [99, 101], Lemma 2.1 of [66] and Lemma 2.1 of [113]), we can prove that (12.3.1) is locally solvable in classical sense, which is stated as the following lemma.

Lemma 12.3.2 *Assume that the initial data u_0 and v_0 satisfy (12.3.8) and the diffusion function $D(u)$ satisfies (12.3.2) and (12.3.4). Then there exist $T_{max} \in (0, \infty]$ and a classical solution (u, v) of (12.3.1) in $\Omega \times (0, T_{max})$ such that*

$$
\begin{cases}
u \in C^0(\bar{\Omega} \times [0, T_{max})) \cap C^{2,1}(\bar{\Omega} \times (0, T_{max})),\\
v \in C^0(\bar{\Omega} \times [0, T_{max})) \cap C^{2,1}(\bar{\Omega} \times (0, T_{max})),
\end{cases}
\tag{12.3.18}
$$

classically solving (12.3.1) in $\Omega \times [0, T_{max})$. Moreover, u and v are nonnegative in $\Omega \times (0, T_{max})$, and

$$
\|u(\cdot, t)\|_{L^\infty(\Omega)} + \|v(\cdot, t)\|_{L^\infty(\Omega)} \to \infty \ \ as \ \ t \to T_{max}. \tag{12.3.19}
$$

Lemma 12.3.3 *The solution (u, v) of (12.3.1) satisfies the following properties*

$$
\|u(\cdot, t)\|_{L^1(\Omega)} = \|u_0\|_{L^1(\Omega)} \ \ for\ all \ \ t \in (0, T_{max}) \tag{12.3.20}
$$

and

$$
\|v(\cdot, t)\|_{L^\infty(\Omega)} \le \|v_0\|_{L^\infty(\Omega)} \ \ for\ all \ \ t \in (0, T_{max}). \tag{12.3.21}
$$

12.3.1 A Priori Estimates

In this section, our main purpose is to establish some a priori estimates for the solutions of model (12.3.1). The iteration depends on a series of a priori estimates.

Let us first recall an energy inequality which is a consequence of Lemmas 3.2–3.4 of [107] (see also Lemma 2.3 of [81]).

Lemma 12.3.4 *Assume that the initial data* u_0 *and* v_0 *satisfy* (12.3.8) *and the diffusion function* $D(u)$ *satisfies* (12.3.2)–(12.3.4)*, then the solution of* (12.3.1) *satisfies*

$$\frac{d}{dt}\left\{\int_\Omega u\ln u + 2\int_\Omega |\nabla\sqrt{v}|^2\right\} + \int_\Omega u^{m-2}|\nabla u|^2 + \int_\Omega v|D^2\ln v|^2 + \frac{1}{2}\int_\Omega u\frac{|\nabla v|^2}{v} \le 0 \tag{12.3.22}$$

for all $t \in (0, T_{max})$*. Moreover, for each* $t \in (0, T_{max})$*, one can find a constant* $C > 0$ *such that*

$$\int_0^t\int_\Omega u^{m-2}|\nabla u|^2 \le C \tag{12.3.23}$$

and

$$\int_0^t\int_\Omega |\nabla v|^4 \le C \tag{12.3.24}$$

as well as

$$\int_\Omega |\nabla v(\cdot,t)|^2 \le C. \tag{12.3.25}$$

Proof First, testing the first equation in (12.3.1) by $\ln u$ yields

$$\frac{d}{dt}\left(\int_\Omega u\ln u + 2\int_\Omega |\nabla\sqrt{v}|^2\right) + \int_\Omega \frac{D(u)}{u}|\nabla u|^2 + \int_\Omega v|D^2\ln v|^2 + \frac{1}{2}\int_\Omega u\frac{|\nabla v|^2}{v} = \frac{1}{2}\int_{\partial\Omega}\frac{1}{v}\frac{\partial|\nabla v|^2}{\partial v} \quad \text{for all } t \in (0, T_{\max}). \tag{12.3.26}$$

Since Ω is convex and $\frac{\partial v}{\partial v} = 0$ on $\partial\Omega$, $\frac{\partial|\nabla v|^2}{\partial v} \le 0$ on $\partial\Omega$ (see [78]), so that also using (12.3.3) relation (12.3.22) is shown. Hence, by (12.3.3), we can get (12.3.23). Finally, due to the fact $\int_\Omega \frac{|\nabla v|^4}{v^3} \le (2+\sqrt{N})\int_\Omega v|D^2\ln v|^2$ (see, e.g., Lemma 3.3 of [107]) and $|\nabla\sqrt{v}|^2 = \frac{|\nabla v|^2}{4v}$, by (12.3.21) and (12.3.22), we can derive (12.3.24)–(12.3.25). ◻

Lemma 12.3.5 *Let $p > 1$ and $q > 1$. Assume that the initial data u_0 and v_0 satisfy* (12.3.8) *and the diffusion function $D(u)$ satisfies* (12.3.2)–(12.3.4)*. Then, there exists a constant $C > 0$ such that*

$$\begin{aligned}
&\frac{d}{dt}\left(\|u\|^p_{L^p(\Omega)} + \|\nabla v\|^{2q}_{L^{2q}(\Omega)}\right) + \frac{2C_D p(p-1)}{(m+p-1)^2}\int_\Omega |\nabla u^{\frac{m+p-1}{2}}|^2 \\
&+\frac{2(q-1)}{q}\int_\Omega \left|\nabla|\nabla v|^q\right|^2 + \frac{q}{2}\int_\Omega |\nabla v|^{2q-2}|D^2 v|^2 \\
&\leq C\left(\int_\Omega u^{p+1-m}|\nabla v|^2 + \int_\Omega u^2|\nabla v|^{2q-2}\right) \text{ for all } t \in (0, T_{max}),
\end{aligned} \tag{12.3.27}$$

where $C := C(p, q, N, C_D, \|v_0\|_{L^\infty(\Omega)})$ is a positive constant.

Proof Multiplying the first equation of (12.3.1) by u^{p-1}, combining with the Young inequality, we conclude that

$$\begin{aligned}
&\frac{1}{p}\frac{d}{dt}\|u\|^p_{L^p(\Omega)} + (p-1)\int_\Omega u^{p-2}D(u)|\nabla u|^2 \\
&\leq (p-1)\int_\Omega u^{p-1}|\nabla u||\nabla v| \\
&\leq \frac{(p-1)C_D}{2}\int_\Omega u^{m+p-3}|\nabla u|^2 + \frac{(p-1)}{2C_D}\int_\Omega u^{p+1-m}|\nabla v|^2,
\end{aligned} \tag{12.3.28}$$

which together with (12.3.3) implies that

$$\frac{1}{p}\frac{d}{dt}\|u\|^p_{L^p(\Omega)} + \frac{(p-1)C_D}{2}\int_\Omega u^{m+p-3}|\nabla u|^2 \leq \frac{(p-1)}{2C_D}\int_\Omega u^{p+1-m}|\nabla v|^2. \tag{12.3.29}$$

Observing that $\nabla v \cdot \nabla\Delta v = \frac{1}{2}\Delta|\nabla v|^2 - |D^2 v|^2$, by a straightforward computation using the second equation in (12.3.1) and several integrations by parts, we find that

$$\begin{aligned}
\frac{1}{2q}\frac{d}{dt}\|\nabla v\|^{2q}_{L^{2q}(\Omega)} &= \int_\Omega |\nabla v|^{2q-2}\nabla v \cdot \nabla(\Delta v - uv) \\
&= \frac{1}{2}\int_\Omega |\nabla v|^{2q-2}\Delta|\nabla v|^2 - \int_\Omega |\nabla v|^{2q-2}|D^2 v|^2 \\
&\quad + \int_\Omega uv\nabla\cdot(|\nabla v|^{2q-2}\nabla v) \\
&= -\frac{q-1}{2}\int_\Omega |\nabla v|^{2q-4}\left|\nabla|\nabla v|^2\right|^2 + \frac{1}{2}\int_{\partial\Omega}|\nabla v|^{2q-2}\frac{\partial|\nabla v|^2}{\partial\nu} \\
&\quad - \int_\Omega |\nabla v|^{2q-2}|D^2 v|^2 + \int_\Omega uv|\nabla v|^{2q-2}\Delta v \\
&\quad + \int_\Omega uv\nabla v\cdot\nabla(|\nabla v|^{2q-2})
\end{aligned}$$

$$= -\frac{2(q-1)}{q^2}\int_\Omega \left|\nabla|\nabla v|^q\right|^2 + \frac{1}{2}\int_{\partial\Omega}|\nabla v|^{2q-2}\frac{\partial|\nabla v|^2}{\partial \nu}$$
$$- \int_\Omega |\nabla v|^{2q-2}|D^2 v|^2 + \int_\Omega uv|\nabla v|^{2q-2}\Delta v$$
$$+ \int_\Omega uv\nabla v\cdot\nabla(|\nabla v|^{2q-2}) \tag{12.3.30}$$

for all $t \in (0, T_{\max})$. Now, by $|\Delta v| \le \sqrt{N}|D^2 v|$ and the Young inequality, we can get

$$\int_\Omega uv|\nabla v|^{2q-2}\Delta v \le \sqrt{N}\|v_0\|_{L^\infty(\Omega)}\int_\Omega u|\nabla v|^{2q-2}|D^2 v|$$
$$\le \frac{1}{4}\int_\Omega |\nabla v|^{2q-2}|D^2 v|^2 + N\|v_0\|^2_{L^\infty(\Omega)}\int_\Omega u^2|\nabla v|^{2q-2}. \tag{12.3.31}$$

Now, by the Cauchy–Schwarz inequality, we have

$$\int_\Omega uv\nabla v\cdot\nabla(|\nabla v|^{2q-2}) = (q-1)\int_\Omega uv|\nabla v|^{2(q-2)}\nabla v\cdot\nabla|\nabla v|^2$$
$$\le \frac{q-1}{8}\int_\Omega |\nabla v|^{2q-4}\left|\nabla|\nabla v|^2\right|^2 + 2(q-1)\|v_0\|^2_{L^\infty(\Omega)}$$
$$\times\int_\Omega |u|^2|\nabla v|^{2q-2} \le \frac{(q-1)}{2q^2}\int_\Omega \left|\nabla|\nabla v|^q\right|^2$$
$$+2(q-1)\|v_0\|^2_{L^\infty(\Omega)}\int_\Omega |u|^2|\nabla v|^{2q-2}. \tag{12.3.32}$$

On the other hand, collecting (12.3.29)–(12.3.32) yields (12.3.27) after obvious rearrangements. This completes the proof of Lemma 12.3.5. □

Our main result on global existence and boundedness thereby becomes a straightforward consequence of Lemma 12.2.1 and Lemma 12.3.5. Therefore we omit it.

12.4 The (Quasilinear) Chemotaxis–Haptotaxis Model

In order to describe the cancer invasion mechanism, in 2005, Chaplain and Lolas ([12]) extended the classical Keller–Segel model where, in addition to random diffusion, cancer cells bias their movement towards a gradient of a diffusible matrix-degrading enzyme (MDE) secreted by themselves, as well as a gradient of a static tissue, referred to as extracellular matrix (ECM), by detecting matrix molecules such as vitronectin adhered therein. The latter type of directed migration of cancer cells is usually referred to as haptotaxis (see Chaplain and Lolas [11, 13]). According to the model proposed in [11–13, 36, 72], in this section, we consider the chemotaxis–haptotaxis system with (generalized) logistic source

$$\begin{cases} u_t = \Delta u - \chi \nabla \cdot (u \nabla v) - \xi \nabla \cdot (u \nabla w) + u(1 - u^{r-1} - w), & x \in \Omega, t > 0, \\ v_t = \Delta v + u - v, \quad x \in \Omega, t > 0, \\ w_t = -vw, \quad x \in \Omega, t > 0, \\ \frac{\partial u}{\partial \nu} = \frac{\partial v}{\partial \nu} = \frac{\partial w}{\partial \nu} = 0, \quad x \in \partial\Omega, t > 0, \\ u(x,0) = u_0(x), v(x,0) = v_0(x), w(x,0) = w_0(x), \quad x \in \Omega, \end{cases} \tag{12.4.1}$$

where $r > 1, \Omega \subset \mathbb{R}^N (N \geq 1)$ is a bounded domain with smooth boundary, $\frac{\partial}{\partial \nu}$ denotes the outward normal derivative on $\partial\Omega$, the three variables u, v, and w represent the cancer cell density, the MDE concentration, and the ECM density, respectively. The parameters χ, ξ, and μ are positive which measure the chemotactic, haptotactic sensitivities, and the proliferation rate of the cells, respectively. As is pointed out by [3] (see also Meral et al. [58], Tao and Winkler [77], Zheng [126]), in this modeling context the cancer cells are also usually assumed to follow a generalized logistic growth $u(1-u^{r-1}-w)$ $(r > 1)$, which denotes the proliferation rate of the cells and competing for space with healthy tissue. And the initial data (u_0, v_0, w_0) is supposed to be satisfied the following conditions:

$$\begin{cases} u_0 \in C(\bar{\Omega}) \text{ with } u_0 \geq 0 \text{ in } \Omega \text{ and } u_0 \not\equiv 0, \\ v_0 \in W^{1,\infty}(\Omega) \text{ with } v_0 \geq 0 \text{ in } \Omega, \\ w_0 \in C^{2+\vartheta}(\bar{\Omega}) \text{ with } w_0 \geq 0 \text{ in } \bar{\Omega} \text{ and } \frac{\partial w_0}{\partial \nu} = 0 \text{ on } \partial\Omega \end{cases} \tag{12.4.2}$$

with some $\vartheta \in (0, 1)$.

In order to better understand model (12.4.1), let us mention the following quasi-linear chemotaxis–haptotaxis system, which is a closely related variant of (12.4.1)

$$\begin{cases} u_t = \nabla \cdot (\phi(u) \nabla u) - \chi \nabla \cdot (u \nabla v) - \xi \nabla \cdot (u \nabla w) + \mu u(1 - u^{r-1} - w), & x \in \Omega, t > 0, \\ \tau v_t = \Delta v + u - v, \quad x \in \Omega, t > 0, \\ w_t = -vw, \quad x \in \Omega, t > 0, \\ \frac{\partial u}{\partial \nu} = \frac{\partial v}{\partial \nu} = \frac{\partial w}{\partial \nu} = 0, \quad x \in \partial\Omega, t > 0, \\ u(x,0) = u_0(x), v(x,0) = v_0(x), w(x,0) = w_0(x), \quad x \in \Omega, \end{cases} \tag{12.4.3}$$

where $\mu \geq 0$, $\tau \in \{0, 1\}$, the function $\phi(u)$ fulfills

$$\phi \in C^2([0, \infty)) \tag{12.4.4}$$

and there exist constants $m \geq 1$ and C_ϕ such that

$$\phi(u) \geq C_\phi (u + 1)^{m-1} \text{ for all } u \geq 0. \tag{12.4.5}$$

When $w \equiv 0$, (12.4.3) is reduced to the chemotaxis-only system with (generalized) logistic source (see Xiang [115], Zheng et al. [120, 121, 124, 128, 130, 132]). Global existence, boundedness, and asymptotic behavior of solution were studied in [53, 57, 89, 123]. From a theoretical point of view, due to the fact that the chemotaxis and haptotaxis terms require different L^p-estimate techniques, the problem related to the chemotaxis–haptotaxis models of cancer invasion presents an important mathematical challenging. There are only few results on the mathematical analysis of this (quasilinear) chemotaxis–haptotaxis system (12.4.3) (Cao [8], Zheng et al. [55], Tao et al. [74, 75, 77, 82–84], Wang et al. [55, 90, 92, 130]). Indeed, if MDEs diffuse much faster than cells (see [42, 83]), ($\tau = 0$ in the second equation of (12.4.3)), (12.4.3) is reduced to the parabolic–ODE–elliptic chemotaxis–haptotaxis system, the (generalized) logistic source. To the best of our knowledge, there exist some boundedness and stabilization results on the simplified parabolic–elliptic–ODE chemotaxis–haptotaxis model [75, 82, 83]. When $r = 2$ in the first equation of (12.4.3), the global boundedness of solutions to the chemotaxis–haptotaxis system with the standard logistic source has been proved for any $\mu > 0$ in two dimensions and for large μ (compared to the chemotactic sensitivity χ) in three dimensions (see Tao and Wang [75]). In [83], Tao and Winkler studied the global boundedness for model (12.4.3) under the condition $\boldsymbol{\mu} > \frac{\mathbf{(N-2)^+}}{\mathbf{N}}\boldsymbol{\chi}$, moreover, in addition to explicit smallness on w_0, they gave the exponential decay of w in the large time limit. While if $r > 1$ (the (generalized) logistic source), one can see [130].

As for parabolic–ODE–parabolic system (12.4.3), if $r = 2$, there has been some progress made in two or three dimensions (see Cao [8] Tao and Winkler [73, 75, 83]). In fact, when $\phi \equiv 1$, Cao ([8]) and Tao ([74]) proved that (12.4.3) admits a unique, smooth, and bounded solution if $\mu > 0$ on $N = 2$ and μ is **large enough** on $N = 3$. Recently, assume that μ is **large enough** and $3 \leq N \leq 8$, the boundedness of the global solution of system (12.4.3) is obtained by Wang and Ke in [92]. However, they did not give the lower bound estimation for the logistic source. Note that the global existence and boundedness of solutions to (12.4.3) is still open in three dimensions for **small** $\mu > 0$ and in higher dimensions.

The main objective of the present section is to address the boundedness to solutions of (12.4.1) without any restriction on the **space dimension**. Our main result is the following.

Taking into account all these processes, Tao and Winkler ([77]) proposed the model generalizing the prototypes

$$\begin{cases} u_t = \nabla\cdot(\phi(u)\nabla u) - \chi\nabla\cdot(u\nabla v) - \xi\nabla\cdot(u\nabla w) + \mu u(1-u-w), & x\in\Omega, t>0,\\ v_t = \Delta v + u - v, \quad x\in\Omega, t>0,\\ w_t = -vw, \quad x\in\Omega, t>0,\\ \frac{\partial u}{\partial\nu} = \frac{\partial v}{\partial\nu} = \frac{\partial w}{\partial\nu} = 0, \quad x\in\partial\Omega, t>0,\\ u(x,0)=u_0(x), v(x,0)=v_0(x), w(x,0)=w_0(x), \quad x\in\Omega, \end{cases} \tag{12.4.6}$$

where Ω is a bounded domain in $\mathbb{R}^N (N \geq 2)$ with smooth boundary $\partial\Omega$, $\Delta = \sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2}$, $\frac{\partial}{\partial \nu}$ denotes the outward normal derivative on $\partial\Omega$, $\chi > 0$ is a parameter referred to as chemosensitivity. $\phi(u)$ describes the density-dependent motility of cancer cells through the ECM, and μ is the proliferation rate of the cells. The variables u, v, and w describe the density of the cancer cell population, the concentration of a matrix-degrading enzyme (MDE), and the concentration of extracellular matrix (ECM), respectively. In order to prove our results, we assume that the function $\phi(u)$ fulfill

$$\phi \in C^2([0, \infty)) \text{ for all } u \geq 0 \tag{12.4.7}$$

and there exist constants $m \geq 1$ and C_ϕ such that

$$\phi(u) \geq C_\phi (u+1)^{m-1} \text{ for all } u \geq 0. \tag{12.4.8}$$

Theorem 12.4.1 *Assume that ϕ satisfy (12.4.7)–(12.2.5) and the initial data (u_0, v_0, w_0) fulfills (12.4.2) with some $\vartheta \in (0, 1)$. If $m > \frac{2N}{N+2}$, then there exists a triple $(u, v, w) \in (C^0(\bar{\Omega} \times [0, \infty)) \cap C^{2,1}(\bar{\Omega} \times (0, \infty)))^3$ which solves (12.4.6) in the classical sense. Moreover, u, v, and w are bounded in $\Omega \times (0, \infty)$.*

In this section, we are going to establish an iteration step to develop the main ingredient of our result. The iteration depends on a series of a priori estimates. The following lemma can be proved by Lemma 3.4 of [77] and Lemma 2.2 of [90] (see also Lemma 2.1 of [104] and [127]).

Lemma 12.4.1 *There exists $C > 0$ such that the solution of (12.4.6) satisfies*

$$\int_\Omega u + \int_\Omega |\nabla v|^2 + \int_\Omega |\nabla v|^l \leq C \; \textit{for all} \; t \in (0, T_{max}) \tag{12.4.9}$$

with $l \in [1, \frac{N}{N-1})$.

Remark 12.4.1 The bounded of $\int_\Omega |\nabla v|^2$ plays an important role in getting the boundedness of $\int_\Omega |u|^k + \int_\Omega |\nabla v|^{2\beta}$ for some large k and β.

Next, we are in a position to improve the regularity of u in a higher L^p space. Firstly, we give the following Lemma which plays an important role in obtaining the main results. The proof is almost same as that of Lemma 3.2 in [90]. Consequently, we omit the proof here.

Lemma 12.4.2 *Let (u, v, w) be a solution to (12.4.6) on $(0, T_{max})$. Then for any $k > 1$, there exists a positive constant C_1 which is independent of k such that*

$$-\xi \int_\Omega (u+1)^{k-1} \nabla \cdot (u \nabla w) \leq C_1 \left(\int_\Omega (u+1)^k (v+1) + k \int_\Omega (u+1)^{k-1} |\nabla u| \right). \tag{12.4.10}$$

Proof The third equation of (12.4.6) can be solved according to

$$w(x,t) = w_0(x)e^{-\int_0^t v(x,s)ds}, \quad (x,t) \in \Omega \times (0, T_{\max}), \tag{12.4.11}$$

from which we can obtain

$$\nabla w(x,t) = e^{-\int_0^t v(x,s)ds}(\nabla w_0(x) - w_0(x)\int_0^t \nabla v(x,s)ds \tag{12.4.12}$$

and

$$\begin{aligned}\Delta w(x,t) &\geq \Delta w_0(x)e^{-\int_0^t v(x,s)ds} \\ &\quad - 2e^{-\int_0^t v(x,s)ds} \cdot \int_0^t \nabla v(x,s)ds - w_0(x)e^{-\int_0^t v(x,s)ds}\int_0^t \Delta v(x,s)ds.\end{aligned} \tag{12.4.13}$$

For any $k \geq 1$, we integrate the left-hand side of (12.4.10) and then get

$$\begin{aligned}&-\xi\int_\Omega (u+1)^{k-1}\nabla\cdot(u\nabla w)dx \\ &= (k-1)\xi\int_\Omega (u+1)^{k-2}u\nabla u\cdot\nabla w dx \\ &= (k-1)\xi\int_\Omega \nabla\varphi(u)\cdot\nabla w dx \\ &= -(k-1)\xi\int_\Omega \varphi(u)\Delta w dx \\ &\leq -(k-1)\xi\int_\Omega \varphi(u)(\Delta w_0(x)e^{-\int_0^t v(x,s)ds} - 2e^{-\int_0^t v(x,s)ds}\nabla w_0(x) \\ &\quad \times \textstyle\int_0^t \nabla v(x,s)ds - w_0(x)e^{-\int_0^t v(x,s)ds}\displaystyle\int_0^t \Delta v(x,s)ds)dx \\ &= A_1 + A_2 + A_3,\end{aligned} \tag{12.4.14}$$

where

$$\varphi(u) := \int_0^u \tau(\tau+1)^{k-2}d\tau,$$

$$A_1 := -(k-1)\xi\int_\Omega \varphi(u)\Delta w_0(x)e^{-\int_0^t v(x,s)ds}dx,$$

$$A_2 := 2(k-1)\xi\int_\Omega \varphi(u)e^{-\int_0^t v(x,s)ds}\nabla w_0(x)\cdot\int_0^t \nabla v(x,s)dsdx$$

and

$$A_3 := (k-1)\xi \int_\Omega \varphi(u) w_0(x) e^{-\int_0^t v(x,s)ds} \int_0^t \Delta v(x,s) ds dx.$$

Next, due to $u, v \geq 0$, the Young inequality, we have

$$\begin{aligned} A_1 &= -(k-1)\xi \int_\Omega \varphi(u)\Delta w_0(x) e^{-\int_0^t v(x,s)ds} dx \\ &\leq (k-1)\xi \|\Delta w_0\|_{L^\infty(\Omega)} \int_\Omega \varphi(u) dx \\ &\leq C_0 \int_\Omega (u+1)^{k-1} dx, \end{aligned} \tag{12.4.15}$$

$$\begin{aligned} A_2 &= 2(k-1)\xi \int_\Omega \varphi(u) e^{-\int_0^t v(x,s)ds} \nabla w_0(x) \int_0^t \nabla v(x,s) ds dx \\ &= -2(k-1)\xi \int_\Omega \varphi(u)\nabla w_0(x) \cdot \nabla e^{-\int_0^t v(x,s)ds} dx \\ &= 2(k-1)\xi \int_\Omega \varphi(u)(u+1)^{k-2}\nabla u \cdot \nabla w_0(x) e^{-\int_0^t v(x,s)ds} dx \\ &\quad +2(k-1)\xi \int_\Omega \varphi(u)\Delta w_0(x) e^{-\int_0^t v(x,s)ds} dx \\ &\leq 2C_0(k-1) \int_\Omega (u+1)^{k-1}|\nabla u| dx + 2C_0 \int_\Omega (u+1)^k dx \end{aligned} \tag{12.4.16}$$

and

$$\begin{aligned} A_3 &= (k-1)\xi \int_\Omega \varphi(u) w_0(x) e^{-\int_0^t v(x,s)ds} \int_0^t (v_s(x,s) + v(x,s) - u(x,s)) ds dx \\ &\leq (k-1)\xi \int_\Omega \varphi(u) w_0(x) e^{-\int_0^t v(x,s)ds} (v(x,t) - v_0(x) + \int_0^t v(x,s) ds) dx \\ &\leq C_0 \int_\Omega (u+1)^k v dx + C_0 \int_\Omega (u+1)^k dx. \end{aligned} \tag{12.4.17}$$

Here C_0 depends on $\|\Delta w_0\|_{L^\infty(\Omega)}$, $\|\nabla w_0\|_{L^\infty(\Omega)}$, ξ and $\|w_0\|_{L^\infty(\Omega)}$. Finally, inserting (12.4.15)–(12.4.17) into (12.4.14), we can get the results. □

Lemma 12.4.3 *Let $\Omega \subset \mathbb{R}^N (N \geq 2)$ be a bounded domain with smooth boundary. Furthermore, assume that ϕ satisfies (12.4.7)–(12.4.8) and $m > \frac{2N}{N+2}$. Then for all $k > 1$, there exists $C > 0$ such that*

$$\begin{aligned} &\frac{d}{dt}\left(\frac{1}{k}\|u+1\|^k_{L^k(\Omega)} + \frac{1}{2\beta}\|\nabla v\|^{2\beta}_{L^{2\beta}(\Omega)}\right) + \frac{(\beta-1)}{\beta^2}\int_\Omega \left|\nabla|\nabla v|^\beta\right|^2 + \frac{\mu}{2}\int_\Omega (u+1)^{k+1} \\ &+\frac{1}{2}\int_\Omega |\nabla v|^{2\beta-2}|D^2 v|^2 + \int_\Omega |\nabla v|^{2\beta} + \frac{(k-1)C_\phi}{4}\int_\Omega (u+1)^{m+k-3}|\nabla u|^2 \\ &\leq C\left(\frac{\chi^2(k-1)}{2C_\phi}\int_\Omega (u+1)^{k+1-m}|\nabla v|^2 + \int_\Omega u^2|\nabla v|^{2\beta-2} + \int_\Omega v^{k+1}\right) + C \\ &:= J_1 + J_2 + J_3 + C \ \ \textit{for all} \ \ t \in (0, T_{max}). \end{aligned} \tag{12.4.18}$$

where

$$J_1 := \frac{C\chi^2(k-1)}{2C_\phi}\int_\Omega (u+1)^{k+1-m}|\nabla v|^2,$$

$$J_2 := C\int_\Omega u^2|\nabla v|^{2\beta-2}$$

and

$$J_3 := C\int_\Omega v^{k+1}.$$

Proof Multiplying (12.4.6)$_1$ (the first equation of (12.4.6)) by $(u+1)^{k-1}$ and integrating over Ω, we get

$$\begin{aligned}
&\frac{1}{k}\frac{d}{dt}\|u+1\|^k_{L^k(\Omega)} + (k-1)\int_\Omega (u+1)^{k-2}\phi(u)|\nabla u|^2\\
&= -\chi\int_\Omega \nabla\cdot(u\nabla v)(u+1)^{k-1} - \xi\int_\Omega \nabla\cdot(u\nabla w)(u+1)^{k-1}\\
&\quad +\mu\int_\Omega (u+1)^{k-1}u(1-u-w),
\end{aligned} \tag{12.4.19}$$

which implies that

$$\begin{aligned}
&\frac{1}{k}\frac{d}{dt}\|u+1\|^k_{L^k(\Omega)} + (k-1)\int_\Omega (u+1)^{k-2}\phi(u)|\nabla u|^2\\
&\quad +\mu\int_\Omega (u+1)^{k+1} + 2\mu\int_\Omega (u+1)^{k-1} \le -\chi\int_\Omega \nabla\cdot(u\nabla v)(u+1)^{k-1}\\
&\quad -\xi\int_\Omega \nabla\cdot(u\nabla w)(u+1)^{k-1} + 3\mu\int_\Omega (u+1)^k.
\end{aligned} \tag{12.4.20}$$

By (12.2.5), we obtain

$$(k-1)\int_\Omega (u+1)^{m+k-3}|\nabla u|^2 \le (k-1)\int_\Omega (u+1)^{k-2}\phi(u)|\nabla u|^2. \tag{12.4.21}$$

Integrating by parts to the first term on the right-hand side of (12.4.20) and from the second equation in (12.4.6) we obtain

$$\begin{aligned}
&-\chi\int_\Omega \nabla\cdot(u\nabla v)(u+1)^{k-1}\\
&= (k-1)\chi\int_\Omega u(u+1)^{k-2}\nabla u\cdot\nabla v\\
&\le \frac{(k-1)C_\phi}{2}\int_\Omega (u+1)^{m+k-3}|\nabla u|^2 + \frac{\chi^2(k-1)}{2C_\phi}\int_\Omega (u+1)^{k+1-m}|\nabla v|^2.
\end{aligned} \tag{12.4.22}$$

On the other hand, due to (12.4.2), we have

$$\begin{aligned}
&-\xi \int_\Omega \nabla \cdot (u\nabla w)(u+1)^{k-1} \\
&\leq C_1 \int_\Omega (u+1)^k(v+1) + C_1 k \int_\Omega (u+1)^{k-1}|\nabla u| \\
&\leq C_1 \int_\Omega (u+1)^k(v+1) + \frac{(k-1)C_\phi}{4}\int_\Omega (u+1)^{m+k-3}|\nabla u|^2 \\
&\quad + \frac{2k^2C_1^2}{C_\phi}\int_\Omega (u+1)^{k+1-m}.
\end{aligned} \tag{12.4.23}$$

Furthermore, inserting (12.4.21)–(12.4.23) into (12.4.20), we have

$$\begin{aligned}
&\frac{1}{k}\frac{d}{dt}\|u+1\|^k_{L^k(\Omega)} + \frac{(k-1)C_\phi}{4}\int_\Omega (u+1)^{m+k-3}|\nabla u|^2 + \frac{\mu}{2}\int_\Omega (u+1)^{k+1} \\
&\leq \frac{\chi^2(k-1)}{2C_\phi}\int_\Omega (u+1)^{k+1-m}|\nabla v|^2 + C_1\int_\Omega (u+1)^k(v+1) \\
&\quad + \frac{2k^2C_1^2}{C_\phi}\int_\Omega (u+1)^{k+1-m} + \mu\int_\Omega (u+1)^k.
\end{aligned} \tag{12.4.24}$$

Hence, with the help of the Young inequality, we can derive that

$$\begin{aligned}
&\frac{1}{k}\frac{d}{dt}\|u+1\|^k_{L^k(\Omega)} + \frac{(k-1)C_\phi}{4}\int_\Omega (u+1)^{m+k-3}|\nabla u|^2 + \frac{\mu}{2}\int_\Omega (u+1)^{k+1} \\
&\leq \frac{\chi^2(k-1)}{2C_\phi}\int_\Omega (u+1)^{k+1-m}|\nabla v|^2 + C_2\int_\Omega v^{k+1} + C_2,
\end{aligned} \tag{12.4.25}$$

where $C_2 > 0$ depends on $k, m, C_1, \mu, |\Omega|$, and C_ϕ.

Using that $\nabla v \cdot \nabla\Delta v = \frac{1}{2}\Delta|\nabla v|^2 - |D^2v|^2$, by a straightforward computation using the second equation in (12.4.6) and several integrations by parts, we find that

$$\begin{aligned}
\frac{1}{2\beta}\frac{d}{dt}\|\nabla v\|^{2\beta}_{L^{2\beta}(\Omega)} &= \int_\Omega |\nabla v|^{2\beta-2}\nabla v\cdot\nabla(\Delta v - v + u) \\
&= \frac{1}{2}\int_\Omega |\nabla v|^{2\beta-2}\Delta|\nabla v|^2 - \int_\Omega |\nabla v|^{2\beta-2}|D^2v|^2 \\
&\quad - \int_\Omega |\nabla v|^{2\beta} - \int_\Omega u\nabla\cdot(|\nabla v|^{2\beta-2}\nabla v) \\
&= -\frac{\beta-1}{2}\int_\Omega |\nabla v|^{2\beta-4}\left|\nabla|\nabla v|^2\right|^2 + \frac{1}{2}\int_{\partial\Omega}|\nabla v|^{2\beta-2}\frac{\partial|\nabla v|^2}{\partial\nu} \\
&\quad - \int_\Omega |\nabla v|^{2\beta} - \int_\Omega |\nabla v|^{2\beta-2}|D^2v|^2
\end{aligned}$$

$$
\begin{aligned}
&- \int_\Omega u|\nabla v|^{2\beta-2}\Delta v - \int_\Omega u\nabla v\cdot\nabla(|\nabla v|^{2\beta-2}) \\
=&- \frac{2(\beta-1)}{\beta^2}\int_\Omega \left|\nabla|\nabla v|^\beta\right|^2 \\
&+ \frac{1}{2}\int_{\partial\Omega}|\nabla v|^{2\beta-2}\frac{\partial|\nabla v|^2}{\partial\nu} - \int_\Omega |\nabla v|^{2\beta-2}|D^2v|^2 \\
&- \int_\Omega u|\nabla v|^{2\beta-2}\Delta v - \int_\Omega u\nabla v\cdot\nabla(|\nabla v|^{2\beta-2}) - \int_\Omega |\nabla v|^{2\beta}
\end{aligned}
\tag{12.4.26}
$$

for all $t \in (0, T_{\max})$. Here, since $|\Delta v| \le \sqrt{N}|D^2v|$, by the Young inequality, we can estimate

$$
\begin{aligned}
\int_\Omega u|\nabla v|^{2\beta-2}\Delta v &\le \sqrt{N}\int_\Omega u|\nabla v|^{2\beta-2}|D^2v| \\
&\le \frac{1}{4}\int_\Omega |\nabla v|^{2\beta-2}|D^2v|^2 + N\int_\Omega u^2|\nabla v|^{2\beta-2}
\end{aligned}
\tag{12.4.27}
$$

for all $t \in (0, T_{\max})$. As moreover by the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
-\int_\Omega u\nabla v\cdot\nabla(|\nabla v|^{2\beta-2}) &= -(\beta-1)\int_\Omega u|\nabla v|^{2(\beta-2)}\nabla v\cdot\nabla|\nabla v|^2 \\
&\le \frac{\beta-1}{8}\int_\Omega |\nabla v|^{2\beta-4}\left|\nabla|\nabla v|^2\right|^2 \\
&\quad +2(\beta-1)\int_\Omega |u|^2|\nabla v|^{2\beta-2} \\
&\le \frac{(\beta-1)}{2\beta^2}\int_\Omega \left|\nabla|\nabla v|^\beta\right|^2 + 2(\beta-1)\int_\Omega |u|^2|\nabla v|^{2\beta-2}.
\end{aligned}
\tag{12.4.28}
$$

Next we deal with the integration on $\partial\Omega$. We see from Lemma 12.1.3 that

$$
\begin{aligned}
&\int_{\partial\Omega}\frac{\partial|\nabla v|^2}{\partial\nu}|\nabla v|^{2\beta-2} \\
&\le C_\Omega\int_{\partial\Omega}|\nabla v|^{2\beta} \\
&= C_\Omega||\nabla v|^\beta|^2_{L^2(\partial\Omega)}.
\end{aligned}
\tag{12.4.29}
$$

Let us take $r \in (0, \frac{1}{2})$. By the embedding $W^{r+\frac{1}{2},2}(\Omega) \hookrightarrow L^2(\partial\Omega)$ is compact (see, e.g., Haroske and Triebel [29]), we have

$$
\||\nabla v|^\beta\|^2_{L^2(\partial\Omega)} \le C_3\||\nabla v|^\beta\|^2_{W^{r+\frac{1}{2},2}(\Omega)}. \tag{12.4.30}
$$

Let us pick $a \in (0,1)$ satisfying

$$a = \frac{\frac{1}{2N} + \frac{\beta}{l} + \frac{\gamma}{N} - \frac{1}{2}}{\frac{1}{N} + \frac{\beta}{l} - \frac{1}{2}}.$$

Noting that $\gamma \in (0, \frac{1}{2})$ and $\beta > 1$ imply that $\gamma + \frac{1}{2} \leq a < 1$, we see from the fractional Gagliardo–Nirenberg inequality (Lemma 12.1.3) and boundedness of $|\nabla v|^l$ that

$$\begin{aligned}
&||\nabla v|^{\beta}|^2_{W^{r+\frac{1}{2},2}(\Omega)}\\
&\leq c_0\|\nabla|\nabla v|^{\beta}\|^a_{L^2(\Omega)}\||\nabla v|^{\beta}\|^{1-a}_{L^{\frac{l}{\beta}}(\Omega)} + c_0'\||\nabla v|^{\beta}\|_{L^{\frac{l}{\beta}}(\Omega)}\\
&\leq C_4\|\nabla|\nabla v|^{\beta}\|^a_{L^2(\Omega)} + C_4.
\end{aligned} \tag{12.4.31}$$

Combining (12.4.29) and (12.4.30) with (12.4.31), we obtain

$$\int_{\partial\Omega} \frac{\partial|\nabla v|^2}{\partial\nu}|\nabla v|^{2\beta-2} \leq C_5|\nabla|\nabla v|^{\beta}|^a_{L^2(\Omega)} + C_5. \tag{12.4.32}$$

Now, inserting (12.4.32)–(12.4.28) into (12.4.26) and using the Young inequality we can get

$$\begin{aligned}
&\frac{1}{2\beta}\frac{d}{dt}\|\nabla v\|^{2\beta}_{L^{2\beta}(\Omega)} + \frac{3(\beta-1)}{4\beta^2}\int_{\Omega}\left|\nabla|\nabla v|^{\beta}\right|^2\\
&\quad+\frac{1}{2}\int_{\Omega}|\nabla v|^{2\beta-2}|D^2v|^2 + \int_{\Omega}|\nabla v|^{2\beta}\\
&\leq C_6\int_{\Omega}u^2|\nabla v|^{2\beta-2} + C_6 \text{ for all } t \in (0, T_{\max}),
\end{aligned} \tag{12.4.33}$$

which together with (12.4.25) implies the results. ◻

Lemma 12.4.4 *Let $\Omega \subset \mathbb{R}^N (N \geq 2)$ be a bounded domain with smooth boundary. Furthermore, assume that ϕ satisfies (12.4.7)–(12.2.5) and $m > \frac{2N}{N+2}$. There exist $\bar{\beta} > 1$ and $C > 0$ such that*

$$\|u(\cdot,t)\|_{L^k(\Omega)} + \|\nabla v(\cdot,t)\|_{L^{2\beta}(\Omega)} \leq C \textit{ for all } t \in (0, T_{max}) \tag{12.4.34}$$

for all $\beta \geq \bar{\beta}$ and $k > 1$.

Proof In order to avoid common sense errors, we will prove the Lemma for two cases. Case $N \geq 3$; let

$$\bar{\beta} = \max\{\frac{(N+2)(2m-1)}{N}, \frac{7N+4}{2}, \frac{4(2Nm+2-3N)}{(m-1)(N+2)-(N-2)}\}$$

and

$$k_0(\beta) = \frac{2m-2}{N-2}(N\beta - N + 2) + 2 - \frac{2}{N} - m.$$

Then due to $m > \frac{2N}{N+2}$ and $N > 2$, we have

$$\begin{aligned} \min\{k_0(\beta), 2\beta - 1\} &> \frac{\frac{\beta}{4} - 1}{\frac{\beta}{4} + N\frac{\beta}{8} - N}(N\beta - N + 2) + 2 - \frac{2}{N} - m \\ &\geq \max\left\{1 - m + \frac{N-2}{4N}\beta, m - \frac{2}{N}, 2 - \frac{2}{N} - m\right\} \end{aligned} \tag{12.4.35}$$

for all $\beta \geq \bar{\beta}$. Therefore, we can choose

$$k \in \left(\frac{\frac{\beta}{4} - 1}{\frac{\beta}{4} + N\frac{\beta}{8} - N}(N\beta - N + 2) + 2 - \frac{2}{N} - m, \min\{k_0(\beta), 2\beta - 1\}\right). \tag{12.4.36}$$

Now for the above k, by the Hölder inequality, we have

$$\begin{aligned} J_1 &= \frac{C_6\chi^2(k-1)}{2C_\phi}\int_\Omega (u+1)^{k+1-m}|\nabla v|^2 \\ &\leq \frac{C_6\chi^2(k-1)}{2C_\phi}\left(\int_\Omega (u+1)^{\frac{N}{N-2}(k+1-m)}\right)^{\frac{N-2}{N}}\left(\int_\Omega |\nabla v|^N\right)^{\frac{2}{N}} \\ &= \frac{C_6\chi^2(k-1)}{2C_\phi}\|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{\frac{2N}{N-2}(k+1-m)}{k+m-1}}(\Omega)}^{\frac{2(k+1-m)}{k+m-1}}\|\nabla v\|_{L^N(\Omega)}^2. \end{aligned} \tag{12.4.37}$$

Since, $m \geq 1$ and $N > 2$, we have

$$\frac{1}{k+m-1} \leq \frac{k+1-m}{k+m-1}\frac{\frac{N}{2}}{\frac{N}{2}-1} \leq \frac{N}{N-2},$$

which together with Lemma 12.1.2 implies that

$$\begin{aligned} &\frac{C_6\chi^2(k-1)}{2C_\phi}\|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{\frac{2N}{N-2}(k+1-m)}{k+m-1}}(\Omega)}^{\frac{2(k+1-m)}{k+m-1}} \\ &\leq C_7\Bigg(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|_{L^2(\Omega)}^{\mu_1}\|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{2}{k+m-1}}(\Omega)}^{1-\mu_1} \\ &\quad + \|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{2}{k+m-1}}(\Omega)}\Bigg)^{\frac{2(k+1-m)}{k+m-1}} \leq C_8\left(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|_{L^2(\Omega)}^{\frac{2(k+1-m)\mu_1}{k+m-1}} + 1\right) \end{aligned} \tag{12.4.38}$$

with some positive constants C_{10}, C_8 and

$$\mu_1 = \frac{\frac{N[k+m-1]}{2} - \frac{N(k+m-1)}{\frac{2N}{N-2}(k+1-m)}}{1 - \frac{N}{2} + \frac{N[k+m-1]}{2}} = [k+m-1]\frac{\frac{N}{2} - \frac{N}{\frac{2N}{N-2}(k+1-m)}}{1 - \frac{N}{2} + \frac{N[k+m-1]}{2}} \in (0,1).$$

On the other hand, due to Lemma 12.1.2 and the fact that $\beta \geq \bar{\beta} > \frac{N}{2}$ and $N > 2$, we have

$$\begin{aligned}
\|\nabla v\|^2_{L^N(\Omega)} &= \||\nabla v|^\beta\|^{\frac{2}{\beta}}_{L^{\frac{N}{\beta}}(\Omega)} \\
&\leq C_9\left(\|\nabla|\nabla v|^\beta\|^{\frac{2\mu_2}{\beta}}_{L^2(\Omega)}\||\nabla v|^\beta\|^{\frac{2(1-\mu_2)}{\beta}}_{L^{\frac{2}{\beta}}(\Omega)} + \||\nabla v|^\beta\|^{\frac{2}{\beta}}_{L^{\frac{2}{\beta}}(\Omega)}\right) \\
&\leq C_{10}\left(\|\nabla|\nabla v|^\beta\|^{\frac{2\mu_2}{\beta}}_{L^2(\Omega)} + 1\right),
\end{aligned} \tag{12.4.39}$$

where some positive constants C_9, C_{10} and

$$\mu_2 = \frac{\frac{N\beta}{2} - \frac{N\beta}{N}}{1 - \frac{N}{2} + \frac{N\beta}{2}} = \beta\frac{\frac{N}{2} - \frac{N}{N}}{1 - \frac{N}{2} + \frac{N\beta}{2}} \in (0,1).$$

Inserting (12.4.38)–(12.4.39) into (12.4.49) and using the Young inequality, $\beta \geq \bar{\beta}$ and (12.4.36), we have

$$\begin{aligned}
J_1 &\leq C_{11}\left(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|^{\frac{2(k+1-m)\mu_1}{k+m-1}}_{L^2(\Omega)} + 1\right)\left(\|\nabla|\nabla v|^\beta\|^{\frac{2\mu_2}{\beta}}_{L^2(\Omega)} + 1\right) \\
&= C_{11}\left(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|^{\frac{N(k+1-m-\frac{N-2}{N})}{1-\frac{N}{2}+\frac{N[k+m-1]}{2}}}_{L^2(\Omega)} + 1\right)\left(\|\nabla|\nabla v|^\beta\|^{\frac{N(\frac{2}{2}-\frac{1}{\frac{N}{2}})}{1-\frac{N}{2}+\frac{N\beta}{2}}}_{L^2(\Omega)} + 1\right) \\
&\leq \delta\int_\Omega |\nabla(u+1)^{\frac{m+k-1}{2}}|^2 + \delta\|\nabla|\nabla v|^\beta\|^2_{L^2(\Omega)} + C_{12} \text{ for all } t \in (0, T_{\max}).
\end{aligned} \tag{12.4.40}$$

Next, due to the Hölder inequality and $\beta \geq \bar{\beta} > 8$, we have

$$\begin{aligned}
J_2 &= C_6\int_\Omega u^2|\nabla v|^{2\beta-2} \\
&\leq C_6\left(\int_\Omega u^{\frac{\beta}{4}}\right)^{\frac{8}{\beta}}\left(\int_\Omega |\nabla v|^{\frac{\beta(2\beta-2)}{\beta-8}}\right)^{\frac{\beta-8}{\beta}} \\
&\leq C_6\left(\int_\Omega (u+1)^{\frac{\beta}{4}}\right)^{\frac{8}{\beta}}\left(\int_\Omega |\nabla v|^{\frac{\beta(2\beta-2)}{\beta-8}}\right)^{\frac{\beta-8}{\beta}} \\
&= C_6\|(u+1)^{\frac{k+m-1}{2}}\|^{\frac{4}{k+m-1}}_{L^{\frac{\frac{\beta}{2}}{k+m-1}}(\Omega)}\|\nabla v\|^{(2\beta-2)}_{L^{\frac{\beta(2\beta-2)}{\beta-8}}(\Omega)}.
\end{aligned} \tag{12.4.41}$$

On the other hand, with the help of $k > 1 - m + \frac{N-2}{4N}\beta$, $\beta \geq \bar{\beta} \geq 4$ and Lemma 12.1.2 we conclude that

$$
\begin{aligned}
& C_6 \|(u+1)^{\frac{k+m-1}{2}}\|^{\frac{4}{k+m-1}}_{L^{\frac{\frac{\beta}{2}}{k+m-1}}(\Omega)} \\
& \leq C_{13}\left(\|\nabla (u+1)^{\frac{k+m-1}{2}}\|^{\mu_3}_{L^2(\Omega)} \|(u+1)^{\frac{k+m-1}{2}}\|^{(1-\mu_3)}_{L^{\frac{2}{k+m-1}}(\Omega)}\right. \\
& \quad \left. + \|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{2}{k+m-1}}(\Omega)}\right)^{\frac{4}{k+m-1}} \leq C_{14}\left(\|\nabla (u+1)^{\frac{k+m-1}{2}}\|^{\frac{4\mu_3}{k+m-1}}_{L^2(\Omega)} + 1\right)
\end{aligned} \tag{12.4.42}
$$

with some positive constants C_{13}, C_{14} and

$$
\mu_3 = \frac{\frac{N[k+m-1]}{2} - \frac{N(k+m-1)}{\frac{\beta}{2}}}{1 - \frac{N}{2} + \frac{N[k+m-1]}{2}} = [k+m-1]\frac{\frac{N}{2} - \frac{N}{\frac{\beta}{2}}}{1 - \frac{N}{2} + \frac{N[k+m-1]}{2}} \in (0, 1).
$$

On the other hand, it then follows from $\beta \geq \bar{\beta} > \frac{7N+2}{2}$ and Lemma 12.1.2 that

$$
\begin{aligned}
\|\nabla v\|^{(2\beta-2)}_{L^{\frac{\beta(2\beta-2)}{\beta-8}}(\Omega)} &= \||\nabla v|^{\beta}\|^{\frac{2\beta-2}{\beta}}_{L^{\frac{\frac{\beta(2\beta-2)}{\beta-8}}{\beta}}(\Omega)} \\
&\leq C_{15}(\|\nabla |\nabla v|^{\beta}\|^{\frac{(2\beta-2)\mu_4}{\beta}}_{L^2(\Omega)} \||\nabla v|^{\beta}\|^{\frac{(2\beta-2)(1-\mu_4)}{\beta}}_{L^{\frac{2}{\beta}}(\Omega)} + \||\nabla v|^{\beta}\|^{\frac{(2\beta-2)}{\beta}}_{L^{\frac{2}{\beta}}(\Omega)}) \\
&\leq C_{16}(\|\nabla |\nabla v|^{\beta}\|^{\frac{(2\beta-2)\mu_4}{\beta}}_{L^2(\Omega)} + 1),
\end{aligned} \tag{12.4.43}
$$

where some positive constants C_{15}, C_{16} and

$$
\mu_4 = \frac{\frac{N\beta}{2} - \frac{N\beta}{\frac{\beta(2\beta-2)}{\beta-8}}}{1 - \frac{N}{2} + \frac{N\beta}{2}} = \beta\frac{\frac{N}{2} - \frac{N}{\frac{\beta(2\beta-2)}{\beta-8}}}{1 - \frac{N}{2} + \frac{N\beta}{2}} \in (0, 1).
$$

Inserting (12.4.42)–(12.4.43) into (12.4.41) and using $k > \frac{\frac{\beta}{4}-1}{\frac{\beta}{4}+N\frac{\beta}{8}-N}(N\beta - N + 2) + 2 - \frac{2}{N} - m$ and Lemma 12.1.2, we have

$$
\begin{aligned}
J_2 &\leq C_{17}\left(\|\nabla (u+1)^{\frac{k+m-1}{2}}\|^{\frac{N(2-\frac{8}{\beta})}{1-\frac{N}{2}+\frac{N[k+m-1]}{2}}}_{L^2(\Omega)} + 1\right)\left(\|\nabla |\nabla v|^{\beta}\|^{\frac{N(\frac{2\beta-2}{2}-\frac{\beta-8}{\beta})}{1-\frac{N}{2}+\frac{N\beta}{2}}}_{L^2(\Omega)} + 1\right) \\
&\leq \delta \int_{\Omega} |\nabla (u+1)^{\frac{m+k-1}{2}}|^2 + \delta \|\nabla |\nabla v|^{\beta}\|^2_{L^2(\Omega)} + C_{18} \text{ for all } t \in (0, T_{\max}).
\end{aligned} \tag{12.4.44}
$$

Finally, with the help of (12.4.9) and by the Sobolev inequality, the Young inequality and (12.4.36), we have

$$\begin{aligned} J_3 &= C_6 \int_\Omega v^{k+1} \\ &\le C_{19} \|v\|^{k+1}_{L^\infty(\Omega)} \\ &\le C_{20} \left(\|\nabla v\|^{k+1}_{L^{N+1}(\Omega)} + 1 \right) \\ &\le C_{21} \left(\|\nabla v\|^{k+1}_{L^{2\beta}(\Omega)} + 1 \right) \\ &\le \frac{1}{2} \|\nabla v\|^{2\beta}_{L^{2\beta}(\Omega)} + C_{22} \end{aligned} \tag{12.4.45}$$

with $\beta \ge \bar{\beta} > \frac{N+1}{2}$. Now, inserting (12.4.40), (12.4.44)–(12.4.45) into (12.4.33) and using the Young inequality and choosing δ small enough yields to

$$\begin{aligned} &\frac{d}{dt} \left(\frac{1}{k} \|u+1\|^k_{L^k(\Omega)} + \frac{1}{2\beta} \|\nabla v\|^{2\beta}_{L^{2\beta}(\Omega)} \right) \\ &\quad + \frac{3(\beta-1)}{8\beta^2} \int_\Omega \left| \nabla |\nabla v|^\beta \right|^2 + \frac{\mu}{2} \int_\Omega (u+1)^{k+1} \\ &\quad + \frac{1}{2} \int_\Omega |\nabla v|^{2\beta-2} |D^2 v|^2 + \frac{1}{2} \int_\Omega |\nabla v|^{2\beta} + \frac{(k-1)C_\phi}{8} \int_\Omega (u+1)^{m+k-3} |\nabla u|^2 \\ &\le C_{23} \quad \text{for all} \quad t \in (0, T_{\max}). \end{aligned} \tag{12.4.46}$$

Therefore, letting $y := \int_\Omega (u+1)^k + \int_\Omega |\nabla v|^{2\beta}$ in (12.4.46) yields to

$$\frac{d}{dt} y(t) + C_{24} y(t) \le C_{25} \quad \text{for all} \quad t \in (0, T_{\max}).$$

Thus a standard ODE comparison argument implies boundedness of $y(t)$ for all $t \in (0, T_{\max})$. Clearly, $\|(u+1)(\cdot,t)\|_{L^k(\Omega)}$ and $\|\nabla v(\cdot,t)\|_{L^{2\beta}(\Omega)}$ are bounded for all $t \in (0, T_{\max})$. Obviously, $\lim_{\beta\to+\infty} \frac{\frac{\beta}{4}-1}{\frac{\beta}{4}+N\frac{\beta}{8}-N}(N\beta - N + 2) + 2 - \frac{2}{N} - m = \lim_{\beta\to+\infty} \min\{k_0(\beta), 2\beta - 1\} = +\infty$, hence, the boundedness of $\|(u+1)(\cdot,t)\|_{L^k(\Omega)}$ and the Hölder inequality implies the results.

Case $N = 2$; let $\tilde{\beta} = \max\{2, 2m, m - \frac{1}{2}\}$, then due to $m > 1$, we have

$$(4m-3)\beta + 1 - m > \beta + 1 - m \ge \max\left\{ \frac{7}{3} - m, m - 1 \right\} \quad \text{for all} \quad \beta \ge \tilde{\beta}. \tag{12.4.47}$$

Therefore, we can choose

$$k \in (\beta + 1 - m, (4m-3)\beta + 1 - m). \tag{12.4.48}$$

By the Hölder inequality, we have

$$
\begin{aligned}
J_1 &= \frac{C_6\chi^2(k-1)}{2C_\phi}\int_\Omega (u+1)^{k+1-m}|\nabla v|^2 \\
&\le \frac{C_6\chi^2(k-1)}{2C_\phi}\left(\int_\Omega (u+1)^{2(k+1-m)}\right)^{\frac{1}{2}}\left(\int_\Omega |\nabla v|^4\right)^{\frac{1}{2}} \\
&= \frac{C_6\chi^2(k-1)}{2C_\phi}\|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{4(k+1-m)}{k+m-1}}(\Omega)}^{\frac{2(k+1-m)}{k+m-1}}\|\nabla v\|_{L^4(\Omega)}^2.
\end{aligned}
\tag{12.4.49}
$$

Since $m \ge 1$ and $k \ge m - \frac{1}{2}$, we have

$$
\frac{1}{k+m-1} \le \frac{2(k+1-m)}{k+m-1} < +\infty,
$$

which together with Lemma 12.1.2 implies that

$$
\begin{aligned}
&\frac{C_6\chi^2(k-1)}{2C_\phi}\|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{4(k+1-m)}{k+m-1}}(\Omega)}^{\frac{2(k+1-m)}{k+m-1}} \\
&\le C_{26}\left(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|_{L^2(\Omega)}^{\mu_5}\|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{2}{k+m-1}}(\Omega)}^{1-\mu_5}\right. \\
&\quad \left.+\|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{2}{k+m-1}}(\Omega)}\right)^{\frac{2(k+1-m)}{k+m-1}} \le C_{27}\left(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|_{L^2(\Omega)}^{\frac{2(k+1-m)\mu_5}{k+m-1}}+1\right)
\end{aligned}
\tag{12.4.50}
$$

with some positive constants C_{26}, C_{27} and

$$
\mu_5 = \frac{(k+m-1)-\frac{(k+m-1)}{2(k+1-m)}}{(k+m-1)} = 1 - \frac{1}{2(k+1-m)} \in (0,1).
$$

On the other hand, due to Lemma 12.1.2 and the fact that $\frac{2}{\beta} < +\infty$, we have

$$
\begin{aligned}
\|\nabla v\|_{L^4(\Omega)}^2 &= \||\nabla v|^\beta\|_{L^{\frac{4}{\beta}}(\Omega)}^{\frac{2}{\beta}} \\
&\le C_{28}\left(\|\nabla|\nabla v|^\beta\|_{L^2(\Omega)}^{\frac{2\mu_6}{\beta}}\||\nabla v|^\beta\|_{L^{\frac{2}{\beta}}(\Omega)}^{\frac{2(1-\mu_6)}{\beta}} + \||\nabla v|^\beta\|_{L^{\frac{2}{\beta}}(\Omega)}^{\frac{2}{\beta}}\right) \\
&\le C_{29}\left(\|\nabla|\nabla v|^\beta\|_{L^2(\Omega)}^{\frac{2\mu_6}{\beta}} + 1\right),
\end{aligned}
\tag{12.4.51}
$$

where some positive constants C_{28}, C_{29} and

$$\mu_6 = \frac{\beta - \frac{\beta}{2}}{\beta} = \frac{1}{2} \in (0, 1).$$

Inserting (12.4.50)–(12.4.51) into (12.4.49) and using the Young inequality, we have

$$\begin{aligned} J_1 &\le C_{30}\left(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|_{L^2(\Omega)}^{\frac{2(k+1-m)\mu_5}{k+m-1}} + 1\right)\left(\|\nabla|\nabla v|^{\beta}\|_{L^2(\Omega)}^{\frac{2\mu_6}{\beta}} + 1\right) \\ &= C_{30}\left(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|_{L^2(\Omega)}^{\frac{2(k+1-m-\frac{1}{2})}{(k+m-1)}} + 1\right)\left(\|\nabla|\nabla v|^{\beta}\|_{L^2(\Omega)}^{\frac{2(\frac{2}{2}-\frac{1}{2})}{\beta}} + 1\right) \\ &\le \delta\int_\Omega |\nabla(u+1)^{\frac{m+k-1}{2}}|^2 + \delta\|\nabla|\nabla v|^{\beta}\|_{L^2(\Omega)}^2 + C_{31} \text{ for all } t \in (0, T_{\max}). \end{aligned} \tag{12.4.52}$$

Here we have used the fact that $k < (4m-3)\beta + 1 - m$ and $k > 1 - m$.

Now, due to the Hölder inequality, we have

$$\begin{aligned} J_2 &= C_6\int_\Omega u^2|\nabla v|^{2\beta-2} \\ &\le C_{32}\left(\int_\Omega u^4\right)^{\frac{1}{2}}\left(\int_\Omega |\nabla v|^{2(2\beta-2)}\right)^{\frac{1}{2}} \\ &\le C_{32}\left(\int_\Omega (u+1)^4\right)^{\frac{1}{2}}\left(\int_\Omega |\nabla v|^{2(2\beta-2)}\right)^{\frac{1}{2}} \\ &= C_{32}\|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{8}{k+m-1}}(\Omega)}^{\frac{4}{k+m-1}}\|\nabla v\|_{L^{2(2\beta-2)}(\Omega)}^{(2\beta-2)}. \end{aligned} \tag{12.4.53}$$

On the other hand, with the help of $k > 1$, $m > 1$ and Lemma 12.1.2 we conclude that

$$\begin{aligned} & C_6\|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{8}{k+m-1}}(\Omega)}^{\frac{4}{k+m-1}} \\ &\le C_{33}\left(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|_{L^2(\Omega)}^{\mu_7}\|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{2}{k+m-1}}(\Omega)}^{(1-\mu_7)} \right. \\ &\quad \left. + \|(u+1)^{\frac{k+m-1}{2}}\|_{L^{\frac{2}{k+m-1}}(\Omega)}\right)^{\frac{4}{k+m-1}} \le C_{34}\left(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|_{L^2(\Omega)}^{\frac{4\mu_7}{k+m-1}} + 1\right) \end{aligned} \tag{12.4.54}$$

with some positive constants C_{29}, C_{30} and

$$\mu_7 = \frac{[k+m-1] - \frac{(k+m-1)}{4}}{[k+m-1]} = \frac{3}{4} \in (0, 1).$$

On the other hand, it then follows from $\beta \geq 2$ and Lemma 12.1.2 that

$$\begin{aligned}\|\nabla v\|^{(2\beta-2)}_{L^{2(2\beta-2)}(\Omega)} &= \||\nabla v|^{\beta}\|^{\frac{2\beta-2}{\beta}}_{L^{\frac{2(2\beta-2)}{\beta}}(\Omega)} \\ &\leq C_{35}\left(\|\nabla|\nabla v|^{\beta}\|^{\frac{(2\beta-2)\mu_8}{\beta}}_{L^2(\Omega)}\||\nabla v|^{\beta}\|^{\frac{(2\beta-2)(1-\mu_8)}{\beta}}_{L^{\frac{2}{\beta}}(\Omega)} + \||\nabla v|^{\beta}\|^{\frac{(2\beta-2)}{\beta}}_{L^{\frac{2}{\beta}}(\Omega)}\right) \\ &\leq C_{36}\left(\|\nabla|\nabla v|^{\beta}\|^{\frac{(2\beta-2)\mu_8}{\beta}}_{L^2(\Omega)} + 1\right),\end{aligned} \tag{12.4.55}$$

where some positive constants C_{31}, C_{32} and

$$\mu_8 = \frac{\beta - \frac{\beta}{(2\beta-2)}}{\beta} = \beta\frac{1 - \frac{1}{(2\beta-2)}}{\beta} \in (0, 1).$$

Inserting (12.4.54)–(12.4.55) into (12.4.53) and using $k > \beta + 1 - m$ and Lemma 12.1.2, we have

$$\begin{aligned}J_2 &\leq C_{37}\left(\|\nabla(u+1)^{\frac{k+m-1}{2}}\|^{\frac{2(2-\frac{1}{2})}{[k+m-1]}}_{L^2(\Omega)} + 1\right)\left(\|\nabla|\nabla v|^{\beta}\|^{\frac{2(\beta-1-\frac{1}{2})}{\beta}}_{L^2(\Omega)} + 1\right) \\ &\leq \delta\int_{\Omega}|\nabla(u+1)^{\frac{m+k-1}{2}}|^2 + \delta\|\nabla|\nabla v|^{\beta}\|^2_{L^2(\Omega)} + C_{38} \text{ for all } t \in (0, T_{\max}).\end{aligned} \tag{12.4.56}$$

The same argument as in the derivation of (12.4.45)–(12.4.46) then shows the results. □

Note that $\|\nabla v(\cdot, t)\|_{L^{2\beta}(\Omega)}$ (β is big enough)is bounded by (12.4.34), however $\|\nabla w(\cdot, t)\|_{L^{\infty}(\Omega)}$ might become unbounded in light of (12.4.11). Therefore, Lemma A.1 in [78] cannot be directly applied to the first equation in (12.4.6) to get the boundedness of $\|u(\cdot, t)\|_{L^{\infty}(\Omega)}$. In view of Lemma 12.4.4 and using the standard Moser-type iteration procedure (see [1]), we derive that u is uniformly bounded in $\Omega \times (0, T_{\max})$.

Lemma 12.4.5 *Let $T \in (0, T_{max})$, $\chi > 0$, $\xi > 0$, and assume that u_0, v_0, and w_0 satisfy* (12.4.2) *with some $\vartheta \in (0, 1)$. Then there exists $C > 0$ independent of T such that the solution (u, v, w) of* (12.4.6) *satisfies*

$$\|u(\cdot, t)\|_{L^{\infty}(\Omega)} \leq C \text{ for all } t \in (0, T). \tag{12.4.57}$$

Proof Firstly, due to $\|u(\cdot, t)\|_{L^k(\Omega)}$ is bounded for any large k, by the fundamental estimates for Neumann semigroup (see Lemma 4.1 of [39]) or the standard regularity theory of parabolic equation, we immediately get:

$$\|\nabla v(\cdot,t)\|_{L^\infty(\Omega)} \le C_{35} \text{ for all } t \in (0,T), \tag{12.4.58}$$

where C_{39} is a positive constant independent of T. Hence by (12.4.24) and (12.4.58), we conclude that

$$\begin{aligned}&\frac{1}{k}\frac{d}{dt}\|u+1\|^k_{L^k(\Omega)} + \frac{(k-1)C_\phi}{4}\int_\Omega (u+1)^{m+k-3}|\nabla u|^2 + \frac{\mu}{2}\int_\Omega (u+1)^{k+1}\\ &\le C_{40}k\int_\Omega (u+1)^{k+1-m} + C_1\int_\Omega (u+1)^k + \mu\int_\Omega (u+1)^k,\end{aligned} \tag{12.4.59}$$

where $C_{40} > 0$, as all subsequently appearing constants $C_i (i = 41, 42, \ldots)$ are independent of k as well as of T. Now, according to $m \ge 1$ and $u \ge 0$, an obvious rearrangement implies

$$\frac{d}{dt}\|u+1\|^k_{L^k(\Omega)} + \int_\Omega |\nabla(u+1)^{\frac{k}{2}}|^2 + k\int_\Omega (u+1)^{k+1} \le C_{41}k^2\int_\Omega (u+1)^k. \tag{12.4.60}$$

Therefore by means of the Gagliardo–Nirenberg inequality (see Lemma 12.1.2) and the Young inequality, we can get that

$$\frac{d}{dt}\|u+1\|^k_{L^k(\Omega)} + k\int_\Omega (u+1)^k \le C_{42}k^{2+N}\int_\Omega (u+1)^{\frac{k}{2}}. \tag{12.4.61}$$

Now, choosing $k_i = 2^i (i \in \mathbb{N})$ and letting $M_i = \sup_{t\in(0,T)}\int_\Omega (u+1)^{\frac{k_i}{2}}$. Then (12.4.61) implies that

$$\frac{d}{dt}\|u+1\|^{k_i}_{L^{k_i}(\Omega)} + k_i\int_\Omega (u+1)^{k_i} \le \lambda^i \|(u+1)^{\frac{k_i}{2}}\|^2_{L^1(\Omega)} \tag{12.4.62}$$

with some $\lambda > 1$. Now, if $\lambda^i\|(u+1)^{\frac{k_i}{2}}\|^2_{L^1(\Omega)} \le \|1+u_0\|^{k_i}_{L^\infty(\Omega)}$ for infinitely many $i \ge 1$, we get (12.4.57) with $C = \|1+u_0\|_{L^\infty(\Omega)}$. Otherwise, by a straightforward induction (see, e.g., Lemma 3.12 of [84]) we have

$$\begin{aligned}\int_\Omega (1+u)^{k_i} &\le \lambda^i(\lambda^{i-1}M^2_{i-2})^2\\ &= \lambda^{i+2(i-1)}M^{2^2}_{i-2}\\ &\le \lambda^{i+\Sigma^i_{j=2}(j-1)}M_0^{2^i}.\end{aligned} \tag{12.4.63}$$

Taking k_i-th roots on both sides of (12.4.63), using the fact that $\ln(1+z) \le z$ for all $z \ge 0$, we can easily get (12.4.57). □

Collecting the above Lemmas, we can prove Theorem 12.4.1.

The Proof of Theorem 12.4.1 In view of (12.4.58) and Lemma 12.4.5, we obtain that u and ∇v are uniformly bounded in $\Omega \times (0, T_{\max})$. According to Lemma 12.3.2, this entails that (u, v, w) is global in time, and that u is bounded in $\Omega \times (0, \infty)$.

Lemma 12.4.6 *Under the assumptions in theorem 12.4.1, we derive that there exists a positive constant C such that the solution of* (12.4.1) *satisfies*

$$\int_\Omega u(x,t) + \int_\Omega v^2(x,t) + \int_\Omega |\nabla v(x,t)|^2 \le C \text{ for all } t \in (0, T_{max}). \quad (12.4.64)$$

Moreover, for each $T \in (0, T_{max})$, one can find a constant $C > 0$ independent of ε such that

$$\int_0^T \int_\Omega \left[|\nabla v|^2 + u^r + |\Delta v|^2 \right] \le C. \quad (12.4.65)$$

Now, applying almost exactly the same arguments as in the proof of Lemma 3.2 of [90] (the minor necessary changes are left as an easy exercise to the reader), we conclude the following Lemma:

Lemma 12.4.7 *Let (u, v, w) be a solution to* (12.4.1) *on $(0, T_{max})$. Then for any $k > 1$, there exists a positive constant $C_\beta := C(\xi, \|w_0\|_{L^\infty(\Omega)}, \beta)$ which depends on ξ, $\|w_0\|_{L^\infty(\Omega)}$ and β such that*

$$-\xi \int_\Omega u^{k-1} \nabla \cdot (u \nabla w) \le C_\beta \left(\frac{k-1}{k} \int_\Omega u^k (v+1) + k \int_\Omega u^{k-1} |\nabla u|\right). \quad (12.4.66)$$

where β is the same as (12.2.19).

Proof Here and throughout the proof of Lemma 12.4.7, we shall denote by $M_i (i \in N)$ several positive constants independent of k. Firstly, observing that the third equation of (12.4.1) is an ODE, we derive that

$$w(x,t) = w(x, s_0) e^{-\int_0^t v(x,s)ds}, \quad (x,t) \in \Omega \times (0, T_{\max}). \quad (12.4.67)$$

Hence, by a basic calculation, we conclude that

$$\nabla w(x,t) = \nabla w(x, s_0) e^{-\int_0^t v(x,s)ds} - w(x, s_0) e^{-\int_0^t v(x,s)ds} \int_0^t \nabla v(x,s)ds, \quad (x,t) \in \Omega \times (0, T_{\max}). \quad (12.4.68)$$

and

$$\begin{aligned}
&\Delta w(x,t)\\
&\geq \Delta w(x,s_0)e^{-\int_0^t v(x,s)ds} - 2\nabla w(x,s_0)\cdot\int_0^t \nabla v(x,s)ds e^{-\int_0^t v(x,s)ds}\\
&\quad -w(x,s_0)e^{-\int_0^t \Delta v(x,s)ds}\int_0^t \Delta v(x,s)ds.
\end{aligned} \tag{12.4.69}$$

On the other hand, for any $k \geq 1$, integrating by parts yields

$$\begin{aligned}
&-\xi\int_\Omega u^{k-1}\nabla\cdot(u\nabla w)\\
&= -\xi\frac{k-1}{k}\int_\Omega u^k\Delta w\\
&\leq \xi\frac{k-1}{k}\int_\Omega u^k\Big(-\Delta w(x,s_0)e^{-\int_0^t v(x,s)ds}\\
&\quad +2\nabla w(x,s_0)\cdot\int_0^t \nabla v(x,s)ds e^{-\int_0^t v(x,s)ds}\Big)\\
&\quad +\xi\frac{k-1}{k}\int_\Omega u^k w(x,s_0)e^{-\int_0^t \Delta v(x,s)ds}\int_0^t \Delta v(x,s)ds\\
&:= J_1.
\end{aligned} \tag{12.4.70}$$

Now, using $v \geq 0$ and the Young inequality, we have

$$\begin{aligned}
J_1 &\leq -\xi\frac{k-1}{k}\int_\Omega u^k\Delta w(x,s_0)e^{-\int_0^t v(x,s)ds} + \xi\frac{k-1}{k}\int_\Omega u^k w(x,s_0)e^{-\int_0^t v(x,s)ds}\\
&\quad\times\int_0^t \Delta v(x,s)ds + \xi\frac{2(k-1)}{k}\int_\Omega u^k\nabla w(x,s_0)\cdot\int_0^t \nabla v(x,s)ds e^{-\int_0^t v(x,s)ds}\\
&= -\xi\frac{k-1}{k}\int_\Omega u^k\Delta w(x,s_0)e^{-\int_0^t v(x,s)ds} + \xi\frac{k-1}{k}\int_\Omega u^k w(x,s_0)e^{-\int_0^t v(x,s)ds}\\
&\quad\times\int_0^t \Delta v(x,s)ds - \xi\frac{2(k-1)}{k}\int_\Omega u^k\nabla w(x,s_0)\cdot\nabla e^{-\int_0^t v(x,s)ds}\\
&\leq \xi\beta\int_\Omega u^k + \xi\frac{k-1}{k}\int_\Omega u^k w(x,s_0)e^{-\int_0^t v(x,s)ds}\int_0^t \Delta v(x,s)ds + 2\xi(k-1)\\
&\quad\times\int_\Omega u^{k-1}\nabla u\cdot\nabla w(x,s_0)e^{-\int_0^t v(x,s)ds} + \frac{2(k-1)}{k}\int_\Omega u^k\Delta w(x,s_0)e^{-\int_0^t v(x,s)ds}\\
&\leq M_1\int_\Omega u^k + \xi\frac{k-1}{k}\int_\Omega u^k w(x,s_0)e^{-\int_0^t v(x,s)ds}\int_0^t \Delta v(x,s)ds\\
&\quad +M_2(k\int_\Omega u^{k-1}|\nabla u| + \int_\Omega u^k).
\end{aligned} \tag{12.4.71}$$

Next, due to the second equality of (12.4.1) and $u \geq 0$, we conclude that

$$\begin{aligned}
&\xi\frac{k-1}{k}\int_\Omega u^k w(x,s_0)e^{-\int_0^t v(x,s)ds}\int_0^t \Delta v(x,s)ds\\
&=\xi\frac{k-1}{k}\int_\Omega u^k w(x,s_0)e^{-\int_0^t v(x,s)ds}\int_0^t (v_s(x,s)+v(x,s)-u(x,s)))ds\\
&\leq\xi\frac{k-1}{k}\int_\Omega u^k w(x,s_0)e^{-\int_0^t v(x,s)ds}(v(x,t)-v_0(x)+\int_0^t v(x,s)ds\\
&\leq\xi\frac{k-1}{k}\|w_0\|_{L^\infty(\Omega)}\int_\Omega u^k(v+1).
\end{aligned} \tag{12.4.72}$$

Here we have used the fact that $\frac{t}{e^t} \leq 1$ (for all $t \geq 0$). Collecting (12.4.71) with (12.4.72), we can get the result. □

Lemma 12.4.8 *Let*

$$A_1 = \frac{1}{\delta+1}\left[\frac{\delta+1}{\delta}\right]^{-\delta}\left(\frac{\delta-1}{\delta}\right)^{\delta+1}, \tag{12.4.73}$$

$H(y) = y + A_1 y^{-\delta}\chi^{\delta+1}C_{\delta+1}$ *and* $\tilde{H}(y) = y + A_1 y^{-\delta}C_\beta^{\delta+1}C_{\delta+1}$ *for* $y > 0$. *For any fixed* $\delta \geq 1$, $\chi, C_\beta, C_{\delta+1} > 0$, *Then*

$$\min_{y>0} H(y) = \frac{(\delta-1)}{\delta}C_{\delta+1}^{\frac{1}{\delta+1}}\chi$$

and

$$\min_{y>0}\tilde{H}(y) = \frac{(\delta-1)}{\delta}C_{\delta+1}^{\frac{1}{\delta+1}}C_\beta. \tag{12.4.74}$$

Proof It is easy to verify that

$$H'(y) = 1 - A_1\delta C_{\delta+1}\left(\frac{\chi}{y}\right)^{\delta+1}.$$

Let $H'(y) = 0$, we have

$$y = (A_1 C_{\delta+1}\delta)^{\frac{1}{\delta+1}}\chi.$$

On the other hand, by $\lim_{y\to 0^+} H(y) = +\infty$ and $\lim_{y\to+\infty} H(y) = +\infty$, we have

$$\begin{aligned}
\min\nolimits_{y>0} H(y) &= H\left[(A_1C_{\delta+1}\delta)^{\frac{1}{\delta+1}}\chi\right] = (A_1C_{\delta+1})^{\frac{1}{\delta+1}}(\delta^{\frac{1}{\delta+1}}+\delta^{-\frac{\delta}{\delta+1}})\chi\\
&= \frac{(\delta-1)}{\delta}C_{\delta+1}^{\frac{1}{\delta+1}}\chi.
\end{aligned} \tag{12.4.75}$$

Employing the same arguments as in the proof of (12.4.75), we conclude (12.4.74). □

Lemma 12.4.9 *Let* $r = 2$ *and* (u, v, w) *be a solution to* (12.4.1) *on* $(0, T_{max})$. *If*

$$\mu > \frac{(N-2)_+}{N}(\chi + C_\beta)C_{\frac{N}{2}+1}^{\frac{1}{\frac{N}{2}+1}}, \tag{12.4.76}$$

then for all $p > 1$, *there exists a positive constant* $C := C(p, |\Omega|, \mu, \chi, \xi, \beta)$ *such that*

$$\int_\Omega u^p(x,t)dx \le C \quad \textit{for all} \;\; t \in (0, T_{max}). \tag{12.4.77}$$

Proof Let $l > 1$. Multiplying the first equation of (12.4.1) by u^{l-1} and integrating over Ω, we get

$$\begin{aligned}
&\frac{1}{l}\frac{d}{dt}\|u\|^l_{L^l(\Omega)} + (l-1)\int_\Omega u^{l-2}|\nabla u|^2 dx \\
&= -\chi\int_\Omega \nabla\cdot(u\nabla v)u^{l-1}dx - \xi\int_\Omega \nabla\cdot(u\nabla w)u^{l-1}dx + \int_\Omega u^{l-1}(au - \mu u^2)dx,
\end{aligned} \tag{12.4.78}$$

that is,

$$\begin{aligned}
&\frac{1}{l}\frac{d}{dt}\|u\|^l_{L^l(\Omega)} + (l-1)\int_\Omega u^{l-2}|\nabla u|^2 dx \\
&\le -\frac{l+1}{l}\int_\Omega u^l dx - \chi\int_\Omega \nabla\cdot(u\nabla v)u^{l-1}dx \\
&\quad -\xi\int_\Omega \nabla\cdot(u\nabla w)u^{l-1}dx + \int_\Omega\left(\frac{l+1}{l}u^l + u^{l-1}(au - \mu u^2)\right)dx.
\end{aligned} \tag{12.4.79}$$

Hence, by Young inequality, it reads that

$$\begin{aligned}
&\int_\Omega\left(\frac{l+1}{l}u^l + u^{l-1}(au - \mu u^2)\right)dx \\
&\le \frac{l+1}{l}\int_\Omega u^l dx + a\int_\Omega u^l dx - \mu\int_\Omega u^{l+1}dx \\
&\le (\varepsilon_1 - \mu)\int_\Omega u^{l+1}dx + C_1(\varepsilon_1, l),
\end{aligned} \tag{12.4.80}$$

where

$$C_1(\varepsilon_1, l) = \frac{1}{l+1}\left(\varepsilon_1\frac{l+1}{l}\right)^{-l}\left(\frac{l+1}{l} + a\right)^{l+1}|\Omega|.$$

Next, integrating by parts to the first term on the right-hand side of (12.4.78), we obtain

$$
\begin{aligned}
&-\chi \int_\Omega \nabla \cdot (u\nabla v) u^{l-1} dx \\
&= (l-1)\chi \int_\Omega u^{l-1} \nabla u \cdot \nabla v dx \\
&\leq \frac{l-1}{l} \chi \int_\Omega u^l |\Delta v| dx.
\end{aligned}
\tag{12.4.81}
$$

Next, due to (12.4.66) and the Young inequality, we derive that there exist positive constant $C_2 := (\frac{1}{2}\frac{1}{l-1}C_\beta^2 l^2 + C_\beta)$, and $C_3 := \frac{1}{l+1}(\varepsilon_3 \frac{l+1}{l})^{-l} C_2^{l+1}$ such that

$$
\begin{aligned}
-\xi \int_\Omega \nabla \cdot (u\nabla w) u^{l-1} dx &\leq C_\beta \left(\frac{l-1}{l} \int_\Omega u^l (v+1) + l \int_\Omega u^{l-1} |\nabla u| \right) \\
&\leq \frac{l-1}{2} \int_\Omega u^{l-2} |\nabla u|^2 + \left(\frac{1}{2} \frac{1}{l-1} C_\beta^2 l^2 + \frac{l-1}{l} C_\beta \right) \\
&\quad \times \int_\Omega u^l + C_\beta \frac{l-1}{l} \int_\Omega u^l v \\
&\leq \frac{l-1}{2} \int_\Omega u^{l-2} |\nabla u|^2 + C_2 \int_\Omega u^l + C_\beta \frac{l-1}{l} \int_\Omega u^l v \\
&\leq \frac{l-1}{2} \int_\Omega u^{l-2} |\nabla u|^2 + (\varepsilon_2 + \varepsilon_3) \int_\Omega u^{l+1} \\
&\quad + \frac{1}{l+1} (\varepsilon_2 \frac{l+1}{l})^{-l} C_\beta^{l+1} \left(\frac{l-1}{l} \right)^{l+1} \int_\Omega v^{l+1} + C_3
\end{aligned}
\tag{12.4.82}
$$

for any positive constant ε_2 and ε_3.

Now, let

$$
\lambda_0 := (A_1 C_{l+1} l)^{\frac{1}{l+1}} \chi \tag{12.4.83}
$$

and

$$
\tilde{\lambda}_0 := (A_1 C_{l+1} l)^{\frac{1}{l+1}} C_\beta, \tag{12.4.84}
$$

where A_1 is given by (12.4.73). While from (12.4.81) and the Young inequality, we have

$$
\begin{aligned}
&-\chi \int_\Omega \nabla \cdot (u\nabla v) u^{l-1} dx \\
&\leq \lambda_0 \int_\Omega u^{l+1} dx + \frac{1}{l+1} \left[\lambda_0 \frac{l+1}{l} \right]^{-l} \left(\frac{l-1}{l} \chi \right)^{l+1} \int_\Omega |\Delta v|^{l+1} dx \\
&= \lambda_0 \int_\Omega u^{l+1} dx + A_1 \lambda_0^{-l} \chi^{l+1} \int_\Omega |\Delta v|^{l+1} dx,
\end{aligned}
\tag{12.4.85}
$$

where A_1 is given by (12.4.73). Thus, inserting (12.4.80), (12.4.82), and (12.4.85) into (12.4.79), we get

$$\begin{aligned}\frac{1}{l}\frac{d}{dt}\|u\|^l_{L^l(\Omega)} + \tfrac{l-1}{2}\textstyle\int_\Omega u^{l-2}|\nabla u|^2 &\le (\varepsilon_1+\varepsilon_2+\lambda_0-\mu)\int_\Omega u^{l+1}dx - \frac{l+1}{l}\int_\Omega u^l dx\\ &+A_1\lambda_0^{-l}\chi^{l+1}\int_\Omega|\Delta v|^{l+1}dx\\ &+A_1\varepsilon_2^{-l}C_\beta^{l+1}\int_\Omega v^{l+1}+C_1(\varepsilon_1,l).\end{aligned}$$

For any $t\in(s_0,T_{\max})$, employing the variation-of-constants formula to the above inequality, we obtain

$$\begin{aligned}&\frac{1}{l}\|u(t)\|^l_{L^l(\Omega)}\\ &\le \frac{1}{l}e^{-(l+1)(t-s_0)}\|u(s_0)\|^l_{L^l(\Omega)}+(\varepsilon_1+\varepsilon_2+\lambda_0-\mu)\int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega u^{l+1}dxds\\ &\quad+A_1\lambda_0^{-l}\chi^{l+1}\int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega|\Delta v|^{l+1}dxds\\ &\quad+C_1(\varepsilon_1,l)\int_{s_0}^t e^{-(l+1)(t-s)}ds+A_1\varepsilon_2^{-l}C_\beta^{l+1}\int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega v^{l+1}dxds\\ &\le(\varepsilon_1+\varepsilon_2+\lambda_0-\mu)\int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega u^{l+1}dxds+A_1\lambda_0^{-l}\chi^{l+1}\int_{s_0}^t e^{-(l+1)(t-s)}\\ &\quad\times\int_\Omega|\Delta v|^{l+1}dxds+A_1\varepsilon_2^{-l}C_\beta^{l+1}\int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega v^{l+1}dxds+C_2(\varepsilon_1,l),\end{aligned}\tag{12.4.86}$$

where

$$C_2:=C_2(\varepsilon_1,l)=\frac{1}{l}\|u(s_0)\|^l_{L^l(\Omega)}+C_1(\varepsilon_1,l)\int_{s_0}^t e^{-(l+1)(t-s)}ds.$$

Now, choosing $\varepsilon_2:=\tilde{\lambda}_0$ in (12.4.86), by (12.4.84), Lemma 12.1.5 and the second equation of (12.4.1), we have

$$\begin{aligned}&A_1\lambda_0^{-l}\chi^{l+1}\int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega|\Delta v|^{l+1}dxds\\ &=A_1\lambda_0^{-l}\chi^{l+1}e^{-(l+1)t}\int_{s_0}^t e^{(l+1)s}\int_\Omega|\Delta v|^{l+1}dxds\\ &\le A_1\lambda_0^{-l}\chi^{l+1}e^{-(l+1)t}C_{l+1}\left(\int_{s_0}^t\int_\Omega e^{(l+1)s}u^{l+1}dxds+e^{(l+1)s_0}\|v(s_0,t)\|^{l+1}_{W^{2,l+1}}\right)\end{aligned}\tag{12.4.87}$$

and

$$\begin{aligned}&A_1\tilde{\lambda}_0^{-l}C_\beta^{l+1}\int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega v^{l+1}dxds\\&=A_1\tilde{\lambda}_0^{-l}C_\beta^{l+1}e^{-(l+1)t}\int_{s_0}^t e^{(l+1)s}\int_\Omega v^{l+1}dxds\\&\le A_1\tilde{\lambda}_0^{-l}C_\beta^{l+1}e^{-(l+1)t}C_{l+1}\left(\int_{s_0}^t\int_\Omega e^{(l+1)s}u^{l+1}dxds+e^{(l+1)s_0}\|v(s_0,t)\|_{W^{2,l+1}}^{l+1}\right)\end{aligned} \tag{12.4.88}$$

for all $t \in (s_0, T_{\max})$. By substituting (12.4.87)–(12.4.88) into (12.4.86), using (12.4.83) and Lemma 12.4.8, we get

$$\begin{aligned}&\frac{1}{l}\|u(t)\|_{L^l(\Omega)}^l\\&\le\left(\varepsilon_1+\tilde{\lambda}_0+A_1\tilde{\lambda}_0^{-l}C_\beta^{l+1}C_{l+1}+\lambda_0+A_1\lambda_0^{-l}\chi^{l+1}C_{l+1}-\mu\right)\\&\quad\times\int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega u^{l+1}dxds\\&\quad+A_1\left(\lambda_0^{-l}\chi^{l+1}+\tilde{\lambda}_0^{-l}C_\beta^{l+1}\right)e^{-(l+1)(t-s_0)}C_{l+1}\|v(s_0,t)\|_{W^{2,l+1}}^{l+1}+C_2(l,\varepsilon_1)\\&=\left(\varepsilon_1+\varepsilon_3+\frac{(l-1)}{l}C_{l+1}^{\frac{1}{l+1}}C_\beta+\frac{(l-1)}{l}C_{l+1}^{\frac{1}{l+1}}\chi-\mu\right)\int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega u^{l+1}dxds\\&\quad+A_1\left(\lambda_0^{-l}\chi^{l+1}+\tilde{\lambda}_0^{-l}C_\beta^{l+1}\right)e^{-(l+1)(t-s_0)}C_{l+1}\|v(s_0,t)\|_{W^{2,l+1}}^{l+1}+C_2(l,\varepsilon_1).\end{aligned} \tag{12.4.89}$$

Since $\mu > \frac{(N-2)_+}{N}(C_\beta+\chi)C_{\frac{N}{2}+1}^{\frac{1}{\frac{N}{2}+1}}$, we may choose $\frac{N+2}{(N-2)_+} > q_0 > \frac{N}{2}$ in (12.4.89) such that

$$\mu > \frac{q_0-1}{q_0}(C_\beta+\chi)C_{q_0+1}^{\frac{1}{q_0+1}},$$

thus, pick ε_1 and ε_3 appropriating small such that

$$0<\varepsilon_1+\varepsilon_3<\mu-\frac{q_0-1}{q_0}(C_\beta+\chi)C_{q_0+1}^{\frac{1}{q_0+1}}. \tag{12.4.90}$$

Observe $q_0 < \frac{N+2}{(N-2)_+}$, in light of (12.4.64), we may derive that there exists a positive constant C_4 such that

$$\int_\Omega v^{q_0+1}(x,t)\le C_4 \quad \text{for all } t\in(0,T_{max}). \tag{12.4.91}$$

Collecting (12.4.90) and (12.4.91) then in light of (12.4.89), we derive that there exists a positive constant C_5 such that

$$\int_\Omega u^{q_0}(x,t)dx \le C_5 \quad \text{for all} \quad t \in (s_0, T_{\max}). \tag{12.4.92}$$

Next, we fix $q < \frac{Nq_0}{(N-q_0)^+}$ and choose some $\alpha > \frac{1}{2}$ such that

$$q < \frac{1}{\frac{1}{q_0} - \frac{1}{N} + \frac{2}{N}(\alpha - \frac{1}{2})} \le \frac{Nq_0}{(N-q_0)^+}. \tag{12.4.93}$$

Now, involving the variation-of-constants formula for v, we have

$$v(t) = e^{-\tau(A+1)}v(s_0) + \int_{s_0}^t e^{-(t-s)(A+1)}u(s)ds, \quad t \in (s_0, T_{\max}). \tag{12.4.94}$$

Hence, it follows from (12.2.19) and (12.4.94) that

$$\begin{aligned}
&\|(A+1)^\alpha v(t)\|_{L^q(\Omega)} \\
&\le C_6 \int_{s_0}^t (t-s)^{-\alpha - \frac{N}{2}(\frac{1}{q_0} - \frac{1}{q})} e^{-\mu(t-s)} \|u(s)\|_{L^{q_0}(\Omega)} ds \\
&\quad + C_6 s_0^{-\alpha - \frac{N}{2}(1 - \frac{1}{q})} \|v(s_0, t)\|_{L^1(\Omega)} \\
&\le C_6 \int_0^{+\infty} \sigma^{-\alpha - \frac{N}{2}(\frac{1}{q_0} - \frac{1}{q})} e^{-\mu\sigma} d\sigma + C_6 s_0^{-\alpha - \frac{N}{2}(1 - \frac{1}{q})} \beta.
\end{aligned} \tag{12.4.95}$$

Hence, due to (12.4.93) and (12.4.95), we have

$$\int_\Omega |\nabla v(t)|^q \le C_7 \quad \text{for all} \quad t \in (s_0, T_{\max}) \tag{12.4.96}$$

and $q \in [1, \frac{Nq_0}{(N-q_0)^+})$. Finally, in view of (12.2.19) and (12.4.96), we can get

$$\int_\Omega |\nabla v(t)|^q \le C_8 \quad \text{for all} \quad t \in (0, T_{\max}) \quad \text{and} \quad q \in [1, \frac{Nq_0}{(N-q_0)^+}) \tag{12.4.97}$$

with some positive constant C_8.

Multiplying both sides of the first equation in (12.4.1) by u^{p-1}, integrating over Ω and integrating by parts, we arrive at

$$\begin{aligned}
&\frac{1}{p}\frac{d}{dt}\|u\|^p_{L^p(\Omega)} + (p-1)\int_\Omega u^{p-2}|\nabla u|^2 dx \\
&= -\chi \int_\Omega \nabla \cdot (u\nabla v) u^{p-1} dx - \xi \int_\Omega \nabla \cdot (u \nabla w) u^{p-1} dx + \int_\Omega u^{p-1}(au - \mu u^2) dx \\
&= \chi(p-1)\int_\Omega u^{p-1}\nabla u \cdot \nabla v dx + \xi(p-1)\int_\Omega u^{p-1}\nabla u \cdot \nabla w dx \\
&+ \int_\Omega u^{p-1}(au - \mu u^2) dx,
\end{aligned} \tag{12.4.98}$$

which together with the Young inequality and (12.4.66) implies that

$$\begin{aligned}&\frac{1}{p}\frac{d}{dt}\|u\|^p_{L^p(\Omega)}+(p-1)\int_\Omega u^{p-2}|\nabla u|^2dx\\&\le\frac{p-1}{2}\int_\Omega u^{p-2}|\nabla u|^2dx+\frac{\chi^2(p-1)}{2}\int_\Omega u^p|\nabla v|^2dx+C_9\int_\Omega v^{p+1}\\&-\frac{\mu}{2}\int_\Omega u^{p+1}dx+C_{10}\end{aligned}\tag{12.4.99}$$

for some positive constants C_9 and C_{10}. Now, in light of $q_0>\frac{N}{2}$, due to (12.4.97) and the Sobolev imbedding theorem, we derive that there exists a positive constant C_{11} such that

$$C_9\int_\Omega v^{p+1}(x,t)\le C_{11}\ \text{ for all }\ t\in(0,T_{\max})\ \text{ and }\ p>1.\tag{12.4.100}$$

Since $q_0>\frac{N}{2}$ yields $q_0<\frac{Nq_0}{2(N-q_0)^+}$, in light of the Hölder inequality, (12.2.19) and (12.4.97), we derive that

$$\begin{aligned}\frac{\chi^2(p-1)}{2}\int_\Omega u^p|\nabla v|^2&\le\frac{\chi^2(p-1)}{2}\left(\int_\Omega u^{\frac{q_0}{q_0-1}p}\right)^{\frac{q_0-1}{q_0}}\left(\int_\Omega|\nabla v|^{2q_0}\right)^{\frac{1}{q_0}}\\&\le C_{12}\|u^{\frac{p}{2}}\|^2_{L^{2\frac{q_0}{q_0-1}}(\Omega)},\end{aligned}\tag{12.4.101}$$

where C_{10} is a positive constant. Since $q_0>\frac{N}{2}$ and $p>q_0-1$, we have

$$\frac{q_0}{p}\le\frac{q_0}{q_0-1}\le\frac{N}{N-2},$$

which together with the Gagliardo–Nirenberg inequality implies that

$$\begin{aligned}C_{12}\|u^{\frac{p}{2}}\|^2_{L^{2\frac{q_0}{q_0-1}}(\Omega)}&\le C_{13}\left(\|\nabla u^{\frac{p}{2}}\|^{\mu_1}_{L^2(\Omega)}\|u^{\frac{p}{2}}\|^{1-\mu_1}_{L^{\frac{2q_0}{p}}(\Omega)}+\|u^{\frac{p}{2}}\|_{L^{\frac{2q_0}{p}}(\Omega)}\right)^2\\&\le C_{14}\left(\|\nabla u^{\frac{p}{2}}\|^{2\mu_1}_{L^2(\Omega)}+1\right)\\&=C_{14}\left(\|\nabla u^{\frac{p}{2}}\|^{\frac{2N(p-q_0+1)}{Np+2q_0-Nq_0}}_{L^2(\Omega)}+1\right)\end{aligned}\tag{12.4.102}$$

with some positive constants C_{13}, C_{14} and

$$\mu_1=\frac{\frac{Np}{2q_0}-\frac{Np}{2\frac{q_0}{q_0-1}p}}{1-\frac{N}{2}+\frac{Np}{2q_0}}=p\frac{\frac{N}{2q_0}-\frac{N}{2\frac{q_0}{q_0-1}p}}{1-\frac{N}{2}+\frac{Np}{2q_0}}\in(0,1).$$

Now, in view of the Young inequality, we derive that

$$\frac{\chi^2(p-1)}{2}\int_\Omega u^p|\nabla v|^2dx \le \frac{p-1}{4}\int_\Omega u^{p-2}|\nabla u|^2dx + C_{15}. \tag{12.4.103}$$

Inserting (12.4.103) into (12.4.104), we conclude that

$$\frac{1}{p}\frac{d}{dt}\|u\|^p_{L^p(\Omega)} + \frac{p-1}{4}\int_\Omega u^{p-2}|\nabla u|^2dx + \frac{\mu}{2}\int_\Omega u^{p+1}dx \le C_{16}. \tag{12.4.104}$$

Therefore, integrating the above inequality with respect to t yields

$$\|u(\cdot,t)\|_{L^p(\Omega)} \le C_{17} \text{ for all } p \ge 1 \text{ and } t \in (0, T_{\max}) \tag{12.4.105}$$

for some positive constant C_{17}. The proof Lemma 12.4.9 is complete. □

Lemma 12.4.10 *Let* (u, v, w) *be a solution to* (12.4.1) *on* $(0, T_{max})$. *Assume that* $r > 2$. *Then for all* $p > 1$, *there exists a positive constant* $C := C(p, |\Omega|, r, \mu, \xi, \chi, \beta)$ *such that*

$$\int_\Omega u^p(x,t)dx \le C \quad \textit{for all } t \in (0, T_{max}). \tag{12.4.106}$$

Proof Firstly, multiplying the first equation of (12.4.1) by u^{l-1} and integrating over Ω, we get

$$\begin{aligned}&\frac{1}{l}\frac{d}{dt}\|u\|^l_{L^l(\Omega)} + (l-1)\int_\Omega u^{l-2}|\nabla u|^2dx\\&= -\chi\int_\Omega \nabla\cdot(u\nabla v)u^{l-1}dx - \xi\int_\Omega \nabla\cdot(u\nabla w)u^{l-1}dx + \int_\Omega u^{l-1}(au-\mu u^r)dx.\end{aligned} \tag{12.4.107}$$

In light of the Young inequality and $r > 2$, it reads that there exists a positive constant C_1 such that

$$\begin{aligned}&\int_\Omega\left(\frac{l+1}{l}u^l + u^{l-1}(au-\mu u^r)\right)dx\\&\le \frac{l+1}{l}\int_\Omega u^l dx + a\int_\Omega u^l dx - \mu\int_\Omega u^{l+r-1}dx\\&\le -\frac{7\mu}{8}\int_\Omega u^{l+r-1}dx + C_1.\end{aligned} \tag{12.4.108}$$

Next, integrating by parts to the first term on the right-hand side of (12.4.107) and using the Young inequality, we obtain

$$\begin{aligned} &-\chi \int_\Omega \nabla\cdot(u\nabla v)u^{l-1}dx \\ &\le \frac{l-1}{l}\chi \int_\Omega u^l|\Delta v|dx \\ &\le \frac{\mu}{8}\int_\Omega u^{l+r-1}dx + C_2\int_\Omega |\Delta v|^{\frac{r+l-1}{r-1}}dx \\ &\le \frac{\mu}{8}\int_\Omega u^{l+r-1}dx + \int_\Omega |\Delta v|^{l+1}dx + C_3. \end{aligned} \tag{12.4.109}$$

Next, due to (12.4.66) and the Young inequality, we derive that there exist positive constant C_4, C_5 and C_6 such that

$$\begin{aligned} -\xi\int_\Omega \nabla\cdot(u\nabla w)u^{l-1}dx &\le C_\beta\left(\frac{l-1}{l}\int_\Omega u^l(v+1) + l\int_\Omega u^{l-1}|\nabla u|\right) \\ &\le C_4\left(\int_\Omega u^l(v+1) + l\int_\Omega u^{l-1}|\nabla u|\right) \\ &\le \frac{l-1}{2}\int_\Omega u^{l-2}|\nabla u|^2 + C_5\int_\Omega u^l + C_5\int_\Omega u^l v \\ &\le \frac{l-1}{2}\int_\Omega u^{l-2}|\nabla u|^2 + \frac{\mu}{8}\int_\Omega u^{l+r-1} + \int_\Omega v^{l+1} + C_6. \end{aligned} \tag{12.4.110}$$

Thus, inserting (12.4.108)–(12.4.85) into (12.4.107), we get

$$\begin{aligned} \frac{1}{l}\frac{d}{dt}\|u\|^l_{L^l(\Omega)} + \tfrac{l-1}{2}\textstyle\int_\Omega u^{l-2}|\nabla u|^2 \le -\frac{5\mu}{8}\int_\Omega u^{l+r-1}dx - \frac{l+1}{l}\int_\Omega u^l dx \\ + \int_\Omega |\Delta v|^{l+1}dx + \int_\Omega v^{l+1} + C_7. \end{aligned} \tag{12.4.111}$$

For any $t\in(s_0,T_{\max})$, apply the variation-of-constants formula to (12.4.111), we get

$$\begin{aligned} &\frac{1}{l}\|u(t)\|^l_{L^l(\Omega)} \\ &\le \frac{1}{l}e^{-(l+1)(t-s_0)}\|u(s_0)\|^l_{L^l(\Omega)} - \frac{5\mu}{8}\int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega u^{l+r-1}dxds \\ &\quad + \int_{s_0}^t e^{-(l+1)(t-s)}\int_\Omega |\Delta v|^{l+1}dxds + C_7\int_{s_0}^t e^{-(l+1)(t-s)}ds \end{aligned}$$

$$
\begin{aligned}
&+\int_{s_0}^{t} e^{-(l+1)(t-s)} \int_{\Omega} v^{l+1} dxds \\
&\leq -\frac{5\mu}{8} \int_{s_0}^{t} e^{-(l+1)(t-s)} \int_{\Omega} u^{l+r-1} dxds \\
&+\int_{s_0}^{t} e^{-(l+1)(t-s)} \int_{\Omega} \left(|\Delta v|^{l+1} + v^{l+1}\right) dxds + C_8,
\end{aligned}
\tag{12.4.112}
$$

where

$$
C_8 := \frac{1}{l} \|u(s_0)\|_{L^l(\Omega)}^{l} + C_7 \int_{s_0}^{t} e^{-(l+1)(t-s)} ds.
$$

Now, by Lemma 12.1.5, we have

$$
\begin{aligned}
&\int_{s_0}^{t} e^{-(l+1)(t-s)} \int_{\Omega} \left(|\Delta v|^{l+1} + |v|^{l+1}\right) dxds \\
&= e^{-(l+1)t} \int_{s_0}^{t} e^{(l+1)s} \int_{\Omega} \left(|\Delta v|^{l+1} + |v|^{l+1}\right) dxds \\
&\leq e^{-(l+1)t} C_{l+1} \left(\int_{s_0}^{t} \int_{\Omega} e^{(l+1)s} u^{l+1} dxds + e^{(l+1)s_0} \|v(s_0, t)\|_{W^{2,l+1}}^{l+1} \right)
\end{aligned}
\tag{12.4.113}
$$

for all $t \in (s_0, T_{\max})$. By substituting (12.4.113) into (12.4.112) and using the Young inequality, we get

$$
\begin{aligned}
&\frac{1}{l} \|u(t)\|_{L^l(\Omega)}^{l} \\
&\leq -\frac{5\mu}{8} \int_{s_0}^{t} e^{-(l+1)(t-s)} \int_{\Omega} u^{l+r-1} dxds + C_{l+1} \int_{s_0}^{t} e^{-(l+1)(t-s)} \int_{\Omega} u^{l+1} dxds \\
&+ e^{-(l+1)(t-s_0)} C_{l+1} \|v(s_0, t)\|_{W^{2,l+1}}^{l+1} + C_8 \\
&\leq -\frac{\mu}{2} \int_{s_0}^{t} e^{-(l+1)(t-s)} \int_{\Omega} u^{l+r-1} dxds + C_9
\end{aligned}
\tag{12.4.114}
$$

with

$$
\begin{aligned}
C_9 = {} & e^{-(l+1)(t-s_0)} C_{l+1} \|v(s_0, t)\|_{W^{2,l+1}}^{l+1} + \frac{r-2}{r+l-1} \left(\frac{\mu}{8} \frac{r+l-1}{l+1} \right)^{-\frac{l+1}{r-2}} \\
& \times C_{l+1}^{\frac{r+l-1}{r-2}} \frac{\mu}{8} |\Omega| \int_{s_0}^{t} e^{-(l+1)(t-s)} ds + C_8.
\end{aligned}
$$

Therefore, integrating (12.4.114) with respect to t and using (12.2.19) yields

$$\|u(\cdot,t)\|_{L^l(\Omega)} \le C_{11} \text{ for all } l \ge 1 \text{ and } t \in (0, T_{\max}) \tag{12.4.115}$$

for some positive constant C_{11}. The proof Lemma 12.4.9 is complete. □

Our main result on global existence and boundedness thereby becomes a straightforward consequence of Lemmata 12.4.9–12.4.10 and Lemma 12.3.2. Indeed, collecting the above Lemmata, in the following, by invoking a Moser-type iteration (see Lemma A.1 in [78]) and the standard estimate for Neumann semigroup (or the standard parabolic regularity arguments), we will prove Theorem 12.4.1.

The Proof of Theorem 12.4.1 Firstly, due to Lemmata 12.4.9–12.4.10, we derive that there exist positive constants $q_0 > N$ and C_1 such that

$$\|u(\cdot,t)\|_{L^{q_0}(\Omega)} \le C_1 \text{ for all } t \in (0, T_{\max}). \tag{12.4.116}$$

Next, employing the standard estimate for Neumann semigroup provides C_2 and $C_3 > 0$ such that

$$\begin{aligned}
&\|\nabla v(t)\|_{L^\infty(\Omega)}\\
&\le C_2 \int_{s_0}^{t} (t-s)^{-\alpha-\frac{N}{2q_0}} e^{-\mu(t-s)} \|u(s)\|_{L^{q_0}(\Omega)} ds + C_2 s_0^{-\alpha} \|v(s_0,t)\|_{L^\infty(\Omega)}\\
&\le C_2 \int_0^{+\infty} \sigma^{-\alpha-\frac{N}{2q_0}} e^{-\mu\sigma} d\sigma + C_2 s_0^{-\alpha} \beta\\
&\le C_3 \text{ for all } t \in (0, T_{\max}).
\end{aligned} \tag{12.4.117}$$

Multiplying both sides of the first equation in (12.4.1) by u^{p-1}, integrating over Ω and integrating by parts, we conclude that

$$\begin{aligned}
&\frac{1}{p}\frac{d}{dt}\|u\|^p_{L^p(\Omega)} + (p-1)\int_\Omega u^{p-2}|\nabla u|^2 dx\\
&= -\chi \int_\Omega \nabla\cdot(u\nabla v) u^{p-1} dx - \xi \int_\Omega \nabla\cdot(u\nabla w) u^{p-1} dx + \int_\Omega u^{p-1}(au - \mu u^r) dx\\
&= \chi(p-1)\int_\Omega u^{p-1}\nabla u\cdot\nabla v dx - \xi \int_\Omega \nabla\cdot(u\nabla w) u^{p-1} dx\\
&\quad + \int_\Omega u^{p-1}(au-\mu u^r) dx.
\end{aligned} \tag{12.4.118}$$

Due to (12.4.66) and (12.4.117) and the Young inequality, we derive that there exist positive constants C_4, C_5, C_6 and C_7 independent of p such that

$$\begin{aligned}
\chi(p-1)\int_\Omega u^{p-1}\nabla u\cdot\nabla v dx &\le \chi(p-1)C_4 \int_\Omega u^{p-1}|\nabla u| dx\\
&\le \frac{p-1}{4}\int_\Omega u^{p-2}|\nabla u|^2 + C_5 p \int_\Omega u^p
\end{aligned} \tag{12.4.119}$$

and

$$-\xi \int_\Omega u^{p-1}\nabla\cdot(u\nabla w) \le C_6\left(\int_\Omega u^p(v+1) + p\int_\Omega u^{p-1}|\nabla u|\right) \le \frac{p-1}{4}\int_\Omega u^{p-2}|\nabla u|^2 + C_7 p\int_\Omega u^p. \quad (12.4.120)$$

Hence by (12.4.118)–(12.4.120), we conclude that there exist positive constants C_8 and C_9 independent of p such that

$$\frac{d}{dt}\|u\|^p_{L^p(\Omega)} + C_8\int_\Omega |\nabla u^{\frac{p}{2}}|^2 dx + \int_\Omega u^p \le C_9 p^2 \int_\Omega u^p. \quad (12.4.121)$$

Here and throughout the proof of Theorem 12.4.1, we shall denote by $C_i (i \in \mathbb{N})$ several positive constants independent of p. Next, with the help of the Gagliardo–Nirenberg inequality, we derive that

$$\begin{aligned} C_9 p^2\int_\Omega u^p &= C_9 p^2 \|u^{\frac{p}{2}}\|^2_{L^2(\Omega)} \\ &\le C_9 p^2\left(\|\nabla u^{\frac{p}{2}}\|^{2\varsigma_1}_{L^2(\Omega)}\|u^{\frac{p}{2}}\|^{2(1-\varsigma_1)}_{L^1(\Omega)} + \|u^{\frac{p}{2}}\|^2_{L^1(\Omega)}\right) \\ &= C_9 p^2\left(\|\nabla u^{\frac{p}{2}}\|^{\frac{2N}{N+2}}_{L^2(\Omega)}\|u^{\frac{p}{2}}\|^{\frac{4}{N+2}}_{L^1(\Omega)} + \|u^{\frac{p}{2}}\|^2_{L^1(\Omega)}\right) \\ &\le C_9 p^2\left(\|\nabla u^{\frac{p}{2}}\|^{\frac{2N}{N+2}}_{L^2(\Omega)}\|u^{\frac{p}{2}}\|^{\frac{4}{N+2}}_{L^1(\Omega)} + \|u^{\frac{p}{2}}\|^2_{L^1(\Omega)}\right) \\ &\le C_8\|\nabla u^{\frac{p}{2}}\|^2_{L^2(\Omega)} + C_{10}p^{N+2}\|u^{\frac{p}{2}}\|^2_{L^1(\Omega)} + C_9 p^2\|u^{\frac{p}{2}}\|^2_{L^1(\Omega)} \\ &\le C_8\|\nabla u^{\frac{p}{2}}\|^2_{L^2(\Omega)} + C_{11}p^{N+2}\|u^{\frac{p}{2}}\|^2_{L^1(\Omega)}, \end{aligned} \quad (12.4.122)$$

where

$$0 < \varsigma_1 = \frac{N - \frac{N}{2}}{1 - \frac{N}{2} + N} = \frac{N}{N+2} < 1,$$

C_{10} and C_{11} are positive constants independent of p. Therefore, inserting (12.4.122) into (12.4.121), we derive that

$$\frac{d}{dt}\|u\|^p_{L^p(\Omega)} + \int_\Omega u^p \le C_{11}p^{2+N}\|u^{\frac{p}{2}}\|^2_{L^1(\Omega)} \le C_{11}p^{2+N}\left(\max\{1, \|u^{\frac{p}{2}}\|^2_{L^1(\Omega)}\right)^2. \quad (12.4.123)$$

Now, choosing $p_i = 2^i$ and letting $M_i = \max\{1, \sup_{t\in(0,T)} \int_\Omega u^{\frac{p_i}{2}}\}$ for $T \in (0, T_{max})$ and $i = 1, 2, \cdots$. Then (12.4.123) implies that

$$\frac{d}{dt}\|u\|_{L^{p_i}(\Omega)}^{p_i} + \int_\Omega u^{p_i} \le C_{11} p_i^{2+N} M_{i-1}^2(T), \tag{12.4.124}$$

which together with the comparison argument entails that there exists a $\lambda > 1$ independent of i such that

$$M_i(T) \le \max\{\lambda^i M_{i-1}^2(T), |\Omega|\|u_0\|_{L^\infty(\Omega)}^{p_i}\}. \tag{12.4.125}$$

Now, if $\lambda^i M_{i-1}^2(T) \le |\Omega|\|u_0\|_{L^\infty(\Omega)}^{p_i}$ for infinitely many $i \ge 1$, we get

$$\|u(\cdot, t)\|_{L^\infty(\Omega)} \le C \quad \text{for all} \quad t \in (0, T) \tag{12.4.126}$$

with $C = \|u_0\|_{L^\infty(\Omega)}$. Otherwise, if $\lambda^i M_{i-1}^2(T) > |\Omega|\|u_0\|_{L^\infty(\Omega)}^{p_i}$ for all sufficiently large i, then by (12.4.125), we derive that

$$M_i(T) \le \lambda^i M_{i-1}^2(T) \quad \text{for all sufficiently large} \quad i. \tag{12.4.127}$$

Hence, we may choose λ large enough such that

$$M_i(T) \le \lambda^i M_{i-1}^2(T) \quad \text{for all} \quad i \ge 1. \tag{12.4.128}$$

Therefore, in light of a straightforward induction (see, e.g., Lemma 3.12 of [84]) we have

$$\begin{aligned} M_i(T) &\le \lambda^i \left(\lambda^{i-1} M_{i-2}^2\right)^2 \\ &= \lambda^{i+2(i-1)} M_{i-2}^{2^2} \\ &\le \lambda^{i+\Sigma_{j=2}^i (j-1)} M_0^{2^i}. \end{aligned} \tag{12.4.129}$$

Taking p_i-th roots on both sides of (12.4.129), with some basic calculation and by taking $T \nearrow T_{\max}$, we can finally conclude that

$$\|u(\cdot, t)\|_{L^\infty(\Omega)} \le C \quad \text{for all} \quad t \in (0, T_{\max}). \tag{12.4.130}$$

Now, with the above estimate in hand, using (12.4.68), we may establish

$$\|\nabla w(\cdot, t)\|_{L^\infty(\Omega)} \le C \ \text{ for all } \ t \in (0, T_{\max}). \tag{12.4.131}$$

Finally, according to Lemma 12.3.2, this together with (12.4.117) and (12.4.130) entails that (u, v, w) is global in time, and that u is bounded in $\Omega \times (0, \infty)$. □

12.5 The (Quasilinear) Keller–Segel–Navier–Stokes System

Chemotaxis is a biological process in which cells move toward a chemically more favorable environment (see Hillen and Painter [35]). In 1970, Keller and Segel (see Keller and Segel [43, 44]) proposed a mathematical model for chemotaxis phenomena through a system of parabolic equations (see, e.g., Winkler et al. [3, 39, 108], Osaki and Yagi [64], Horstmann [37]). To describe chemotaxis of cell populations, the signal is produced by the cells, an important variant of the quasilinear chemotaxis model

$$\begin{cases} n_t = \nabla \cdot (D(n)\nabla n) - \chi \nabla \cdot (S(n)\nabla c), \\ c_t = \Delta c - c + n \end{cases} \tag{12.5.1}$$

was initially proposed by Painter and Hillen ([66], see also Winkler et al. [3, 80]) where n denotes the cell density and c describes the concentration of the chemical signal secreted by cells. The function S measures the chemotactic sensitivity, which may depend on n, $D(n)$ is the diffusion function. The results about the chemotaxis model (12.5.1) appear to be rather complete, which dealt with the problem (12.5.1) whether the solutions are global bounded or blow-up (see Cieślak et al. [14, 15, 18], Hillen [35], Horstmann et al. [38], Ishida et al. [41], Kowalczyk [47], Winkler et al. [78, 108, 114]). In fact, Tao and Winkler ([78]), proved that the solutions of (12.5.1) are global and bounded provided that $\frac{S(n)}{D(n)} \leq c(n+1)^{\frac{2}{N}+\varepsilon}$ for all $n \geq 0$ with some $\varepsilon > 0$ and $c > 0$, and $D(n)$ satisfies some another technical conditions. For the more related works in this direction, we mention that a corresponding quasilinear version, the logistic damping or the signal is consumed by the cells has been deeply investigated by Cieślak and Stinner [15, 17], Tao and Winkler [78, 103, 114] and Zheng et al. [120, 121, 129, 131].

In various situations, however, the migration of bacteria is furthermore substantially affected by changes in their environment (see Winkler et al. [3, 85]). As in the quasilinear Keller–Segel system (12.5.1) where the chemoattractant is produced by cells, the corresponding chemotaxis–fluid model is then quasilinear Keller–Segel–Navier-Stokes system of the form

$$\begin{cases} n_t + u \cdot \nabla n = \nabla \cdot (D(n)\nabla n) - \nabla \cdot (S(n)\nabla c), \quad x \in \Omega, t > 0, \\ c_t + u \cdot \nabla c = \Delta c - c + n, \quad x \in \Omega, t > 0, \\ u_t + \kappa (u \cdot \nabla) u + \nabla P = \Delta u + n \nabla \phi, \quad x \in \Omega, t > 0, \\ \nabla \cdot u = 0, \quad x \in \Omega, t > 0, \end{cases} \tag{12.5.2}$$

where n and c are denoted as before, u and P stand for the velocity of incompressible fluid and the associated pressure, respectively. ϕ is a given potential function and $\kappa \in \mathbb{R}$ denotes the strength of nonlinear fluid convection. Problem (12.5.2) is proposed to describe chemotaxis–fluid interaction in cases when the evolution of the chemoattractant is essentially dominated by production through cells ([3, 35]).

If the signal is consumed, rather than produced, by the cells, Tuval et al. ([87]) proposed the following model

$$\begin{cases} n_t + u \cdot \nabla n = \nabla \cdot (D(n)\nabla n) - \nabla \cdot (nS(c)\nabla c), & x \in \Omega, t > 0, \\ c_t + u \cdot \nabla c = \Delta c - nf(c), & x \in \Omega, t > 0, \\ u_t + \kappa(u \cdot \nabla)u + \nabla P = \Delta u + n\nabla\phi, & x \in \Omega, t > 0, \\ \nabla \cdot u = 0, & x \in \Omega, t > 0. \end{cases} \tag{12.5.3}$$

Here $f(c)$ is the consumption rate of the oxygen by the cells. Approaches based on a natural energy functional, the (quasilinear) chemotaxis-(Navier-)Stokes system (12.5.3) have been studied in the last few years and the main focus is on the solvability result (see, e.g., Chae et al. [10], Francesco et al. [21], Duan et al. [23, 24], Liu and Lorz [54, 56], Tao and Winkler [85, 107, 110, 113], Zhang and Zheng [118] and references therein). For instance, if $\kappa = 0$ in (12.5.3), the model is simplified to the chemotaxis-Stokes equation. In [104], Winkler showed the global weak solutions of (12.5.3) in bounded three-dimensional domains. Other variants of the model of (12.5.3) that include porous medium-type diffusion and S being a chemotactic sensitivity tensor, one can see Winkler ([112]) and Zheng ([119]) and the references therein for details.

In contrast to problem (12.5.3), the mathematical analysis of the Keller–Segel–Stokes system (12.5.2) ($\kappa = 0$) is quite few (Black [4], Wang et al. [52, 94, 95]). Among these results, Wang et al. [94, 95] proved the global boundedness of solutions to the two-dimensional and three-dimensional Keller-Segel–Stokes system (12.5.2) when S is a tensor satisfying some dampening condition with respect to n. However, for the three-dimensional fully Keller–Segel-Navier-Stokes system (12.5.2) ($\kappa \in \mathbb{R}$), to the best our knowledge, there is no result on global solvability. Motivated by the above works, we will investigate the interaction of the fully quasilinear Keller–Segel–Navier–Stokes in this section. Precisely, we shall consider the following initial-boundary problem

$$\begin{cases} n_t + u \cdot \nabla n = \Delta n^m - \nabla \cdot (n\nabla c), & x \in \Omega, t > 0, \\ c_t + u \cdot \nabla c = \Delta c - c + n, & x \in \Omega, t > 0, \\ u_t + \kappa(u \cdot \nabla)u + \nabla P = \Delta u + n\nabla\phi, & x \in \Omega, t > 0, \\ \nabla \cdot u = 0, & x \in \Omega, t > 0, \\ \nabla n \cdot \nu = \nabla c \cdot \nu = 0, u = 0, & x \in \partial\Omega, t > 0, \\ n(x,0) = n_0(x), c(x,0) = c_0(x), u(x,0) = u_0(x), & x \in \Omega, \end{cases} \tag{12.5.4}$$

where $\Omega \subseteq \mathbb{R}^3$ is a bounded domain with smooth boundary.

12.5.1 Preliminaries and Theorems

Due to the strongly nonlinear term $(u \cdot \nabla)u$ and Δn^m, the problem (12.5.4) has no classical solutions in general, and thus we consider its weak solutions in the following sense. We first specify the notion of weak solution to which we will refer in the sequel.

Definition 12.5.1 Let $T > 0$ and (n_0, c_0, u_0) fulfills (12.3.8). Then a triple of functions (n, c, u) is called a weak solution of (12.5.4) if the following conditions are satisfied

$$\begin{cases} n \in L^1_{loc}(\bar{\Omega} \times [0, T)), \\ c \in L^1_{loc}([0, T); W^{1,1}(\Omega)), \\ u \in L^1_{loc}([0, T); W^{1,1}(\Omega)), \end{cases} \tag{12.5.5}$$

where $n \geq 0$ and $c \geq 0$ in $\Omega \times (0, T)$ as well as $\nabla \cdot u = 0$ in the distributional sense in $\Omega \times (0, T)$, moreover,

$$\begin{array}{l} u \otimes u \in L^1_{loc}(\bar{\Omega} \times [0, \infty); \mathbb{R}^{3\times 3}) \text{ and } n^m \text{ belong to } L^1_{loc}(\bar{\Omega} \times [0, \infty)), \\ cu, \ \ nu \text{ and } n|\nabla c| \text{ belong to } L^1_{loc}(\bar{\Omega} \times [0, \infty); \mathbb{R}^3) \end{array} \tag{12.5.6}$$

and

$$-\int_0^T \int_\Omega n\varphi_t - \int_\Omega n_0\varphi(\cdot, 0) = \int_0^T \int_\Omega n^m \Delta\varphi + \int_0^T \int_\Omega n\nabla c \cdot \nabla\varphi + \int_0^T \int_\Omega nu \cdot \nabla\varphi \tag{12.5.7}$$

for any $\varphi \in C_0^\infty(\bar{\Omega} \times [0, T))$ satisfying $\frac{\partial\varphi}{\partial\nu} = 0$ on $\partial\Omega \times (0, T)$ as well as

$$-\int_0^T \int_\Omega c\varphi_t - \int_\Omega c_0\varphi(\cdot, 0) = -\int_0^T \int_\Omega \nabla c \cdot \nabla\varphi - \int_0^T \int_\Omega c\varphi + \int_0^T \int_\Omega n\varphi + \int_0^T \int_\Omega cu \cdot \nabla\varphi \tag{12.5.8}$$

for any $\varphi \in C_0^\infty(\bar{\Omega} \times [0, T))$ and

$$-\int_0^T \int_\Omega u\varphi_t - \int_\Omega u_0\varphi(\cdot, 0) - \kappa \int_0^T \int_\Omega u \otimes u \cdot \nabla\varphi = -\int_0^T \int_\Omega \nabla u \cdot \nabla\varphi - \int_0^T \int_\Omega n\nabla\phi \cdot \varphi \tag{12.5.9}$$

for any $\varphi \in C_0^\infty(\bar{\Omega}\times[0,T);\mathbb{R}^3)$ fulfilling $\nabla\varphi \equiv 0$ in $\Omega\times(0,T)$. If $\Omega\times(0,\infty) \longrightarrow \mathbb{R}^5$ is a weak solution of (12.5.4) in $\Omega\times(0,T)$ for all $T>0$, then we call (n,c,u) a global weak solution of (12.5.4).

Throughout this paper, we assume that

$$\phi \in W^{1,\infty}(\Omega) \tag{12.5.10}$$

and the initial data (n_0, c_0, u_0) fulfills

$$\begin{cases} n_0 \in C^\kappa(\bar{\Omega}) \ \text{ for certain } \ \kappa > 0 \ \text{ with } \ n_0 \geq 0 \ \text{ in } \ \Omega, \\ c_0 \in W^{1,\infty}(\Omega) \ \text{ with } \ c_0 \geq 0 \ \text{ in } \ \bar{\Omega}, \\ u_0 \in D(A_r^\gamma) \ \text{ for } \ \text{ some } \ \gamma \in (\frac{1}{2}, 1) \ \text{ and any } \ r \in (1,\infty), \end{cases} \tag{12.5.11}$$

where A_r denotes the Stokes operator with domain $D(A_r) := W^{2,r}(\Omega)\cap W_0^{1,r}(\Omega)\cap L_\sigma^r(\Omega)$, and $L_\sigma^r(\Omega) := \{\varphi \in L^r(\Omega) | \nabla\cdot\varphi = 0\}$ for $r \in (1,\infty)$ ([71]).

Theorem 12.5.1 *Let $\Omega \subset \mathbb{R}^3$ be a smooth bounded domain. Assume that* (12.5.10) *holds. If*

$$m > 2, \tag{12.5.12}$$

then for any choice of n_0, c_0 and u_0 fulfilling (12.5.11)*, the problem* (12.5.4) *possesses at least one global weak solution (n, c, u, P) in the sense of Definition 12.5.1.*

Remark 12.5.1 From Theorem 12.5.1, we conclude that if the exponent m of nonlinear diffusion is larger than 2, then model (12.5.4) exists a global solution, which implies the nonlinear diffusion term benefits the global of solutions, which seems partly extends the results of Tao and Winkler [85], who proved the possibility of boundedness, in the case that $m = 1$, the coefficient of logistic source suitably large and the strength of nonlinear fluid convection $\kappa = 0$.

Our intention is to construct a global weak solution of (12.5.4) as the limit of smooth solutions of appropriately regularized problems. To this end, in order to deal with the strongly nonlinear term $(u\cdot\nabla)u$ and Δn^m, we need to introduce the following approximating equation of (12.5.4):

$$\begin{cases} n_{\varepsilon t} + u_\varepsilon\cdot\nabla n_\varepsilon = \Delta(n_\varepsilon+\varepsilon)^m - \nabla\cdot(n_\varepsilon\nabla c_\varepsilon), \quad x\in\Omega, t>0, \\ c_{\varepsilon t} + u_\varepsilon\cdot\nabla c_\varepsilon = \Delta c_\varepsilon - c_\varepsilon + n_\varepsilon, \quad x\in\Omega, t>0, \\ u_{\varepsilon t} + \nabla P_\varepsilon = \Delta u_\varepsilon - \kappa(Y_\varepsilon u_\varepsilon\cdot\nabla)u_\varepsilon + n_\varepsilon\nabla\phi, \quad x\in\Omega, t>0, \\ \nabla\cdot u_\varepsilon = 0, \quad x\in\Omega, t>0, \\ \nabla n_\varepsilon\cdot\nu = \nabla c_\varepsilon\cdot\nu = 0, u_\varepsilon = 0, \quad x\in\partial\Omega, t>0, \\ n_\varepsilon(x,0) = n_0(x), c_\varepsilon(x,0) = c_0(x), u_\varepsilon(x,0) = u_0(x), \quad x\in\Omega, \end{cases} \tag{12.5.13}$$

where

$$Y_\varepsilon w := (1+\varepsilon A)^{-1} w \quad \text{for all} \ \ w \in L^2_\sigma(\Omega) \tag{12.5.14}$$

is the standard Yosida approximation. In light of the well-established fixed point arguments (see [112], Lemma 2.1 of [66] and Lemma 2.1 of [113]), we can prove that (12.5.13) is locally solvable in classical sense, which is stated as the following lemma.

Lemma 12.5.1 *Assume that* $\varepsilon \in (0,1)$. *Then there exist* $T_{max,\varepsilon} \in (0,\infty]$ *and a classical solution* $(n_\varepsilon, c_\varepsilon, u_\varepsilon, P_\varepsilon)$ *of* (12.5.13) *in* $\Omega \times (0, T_{max,\varepsilon})$ *such that*

$$\begin{cases} n_\varepsilon \in C^0(\bar\Omega \times [0, T_{max,\varepsilon})) \cap C^{2,1}(\bar\Omega \times (0, T_{max,\varepsilon})), \\ c_\varepsilon \in C^0(\bar\Omega \times [0, T_{max,\varepsilon})) \cap C^{2,1}(\bar\Omega \times (0, T_{max,\varepsilon})), \\ u_\varepsilon \in C^0(\bar\Omega \times [0, T_{max,\varepsilon})) \cap C^{2,1}(\bar\Omega \times (0, T_{max,\varepsilon})), \\ P_\varepsilon \in C^{1,0}(\bar\Omega \times (0, T_{max,\varepsilon})), \end{cases} \tag{12.5.15}$$

classically solving (12.5.13) *in* $\Omega \times [0, T_{max,\varepsilon})$. *Moreover,* n_ε *and* c_ε *are nonnegative in* $\Omega \times (0, T_{max,\varepsilon})$, *and*

$$\|n_\varepsilon(\cdot,t)\|_{L^\infty(\Omega)} + \|c_\varepsilon(\cdot,t)\|_{W^{1,\infty}(\Omega)} + \|A^\gamma u_\varepsilon(\cdot,t)\|_{L^2(\Omega)} \to \infty \ \ \text{as} \ \ t \to T_{max,\varepsilon}, \tag{12.5.16}$$

where γ *is given by* (12.5.11).

12.5.2 A Priori Estimates

In the following, we are going to establish an iteration step to develop the main ingredient of our result. The iteration depends on a series of a priori estimate. The proof of this lemma is very similar to that of Lemmata 2.2 and 2.6 of [85], so we omit its proof here.

Lemma 12.5.2 *There exists* $\lambda > 0$ *independent of* ε *such that the solution of* (12.5.13) *satisfies*

$$\int_\Omega n_\varepsilon + \int_\Omega c_\varepsilon \le \lambda \ \ \text{for all} \ \ t \in (0, T_{max,\varepsilon}). \tag{12.5.17}$$

Lemma 12.5.3 *Let* $m > 2$. *Then there exists* $C > 0$ *independent of* ε *such that the solution of* (12.5.13) *satisfies*

$$\int_\Omega (n_\varepsilon + \varepsilon)^{m-1} + \int_\Omega c_\varepsilon^2 + \int_\Omega |u_\varepsilon|^2 \le C \ \ \text{for all} \ \ t \in (0, T_{max,\varepsilon}). \tag{12.5.18}$$

In addition, for each $T \in (0, T_{max,\varepsilon})$, one can find a constant $C > 0$ independent of ε such that

$$\int_0^T \int_\Omega \left[(n_\varepsilon + \varepsilon)^{2m-4}|\nabla n_\varepsilon|^2 + |\nabla c_\varepsilon|^2 + |\nabla u_\varepsilon|^2\right] \leq C. \tag{12.5.19}$$

Proof Taking c_ε as the test function for the second equation of (12.5.13) and using $\nabla \cdot u_\varepsilon = 0$ and the Young inequality yields that

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|c_\varepsilon\|^2_{L^2(\Omega)} + \int_\Omega |\nabla c_\varepsilon|^2 + \int_\Omega |c_\varepsilon|^2 &= \int_\Omega n_\varepsilon c_\varepsilon \\ &\leq \frac{1}{2}\int_\Omega n_\varepsilon^2 + \frac{1}{2}\int_\Omega c_\varepsilon^2.\end{aligned} \tag{12.5.20}$$

On the other hand, due to the Gagliardo–Nirenberg inequality, (12.5.17), in light of the Young inequality and $m > 2$, we obtain that

$$\begin{aligned}\|n_\varepsilon\|^2_{L^2(\Omega)} &\leq \|n_\varepsilon + \varepsilon\|^2_{L^2(\Omega)} \\ &= \|(n_\varepsilon + \varepsilon)^{m-1}\|^{\frac{2}{m-1}}_{L^{\frac{2}{m-1}}(\Omega)} \\ &\leq C_1 \|\nabla (n_\varepsilon + \varepsilon)^{m-1}\|^{\frac{6}{6m-7}}_{L^2(\Omega)} \|(n_\varepsilon + \varepsilon)^{m-1}\|^{\frac{2}{m-1}-\frac{6}{6m-7}}_{L^{\frac{1}{m-1}}(\Omega)} \\ &\leq C_2 \left(\|\nabla (n_\varepsilon + \varepsilon)^{m-1}\|^{\frac{6}{6m-7}}_{L^2(\Omega)} + 1\right) \\ &\leq \frac{m^2}{2(m-1)^2}\|\nabla (n_\varepsilon + \varepsilon)^{m-1}\|^2_{L^2(\Omega)} + C_3 \ \text{ for all } \ t \in (0, T_{\max,\varepsilon})\end{aligned} \tag{12.5.21}$$

with some positive constants C_1, C_2, and C_3 independent of ε. Hence, in light of (12.5.20) and (12.5.21), we derive that

$$\begin{aligned}\frac{d}{dt}\|c_\varepsilon\|^2_{L^2(\Omega)} + 2\int_\Omega |\nabla c_\varepsilon|^2 + \int_\Omega c_\varepsilon^2 \leq & \frac{m^2}{2(m-1)^2}\|\nabla (n_\varepsilon + \varepsilon)^{m-1}\|^2_{L^2(\Omega)} \\ & + C_3 \text{ for all } t \in (0, T_{\max,\varepsilon})\end{aligned} \tag{12.5.22}$$

and some positive constant C_3 independent of ε. Next, multiply the first equation in (12.5.13) by $(n_\varepsilon + \varepsilon)^{m-2}$ and combining with the second equation, using $\nabla \cdot u_\varepsilon = 0$ and the Young inequality implies that

$$\begin{aligned}&\frac{1}{m-1}\frac{d}{dt}\|n_\varepsilon + \varepsilon\|^{m-1}_{L^{m-1}(\Omega)} + m(m-2)\int_\Omega (n_\varepsilon + \varepsilon)^{2m-4}|\nabla n_\varepsilon|^2 \\ &\leq (m-2)\int_\Omega (n_\varepsilon + \varepsilon)^{m-2}|\nabla n_\varepsilon||\nabla c_\varepsilon| \\ &\leq \frac{m(m-2)}{2}\int_\Omega (n_\varepsilon + \varepsilon)^{2m-4}|\nabla n_\varepsilon|^2 + \frac{(m-2)}{2m}\int_\Omega |\nabla c_\varepsilon|^2.\end{aligned} \tag{12.5.23}$$

Now, multiplying the third equation of (12.5.13) by u_ε, integrating by parts and using $\nabla \cdot u_\varepsilon = 0$, we derive that

$$\frac{1}{2}\frac{d}{dt}\int_\Omega |u_\varepsilon|^2 + \int_\Omega |\nabla u_\varepsilon|^2 = \int_\Omega n_\varepsilon u_\varepsilon \cdot \nabla\phi \quad \text{for all} \quad t \in (0, T_{\max,\varepsilon}). \tag{12.5.24}$$

Here we use the Hölder inequality and (12.5.10) and the continuity of the embedding $W^{1,2}(\Omega) \hookrightarrow L^6(\Omega)$ and to find $C_4 > 0$ and $C_5 > 0$ such that

$$\begin{aligned}\int_\Omega n_\varepsilon u_\varepsilon \cdot \nabla\phi &\le \|\nabla\phi\|_{L^\infty(\Omega)}\|n_\varepsilon\|_{L^{\frac{6}{5}}(\Omega)}\|\nabla u_\varepsilon\|_{L^2(\Omega)}\\ &\le C_4\|n_\varepsilon\|_{L^{\frac{6}{5}}(\Omega)}\|\nabla u_\varepsilon\|_{L^2(\Omega)}\\ &\le \frac{C_4^2}{2}\|n_\varepsilon\|^2_{L^{\frac{6}{5}}(\Omega)} + \frac{1}{2}\|\nabla u_\varepsilon\|^2_{L^2(\Omega)}\\ &\le C_5\|n_\varepsilon\|^2_{L^2(\Omega)} + \frac{1}{2}\|\nabla u_\varepsilon\|^2_{L^2(\Omega)} \quad \text{for all} \quad t \in (0, T_{max,\varepsilon}),\end{aligned} \tag{12.5.25}$$

which in conjunction with (12.5.21) yields

$$\begin{aligned}\int_\Omega n_\varepsilon u_\varepsilon \cdot \nabla\phi \le{}& \frac{m^2}{4(m-1)^2}\|\nabla(n_\varepsilon+\varepsilon)^{m-1}\|^2_{L^2(\Omega)} + \frac{1}{2}\|\nabla u_\varepsilon\|^2_{L^2(\Omega)}\\ &+C_6 \quad \text{for all} \quad t \in (0, T_{\max,\varepsilon}),\end{aligned} \tag{12.5.26}$$

where C_6 is a positive constant independent of ε. Inserting (12.5.26) into (12.5.25) and using the Young inequality and $m > 2$, we conclude that there exists a positive constant C_7 such that

$$\begin{aligned}\frac{d}{dt}\int_\Omega |u_\varepsilon|^2 + \int_\Omega |\nabla u_\varepsilon|^2 \le{}& \frac{m^2}{2(m-1)^2}\|\nabla(n_\varepsilon+\varepsilon)^{m-1}\|^2_{L^2(\Omega)}\\ &+ C_7 \quad \text{for all} \quad t \in (0, T_{\max,\varepsilon}).\end{aligned} \tag{12.5.27}$$

Taking an evident linear combination of the inequalities provided by (12.5.22), (12.5.23), and (12.5.27), we conclude

$$\begin{aligned}&\frac{d}{dt}\left(\|c_\varepsilon\|^2_{L^2(\Omega)} + \frac{2m}{(m-2)(m-1)}\|n_\varepsilon+\varepsilon\|^{m-1}_{L^{m-1}(\Omega)} + \int_\Omega |u_\varepsilon|^2\right) + \int_\Omega |\nabla u_\varepsilon|^2\\ &\quad+\frac{m^2}{(m-1)^2}\|\nabla(n_\varepsilon+\varepsilon)^{m-1}\|^2_{L^2(\Omega)} + \int_\Omega |\nabla c_\varepsilon|^2 + \int_\Omega c_\varepsilon^2\\ &\le C_8 \quad \text{for all} \quad t \in (0, T_{\max,\varepsilon}),\end{aligned} \tag{12.5.28}$$

where C_8 is a positive constant. An elementary calculus entails (12.5.18) and (12.5.19). □

With the help of Lemma 12.5.3, in light of the Gagliardo–Nirenberg inequality and an application of well-known arguments from parabolic regularity theory, we can derive the following Lemma:

Lemma 12.5.4 *Let* $m > 2$. *Then there exists* $C > 0$ *independent of* ε *such that the solution of* (12.5.13) *satisfies*

$$\int_\Omega c_\varepsilon^{\frac{8(m-1)}{3}} \le C \quad \textit{for all} \;\; t \in (0, T_{max,\varepsilon}). \tag{12.5.29}$$

In addition, for each $T \in (0, T_{max,\varepsilon})$, *one can find a constant* $C > 0$ *independent of* ε *such that*

$$\int_0^T \int_\Omega \left[n_\varepsilon^{\frac{8(m-1)}{3}} + c_\varepsilon^{\frac{8m-14}{3}} |\nabla c_\varepsilon|^2 + c_\varepsilon^{\frac{40(m-1)}{9}} \right] \le C. \tag{12.5.30}$$

Proof Firstly, due to (12.5.18) and (12.5.19), in light of the Gagliardo–Nirenberg inequality, for some C_1 and $C_2 > 0$ which are independent of ε, we derive that

$$\begin{aligned}
\int_0^T \int_\Omega (n_\varepsilon + \varepsilon)^{\frac{8(m-1)}{3}} &= \int_0^T \|(n_\varepsilon + \varepsilon)^{m-1}\|^{\frac{8}{3}}_{L^{\frac{8}{3}}(\Omega)} \\
&\le C_1 \int_0^T \Big(\|\nabla (n_\varepsilon + \varepsilon)^{m-1}\|^2_{L^2(\Omega)} \|(n_\varepsilon + \varepsilon)^{m-1}\|^{\frac{2}{3}}_{L^1(\Omega)} \\
&\quad + \|(n_\varepsilon + \varepsilon)^{m-1}\|^{\frac{8}{3}}_{L^1(\Omega)} \Big) \\
&\le C_2(T+1) \;\; \text{for all} \;\; T > 0.
\end{aligned} \tag{12.5.31}$$

Next, taking $c_\varepsilon^{\frac{8m-11}{3}}$ as the test function for the second equation of (12.5.13) and using $\nabla \cdot u_\varepsilon = 0$ and the Young inequality yields that

$$\begin{aligned}
&\frac{3}{8(m-1)} \frac{d}{dt} \|c_\varepsilon\|^{\frac{8(m-1)}{3}}_{L^{\frac{8(m-1)}{3}}(\Omega)} + \frac{8m-11}{3} \int_\Omega c_\varepsilon^{\frac{8m-14}{3}} |\nabla c_\varepsilon|^2 + \int_\Omega c_\varepsilon^{\frac{8(m-1)}{3}} \\
&= \int_\Omega n_\varepsilon c_\varepsilon^{\frac{8m-11}{3}} \\
&\le C_3 \int_\Omega n_\varepsilon^{\frac{8(m-1)}{3}} + \frac{1}{2} \int_\Omega c_\varepsilon^{\frac{8(m-1)}{3}} \quad \text{for all} \;\; t \in (0, T_{\max,\varepsilon})
\end{aligned} \tag{12.5.32}$$

with some positive constant C_3. Hence, due to (12.5.31) and (12.5.32), we can find $C_4 > 0$ such that

$$\int_\Omega c_\varepsilon^{\frac{8(m-1)}{3}} \le C_4 \quad \text{for all} \;\; t \in (0, T_{\max,\varepsilon}) \tag{12.5.33}$$

and

$$\int_0^T \int_\Omega c_\varepsilon^{\frac{8m-14}{3}} |\nabla c_\varepsilon|^2 \le C_4(T+1) \quad \text{for all } \; T \in (0, T_{\max,\varepsilon}). \tag{12.5.34}$$

Now, due to (12.5.33) and (12.5.34), in light of the Gagliardo–Nirenberg inequality, we derive that there exist positive constants C_5 and C_6 such that

$$\begin{aligned}\int_0^T \int_\Omega c_\varepsilon^{\frac{40(m-1)}{9}} &= \int_0^T \|c_\varepsilon^{\frac{4(m-1)}{3}}\|^{\frac{10}{3}}_{L^{\frac{10}{3}}(\Omega)} \\ &\le C_5 \int_0^T \left(\|\nabla c_\varepsilon^{\frac{4(m-1)}{3}}\|^2_{L^2(\Omega)} \|c_\varepsilon^{\frac{4(m-1)}{3}}\|^{\frac{4}{3}}_{L^2(\Omega)} + \|c_\varepsilon^{\frac{4(m-1)}{3}}\|^{\frac{10}{3}}_{L^2(\Omega)} \right) \\ &\le C_6(T+1) \;\; \text{for all } \; T > 0.\end{aligned} \tag{12.5.35}$$

Finally, collecting (12.5.31) with (12.5.33)–(12.5.35), we can get the results. □

Lemma 12.5.5 *There exists a positive constant $C := C(\varepsilon)$ depends on ε such that*

$$\int_\Omega |\nabla u_\varepsilon(\cdot, t)|^2 \le C \;\; \textit{for all} \;\; t \in (0, T_{max,\varepsilon}) \tag{12.5.36}$$

and

$$\int_0^T \int_\Omega |\Delta u_\varepsilon|^2 \le C \;\; \textit{for all} \;\; T \in (0, T_{max,\varepsilon}). \tag{12.5.37}$$

Proof Firstly, due to $D(1+\varepsilon A) := W^{2,2}(\Omega) \cap W^{1,2}_{0,\sigma}(\Omega) \hookrightarrow L^\infty(\Omega)$, by (12.5.18), we derive that for some $C_1 > 0$ and $C_2 > 0$,

$$\begin{aligned}\|Y_\varepsilon u_\varepsilon\|_{L^\infty(\Omega)} &= \|(I+\varepsilon A)^{-1} u_\varepsilon\|_{L^\infty(\Omega)} \\ &\le C_1 \|u_\varepsilon(\cdot,t)\|_{L^2(\Omega)} \le C_2 \;\; \text{for all } \; t \in (0, T_{\max,\varepsilon}).\end{aligned} \tag{12.5.38}$$

Next, testing the projected Stokes equation $u_{\varepsilon t} + Au_\varepsilon = \mathcal{P}[-\kappa(Y_\varepsilon u_\varepsilon \cdot \nabla)u_\varepsilon + n_\varepsilon \nabla\phi]$ by Au_ε, we derive

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\|A^{\frac{1}{2}} u_\varepsilon\|^2_{L^2(\Omega)} + \int_\Omega |Au_\varepsilon|^2 \\ &= \int_\Omega Au_\varepsilon \mathcal{P}(-\kappa(Y_\varepsilon u_\varepsilon \cdot \nabla)u_\varepsilon) + \int_\Omega \mathcal{P}(n_\varepsilon \nabla\phi) Au_\varepsilon \\ &\le \frac{1}{2}\int_\Omega |Au_\varepsilon|^2 + \kappa^2 \int_\Omega |(Y_\varepsilon u_\varepsilon \cdot \nabla)u_\varepsilon|^2 \\ &\quad + \|\nabla\phi\|^2_{L^\infty(\Omega)} \int_\Omega n_\varepsilon^2 \;\; \text{for all } \; t \in (0, T_{\max,\varepsilon}).\end{aligned} \tag{12.5.39}$$

On the other hand, in light of the Gagliardo–Nirenberg inequality, the Young inequality and (12.5.38), there exists a positive constant C_3 such that

$$\begin{aligned}\kappa^2\int_\Omega |(Y_\varepsilon u_\varepsilon\cdot\nabla)u_\varepsilon|^2 &\le \kappa^2\|Y_\varepsilon u_\varepsilon\|^2_{L^\infty(\Omega)}\int_\Omega|\nabla u_\varepsilon|^2\\ &\le \kappa^2\|Y_\varepsilon u_\varepsilon\|^2_{L^\infty(\Omega)}\int_\Omega|\nabla u_\varepsilon|^2\\ &\le C_3\int_\Omega|\nabla u_\varepsilon|^2 \text{ for all } t\in(0,T_{\max,\varepsilon}).\end{aligned} \tag{12.5.40}$$

Here we have the well-known fact that $\|A(\cdot)\|_{L^2(\Omega)}$ defines a norm equivalent to $\|\cdot\|_{W^{2,2}(\Omega)}$ on $D(A)$ (see Theorem 2.1.1 of [71]). Now, recalling that

$$\|A^{\frac12}u_\varepsilon\|^2_{L^2(\Omega)} = \|\nabla u_\varepsilon\|^2_{L^2(\Omega)},$$

inserting the above equation and (12.5.40) into (12.5.39), we can conclude that

$$\begin{aligned}\frac12\frac{d}{dt}\|\nabla u_\varepsilon\|^2_{L^2(\Omega)}+\int_\Omega|\Delta u_\varepsilon|^2 &\le C_4\int_\Omega|\nabla u_\varepsilon|^2\\ &\quad+\|\nabla\phi\|^2_{L^\infty(\Omega)}\int_\Omega n_\varepsilon^2 \text{ for all } t\in(0,T_{\max,\varepsilon})\end{aligned} \tag{12.5.41}$$

with some positive constant C_4. Collecting (12.5.31) and (12.5.41) and applying the Young inequality, we can get the results. □

12.5.3 The Global Solvability of Regularized Problem (12.5.13)

In this section, we will prove the global solvability of regularized problem (12.5.13). To this end, we need to establish some ε-dependent estimates of $(n_\varepsilon, c_\varepsilon, u_\varepsilon)$ firstly.

Lemma 12.5.6 *There exists $C := C(\varepsilon) > 0$ depends on ε such that*

$$\int_\Omega|\nabla c_\varepsilon(\cdot,t)|^2\le C \textit{ for all } t\in(0,T_{max,\varepsilon}) \tag{12.5.42}$$

and

$$\int_0^T\int_\Omega|\Delta c_\varepsilon|^2\le C \textit{ for all } T\in(0,T_{max,\varepsilon}). \tag{12.5.43}$$

Proof Firstly, testing the second equation in (12.5.13) against $-\Delta c_\varepsilon$ and employing the Young inequality yields

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|\nabla c_\varepsilon\|^2_{L^2(\Omega)} &= \int_\Omega -\Delta c_\varepsilon(\Delta c_\varepsilon - c_\varepsilon + n_\varepsilon - u_\varepsilon\cdot\nabla c_\varepsilon)\\ &= -\int_\Omega |\Delta c_\varepsilon|^2 - \int_\Omega |\nabla c_\varepsilon|^2 - \int_\Omega n_\varepsilon \Delta c_\varepsilon - \int_\Omega (u_\varepsilon\cdot\nabla c_\varepsilon)\Delta c_\varepsilon\\ &\le -\frac{1}{2}\int_\Omega |\Delta c_\varepsilon|^2 - \int_\Omega |\nabla c_\varepsilon|^2 + \int_\Omega n_\varepsilon^2 + \int_\Omega |u_\varepsilon|^2|\nabla c_\varepsilon|^2\end{aligned} \tag{12.5.44}$$

for all $t \in (0, T_{\max,\varepsilon})$. Now, applying (12.5.18) and (12.5.37), the Gagliardo–Nirenberg inequality and the Young inequality, we derive there exist positive constants C_1, C_2, and C_3 such that

$$\begin{aligned}\int_\Omega |u_\varepsilon|^2|\nabla c_\varepsilon|^2 =&\|u_\varepsilon\|^2_{L^8(\Omega)}\|\nabla c_\varepsilon\|^2_{L^{\frac{8}{3}}(\Omega)}\\ \le&\|u_\varepsilon\|^2_{L^8(\Omega)}C_1\left(\|\Delta c_\varepsilon\|^{\frac{11}{8}}_{L^2(\Omega)}\|c_\varepsilon\|^{\frac{5}{8}}_{L^2(\Omega)} + \|c_\varepsilon\|^2_{L^2(\Omega)}\right)\\ \le&\|u_\varepsilon\|^2_{L^8(\Omega)}C_2\left(\|\Delta c_\varepsilon\|^{\frac{11}{8}}_{L^2(\Omega)} + 1\right)\\ \le&\frac{1}{4}\|\Delta c_\varepsilon\|^2_{L^2(\Omega)} + C_3\left(\|u_\varepsilon\|^{\frac{32}{5}}_{L^8(\Omega)} + 1\right)\end{aligned} \tag{12.5.45}$$

for all $t \in (0, T_{\max,\varepsilon})$. Now, in view of the Gagliardo–Nirenberg inequality and the well-known fact that $\|A(\cdot)\|_{L^2(\Omega)}$ defines a norm equivalent to $\|\cdot\|_{W^{2,2}(\Omega)}$ on $W^{2,2}(\Omega)\cap W^{1,2}_0(\Omega)$ (see p. 129, Theorem e of [71]), we have

$$\begin{aligned}C_3\|u_\varepsilon\|^{\frac{32}{5}}_{L^8(\Omega)} &\le C_3\|Au_\varepsilon\|^{\frac{4}{5}}_{L^2(\Omega)}\|u_\varepsilon\|^{\frac{28}{5}}_{L^6(\Omega)}\\ &\le C_4\left(\|Au_\varepsilon\|^2_{L^2(\Omega)} + 1\right),\end{aligned} \tag{12.5.46}$$

where C_4 is a positive constant. Hence, together with (12.5.46) and (12.5.37), we conclude that there exists a positive constant C_5 such that for all $T \in (0, T_{\max,\varepsilon})$,

$$C_3\int_0^T \|u_\varepsilon\|^{\frac{32}{5}}_{L^8(\Omega)} \le C_5. \tag{12.5.47}$$

Inserting (12.5.46) and (12.5.45) into (12.5.44) and using (12.5.31) and (12.5.47), we can derive (12.5.42) and (12.5.43). This completes the proof of Lemma 12.5.6.
□

With Lemmata 12.5.3–12.5.6 at hand, we are now in the position to prove the solution of approximate problem (12.5.13) is actually global in time.

Lemma 12.5.7 *Let $m > 2$. Then for all $\varepsilon \in (0,1)$, the solution of* (12.5.13) *is global in time.*

Proof Assuming that $T_{\max,\varepsilon}$ be finite for some $\varepsilon \in (0,1)$. Next, applying almost exactly the same arguments as in the proof of Lemma 3.4 in [122], we may derive the following estimate: the solution of (12.5.13) satisfies that for all $\beta > 1$

$$\begin{aligned}&\frac{1}{2\beta}\frac{d}{dt}\int_\Omega |\nabla c_\varepsilon|^{2\beta} + \int_\Omega |\nabla c_\varepsilon|^{2\beta} + \frac{1}{2}\int_\Omega |\nabla c_\varepsilon|^{2\beta-2}|D^2 c_\varepsilon|^2\\ &\quad+\frac{(\beta-1)}{2\beta^2}\|\nabla|\nabla c_\varepsilon|^\beta\|^2_{L^2(\Omega)}\\ &\le C_1\int_\Omega n_\varepsilon^2|\nabla c_\varepsilon|^{2\beta-2} + \int_\Omega |Du_\varepsilon||\nabla c_\varepsilon|^{2\beta} + C_1 \text{ for all } t\in(0,T_{\max,\varepsilon}),\end{aligned} \tag{12.5.48}$$

where C_1 is a positive constant, as all subsequently appearing constants $C_2, C_3, \ldots$ possibly depend on ε and β. On the other hand, due to (12.5.36), we derive that there exists a positive constant C_2 such that

$$\|Du_\varepsilon(\cdot,t)\|_{L^2(\Omega)} \le C_2 \text{ for all } t\in(0,T_{\max,\varepsilon}). \tag{12.5.49}$$

Hence, in light of the Hölder inequality and the Gagliardo–Nirenberg inequality, (12.5.42) and the Young inequality, we conclude that

$$\begin{aligned}\int_\Omega |Du_\varepsilon||\nabla c_\varepsilon|^{2\beta} &\le C_2\|\nabla c_\varepsilon\|^{2\beta}_{L^{4\beta}(\Omega)}\\ &= C_2\||\nabla c_\varepsilon|^\beta\|^2_{L^4(\Omega)}\\ &= C_2\left(\|\nabla|\nabla c_\varepsilon|^\beta\|^{\frac{6\beta-3}{6\beta-2}}_{L^2(\Omega)}\||\nabla c_\varepsilon|^\beta\|^{\frac{6\beta-1}{6\beta-2}}_{L^{\frac{2}{\beta}}(\Omega)} + \||\nabla c_\varepsilon|^\beta\|^2_{L^{\frac{2}{\beta}}(\Omega)}\right)\\ &\le C_3\left(\|\nabla|\nabla c_\varepsilon|^\beta\|^{\frac{6\beta-3}{6\beta-2}}_{L^2(\Omega)} + 1\right)\\ &\le \frac{(\beta-1)}{8\beta^2}\|\nabla|\nabla c_\varepsilon|^\beta\|^2_{L^2(\Omega)} + C_4 \text{ for all } t\in(0,T_{\max,\varepsilon})\end{aligned} \tag{12.5.50}$$

with some positive constants C_3 and C_4. Now, inserting (12.5.50) into (12.5.48), we derive that there exists a positive constant C_5 such that

$$\begin{aligned}&\frac{1}{2\beta}\frac{d}{dt}\int_\Omega |\nabla c_\varepsilon|^{2\beta} + \int_\Omega |\nabla c_\varepsilon|^{2\beta} + \frac{1}{2}\int_\Omega |\nabla c_\varepsilon|^{2\beta-2}|D^2 c_\varepsilon|^2\\ &+\frac{3(\beta-1)}{8\beta^2}\|\nabla|\nabla c_\varepsilon|^\beta\|^2_{L^2(\Omega)} \le C_1\int_\Omega n_\varepsilon^2|\nabla c_\varepsilon|^{2\beta-2} + C_5 \text{ for all } t\in(0,T_{max,\varepsilon}).\end{aligned} \tag{12.5.51}$$

Next, with the help of the Young inequality, we derive that there exists a positive constant C_6 such that

$$C_1 \int_\Omega n_\varepsilon^2 |\nabla c_\varepsilon|^{2\beta-2} \le \frac{1}{4}\int_\Omega |\nabla c_\varepsilon|^{4(2\beta-2)} + C_6 \int_\Omega n_\varepsilon^{\frac{8}{3}} + C_1. \tag{12.5.52}$$

Now, choosing $\beta = \frac{4}{3}$ in (12.5.51) and (12.5.52), we conclude that

$$\begin{aligned}&\frac{1}{\frac{8}{3}}\frac{d}{dt}\int_\Omega |\nabla c_\varepsilon|^{\frac{8}{3}} + \frac{3}{4}\int_\Omega |\nabla c_\varepsilon|^{\frac{8}{3}} + \frac{1}{2}\int_\Omega |\nabla c_\varepsilon|^{\frac{2}{3}}|D^2 c_\varepsilon|^2 + \frac{3}{16}\|\nabla|\nabla c_\varepsilon|^{\frac{4}{3}}\|_{L^2(\Omega)}^2\\ &\le C_6\int_\Omega n_\varepsilon^{\frac{8}{3}} + C_1.\end{aligned} \tag{12.5.53}$$

Here we have used the fact that $4(2\beta - 2) = 2\beta$. Hence, in light of (12.5.31) and $m > 2$, by (12.5.53), we derive that there exists a positive constant C_7 such that

$$\|\nabla c_\varepsilon(\cdot, t)\|_{L^{\frac{8}{3}}(\Omega)} \le C_7 \text{ for all } t \in (0, T_{\max,\varepsilon}) \tag{12.5.54}$$

Now, employing almost exactly the same arguments as in the proof of Lemma 3.3 in [122], we conclude that the solution of (12.5.13) satisfies that for all $p > 1$,

$$\begin{aligned}&\frac{1}{p}\frac{d}{dt}\|n_\varepsilon + \varepsilon\|_{L^p(\Omega)}^p + \frac{2m(p-1)}{(m+p-1)^2}\int_\Omega |\nabla(n_\varepsilon+\varepsilon)^{\frac{m+p-1}{2}}|^2\\ &\le C_8\int_\Omega (n_\varepsilon+\varepsilon)^{p+1-m}|\nabla c_\varepsilon|^2\end{aligned} \tag{12.5.55}$$

for all $t \in (0, T_{\max,\varepsilon})$ and some positive constant C_7. By the Hölder inequality and (12.5.54) and using $m > 2$ and the Gagliardo–Nirenberg inequality, we derive there exist positive constants C_9, C_{10}, and C_{11} such that

$$\begin{aligned}&\int_\Omega (n_\varepsilon+\varepsilon)^{p+1-m}|\nabla c_\varepsilon|^2\\ &\le \left(\int_\Omega (n_\varepsilon+\varepsilon)^{4(p+1-m)}\right)^{\frac{1}{4}}\left(\int_\Omega |\nabla c_\varepsilon|^{\frac{8}{3}}\right)^{\frac{3}{4}}\\ &\le C_9\|(n_\varepsilon+\varepsilon)^{\frac{p+m-1}{2}}\|_{L^{\frac{8(p+1-m)}{p+m-1}}(\Omega)}^{\frac{2(p+1-m)}{p+m-1}}\\ &\le C_{10}\Big(\|\nabla(n_\varepsilon+\varepsilon)^{\frac{p+m-1}{2}}\|_{L^2(\Omega)}^{b_1}\|(n_\varepsilon+\varepsilon)^{\frac{p+m-1}{2}}\|_{L^{\frac{2}{p+m-1}}(\Omega)}^{1-b_1}\\ &\quad+\|(n_\varepsilon+\varepsilon)^{\frac{p+m-1}{2}}\|_{L^{\frac{2}{p+m-1}}(\Omega)}\Big)^{\frac{2(p+1-m)}{p+m-1}}\end{aligned}$$

$$\leq C_{11}(\|\nabla(n_\varepsilon+\varepsilon)^{\frac{p+m-1}{2}}\|_{L^2(\Omega)}^{\frac{2(p+1-m)b_1}{p+m-1}}+1)$$
$$= C_{11}(\|\nabla(n_\varepsilon+\varepsilon)^{\frac{p+m-1}{2}}\|_{L^2(\Omega)}^{\frac{12p-12m+9}{6p+6m-8}}+1) \text{ for all } t\in(0,T_{\max,\varepsilon}), \tag{12.5.56}$$

where

$$b_1=\frac{\frac{3[p+m-1]}{2}-\frac{3(p+m-1)}{8(p+1-m)}}{-\frac{1}{2}+\frac{3[p+m-1]}{2}}\in(0,1).$$

Since $m>2$ yields to $\frac{12p-12m+9}{6p+6m-8}<2$, in light of (12.5.56) and the Young inequality, we derive that there exists a positive constant C_{12} such that

$$C_8\int_\Omega(n_\varepsilon+\varepsilon)^{p+1-m}|\nabla c_\varepsilon|^2\leq\frac{m(p-1)}{(m+p-1)^2}\|\nabla(n_\varepsilon+\varepsilon)^{\frac{p+m-1}{2}}\|_{L^2(\Omega)}^{\frac{12p-12m+9}{6p+6m-8}}$$
$$+C_{12} \text{ for all } t\in(0,T_{\max,\varepsilon}). \tag{12.5.57}$$

Together with (12.5.55), this yields the desired estimate

$$\frac{1}{p}\frac{d}{dt}\|n_\varepsilon+\varepsilon\|_{L^p(\Omega)}^p+\frac{2m(p-1)}{(m+p-1)^2}\int_\Omega|\nabla(n_\varepsilon+\varepsilon)^{\frac{m+p-1}{2}}|^2$$
$$\leq C_{12} \text{ for all } t\in(0,T_{\max,\varepsilon}). \tag{12.5.58}$$

Now, with some basic analysis, we may derive that for all $p>1$, there exists a positive constant C_{13} such that

$$\|n_\varepsilon(\cdot,t)\|_{L^p(\Omega)}\leq C_{13} \text{ for all } t\in(0,T_{\max,\varepsilon}). \tag{12.5.59}$$

Let $h_\varepsilon(x,t)=\mathcal{P}[-\kappa(Y_\varepsilon u_\varepsilon\cdot\nabla)u_\varepsilon+n_\varepsilon\nabla\phi]$. Then along with (12.5.18) and (12.5.59), there exists a positive constant C_{13} such that $\|h_\varepsilon(\cdot,t)\|_{L^2(\Omega)}\leq C_{14}$ for all $t\in(0,T_{\max,\varepsilon})$. Hence, we pick an arbitrary $\gamma\in(\frac{3}{4},1)$, then in light of the smoothing properties of the Stokes semigroup ([26]), we derive that for some $C_{15}>0$, we have

$$\|A^\gamma u_\varepsilon(\cdot,t)\|_{L^2(\Omega)}\leq\|A^\gamma e^{-tA}u_0\|_{L^2(\Omega)}+\int_0^t\|A^\gamma e^{-(t-\tau)A}h_\varepsilon(\cdot,\tau)d\tau\|_{L^2(\Omega)}d\tau$$
$$\leq C_{15}t^{-\lambda_1(t-1)}\|u_0\|_{L^2(\Omega)}+C_{15}\int_0^t(t-\tau)^{-\gamma}\|h_\varepsilon(\cdot,\tau)\|_{L^2(\Omega)}d\tau$$
$$\leq C_{15}t^{-\lambda_1(t-1)}\|u_0\|_{L^2(\Omega)}+\frac{C_{14}C_{15}T_{\max,\varepsilon}^{1-\gamma}}{1-\gamma} \text{ for all } t\in(0,T_{\max,\varepsilon}). \tag{12.5.60}$$

Observe that $\gamma > \frac{3}{4}$, $D(A^\gamma)$ is continuously embedded into $L^\infty(\Omega)$, therefore, due to (12.5.60), we derive that there exists a positive constant C_{16} such that

$$\|u_\varepsilon(\cdot,t)\|_{L^\infty(\Omega)} \le C_{16} \text{ for all } t \in (0, T_{\max,\varepsilon}). \tag{12.5.61}$$

Now, for any $\beta > 1$, choosing $p > 0$ large enough such that $p > 2\beta$, then due to (12.5.51) and (12.5.59), invoking the Young inequality, we derive that there exists a positive constant C_{17} such that

$$\begin{aligned}&\frac{1}{2\beta}\frac{d}{dt}\int_\Omega |\nabla c_\varepsilon|^{2\beta} + \frac{1}{2}\int_\Omega |\nabla c_\varepsilon|^{2\beta} + \frac{1}{2}\int_\Omega |\nabla c_\varepsilon|^{2\beta-2}|D^2 c_\varepsilon|^2\\ &+\frac{3(\beta-1)}{8\beta^2}\|\nabla|\nabla c_\varepsilon|^\beta\|^2_{L^2(\Omega)} \le C_{17} \text{ for all } t \in (0, T_{\max,\varepsilon}).\end{aligned} \tag{12.5.62}$$

Now, integrating the above inequality in time, we derive that there exists a positive constant C_{18} such that

$$\|\nabla c_\varepsilon(\cdot,t)\|_{L^{2\beta}(\Omega)} \le C_{18} \text{ for all } t \in (0, T_{\max,\varepsilon}) \text{ and } \beta > 1. \tag{12.5.63}$$

In order to get the boundedness of $\|\nabla c_\varepsilon(\cdot,t)\|_{L^\infty(\Omega)}$, we rewrite the variation-of-constants formula for c_ε in the form

$$c_\varepsilon(\cdot,t) = e^{t(\Delta-1)}c_0 + \int_0^t e^{(t-s)(\Delta-1)}(n_\varepsilon - u_\varepsilon\cdot\nabla c_\varepsilon)(\cdot,s)ds \text{ for all } t \in (0, T_{\max,\varepsilon}).$$

Now, we choose $\theta \in (\frac{7}{8}, 1)$, then the domain of the fractional power $D((-\Delta+1)^\theta) \hookrightarrow W^{1,\infty}(\Omega)$ ([117]). Hence, in view of L^p-L^q estimates associated heat semigroup, (12.5.11), (12.5.59), (12.5.61) and (12.5.63), we derive that there exist positive constants C_{19}, C_{20}, and C_{21} such that

$$\begin{aligned}&\|\nabla c_\varepsilon(\cdot,t)\|_{W^{1,\infty}(\Omega)}\\ \le\ & C_{19}t^{-\theta}e^{-\lambda t}\|c_0\|_{L^4(\Omega)}\\ &+\int_0^t (t-s)^{-\theta}e^{-\lambda(t-s)}\|(n_\varepsilon - u_\varepsilon\cdot\nabla c_\varepsilon)(s)\|_{L^4(\Omega)}ds\\ \le\ & C_{20}\tau^{-\theta} + C_{20}\int_0^t (t-s)^{-\theta}e^{-\lambda(t-s)}\\ &+C_{20}\int_0^t (t-s)^{-\theta}e^{-\lambda(t-s)}[\|n_\varepsilon(s)\|_{L^4(\Omega)} + \|\nabla c_\varepsilon(s)\|_{L^4(\Omega)}]ds\\ \le\ & C_{21} \text{ for all } t \in (\tau, T_{\max,\varepsilon})\end{aligned} \tag{12.5.64}$$

with $\tau \in (0, T_{\max,\varepsilon})$. Next, using the outcome of (12.5.55) with suitably large p as a starting point, we may employ a Moser-type iteration (see, e.g., Lemma A.1 of [78]) applied to the first equation of (12.5.13) to get that

$$\|n_\varepsilon(\cdot,t)\|_{L^\infty(\Omega)} \le C_{22} \text{ for all } t \in (\tau, T_{\max,\varepsilon}) \tag{12.5.65}$$

and some positive constant C_{22}. In view of (12.5.61), (12.5.64), and (12.5.65), we apply Lemma 12.5.1 to reach a contradiction. □

12.5.3.1 Regularity Properties of Time Derivatives

In this subsection, we provide some time-derivatives uniform estimates of solutions to the system (12.5.13). The estimate is used in this section to construct the weak solution of the Eq. (12.5.4). This will be the purpose of the following lemmata:

Lemma 12.5.8 *Let $m > 2$, (12.5.10) and (12.5.11) hold. Then for any $T > 0$, one can find $C > 0$ independent if ε such that*

$$\int_0^T \|\partial_t n_\varepsilon^{m-1}(\cdot,t)\|_{(W^{2,4}(\Omega))^*} dt \le C(T+1) \tag{12.5.66}$$

as well as

$$\int_0^T \|\partial_t c_\varepsilon(\cdot,t)\|^{\frac{5}{3}}_{(W^{1,\frac{5}{2}}(\Omega))^*} dt \le C(T+1) \tag{12.5.67}$$

and

$$\int_0^T \|\partial_t u_\varepsilon(\cdot,t)\|^2_{(W^{1,2}(\Omega))^*} dt \le C(T+1). \tag{12.5.68}$$

Proof Firstly, due to (12.5.18), (12.5.19) and (12.5.31), employing the Hölder inequality (with two exponents $\frac{4m-1}{4(m-1)}$ and $\frac{4(m-1)}{3}$) and the Gagliardo–Nirenberg inequality, we conclude that there exist positive constants C_1, C_2, C_3 and C_4 such that

$$\begin{aligned}\int_0^T\int_\Omega |m(n_\varepsilon+\varepsilon)^{m-1}\nabla n_\varepsilon|^{\frac{8(m-1)}{4m-1}} &\le C_1\left[\int_0^T\int_\Omega (n_\varepsilon+\varepsilon)^{2m-4}|\nabla n_\varepsilon|^2\right]^{\frac{4(m-1)}{4m-1}}\\ &\quad\times\left[\int_0^T\int_\Omega [n_\varepsilon+\varepsilon]^{\frac{8(m-1)}{3}}\right]^{\frac{3}{4m-1}}\\ &\le C_2(T+1)\ \text{ for all }\ T>0\end{aligned} \tag{12.5.69}$$

and

$$\begin{aligned}\int_0^T\int_\Omega |u_\varepsilon|^{\frac{10}{3}} &= \int_0^T \|u_\varepsilon\|^{\frac{10}{3}}_{L^{\frac{10}{3}}(\Omega)}\\ &\le C_3\int_0^T\left(\|\nabla u_\varepsilon\|^2_{L^2(\Omega)}\|u_\varepsilon\|^{\frac{4}{3}}_{L^2(\Omega)}+\|u_\varepsilon\|^{\frac{10}{3}}_{L^2(\Omega)}\right)\\ &\le C_4(T+1)\ \text{ for all }\ T>0.\end{aligned} \tag{12.5.70}$$

Next, testing the first equation of (12.5.13) by certain $(m-1)n_\varepsilon^{m-2}\varphi \in C^\infty(\bar{\Omega})$, we have

$$
\begin{aligned}
&\left|\int_\Omega (n_\varepsilon^{m-1})_t\varphi\right| \\
=&\left|\int_\Omega \left[\Delta(n_\varepsilon+\varepsilon)^m - \nabla\cdot(n_\varepsilon\nabla c_\varepsilon) - u_\varepsilon\cdot\nabla n_\varepsilon\right]\cdot(m-1)n_\varepsilon^{m-2}\varphi\right| \\
\le&\left|-(m-1)\int_\Omega \left[m(n_\varepsilon+\varepsilon)^{m-1}n_\varepsilon^{m-2}\nabla n_\varepsilon\cdot\nabla\varphi + (m-2)(n_\varepsilon+\varepsilon)^{m-1}n_\varepsilon^{m-3}|\nabla n_\varepsilon|^2\varphi\right]\right| \\
&+(m-1)\left|\int_\Omega [(m-2)n_\varepsilon^{m-2}\nabla n_\varepsilon\cdot\nabla c_\varepsilon\varphi + n_\varepsilon^{m-1}\nabla c_\varepsilon\cdot\nabla\varphi]\right| + \left|\int_\Omega n_\varepsilon^{m-1}u_\varepsilon\cdot\nabla\varphi\right| \\
\le&\, m(m-1)\left\{\int_\Omega \left[(n_\varepsilon+\varepsilon)^{m-1}n_\varepsilon^{m-2}|\nabla n_\varepsilon| + (n_\varepsilon+\varepsilon)^{m-1}n_\varepsilon^{m-3}|\nabla n_\varepsilon|^2\right]\right\}\|\varphi\|_{W^{1,\infty}(\Omega)} \\
&+(m-1)^2\left\{\int_\Omega [n_\varepsilon^{m-2}|\nabla n_\varepsilon||\nabla c_\varepsilon| + n_\varepsilon^{m-1}|\nabla c_\varepsilon| + n_\varepsilon^{m-1}|u_\varepsilon|]\right\}\|\varphi\|_{W^{1,\infty}(\Omega)}
\end{aligned}
\tag{12.5.71}
$$

for all $t > 0$. Hence, observe that the embedding $W^{2,4}(\Omega) \hookrightarrow W^{1,\infty}(\Omega)$, due to (12.5.19), (12.5.31) and (12.5.70), applying $m > 2$ and the Young inequlity, we deduce C_1, C_2 and C_3 such that

$$
\begin{aligned}
&\int_0^T \|\partial_t n_\varepsilon^{m-1}(\cdot,t)\|_{(W^{2,4}(\Omega))^*}dt \\
&\le C_1\left\{\int_0^T\int_\Omega (n_\varepsilon+\varepsilon)^{2m-4}|\nabla n_\varepsilon|^2 + \int_0^T\int_\Omega n_\varepsilon^{2m-2}\right. \\
&\qquad \left. + \int_0^T\int_\Omega |\nabla c_\varepsilon|^2 + \int_0^T\int_\Omega |u_\varepsilon|^2\right\} \\
&\le C_2\left\{\int_0^T\int_\Omega (n_\varepsilon+\varepsilon)^{2m-4}|\nabla n_\varepsilon|^2 + \int_0^T\int_\Omega |\nabla c_\varepsilon|^2\right. \\
&\qquad \left. + \int_0^T\int_\Omega n_\varepsilon^{\frac{8(m-1)}{3}} + \int_0^T\int_\Omega |u_\varepsilon|^{\frac{10}{3}} + T\right\} \\
&\le C_3(T+1) \text{ for all } T > 0,
\end{aligned}
\tag{12.5.72}
$$

which implies (12.5.66).

Likewise, given any $\varphi \in C^\infty(\bar{\Omega})$, we may test the second equation in (12.5.13) against φ to conclude that

$$
\begin{aligned}
\left|\int_\Omega \partial_t c_\varepsilon(\cdot,t)\varphi\right| &= \left|\int_\Omega [\Delta c_\varepsilon - c_\varepsilon + n_\varepsilon - u_\varepsilon\cdot\nabla c_\varepsilon]\cdot\varphi\right| \\
&= \left|-\int_\Omega \nabla c_\varepsilon\cdot\nabla\varphi - \int_\Omega c_\varepsilon\varphi + \int_\Omega n_\varepsilon\varphi + \int_\Omega c_\varepsilon u_\varepsilon\cdot\nabla\varphi\right|
\end{aligned}
$$

$$\leq \left\{ \|\nabla c_\varepsilon\|_{L^{\frac{5}{3}}(\Omega)} + \|c_\varepsilon\|_{L^{\frac{5}{3}}(\Omega)} + \|n_\varepsilon\|_{L^{\frac{5}{3}}(\Omega)} + \|c_\varepsilon u_\varepsilon\|_{L^{\frac{5}{3}}(\Omega)} \right\} \|\varphi\|_{W^{1,\frac{5}{2}}(\Omega)} \quad \text{for all } t > 0. \tag{12.5.73}$$

Thus, due to (12.5.19), (12.5.30)–(12.5.31) and (12.5.70), in light of $m > 2$, we invoke the Young inequality again and obtain that there exist positive constant C_8 and C_9 such that

$$\begin{aligned} &\int_0^T \|\partial_t c_\varepsilon(\cdot,t)\|^{\frac{5}{3}}_{(W^{1,\frac{5}{2}}(\Omega))^*} dt \\ &\leq C_8 \left(\int_0^T \int_\Omega |\nabla c_\varepsilon|^2 + \int_0^T \int_\Omega n_\varepsilon^{\frac{8(m-1)}{3}} + \int_0^T \int_\Omega c_\varepsilon^{\frac{40(m-1)}{9}} + \int_0^T \int_\Omega |u_\varepsilon|^{\frac{10}{3}} + T \right) \\ &\leq C_9(T+1) \quad \text{for all } T > 0. \end{aligned} \tag{12.5.74}$$

Hence, (12.5.67) is proved.

Finally, for any given $\varphi \in C^\infty_{0,\sigma}(\Omega;\mathbb{R}^3)$, we infer from the third equation in (12.5.13) that

$$\begin{aligned} \left| \int_\Omega \partial_t u_\varepsilon(\cdot,t)\varphi \right| = \Bigg| &- \int_\Omega \nabla u_\varepsilon \cdot \nabla\varphi - \kappa \int_\Omega (Y_\varepsilon u_\varepsilon \otimes u_\varepsilon) \cdot \nabla\varphi \\ &+ \int_\Omega n_\varepsilon \nabla\phi \cdot \varphi \Bigg| \quad \text{for all } t > 0. \end{aligned} \tag{12.5.75}$$

Now, by virtue of (12.5.19), (12.5.30) and (12.5.38), we also get that there exist positive constants C_{10}, C_{11} and C_{12} such that

$$\begin{aligned} &\int_0^T \|\partial_t u_\varepsilon(\cdot,t)\|^2_{(W^{1,2}(\Omega))^*} dt \\ &\leq C_{10} \left(\int_0^T \int_\Omega |\nabla u_\varepsilon|^2 + \int_0^T \int_\Omega |Y_\varepsilon u_\varepsilon \otimes u_\varepsilon|^2 + \int_0^T \int_\Omega n_\varepsilon^2 \right) \\ &\leq C_{11} \left(\int_0^T \int_\Omega |\nabla u_\varepsilon|^2 + \int_0^T \int_\Omega |Y_\varepsilon u_\varepsilon|^2 + \int_0^T \int_\Omega n_\varepsilon^{\frac{8(m-1)}{3}} + T \right) \\ &\leq C_{12}(T+1) \quad \text{for all } T > 0. \end{aligned} \tag{12.5.76}$$

Hence, (12.5.68) is hold. □

In order to prove the limit functions n and c gained below, we will rely on an additional regularity estimate for $n_\varepsilon \nabla c_\varepsilon$ and $u_\varepsilon \cdot \nabla c_\varepsilon$.

Lemma 12.5.9 *Let $m > 2$, (12.5.10) and (12.5.11) hold. Then for any $T > 0$, one can find $C > 0$ independent of ε such that*

$$\int_0^T \int_\Omega |n_\varepsilon \nabla c_\varepsilon|^{\frac{8(m-1)}{4m-1}} \leq C(T+1) \tag{12.5.77}$$

and

$$\int_0^T \int_\Omega |u_\varepsilon \cdot \nabla c_\varepsilon|^{\frac{5}{4}} \leq C(T+1). \tag{12.5.78}$$

Proof In light of (12.5.19), (12.5.31), (12.5.70) and the Young inequality, we derive that there exist positive constants C_1 and C_2 such that

$$\begin{aligned}\int_0^T \int_\Omega |n_\varepsilon \nabla c_\varepsilon|^{\frac{8(m-1)}{4m-1}} &\leq \left(\int_0^T \int_\Omega |\nabla c_\varepsilon|^2\right)^{\frac{3}{4m-1}} \left(\int_0^T \int_\Omega n_\varepsilon^{\frac{8(m-1)}{3}}\right)^{\frac{4(m-1)}{4m-1}} \\ &\leq C_1(T+1) \text{ for all } T > 0\end{aligned} \tag{12.5.79}$$

and

$$\begin{aligned}\int_0^T \int_\Omega |u_\varepsilon \cdot \nabla c_\varepsilon|^{\frac{5}{4}} &\leq \left(\int_0^T \int_\Omega |\nabla c_\varepsilon|^2\right)^{\frac{5}{8}} \left(\int_0^T \int_\Omega |u_\varepsilon|^{\frac{10}{3}}\right)^{\frac{3}{8}} \\ &\leq C_2(T+1) \text{ for all } T > 0.\end{aligned} \tag{12.5.80}$$

These readily establish (12.5.77) and (12.5.78).

□

12.5.3.2 Passing to the Limit: Proof of Theorem 12.5.1

With the above compactness properties at hand, by means of a standard extraction procedure we can now derive the following lemma which actually contains our main existence result already.

The Proof of Theorem 12.5.1 Firstly, in light of Lemmata 12.5.3–12.5.4 and 12.5.8, we conclude that there exists a positive constant C_1 such that

$$\begin{aligned}&\|n_\varepsilon^{m-1}\|_{L^2_{loc}([0,\infty);W^{1,2}(\Omega))} \leq C_1(T+1) \\ &\text{and } \|\partial_t n_\varepsilon^{m-1}\|_{L^1_{loc}([0,\infty);(W^{2,4}(\Omega))^*)} \leq C_2(T+1)\end{aligned} \tag{12.5.81}$$

as well as

$$\begin{aligned}&\|c_\varepsilon\|_{L^2_{loc}([0,\infty);W^{1,2}(\Omega))} \leq C_1(T+1) \\ &\text{and } \|\partial_t c_\varepsilon\|_{L^1_{loc}([0,\infty);(W^{1,\frac{5}{2}}(\Omega)))^*)} \leq C_1(T+1)\end{aligned} \tag{12.5.82}$$

and

$$\|u_\varepsilon\|_{L^2_{loc}([0,\infty);W^{1,2}(\Omega))} \le C_1(T+1) \\ \text{and} \quad \|\partial_t u_\varepsilon\|_{L^1_{loc}([0,\infty);(W^{1,2}(\Omega))^*)} \le C_1(T+1). \tag{12.5.83}$$

Hence, collecting (12.5.82)–(12.5.83) and employing the Aubin–Lions lemma (see, e.g., [70]), we conclude that

$$(c_\varepsilon)_{\varepsilon\in(0,1)} \quad \text{is strongly precompact in} \quad L^2_{loc}(\bar{\Omega}\times[0,\infty)) \tag{12.5.84}$$

and

$$(u_\varepsilon)_{\varepsilon\in(0,1)} \quad \text{is strongly precompact in} \quad L^2_{loc}(\bar{\Omega}\times[0,\infty)). \tag{12.5.85}$$

Therefore, there exists a subsequence $\varepsilon = \varepsilon_j \subset (0,1)_{j\in\mathbb{N}}$ and the limit functions c and u such that

$$c_\varepsilon \to c \;\text{ in }\; L^2_{loc}(\bar{\Omega}\times[0,\infty)) \;\text{ and a.e. in }\; \Omega\times(0,\infty), \tag{12.5.86}$$

$$u_\varepsilon \to u \;\text{ in }\; L^2_{loc}(\bar{\Omega}\times[0,\infty)) \;\text{ and a.e. in }\; \Omega\times(0,\infty) \tag{12.5.87}$$

as well as

$$\nabla c_\varepsilon \rightharpoonup \nabla c \;\text{ in }\; L^2_{loc}(\bar{\Omega}\times[0,\infty)) \tag{12.5.88}$$

and

$$\nabla u_\varepsilon \rightharpoonup \nabla u \;\text{ in }\; L^2_{loc}(\bar{\Omega}\times[0,\infty)). \tag{12.5.89}$$

Next, in view of (12.5.81), an Aubin–Lions lemma (see, e.g., [70]) applies to yield strong precompactness of $(n_\varepsilon^{m-1})_{\varepsilon\in(0,1)}$ in $L^2(\Omega\times(0,T))$, whence along a suitable subsequence we may derive that $n_\varepsilon^{m-1} \to z_1^{m-1}$ and hence $n_\varepsilon \to z_1$ a.e. in $\Omega\times(0,\infty)$ for some nonnegative measurable $z_1 : \Omega\times(0,\infty)\to\mathbb{R}$. Now, with the help of the Egorov theorem, we conclude that necessarily $z_1 = n$, thus

$$n_\varepsilon \to n \;\text{ a.e. in }\; \Omega\times(0,\infty). \tag{12.5.90}$$

Therefore, observing that $\frac{8(m-1)}{4m-1} > 1, \frac{8(m-1)}{3} > 1$, due to (12.5.69)–(12.5.70), (12.5.31), there exists a subsequence $\varepsilon = \varepsilon_j \subset (0,1)_{j\in\mathbb{N}}$ such that $\varepsilon_j \searrow 0$ as $j\to\infty$

$$(n_\varepsilon+\varepsilon)^{m-1}\nabla n_\varepsilon \rightharpoonup n^{m-1}\nabla n \;\text{ in }\; L_{loc}^{\frac{8(m-1)}{4m-1}}(\bar{\Omega}\times[0,\infty)) \tag{12.5.91}$$

as well as

$$u_\varepsilon \rightharpoonup u \ \text{ in } \ L^{\frac{10}{3}}_{loc}(\bar{\Omega} \times [0, \infty)) \tag{12.5.92}$$

and

$$n_\varepsilon \rightharpoonup n \ \text{ in } \ L^{\frac{8(m-1)}{3}}_{loc}(\bar{\Omega} \times [0, \infty)). \tag{12.5.93}$$

Next, let $g_\varepsilon(x, t) := -c_\varepsilon + n_\varepsilon - u_\varepsilon \cdot \nabla c_\varepsilon$. Therefore, recalling (12.5.19), (12.5.31) and (12.5.78), we conclude that $c_{\varepsilon t} - \Delta c_\varepsilon = g_\varepsilon$ is bounded in $L^{\frac{5}{4}}(\Omega \times (0, T))$ for any $\varepsilon \in (0, 1)$, we may invoke the standard parabolic regularity theory to infer that $(c_\varepsilon)_{\varepsilon\in(0,1)}$ is bounded in $L^{\frac{5}{4}}((0, T); W^{2,\frac{5}{4}}(\Omega))$. Thus, by virtue of (12.5.67) and the Aubin–Lions lemma we derive that the relative compactness of $(c_\varepsilon)_{\varepsilon\in(0,1)}$ in $L^{\frac{5}{4}}((0, T); W^{1,\frac{5}{4}}(\Omega))$. We can pick an appropriate subsequence which is still written as $(\varepsilon_j)_{j\in\mathbb{N}}$ such that $\nabla c_{\varepsilon_j} \to z_2$ in $L^{\frac{5}{4}}(\Omega \times (0, T))$ for all $T \in (0, \infty)$ and some $z_2 \in L^{\frac{5}{4}}(\Omega \times (0, T))$ as $j \to \infty$, hence $\nabla c_{\varepsilon_j} \to z_2$ a.e. in $\Omega \times (0, \infty)$ as $j \to \infty$. In view of (12.5.88) and the Egorov theorem we conclude that $z_2 = \nabla c$, and whence

$$\nabla c_\varepsilon \to \nabla c \quad \text{a.e. in } \ \Omega \times (0, \infty) \ \text{ as } \ \varepsilon = \varepsilon_j \searrow 0. \tag{12.5.94}$$

In the following, we shall prove (n, c, u) is a weak solution of problem (12.5.4) in Definition 12.5.1. In fact, with the help of (12.5.86)–(12.5.89), (12.5.93), we can derive (12.5.5). Now, by the nonnegativity of n_ε and c_ε, we derive $n \geq 0$ and $c \geq 0$. Next, due to (12.5.89) and $\nabla \cdot u_\varepsilon = 0$, we conclude that $\nabla \cdot u = 0$ a.e. in $\Omega \times (0, \infty)$. On the other hand, in view of (12.5.19) and (12.5.31), we can infer from (12.5.77) that

$$n_\varepsilon \nabla c_\varepsilon \rightharpoonup z_3 \quad \text{ in } \ L^{\frac{8(m-1)}{3}}(\Omega \times (0, T)) \ \text{ for each } \ T \in (0, \infty).$$

Next, due to (12.5.86), (12.5.90) and (12.5.94), we derive that

$$n_\varepsilon \nabla c_\varepsilon \to n \nabla c \quad \text{a.e. in } \ \Omega \times (0, \infty) \ \text{ as } \ \varepsilon = \varepsilon_j \searrow 0. \tag{12.5.95}$$

Therefore, we invoke the Egorov theorem again and obtain $z_3 = n\nabla c$, and hence

$$n_\varepsilon \nabla c_\varepsilon \rightharpoonup n \nabla c \quad \text{ in } \ L^{\frac{8(m-1)}{3}}(\Omega \times (0, T)) \ \text{ for each } \ T \in (0, \infty). \tag{12.5.96}$$

Next, due to $\frac{3}{8(m-1)} + \frac{3}{10} < \frac{3}{4}$, in view of (12.5.92) and (12.5.93), we also infer that for each $T \in (0, \infty)$

$$n_\varepsilon u_\varepsilon \rightharpoonup z_4 \ \text{ in } \ L^{\frac{4}{3}}(\Omega \times (0, T)) \ \text{ as } \ \varepsilon = \varepsilon_j \searrow 0,$$

and moreover, (12.5.87) and (12.5.90) imply that

$$n_\varepsilon u_\varepsilon \to nu \text{ a.e. in } \Omega \times (0, \infty) \text{ as } \varepsilon = \varepsilon_j \searrow 0, \tag{12.5.97}$$

which along with the Egorov theorem implies that

$$n_\varepsilon u_\varepsilon \rightharpoonup nu \text{ in } L^{\frac{4}{3}}(\Omega \times (0, T)) \text{ as } \varepsilon = \varepsilon_j \searrow 0 \tag{12.5.98}$$

for each $T \in (0, \infty)$. As a straightforward consequence of (12.5.86) and (12.5.87), it holds that

$$c_\varepsilon u_\varepsilon \to cu \text{ in } L^1_{loc}(\bar{\Omega} \times (0, \infty)) \text{ as } \varepsilon = \varepsilon_j \searrow 0. \tag{12.5.99}$$

Next, by (12.5.87) and using the fact that $\|Y_\varepsilon \varphi\|_{L^2(\Omega)} \le \|\varphi\|_{L^2(\Omega)} (\varphi \in L^2_\sigma(\Omega))$and $Y_\varepsilon \varphi \to \varphi$ in $L^2(\Omega)$ as $\varepsilon \searrow 0$, we derive that there exists a positive constant C_2 such that

$$\begin{aligned} \|Y_\varepsilon u_\varepsilon(\cdot, t) - u(\cdot, t)\|_{L^2(\Omega)} &\le \|Y_\varepsilon[u_\varepsilon(\cdot, t) - u(\cdot, t)]\|_{L^2(\Omega)} \\ &\quad + \|Y_\varepsilon u(\cdot, t) - u(\cdot, t)\|_{L^2(\Omega)} \\ &\le \|u_\varepsilon(\cdot, t) - u(\cdot, t)\|_{L^2(\Omega)} + \|Y_\varepsilon u(\cdot, t) - u(\cdot, t)\|_{L^2(\Omega)} \\ &\to 0 \text{ as } \varepsilon = \varepsilon_j \searrow 0 \end{aligned} \tag{12.5.100}$$

and

$$\begin{aligned} \|Y_\varepsilon u_\varepsilon(\cdot, t) - u(\cdot, t)\|^2_{L^2(\Omega)} &\le \left(\|Y_\varepsilon u_\varepsilon(\cdot, t)|\|_{L^2(\Omega)} + \|u(\cdot, t)|\|_{L^2(\Omega)}\right)^2 \\ &\le \left(\|u_\varepsilon(\cdot, t)|\|_{L^2(\Omega)} + \|u(\cdot, t)|\|_{L^2(\Omega)}\right)^2 \\ &\le C_2 \text{ for all } t \in (0, \infty) \text{ and } \varepsilon \in (0, 1). \end{aligned} \tag{12.5.101}$$

Now, thus, by (12.5.87), (12.5.100) and (12.5.101) and the dominated convergence theorem, we derive that

$$\int_0^T \|Y_\varepsilon u_\varepsilon(\cdot, t) - u(\cdot, t)\|^2_{L^2(\Omega)} dt \to 0 \text{ as } \varepsilon = \varepsilon_j \searrow 0 \text{ for all } T > 0, \tag{12.5.102}$$

which implies that

$$Y_\varepsilon u_\varepsilon \to u \text{ in } L^2_{loc}([0, \infty); L^2(\Omega)). \tag{12.5.103}$$

Now, combining (12.5.87) with (12.5.103), we derive

$$Y_\varepsilon u_\varepsilon \otimes u_\varepsilon \to u \otimes u \text{ in } L^1_{loc}(\bar{\Omega} \times [0, \infty)) \text{ as } \varepsilon = \varepsilon_j \searrow 0. \tag{12.5.104}$$

Therefore, by (12.5.96)–(12.5.99) and (12.5.104) we conclude that the integrability of $n\nabla c, nu$ and $cu, u \otimes u$ in (12.5.6). Finally, according to (12.5.86)–(12.5.99) and (12.5.103)–(12.5.104), we may pass to the limit in the respective weak formulations associated with the regularized system (12.5.13) and get the integral identities (12.5.7)–(12.5.9).

12.6 Open Problem

This paper has proposed an overview and critical analysis on the qualitative study of mathematical problems for models which is related to Chemotaxis-(Navier)-Stokes System. These new models appear to be of interest for the applications in various fields of biology. However, this final section is devoted to the indication of research perspectives. This aim is pursued by bringing to the reader's attention four key questions. More in detail, the following questions need to be discussed:

(i) Is the b^* best or not in (12.2.14)? Can one rigorously prove the finite-time blow-up of solutions to problem (12.2.14).
(ii) Is the b^* best or not in (12.3.1)? Can one rigorously prove the finite-time blow-up of solutions to problem (12.3.1).
(iii) Is it possible to determine whether or not in the spatial three-dimensional case some unbounded solutions to the chemotaxis-growth model (12.2.63) may exist when $f(u) = au - bu^2$ and $b < \frac{(N-2)_+}{N}\chi C_{\frac{N}{2}+1}^{\frac{1}{\frac{N}{2}+1}}$?
(iv) Is it possible to get the global existence or even boundedness to chemotaxis-haptotaxis model (12.4.1) with remodeling of non-diffusible attractant **in higher dimensions (**$N \geq 3$**)**.

Acknowledgements This work is partially supported by the Shandong Provincial Science Foundation for Outstanding Youth (No. ZR2018JL005), the National Natural Science Foundation of China (No. 11601215), the Natural Science Foundation of Shandong Province of China (No. ZR2016AQ17), and the Doctor Start-up Funding of Ludong University (No. LA2016006).

References

1. N.D. Alikakos, L^p bounds of solutions of reaction-diffusion equations. Commun. Partial Differ. Equ. **4**, 827–868 (1979)
2. X. Bai, M. Winkler, Equilibration in a fully parabolic two-species chemotaxis system with competitive kinetics. Ind. Univ. Math. J. **65**, 553–583 (2016)
3. N. Bellomo, A. Belloquid, Y. Tao, M. Winkler, Toward a mathematical theory of Keller–Segel models of pattern formation in biological tissues. Math. Models Methods Appl. Sci. **25**(9), 1663–1763 (2015)
4. T. Black, Sublinear signal production in a two-dimensional Keller–Segel–Stokes system. Nonlinear Anal. Real World Appl. **31**, 593–609 (2016)

5. M. Burger, M. Di Francesco, Y. Dolak-Struss, The Keller–Segel model for chemotaxis with prevention of overcrowding: linear vs nonlinear diffusion. SIAM J. Math. Anal. **38**, 1288–1315 (2007)
6. V. Calvez, J.A. Carrillo, Volume effects in the Keller–Segel model: energy estimates preventing blow-up. J. Math. Pures Appl. **9**(86), 155–175 (2006)
7. X. Cao, Boundedness in a quasilinear parabolic–parabolic Keller–Segel system with logistic source. J. Math. Anal. Appl. **412**, 181–188 (2014)
8. X. Cao, Boundedness in a three-dimensional chemotaxis–haptotaxis model. Z. Angew. Math. Phys. **67**(1), 1–13 (2016)
9. X. Cao, S. Zheng, Boundedness of solutions to a quasilinear parabolic–elliptic Keller–Segel system with logistic source. Math. Methods Appl. Sci. **37**, 2326–2330 (2014)
10. M. Chae, K. Kang, J. Lee, Global existence and temporal decay in Keller–Segel models coupled to fluid equations. Commun. Partial Differ. Equ. **39**, 1205–1235 (2014)
11. M.A.J. Chaplain, A.R.A. Anderson, Mathematical modelling of tissue invasion, in *Cancer Modelling and Simulation*, ed. by L. Preziosi (Chapman Hall/CRC, Boca Raton, 2003), pp. 269–297
12. M.A.J. Chaplain, G. Lolas, Mathematical modelling of cancer invasion of tissue: the role of the urokinase plasminogen activation system. Math. Models Methods Appl. Sci. **11**, 1685–1734 (2005)
13. M.A.J. Chaplain, G. Lolas, Mathematical modelling of cancer invasion of tissue: dynamic heterogeneity. Netw. Heterog. Media **1**, 399–439 (2006)
14. T. Cieślak, P. Laurençot, Finite time blow-up for a one-dimensional quasilinear parabolic–parabolic chemotaxis system. Ann. Inst. Henri Poincaré Non Linear Anal. **27**, 437–446 (2010)
15. T. Cieślak, C. Stinner, Finite-time blowup and global-in-time unbounded solutions to a parabolic–parabolic quasilinear Keller–Segel system in higher dimensions. J. Differ. Equ. **252**, 5832–5851 (2012)
16. T. Cieślak, C. Stinner, Finite-time blow-up in a supercritical quasilinear parabolic–parabolic Keller–Segel system in dimension 2. Acta Appl. Math. **129**, 135–146 (2014)
17. T. Cieślak, C. Stinner, New critical exponents in a fully parabolic quasilinear Keller–Segel system and applications to volume filling models. J. Differ. Equ. **258**, 2080–2113 (2015)
18. T. Cieślak, M. Winkler, Finite-time blow-up in a quasilinear system of chemotaxis. Nonlinearity **21**, 1057–1076 (2008)
19. L. Corrias, B. Perthame, H. Zaag, A chemotaxis model motivated by angiogenesis. C. R. Math. **336**, 141–146 (2003)
20. L. Corrias, B. Perthame, H. Zaag, Global solutions of some chemotaxis and angiogenesis systems in high space dimensions. Milan J. Math. **72**, 1–28 (2004)
21. M. Di Francesco, A. Lorz, P. Markowich, Chemotaxis–fluid coupled model for swimming bacteria with nonlinear diffusion: global existence and asymptotic behavior. Discrete Contin. Dyn. Syst. **28**, 1437–1453 (2010)
22. C. Dombrowski, L. Cisneros, S. Chatkaew, R.E. Goldstein, J.O. Kessler, Self-concentration and large-scale coherence in bacterial dynamics. Phys. Rev. Lett. **93**(9), 098103 (2004)
23. R. Duan, Z.Y. Xiang, A note on global existence for the chemotaxis–Stokes model with nonlinear diffusion. Int. Math. Res. Not. **7**, 1833–1852 (2014)
24. R. Duan, A. Lorz, Markowich, P.A.: Global solutions to the coupled chemotaxis-fluid equations. Commun. Partial Differ. Equ. **35**, 1635–1673 (2010)
25. H. Gajewski, K. Zacharias, Global behavior of a reaction-diffusion system modeling chemotaxis. Math. Nachr. **195**, 177–194 (1998)
26. Y. Giga, Solutions for semilinear parabolic equations in L^p and regularity of weak solutions of the Navier–Stokes system. J. Differ. Equ. **61**, 186–212 (1986)
27. D. Gilbarg, N. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd edn. (Springer, Berlin, 1983)
28. H. Hajaiej, L. Molinet, T. Ozawa, B. Wang, Necessary and sufficient conditions for the fractional Gagliardo–Nirenberg inequalities and applications to Navier–Stokes and generalized boson equations, in *Harmonic Analysis and Nonlinear Partial Differential Equations*, RIMS

Kôkyûroku Bessatsu, vol. B26 (Research Institute for Mathematical Sciences (RIMS), Kyoto, 2011), pp. 159–175
29. D.D. Haroske, H. Triebel, *Distributions, Sobolev Spaces, Elliptic Equations* (European Mathematical Society, Zurich, 2008)
30. M.A. Herrero, J.J.L. Velázquez, Singularity patterns in a chemotaxis model. Math. Ann. **306**, 583–623 (1996)
31. M. Herrero, J. Velázquez, Chemotactic collapse for the Keller–Segel model. J. Math. Biol. **35**, 177–194 (1996)
32. M. Herrero, J. Velázquez, A blow-up mechanism for a chemotaxis model. Ann. Scuola Norm. Super. Pisa Cl. Sci. **24**(4), 633–683 (1997)
33. M. Hieber, J. Prüss, Heat kernels and maximal L^p-L^q estimate for parabolic evolution equations. Commun. Partial Differ. Equ. **22**, 1647–1669 (1997)
34. T. Hillen, K. Painter, Global existence for a parabolic chemotaxis model with prevention of overcrowding. Adv. Appl. Math. **26**, 281–301 (2001)
35. T. Hillen, K.J. Painter, A user's guide to PDE models for chemotaxis. J. Math. Biol. **58**, 183–217 (2009)
36. T. Hillen, K.J. Painter, M. Winkler, Convergence of a cancer invasion model to a logistic chemotaxis model. Math. Models Methods Appl. Sci. **23**, 165–198 (2013)
37. D. Horstmann, From 1970 until present: the Keller–Segel model in chemotaxis and its consequences, I. Jahresberichte der Deutschen Mathematiker-Vereinigung **105**, 103–165 (2003)
38. D. Horstmann, G. Wang, Blow-up in a chemotaxis model without symmetry assumptions. Eur. J. Appl. Math. **12**, 159–177 (2001)
39. D. Horstmann, M. Winkler, Boundedness vs. blow-up in a chemotaxis system. J. Differ. Equ. **215**, 52–107 (2005)
40. S. Ishida, Global existence and boundedness for chemotaxis–Navier–Stokes system with position-dependent sensitivity in $2d$ bounded domains. Discrete Contin. Dyn. Syst. Ser. A **32**, 3463–3482 (2015)
41. S. Ishida, K. Seki, T, Yokota, Boundedness in quasilinear Keller–Segel systems of parabolic–parabolic type on non-convex bounded domains. J. Differ. Equ. **256**, 2993–3010 (2014)
42. W. Jäger, S. Luckhaus, On explosions of solutions to a system of partial differential equations modelling chemotaxis. Trans. Am. Math. Soc. **329**, 819–824 (1992)
43. E. Keller, L. Segel, Model for chemotaxis. J. Theor. Biol. **30**, 225–234 (1970)
44. E. Keller, L. Segel, Initiation of slime mold aggregation viewed as an instability. J. Theor. Biol. **26**, 399–415 (1970)
45. E. Keller, L. Segel, Traveling bands of chemotactic bacteria: a theoretical analysis. J. Theor. Biol. **30**, 377–380 (1971)
46. A. Kiselev, L. Ryzhik, Biomixing by chemotaxis and enhancement of biological reactions. Commun. Partial Differ Equ. **37**(2), 298–318 (2012)
47. R. Kowalczyk, Preventing blow-up in a chemotaxis model. J. Math. Anal. Appl. **305**, 566–585 (2005)
48. O.A. Ladyzenskaja, V.A. Solonnikov, N.N. Ural'eva, *Linear and Quasi-Linear Equations of Parabolic Type*. American Mathematical Society Translation, vol. 23 (AMS, Providence, 1968)
49. J. Lankeit, Eventual smoothness and asymptotics in a three dimensional chemotaxis system with logistic source. J. Differ. Equ. **258**(4), 1158–1191 (2015)
50. T. Li, A. Suen, C. Xue, M. Winkler, Global small-data solutions of a two-dimensional chemotaxis system with rotational ux term. Math. Models Methods Appl. Sci. **25**, 721–746 (2015)
51. X. Li, Z. Xiang, Boundedness in quasilinear Keller-Segel equations with nonlinear sensitivity and logistic source. Discrete Contin. Dyn. Syst. **35**, 3503–3531 (2015)
52. X. Li, Y. Wang, Z. Xiang, Global existence and boundedness in a 2D Keller–Segel-Stokes system with nonlinear diffusion and rotational flux. Commun. Math. Sci. **14**, 1889–1910 (2016)

53. G. Liţanu, C. Morales-Rodrigo, Asymptotic behavior of global solutions to a model of cell invasion. Math. Models Methods Appl. Sci. **20**, 1721–1758 (2010)
54. J.-G. Liu, A. Lorz, A coupled chemotaxis–fluid model: global existence. Ann. Inst. Henri Poincaré Anal. Non Linéaire **28**(5), 643–652 (2011)
55. J. Liu, J. Zheng, Y. Wang, Boundedness in a quasilinear chemotaxis-haptotaxis system with logistic source. Z. Angew. Math. Phys. **67**(2), 1–33 (2016)
56. A. Lorz, Coupled chemotaxis fluid model. Math. Models Methods Appl. Sci. **20**, 987–1004 (2010)
57. A. Marciniak-Czochra, M. Ptashnyk, Boundedness of solutions of a haptotaxis model. Math. Models Methods Appl. Sci. **20**, 449–476 (2010)
58. G. Meral, C. Stinner, C. Surulescu, On a multiscale model involving cell contractivity and its effects on tumor invasion. Discrete Contin. Dyn. Syst. Ser. B **20**, 189–213 (2015)
59. M. Mizukami, Remarks on smallness of chemotactic effect for asymptotic stability in a two-species chemotaxis system. AIMS Math. **1**(3), 156–164 (2016)
60. N. Mizoguchi, P. Souplet, Nondegeneracy of blow-up points for the parabolic Keller–Segel system. Ann. Inst. Henri Poincaré Non Linéaire **31**(4), 851–875 (2014)
61. M. Mizukami, T. Yokota, Global existence and asymptotic stability of solutions to a two-species chemotaxis system with any chemical diffusion. J. Differ. Equ. **261**, 2650–2669 (2016)
62. J. Murray, *Mathematical Biology: II. Spatial Models and Biomedical Applications*, 3rd edn. (Springer, New York, 2003)
63. T. Nagai, T. Senba, K. Yoshida, Application of the Trudinger-Moser inequality to a parabolic system of chemotaxis. Funkc. Ekvacioj Ser. Int. **40**, 411–433 (1997)
64. K. Osaki, A. Yagi, Finite dimensional attractor for one-dimensional Keller–Segel equations. Funkcial. Ekvac. **44**, 441–469 (2001)
65. K. Osaki, T. Tsujikawa, A. Yagi, M. Mimura, Exponential attractor for a chemotaxis growth system of equations. Nonlinear Anal. Theory Methods Appl. **51**, 119–144 (2002)
66. K.J. Painter, T. Hillen, Volume-filling and quorum-sensing in models for chemosensitive movement. Can. Appl. Math. Q. **10**, 501–543 (2002)
67. C.S. Patlak, Random walk with persistence and external bias. Bull. Math. Biophy. 15, 311–338 (1953)
68. B. Perthame, *Transport Equations in Biology* (Birkhäuser Verlag, Basel, 2007)
69. M. Rascle, C. Ziti, Finite time blow-up in some models of chemotaxis. J. Math. Biol. **33**, 388–414 (1995)
70. J. Simon, Compact sets in the space $L^p(O, T; B)$. Annali di Matematica Pura ed Applicata **146**(1), 65–96 (1986)
71. H. Sohr, *The Navier–Stokes Equations, An Elementary Functional Analytic Approach* (Birkhäuser Verlag, Basel, 2001)
72. C. Stinner, C. Surulescu, M. Winkler, Global weak solutions in a PDE-ODE system modeling multiscale cancer cell invasion. SIAM J. Math. Anal. **46**, 1969–2007 (2014)
73. Y. Tao, Global existence of classical solutions to a combined chemotaxis–haptotaxis model with logistic source. J. Math. Anal. Appl. **354**, 60–69 (2009)
74. Y. Tao, Boundedness in a two-dimensional chemotaxis–haptotaxis system. J. Oceanogr. **70**(70), 165–174 (2014)
75. Y. Tao, M. Wang, Global solution for a chemotactic–haptotactic model of cancer invasion. Nonlinearity **21**, 2221–2238 (2014)
76. J.I. Tello, M. Winkler, A chemotaxis system with logistic source. Commun. Partial Differ. Equ. **32**, 849–877 (2007)
77. Y. Tao, M. Winkler, A chemotaxis-haptotaxis model: the roles of nonlinear diffusion and logisticsource. SIAM J. Math. Anal. **43**, 685–704 (2011)
78. Y. Tao, M. Winkler, Boundedness in a quasilinear parabolic–parabolic Keller–Segel system with subcritical sensitivity. J. Differ. Equ. **252**, 692–715 (2012)
79. Y. Tao, M. Winkler, Global existence and boundedness in a Keller–Segel–Stokes model with arbitrary porous medium diffusion. Discrete Contin. Dyn. Syst. Ser. A **32**, 1901–1914 (2012)

80. Y. Tao, M. Wiklner, Eventual smoothness and stabilization of large-data solutions in a three-dimensional chemotaxis system with consumption of chemoattractant. J. Differ. Equ. **252**, 2520–2543 (2012)
81. Y. Tao, M. Winkler, Locally bounded global solutions in a three-dimensional chemotaxis–Stokes system with nonlinear diffusion. Ann. Inst. Henri Poincaré Anal. Non Linéaire **30**, 157–178 (2013)
82. Y. Tao, M. Winkler, Boundedness and stabilization in a multi-dimensional chemotaxis–haptotaxis model. Proc. R. Soc. Edinb. **144**, 1067–1084 (2014)
83. Y. Tao, M. Winkler, Dominance of chemotaxis in a chemotaxis–haptotaxis model. Nonlinearity **27**, 1225–1239 (2014)
84. Y. Tao, M. Winkler, Energy-type estimates and global solvability in a two-dimensional chemotaxis–haptotaxis model with remodeling of non-diffusible attractant. J. Differ. Equ. **257**, 784–815 (2014)
85. Y. Tao, M. Winkler, Boundedness and decay enforced by quadratic degradation in a three-dimensional chemotaxis–fluid system. Z. Angew. Math. Phys. **66**, 2555–2573 (2015)
86. R. Temam, *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*. Applied Mathematical Sciences, 2nd edn. (Springer, New York, 1997)
87. I. Tuval, L. Cisneros, C. Dombrowski, et al., Bacterial swimming and oxygen transport near contact lines. Proc. Natl. Acad. Sci. USA **102**, 2277–2282 (2005)
88. G. Viglialoro, Very weak global solutions to a parabolic-parabolic chemotaxis-system with logistic source. J. Math. Anal. Appl. **439**(1), 197–212 (2016)
89. C. Walker, G.F. Webb, Global existence of classical solutions for a haptotaxis model. SIAM J. Math. Anal. **38**, 1694–1713 (2007)
90. Y. Wang, Boundedness in the higher-dimensional chemotaxis-haptotaxis model with nonlinear diffusion. J. Differ. Equ. **260**(2), 1975–1989 (2016)
91. Y. Wang, Global weak solutions in a three-dimensional Keller-Segel-Navier-Stokes system with subcritical sensitivity. Math. Models Methods Appl. Sci. **27**(14), 2745–2780 (2017)
92. Y. Wang, Y. Ke, Large time behavior of solution to a fully parabolic chemotaxis-haptotaxis model in higher dimensions. J. Differ. Equ. **260**(9), 6960–6988 (2016)
93. Y. Wang, Z. Xiang, Global existence and boundedness in a higher-dimensional quasilinear chemotaxis system. Z. Angew. Math. Phys. **66**(6), 3159–3179 (2015)
94. Y. Wang, Z. Xiang, Global existence and boundedness in a Keller–Segel–Stokes system involving a tensor-valued sensitivity with saturation. J. Differ. Equ. **259**, 7578–7609 (2015)
95. Y. Wang, Z. Xiang, Global existence and boundedness in a Keller–Segel–Stokes system involving a tensor-valued sensitivity with saturation: the 3D case. J. Differ. Equ. **261**, 4944–4973 (2016)
96. Y. Wang, J. Yin, Coexistence periodic solutions of a doubly nonlinear parabolic system with Neumann boundary conditions. J. Math. Anal. Appl. **396**, 704–714 (2012)
97. Z. Wang, M. Winkler, D. Wrzosek, Global regularity vs. infinite-time singularity formation in a chemotaxis model with volume-filling effect and degenerate diffusion. SIAM J. Math. Anal. **44**, 3502–3525 (2012)
98. L. Wang, Y. Li, C. Mu, Boundedness in a parabolic–parabolic quasilinear chemotaxis system with logistic source. Discrete Contin. Dyn. Syst. Ser. A. **34**, 789–802 (2014)
99. L. Wang, C. Mu, S. Zhou, Boundedness in a parabolic–parabolic chemotaxis system with nonlinear diffusion. Z. Angew. Math. Phys. **65**, 1137–1152 (2014)
100. L. Wang, C. Mu, P. Zheng, On a quasilinear parabolic–elliptic chemotaxis system with logistic source. J. Differ. Equ. **256**, 1847–1872 (2014)
101. L. Wang, C. Mu, K. Lin, J. Zhao, Global existence to a higher-dimensional quasilinear chemotaxis system with consumption of chemoattractant. Z. Angew. Math. Phys. **66**(4), 1–16 (2015)
102. M. Winkler, Chemotaxis with logistic source: very weak global solutions and their boundedness properties. J. Math. Anal. Appl. **348**, 708–729 (2008)
103. M. Winkler, Does a volume-filling effect always prevent chemotactic collapse. Math. Methods Appl. Sci. **33**, 12–24 (2010)

104. M. Winkler, Boundedness in the higher-dimensional parabolic–parabolic chemotaxis system with logistic source. Commun. Partial Differ. Equ. **35**, 1516–1537 (2010)
105. M. Winkler, Aggregation vs. global diffusive behavior in the higher-dimensional Keller–Segel model. J. Differ. Equ. **248**, 2889–2905 (2010)
106. M. Winkler, Blow-up in a higher-dimensional chemotaxis system despite logistic growth restriction. J. Math. Anal. Appl. **384**, 261–272 (2011)
107. M. Winkler, Global large-data solutions in a chemotaxis–(Navier–)Stokes system modeling cellular swimming in fluid drops. Commun. Partial Differ. Equ. **37**(2), 319–351 (2012)
108. M. Winkler, Finite-time blow-up in the higher-dimensional parabolic–parabolic Keller–Segel system. J. Math. Pures Appl. **100**, 748–767 (2013)
109. M. Winkler, How far can chemotactic cross-diffusion enforce exceeding carrying capacities? J. Nonlinear Sci. **24**, 809–855 (2014)
110. M. Winkler, Stabilization in a two-dimensional chemotaxis–Navier–Stokes system. Arch. Ration. Mech. Anal. **211**, 455–487 (2014)
111. M. Winkler, Global asymptotic stability of constant equilibriain a fully parabolic chemotaxis system with strong logistic dampening. J. Differ. Equ. **257**, 1056–1077 (2014)
112. M. Winkler, Boundedness and large time behavior in a three-dimensional chemotaxis–Stokes system with nonlinear diffusion and general sensitivity. Cal. Var. Partial Differ. Equ. **54**, 3789–3828 (2015)
113. M. Winkler, Global weak solutions in a three-dimensional chemotaxis–Navier–Stokes system. Ann. Inst. Henri Poincaré Anal. Non Linéaire **33**(5), 1329–1352 (2016)
114. M. Winkler, K.C. Djie, Boundedness and finite-time collapse in a chemotaxis system with volume-filling effect. Nonlinear Anal. Theory Methods Appl. **72**, 1044–1064 (2010)
115. T. Xiang, Boundedness and global existence in the higher-dimensional parabolic–parabolic chemotaxis system with/without growth source. J. Differ. Equ. **258**, 4275–4323 (2015)
116. C. Xue, H.G. Othmer, Multiscale models of taxis-driven patterning in bacterial population. SIAM J. Appl. Math. **70**, 133–167 (2009)
117. Q. Zhang, Y. Li, Global boundedness of solutions to a two-species chemotaxis system. Z. Angew. Math. Phys. **66**(1), 83–93 (2015)
118. Q. Zhang, X. Zheng, Global well-posedness for the two-dimensional incompressible chemotaxis–Navier–Stokes equations. SIAM J. Math. Anal. **46**, 3078–3105 (2014)
119. J. Zheng, A new approach toward locally bounded global solutions to a $3D$ chemotaxis-stokes system with nonlinear diffusion and rotation. arXiv:1701.01334
120. J. Zheng, Boundedness of solutions to a quasilinear parabolic–elliptic Keller–Segel system with logistic source. J. Differ. Equ. **259**(1), 120–140 (2015)
121. J. Zheng, Boundedness of solutions to a quasilinear parabolic–parabolic Keller–Segel system with logistic source. J. Math. Anal. Appl. **431**(2), 867–888 (2015)
122. J. Zheng, Boundedness in a three-dimensional chemotaxis–fluid system involving tensor-valued sensitivity with saturation. J. Math. Anal. Appl. **442**(1), 353–375 (2016)
123. J. Zheng, Boundedness and global asymptotic stability of constant equilibria in a fully parabolic chemotaxis system with nonlinear a logistic source. J. Math. Anal. Appl. **450**, 104–1061 (2017)
124. J. Zheng, A note on boundedness of solutions to a higher-dimensional quasi–linear chemotaxis system with logistic source. Zeitschriftfür Angewandte Mathematik und Mechanik **97**(4), 414–421 (2017)
125. J. Zheng, Boundedness in a two-species quasi-linear chemotaxis system with two chemicals. Topol. Methods Nonlinear Anal. **49**(2), 463–480 (2017)
126. J. Zheng, Boundedness of solution of a higher-dimensional parabolic–ODE–parabolic chemotaxis–haptotaxis model with generalized logistic source. Nonlinearity **30**, 1987–2009 (2017)
127. J. Zheng, Boundedness of solutions to a quasilinear higher-dimensional chemotaxis–haptotaxis model with nonlinear diffusion. Discrete Contin. Dyn. Syst. **37**(1), 627–643 (2017)
128. J. Zheng, Global weak solutions in a three-dimensional Keller-Segel-Navier-Stokes system with nonlinear diffusion. J. Differ. Equ. **263**, 2606–2629 (2017)

129. J. Zheng, Y. Wang, Boundedness and decay behavior in a higher-dimensional quasilinear chemotaxis system with nonlinear logistic source. Comput. Math. Appl. **72**(10), 2604–2619 (2016)
130. J. Zheng, Y. Wang, Boundedness of solutions to a quasilinear chemotaxis–haptotaxis model. Comput. Math. Appl. **71**, 1898–1909 (2016)
131. J. Zheng, Y. Wang, A note on global existence to a higher-dimensional quasilinear chemotaxis system with consumption of chemoattractant. Discrete Contin. Dyn. Syst. Ser. B **22**(2), 669–686 (2017)
132. P. Zheng, C. Mu, X. Song, On the boundedness and decay of solutions for a chemotaxis-haptotaxis system with nonlinear diffusion. Discrete Cont. Dyn. Syst. **36**(3), 1737–1757 (2015)
133. P. Zheng, C. Mu, X. Hu, Boundedness and blow-up for a chemotaxis system with generalized volume-filling effect and logistic source. Discrete Contin. Dyn. Syst. Ser. A **35**, 2299–2323 (2015)
134. J. Zheng, Y. Li, G. Bao, X. Zou, A new result for global existence and boundedness of solutions to a parabolic–parabolic Keller–Segel system with logistic source. J. Math. Anal. Appl. **462**, 1–25 (2018)

Chapter 13
Optimal Control of Quasivariational Inequalities with Applications to Contact Mechanics

Mircea Sofonea

Abstract This chapter deals with the optimal control of a class of elliptic quasi-variational inequalities. We start with an existence and uniqueness result for such inequalities. Then we state an optimal control problem, list the assumptions on the data and prove the existence of optimal pairs. We proceed with a perturbed control problem for which we state and prove a convergence result, under general conditions. Further, we present a relevant particular case for which these conditions are satisfied and, therefore, our convergence result works. Finally, we illustrate the use of these abstract results in the study of a mathematical model which describes the equilibrium of an elastic body in frictional contact with an obstacle, the so-called foundation. The process is static and the contact is modeled with normal compliance and unilateral constraint, associated with the Coulomb's law of dry friction. We prove the existence, uniqueness, and convergence results together with the corresponding mechanical interpretation. We illustrate these results in the study of a one-dimensional example. Finally, we end this chapter with some concluding remarks.

Keywords Quasivariational inequality · Optimal pair · Optimal control · Convergence results · Frictional contact · Unilateral constraint · Weak solution

13.1 Introduction

Variational inequalities represent a powerful mathematical tool used in the study of various nonlinear boundary value problems with partial differential equations. They are usually formulated by using a set of constraints, a nonlinear operator, and a convex function which could be nondifferentiable. Quasivariational inequalities

M. Sofonea (✉)
Laboratoire de Mathématiques et Physique, University of Perpignan Via Domitia, Perpignan, France
e-mail: sofonea@univ-perp.fr

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_13

represent a special class of variational inequalities in which the convex function depends on the solution. The theory of variational inequalities was developed based on arguments of monotonicity and convexity, including properties of the subdifferential of a convex function. Basic references in the field are [4, 10, 17, 22, 25], for instance. Application of variational inequalities in mechanics could be found in the books [14, 19–21, 36], for instance.

The optimal control theory deals with the existence and, when possible, the uniqueness of optimal pairs and optimal control. It also deals with the derivation of necessary conditions of optimality or, better, necessary and sufficient conditions of optimality. This means to find an equation or an inequality which characterizes the optimal control. Basic references for the optimal control of systems governed by partial differential equations are the books [24, 35]. Application of the optimal control theory in mechanics could be found in [1, 2, 13, 31], for instance. Optimal control problems for variational inequalities have been discussed in several works, including [7, 9, 16, 29, 30, 34, 43]. Due to the nonsmooth and nonconvex feature of the functional involved, the treatment of optimal control problems for variational inequalities requires the use of their approximation by smooth optimization problems. And, on this matter, establishing convergence results for the optimal pairs represents a topic of major interest.

Processes of contact between deformable bodies abound in industry and everyday life. A few simple examples are brake pads in contact with wheels, tires on roads, and pistons with skirts. Due to the complex phenomena involved, they lead to strongly nonlinear mathematical models, formulated in terms of various classes of inequalities, including variational and quasivariational inequalities. Because of the importance of contact processes in structural and mechanical systems, considerable effort has been put into their modeling, analysis, and numerical simulations and the literature in the field is extensive. It includes the books [14, 15, 19, 23, 32, 36, 38, 41, 42], for instance. The literature concerning optimal control problems in the study of mathematical models of contact is quite limited. The reason is the strong nonlinearities which arise in the boundary conditions included in such models. The results on optimal control for various contact problems with elastic materials can be found in [3, 6, 8, 11, 12, 26–28, 44] and the references therein.

In the current chapter we consider an optimal control problem for a general class of elliptic quasivariational inequalities. Our motivation is given by the fact that such kind of inequalities arises in the study of frictional contact models and, therefore, their optimal control is important in a large number of engineering applications. The functional framework is the following: let X and Y be real Hilbert spaces endowed with the inner products $(\cdot,\cdot)_X$ and $(\cdot,\cdot)_Y$, respectively, $K \subset X$, $A : X \to X$, $j : X \times X \to \mathbb{R}$, and $\pi : X \to Y$. Then, the inequality problem we consider is the following.

Problem $\mathcal{P}$. Given $f \in Y$, find u such that

$$u \in K, \qquad (Au, v-u)_X + j(u,v) - j(u,u) \geq (f, \pi v - \pi u)_Y \qquad \forall\, v \in K. \tag{13.1}$$

Note that the function j depends on the solution u and, for this reason, we refer to (13.1) as a quasivariational inequality. We assume in what follows that for each $f \in Y$ the quasivariational inequality (13.1) has a unique solution $u = u(f)$. Sufficient conditions on the data which guarantee this assumption will be provided in Theorem 13.2.12. The set of admissible pairs for inequality (13.1) is given by

$$\mathcal{V}_{ad} = \{\, (u, f) \in K \times Y \text{ such that (13.1) holds}\,\}. \tag{13.2}$$

Consider now a cost functional $\mathcal{L} : X \times Y \to \mathbb{R}$, where, here and below, $X \times Y$ represents the product of the Hilbert spaces X and Y, equipped with the canonical inner product. Then, the optimal control problem we study in this chapter is the following.

Problem $\mathcal{Q}$. Find $(u^*, f^*) \in \mathcal{V}_{ad}$ such that

$$\mathcal{L}(u^*, f^*) = \min_{(u,f)\in\mathcal{V}_{ad}} \mathcal{L}(u, f). \tag{13.3}$$

Our aim in this chapter is threefold. The first one is to formulate sufficient assumptions on the data which guarantee the existence of optimal pairs, i.e., elements $(u^*, f^*) \in \mathcal{V}_{ad}$ which solve Problem $\mathcal{Q}$. The answer to this question is provided by Theorem 13.3.1. The second aim is to study the dependence of the optimal pairs with respect to perturbations of the set K, the operator A, and the functional j. The answer to this question is provided by Theorem 13.3.5 which provides a convergence result, under general conditions. This result is completed by Theorem 13.3.8, which holds under specific conditions on the data. Finally, our third aim is to illustrate how these abstract results could be useful in the study of mathematical models of contact. The answer to this question is provided by Theorems 13.4.4–13.4.6 and the corresponding mechanical interpretation.

The rest of this chapter is structured in four sections, as follows: In Sect. 13.2 we provide some preliminary results in the study of Problem $\mathcal{P}$. They concern the existence, uniqueness, and convergence of the solution. Then, in Sect. 13.3 we state and prove the existence of optimal pairs to the control problem $\mathcal{Q}$ as well as a general convergence result. Next, we present a relevant particular case for which our convergence result holds. In Sect. 13.4 we consider a mathematical model of frictional contact with elastic materials. The process is static and the contact is described with normal compliance and unilateral constraint, associated with a version of Coulomb's law of dry friction. We apply our results in Sects. 13.2 and 13.3 in the study of this problem. Moreover, we illustrate them in the study of a one-dimensional example. Finally, we end this chapter with some concluding remarks, presented in Sect. 13.5.

13.2 Quasivariational Inequalities

In this section we provide some results in the study of Problem $\mathcal{P}$ that we need in the rest of this chapter. We first introduce preliminary material from functional analysis, and then we state and prove an existence and uniqueness result, Theorem 13.2.12. Finally, we study the dependence of the solution with respect to the element f and we prove a convergence result, Theorem 13.2.13.

13.2.1 Notation and Preliminaries

All the linear spaces considered in this chapter including abstract normed spaces, Banach spaces, Hilbert spaces, and various function spaces are assumed to be real linear spaces. For a normed space X we denote by $\|\cdot\|_X$ its norm and by 0_X its zero element. In addition, we denote by $\to$ and $\rightharpoonup$ the strong and weak convergence in various normed spaces. For an inner product space X we denote by $(\cdot,\cdot)_X$ its inner product and by $\|\cdot\|_X$ the associated norm. Unless stated otherwise, all the limits, upper and lower limits, below are considered as $n \to \infty$, even if we do not mention it explicitly. The results presented below in this subsection are well known and can be found in many books and survey and, for this reason, we skip their proofs.

Definition 13.2.1 Let X be a normed space. A subset $K \subset X$ is called:

(i) *(strongly) closed* if the limit of each convergent sequence of elements of K belongs to K, that is, $\{u_n\} \subset K,\ u_n \to u$ in $X \implies u \in K$.
(ii) *weakly closed* if the limit of each weakly convergent sequence of elements of K belongs to K, that is, $\{u_n\} \subset K,\ u_n \rightharpoonup u$ in $X \implies u \in K$.
(iii) *convex*, if $u, v \in K \implies (1-t)\,u + t\,v \in K \quad \forall\, t \in [0,1]$.

Evidently, every weakly closed subset of X is (strongly) closed, but the converse is not true, in general. An exception is provided by the class of convex subsets of a Banach space, as shown in the following result.

Theorem 13.2.2 (The Mazur Theorem) *A convex subset of a Banach space is (strongly) closed if and only if it is weakly closed.*

We now recall the following important property which represents a particular case of the well-known Eberlein–Smulyan theorem.

Theorem 13.2.3 *If X is a Hilbert space, then any bounded sequence in X has a weakly convergent subsequence.*

It follows that if X is a Hilbert space and the sequence $\{u_n\} \subset X$ is bounded, that is, $\sup_n \|u_n\|_X < \infty$, then there exists a subsequence $\{u_{n_k}\} \subset \{u_n\}$ and an element $u \in X$ such that $u_{n_k} \rightharpoonup u$ in X. Furthermore, if the limit u is independent of the subsequence, then the whole sequence $\{u_n\}$ converges weakly to u, as stated in the following result.

Theorem 13.2.4 *Let X be a Hilbert space and let $\{u_n\}$ be a bounded sequence of elements in X such that each weakly convergent subsequence of $\{u_n\}$ converges weakly to the same limit $u \in X$. Then $u_n \rightharpoonup u$ in X.*

We now proceed with the definition of some classes of operators.

Definition 13.2.5 Let X be an inner product space and let $A : X \to X$ be an operator. The operator A is said to be:

(i) *monotone*, if $(Au - Av, u - v)_X \geq 0 \quad \forall\, u,\ v \in X$.
(ii) *strongly monotone*, if there exists a constant $m > 0$ such that

$$(Au - Av, u - v)_X \geq m\, \|u - v\|_X^2 \quad \forall\, u,\ v \in X.$$

(iii) *bounded*, if A maps bounded sets into bounded sets.
(iv) *pseudomonotone*, if it is bounded and $u_n \rightharpoonup u$ in X with

$$\limsup_{n\to\infty} (Au_n, u_n - u)_X \leq 0 \tag{13.4}$$

implies

$$\liminf_{n\to\infty} (Au_n, u_n - v)_X \geq (Au, u - v)_X \quad \forall\, v \in X. \tag{13.5}$$

(v) *Lipschitz continuous* if there exists $M > 0$ such that

$$\|Au - Av\|_X \leq M\|u - v\|_X \quad \forall\, u,\ v \in X.$$

(vi) *hemicontinuous* if the real valued function

$$\theta \mapsto (A(u + \theta v), w)_X \quad \text{is continuous on } \mathbb{R}, \quad \forall\, u,\ v,\ w \in X.$$

It is easy to see that a strongly monotone operator $A : X \to X$ is monotone and a Lipschitz continuous operator $A : X \to X$ is bounded and hemicontinuous. Moreover, the following result holds.

Proposition 13.2.6 *Let X be an inner product space and $A : X \to X$ a monotone hemicontinuous operator. Assume that $\{u_n\}$ is a sequence of elements in X which converges weakly to the element $u \in X$ such that (13.4) holds. Then (13.5) holds, too.*

A proof of Proposition 13.2.6 can be found in [41, p. 21]. As a consequence we obtain the following result which will be used later in this chapter.

Corollary 13.2.7 *Let X be an inner product space and $A : X \to X$ a monotone Lipschitz continuous operator. Then A is pseudomonotone.*

Convex lower semicontinuous functions represent a crucial ingredient in the study of variational inequalities. To introduce them, we start with the following definitions.

Definition 13.2.8 Let X be a linear space and let K be a nonempty convex subset of X. A function $\varphi : K \to \mathbb{R}$ is said to be *convex* if

$$\varphi((1-t)u + tv) \leq (1-t)\varphi(u) + t\varphi(v) \tag{13.6}$$

for all $u, v \in K$ and $t \in [0, 1]$. The function φ is *strictly convex* if the inequality in (13.6) is strict for $u \neq v$ and $t \in (0, 1)$.

Definition 13.2.9 Let X be a normed space and let K be a nonempty closed convex subset of X. A function $\varphi : K \to \mathbb{R}$ is said to be *lower semicontinuous (l.s.c.)* at $u \in K$ if

$$\liminf_{n\to\infty} \varphi(u_n) \geq \varphi(u) \tag{13.7}$$

for each sequence $\{u_n\} \subset K$ converging to u in X. The function φ is *l.s.c.* if it is l.s.c. at every point $u \in K$. When inequality (13.7) holds for each sequence $\{u_n\} \subset K$ that converges weakly to u, the function φ is said to be *weakly lower semicontinuous* at u. The function φ *is weakly l.s.c.* if it is weakly l.s.c. at every point $u \in K$.

Since the strong convergence implies the weak convergence, it follows that a weakly lower semicontinuous function is lower semicontinuous. Moreover, the following results hold.

Proposition 13.2.10 *Let X be a Banach space, K a nonempty closed convex subset of X, and $\varphi : K \to \mathbb{R}$ a convex function. Then φ is lower semicontinuous if and only if it is weakly lower semicontinuous.*

The proof of this result is based on Theorem 13.2.2.

13.2.2 Existence and Uniqueness

Everywhere in the rest of this chapter we assume that X is a Hilbert space. Given a subset $K \subset X$, an operator $A : X \to X$, a function $j : X \times X \to \mathbb{R}$, and an element $\widetilde{f} \in X$, we consider the following quasivariational inequality problem: find an element u such that

$$u \in K, \quad (Au, v-u)_X + j(u, v) - j(u, u) \geq \big(\widetilde{f}, v-u\big)_X \quad \forall\, v \in K. \tag{13.8}$$

Quasivariational inequalities of the form (13.8) have been studied by many authors, by using different functional methods, including fixed point and topological degree arguments. The existence and uniqueness results for such inequalities could be found in [12, 33, 40, 41], for instance, under various assumptions on the function j. Here, in this chapter, we consider the following assumptions:

$$K \text{ is a nonempty, closed, convex subset of } X. \tag{13.9}$$

$$\begin{cases} A \text{ is a strongly monotone Lipschitz continuous operator, i.e.,} \\ \text{there exist } m > 0 \text{ and } M > 0 \text{ such that} \\ \text{(a)} \ \ (Au - Av, u - v)_X \geq m\|u - v\|_X^2 \quad \forall\, u,\ v \in X, \\ \text{(b)} \ \ \|Au - Av\|_X \leq M\, \|u - v\|_X \quad \forall\, u,\ v \in X. \end{cases} \tag{13.10}$$

$$\begin{cases} \text{(a) For all } \eta \in X,\ \ j(\eta, \cdot) : X \to \mathbb{R} \text{ is convex and l.s.c.} \\ \text{(b) There exists } \alpha \geq 0 \text{ such that} \\ \qquad j(\eta_1, v_2) - j(\eta_1, v_1) + j(\eta_2, v_1) - j(\eta_2, v_2) \\ \qquad\quad \leq \alpha\, \|\eta_1 - \eta_2\|_X \|v_1 - v_2\|_X \quad \forall\, \eta_1,\ \eta_2,\ v_1,\ v_2 \in X. \end{cases} \tag{13.11}$$

$$m > \alpha. \tag{13.12}$$

We recall the following existence and uniqueness result, which guarantees the unique solvability of Problem $\mathcal{P}$.

Theorem 13.2.11 *Assume that (13.9)–(13.12) hold. Then, for each $\widetilde{f} \in X$ the quasivariational inequality (13.8) has a unique solution.*

A proof of Theorem 13.2.11 can be found in [41, p. 49], based on the Banach fixed point argument. We now turn to the study of Problem $\mathcal{P}$ and, to this end, we consider the following additional assumptions:

$$\begin{cases} \pi \text{ is a linear continuous operator, i.e.,} \\ \quad \text{there exists } c_0 > 0 \text{ such that} \\ \qquad \|\pi v\|_Y \leq c_0\, \|v\|_X \quad \forall\, v \in X. \end{cases} \tag{13.13}$$

$$\begin{cases} \text{There exist } \beta,\ \gamma \geq 0 \text{ such that} \\ \quad j(\eta, v_1) - j(\eta, v_2) \leq (\beta + \gamma \|\eta\|_X)\, \|v_1 - v_2\|_X \\ \qquad \forall\, \eta,\ v_1,\ v_2 \in X. \end{cases} \tag{13.14}$$

$$m > \gamma. \tag{13.15}$$

We have the following result.

Theorem 13.2.12 *Assume that (13.9)–(13.13) hold. Then, for each $f \in Y$, the quasivariational inequality (13.1) has a unique solution. Moreover, if (13.14)* *and (13.15) hold, then the solution satisfies the inequality*

$$\|u\|_X \leq \frac{1}{m-\gamma}\big(\|Au_0\|_X + c_0\|f\|_Y + \gamma\|u_0\|_X + \beta\big) + \|u_0\|_X, \tag{13.16}$$

for any element $u_0 \in K$.

Proof Let $f \in Y$. We use assumption (13.13) to see that the functional $v \mapsto (f, \pi v)_Y$ is linear and continuous on X. Therefore, using the Riesz representation theorem, there exists a unique element $\widetilde{f} \in X$ such that

$$(\widetilde{f}, v)_X = (f, \pi v)_Y \qquad \forall\, v \in X. \tag{13.17}$$

Using now Theorem 13.2.11 we deduce that there exists a unique element u such that

$$u \in K, \qquad (Au, v-u)_X + j(u,v) - j(u,u) \geq \big(\widetilde{f}, v-u\big)_X \qquad \forall\, v \in K. \tag{13.18}$$

The existence and uniqueness part of Theorem 13.2.12 is now a direct consequence of (13.17) and (13.18).

Assume now that (13.14) and (13.15) hold and consider an arbitrary element $u_0 \in K$. Then, taking $v = u_0$ in (13.1) we find that

$$(Au, u-u_0)_X \leq (f, \pi u - \pi u_0)_Y + j(u,u_0) - j(u,u)$$

which implies that

$$(Au - Au_0, u-u_0)_X \leq (Au_0, u_0-u)_X + (f, \pi u - \pi u_0)_Y + j(u,u_0) - j(u,u).$$

We now use assumptions (13.10)(a), (13.13), and (13.14) to deduce that

$$m\,\|u-u_0\|_X^2 \leq \|Au_0\|_X\|u-u_0\|_X$$
$$+c_0\|f\|_Y\|u-u_0\|_X + (\beta + \gamma\|u\|_X)\|u-u_0\|_X.$$

Next, we use the triangle inequality $\|u\|_X \leq \|u-u_0\|_X + \|u_0\|_X$ to deduce that

$$(m-\gamma)\,\|u-u_0\|_X \leq \|Au_0\|_X + c_0\|f\|_Y + \gamma\|u_0\|_X + \beta.$$

This inequality combined with the smallness assumption (13.15) implies the bound (13.16) and concludes the proof. □

13.2.3 A Convergence Result

Theorem 13.2.12 allows us to define the operator $f \mapsto u(f)$ which associates to each element $f \in Y$ the solution $u = u(f) \in K$ of the quasivariational inequality (13.1). An important property of this operator is its weak–strong continuity, which represents a crucial ingredient in the study of the optimal control problem $\mathcal{Q}$. It holds under the following additional assumptions:

$$\left\{\begin{array}{l}\text{For any sequences } \{\eta_k\} \subset X,\ \{u_k\} \subset X \text{ such that} \\ \quad \eta_k \rightharpoonup \eta \in X,\ u_k \rightharpoonup u \in X \quad \text{one has} \\ \qquad \limsup_{k\to\infty}\,[j(\eta_k, v) - j(\eta_k, u_k)] \le j(\eta, v) - j(\eta, u) \quad \forall\, v \in X.\end{array}\right. \tag{13.19}$$

$$\left\{\begin{array}{l}\text{For any sequence } \{v_k\} \subset X \text{ such that} \\ v_k \rightharpoonup v \quad \text{in} \quad X \quad \text{one has} \quad \pi v_k \to \pi v \quad \text{in} \quad Y.\end{array}\right. \tag{13.20}$$

Note that assumption (13.19) implies that for all $\eta \in X$, $j(\eta, \cdot) : X \to \mathbb{R}$ is lower semicontinuous. Indeed, this property can be easily deduced by taking $\eta_k = \eta$ in (13.19). Moreover, assumption (13.20) shows that the operator $\pi : X \to Y$ is completely continuous.

Our main result in this subsection is the following.

Theorem 13.2.13 *Assume that (13.9)–(13.15), (13.19), and (13.20) hold. Then,*

$$f_n \rightharpoonup f \quad \text{in} \quad Y \quad \Longrightarrow \quad u(f_n) \to u(f) \quad \text{in} \quad X, \qquad \text{as } n \to \infty. \tag{13.21}$$

Proof The proof of Theorem 13.2.13 will be carried out in several steps that we present in what follows.

(i) *Weak convergence of a subsequence.* Assume that $\{f_n\}$ is a sequence of elements in Y such that

$$f_n \rightharpoonup f \quad \text{in} \quad Y \quad \text{as } n \to \infty \tag{13.22}$$

and, for simplicity, denote $u(f_n) = u_n$ and $u(f) = u$. Then, it follows from (13.22) that $\{f_n\}$ is a bounded sequence in Y and, therefore, inequality (13.16) implies that $\{u_n\}$ is a bounded sequence in X. Using now Theorem 13.2.3 we deduce that there exists an element $\widetilde{u} \in X$ and a subsequence of $\{u_n\}$, again denoted $\{u_n\}$, such that

$$u_n \rightharpoonup \widetilde{u} \quad \text{in } X \quad \text{as } n \to \infty. \tag{13.23}$$

On the other hand, we recall that K is a closed convex subset of the space X and $\{u_n\} \subset K$. Then, Theorem 13.2.2 and (13.23) imply that

$$\widetilde{u} \in K. \tag{13.24}$$

(ii) *Weak convergence of the whole subsequence.* Let $n \in \mathbb{N}$. We write (13.1) for $f = f_n$ to obtain

$$(Au_n, u_n - v)_X \leq j(u_n, v) - j(u_n, u_n) + (f_n, \pi u_n - \pi v)_Y \qquad \forall\, v \in K, \tag{13.25}$$

then we take $v = \widetilde{u} \in K$ to find that

$$(Au_n, u_n - \widetilde{u})_X \leq j\,(u_n, \widetilde{u}) - j(u_n, u_n) + (f_n, \pi u_n - \pi \widetilde{u})_Y\,.$$

We now pass to the upper limit and use the convergences (13.22), (13.23) and assumptions (13.19), (13.20). As a result we deduce that

$$\lim_{n\to\infty} \sup\, (Au_n, u_n - \widetilde{u})_X \leq 0.$$

Therefore, using assumption (13.10), Corollary 13.2.7, and Definition 13.2.5(iv) we deduce that

$$\liminf_{n\to\infty} (Au_n, u_n - v)_X \geq (A\widetilde{u}, \widetilde{u} - v)_X \qquad \forall\, v \in X. \tag{13.26}$$

On the other hand, passing to the upper limit in inequality (13.25) and using the convergences (13.22), (13.23) and assumptions (13.19), (13.20) yields

$$\limsup_{n\to\infty}\, (Au_n, u_n - \widetilde{u})_X \leq j\,(\widetilde{u}, v) - j\,(\widetilde{u}, \widetilde{u}) + (f, \pi\widetilde{u} - \pi v)_Y \qquad \forall\, v \in K. \tag{13.27}$$

We now combine the inequalities (13.26) and (13.27) to see that

$$(A\widetilde{u}, v - \widetilde{u})_X + j\,(\widetilde{u}, v) - j\,(\widetilde{u}, \widetilde{u}) \geq (f, \pi v - \pi\widetilde{u})_Y \qquad \forall\, v \in K. \tag{13.28}$$

Next, it follows from (13.24) and (13.28) that $\widetilde{u}$ is a solution of inequality (13.1) and, by the uniqueness of the solution of this inequality, guaranteed by Theorem 13.2.12, we obtain that

$$\widetilde{u} = u. \tag{13.29}$$

A careful analysis, based on the arguments above, reveals that u is the weak limit of any weakly convergent subsequence of the sequence $\{u_n\}$. Therefore, using Theorem 13.2.4 we deduce that the whole sequence $\{u_n\}$ converges weakly in X to u as $n \to \infty$, i.e.,

$$u_n \rightharpoonup u \quad \text{in } X \quad \text{as } n \to \infty. \tag{13.30}$$

(iii) *Strong convergence.* Let $n \in \mathbb{N}$ be given. We take $v = u$ in inequality (13.25) to see that

$$(Au_n, u_n - u)_X \le j(u_n, u) - j(u_n, u_n) + (f_n, \pi u_n - \pi u)_Y. \tag{13.31}$$

Next, we use (13.31) and assumption (13.10)(a) to find that

$$\begin{aligned} m\,\|u_n - u\|_X^2 &\le (Au_n - Au, u_n - u)_X \\ &= (Au_n, u_n - u)_X - (Au, u_n - u)_X \\ &\le j(u_n, u) - j(u_n, u_n) + (f_n, \pi u_n - \pi u)_Y - (Au, u_n - u)_X. \end{aligned}$$

We now pass to the upper limit in this inequality and use the convergences (13.22), (13.30) and assumptions (13.19), (13.20) to deduce that

$$\|u_n - u\|_X \to 0 \quad \text{as } n \to \infty.$$

This convergence concludes the proof since, recall, $u_n = u(f_n)$ and $u = u(f)$. □

13.3 Optimal Control of Quasivariational Inequalities

We now move to the study of the optimal control problem $\mathcal{Q}$. We start with an existence result for the optimal pairs, Theorem 13.3.1. We proceed with a convergence result, Theorem 13.3.5. Finally, we consider a relevant particular case for which this convergence result holds, Theorem 13.3.8.

13.3.1 Existence of Optimal Pairs

In the study of Problem $\mathcal{Q}$ we assume that

$$\mathcal{L}(u, f) = g(u) + h(f) \qquad \forall\, u \in X,\ f \in Y, \tag{13.32}$$

where g and h are functions which satisfy the following conditions:

$$\left\{\begin{array}{l} g : X \to \mathbb{R} \text{ is continuous, positive, and bounded, i.e.,} \\ \text{(a)}\ \ v_n \to v \quad \text{in} \quad X \quad \Longrightarrow \quad g(v_n) \to g(v). \\ \text{(b)}\ \ g(v) \ge 0 \quad \forall\, v \in X. \\ \text{(c)}\ \ g \text{ maps bounded sets in } X \text{ into bounded sets in } \mathbb{R}. \end{array}\right. \tag{13.33}$$

$$\begin{cases} h : Y \to \mathbb{R} \text{ is weakly lower semicontinuous and coercive, i.e.,} \\ \text{(a) } f_n \rightharpoonup f \quad \text{in} \quad Y \quad \Longrightarrow \quad \liminf_{n\to\infty} h(f_n) \ge h(f). \\ \text{(b) } \|f_n\|_Y \to \infty \qquad \Longrightarrow \quad h(f_n) \to \infty. \end{cases} \tag{13.34}$$

Our first result in this section is the following.

Theorem 13.3.1 *Assume that (13.9)–(13.15), (13.19), (13.20), and (13.32)–(13.34) hold. Then, there exists at least one solution $(u^*, f^*) \in \mathcal{V}_{ad}$ of Problem $\mathcal{Q}$.*

Proof Let

$$\theta = \inf_{(u,f)\in\mathcal{V}_{ad}} \mathcal{L}(u, f) \in \mathbb{R} \tag{13.35}$$

and let $\{(u_n, f_n)\} \subset \mathcal{V}_{ad}$ be a minimizing sequence for the functional $\mathcal{L}$, i.e.,

$$\lim_{n\to\infty} \mathcal{L}(u_n, f_n) = \theta. \tag{13.36}$$

We claim that the sequence $\{f_n\}$ is bounded in Y. Arguing by contradiction, assume that $\{f_n\}$ is not bounded in Y. Then, passing to a subsequence still denoted $\{f_n\}$, we have

$$\|f_n\|_Y \to +\infty \quad \text{as} \quad n \to +\infty. \tag{13.37}$$

We now use equality (13.32) and assumption (13.33)(b) to see that

$$\mathcal{L}(u_n, f_n) \ge h(f_n).$$

Therefore, passing to the limit as $n \to +\infty$ and using (13.37) combined with assumption (13.34)(b) we deduce that

$$\lim_{n\to+\infty} \mathcal{L}(u_n, f_n) = +\infty. \tag{13.38}$$

Equalities (13.36) and (13.38) imply that $\theta = +\infty$ which is in contradiction with (13.35).

We conclude from above that the sequence $\{f_n\}$ is bounded in Y. Therefore, using Theorem 13.2.3 we deduce that there exists $f^* \in Y$ such that, passing to a subsequence still denoted $\{f_n\}$, we have

$$f_n \rightharpoonup f^* \quad \text{in} \quad Y \quad \text{as} \quad n \to +\infty. \tag{13.39}$$

Let u^* be the solution of the quasivariational inequality (13.1) for $f = f^*$, i.e., $u^* = u(f^*)$. Recall that the existence and uniqueness of this solution is guaranteed by Theorem 13.2.12. Then, by the definition (13.2) of the set $\mathcal{V}_{ad}$ we have

$$(u^*, f^*) \in \mathcal{V}_{ad}. \tag{13.40}$$

Moreover, using (13.39) and (13.21) it follows that

$$u_n \to u^* \quad \text{in} \quad X \quad \text{as} \quad n \to +\infty. \tag{13.41}$$

We now use the convergences (13.39), (13.41) and the weakly lower semicontinuity of the functional $\mathcal{L}$, guaranteed by assumptions (13.33)(a) and (13.34)(a), to deduce that

$$\liminf_{n\to+\infty} \mathcal{L}(u_n, f_n) \geq \mathcal{L}(u^*, f^*). \tag{13.42}$$

It follows from (13.36) and (13.42) that

$$\theta \geq \mathcal{L}(u^*, f^*). \tag{13.43}$$

In addition, (13.40) and (13.35) yield

$$\theta \leq \mathcal{L}(u^*, f^*). \tag{13.44}$$

We now combine inequalities (13.43) and (13.44) to see that (13.3) holds, which concludes the proof. □

Remark 13.3.2 Assume now that $U \subset Y$ is a nonempty weakly closed subset, i.e., it satisfies the following property:

$$\text{for any sequence } \{f_n\} \subset U \text{ such that } \quad f_n \rightharpoonup f \in Y \quad \text{one has } f \in U. \tag{13.45}$$

Then, careful analysis of the previous proof reveals the fact that the statement of Theorem 13.3.1 still remains valid if we replace the definition (13.2) of admissible pairs for inequality (13.1) with the following one:

$$\mathcal{V}_{ad} = \{\, (u, f) \in K \times U \text{ such that (13.1) holds}\,\}. \tag{13.46}$$

Considering the set (13.46) instead of (13.2) leads to a version of Theorem 13.3.1 which could be useful in various applications, when the control f is assumed to satisfy some constraints.

13.3.2 Convergence of Optimal Pairs

In this subsection we focus on the dependence of the solution of the optimal control $\mathcal{Q}$ with respect to the set K, the operator A, and the function j. To this end, we assume in what follows that the hypothesis of Theorem 13.3.1 holds. Moreover, for each $n \in \mathbb{N}$ we consider a perturbation K_n, A_n, and j_n of K, A, and j, respectively, which satisfy the following conditions:

$$K_n \text{ is a nonempty, closed, convex subset of } X. \tag{13.47}$$

$$\begin{cases} A_n \;\text{ is a strongly monotone Lipschitz continuous operator,} \\ \text{i.e., it satisfies condition (13.10) with } m_n > 0 \text{ and } M_n > 0. \end{cases} \tag{13.48}$$

$$\begin{cases} \text{(a) For all } \eta \in X,\; j_n(\eta,\cdot) : X \to \mathbb{R} \text{ is convex.} \\ \text{(b) There exists } \alpha_n \ge 0 \text{ such that} \\ \qquad j_n(\eta_1, v_2) - j_n(\eta_1, v_1) + j_n(\eta_2, v_1) - j_n(\eta_2, v_2) \\ \qquad\quad \le \alpha_n \|\eta_1 - \eta_2\|_X \|v_1 - v_2\|_X \quad \forall\, \eta_1,\, \eta_2,\, v_1,\, v_2 \in X. \\ \text{(c) There exist } \beta_n,\, \gamma_n \ge 0 \text{ such that} \\ \qquad j_n(\eta, v_1) - j_n(\eta, v_2) \le (\beta_n + \gamma_n \|\eta\|_X)\, \|v_1 - v_2\|_X \\ \qquad\quad \forall\, \eta,\, v_1,\, v_2 \in X. \\ \text{(d) For any sequences } \{\eta_k\} \subset X,\; \{u_k\} \subset X \text{ such that} \\ \qquad \eta_k \rightharpoonup \eta \in X,\; u_k \rightharpoonup u \in X \quad \text{one has} \\ \qquad \limsup\limits_{k} \,[j_n(\eta_k, v) - j_n(\eta_k, u_k)] \le j_n(\eta, v) - j_n(\eta, u) \quad \forall\, v \in X. \end{cases} \tag{13.49}$$

$$m_n > \alpha_n. \tag{13.50}$$

$$m_n > \gamma_n. \tag{13.51}$$

We consider the following perturbation of Problem $\mathcal{P}$.

Problem $\mathcal{P}_n$. Given $f_n \in Y$, find u_n such that

$$\begin{aligned} u_n \in K_n, \qquad & (A_n u_n, v - u_n)_X + j_n(u_n, v) - j_n(u_n, u_n) \\ & \ge (f_n, \pi v - \pi u_n)_Y \qquad \forall\, v \in K_n. \end{aligned} \tag{13.52}$$

It follows from Theorem 13.2.12 that for each $f_n \in Y$ there exists a unique solution $u_n = u_n(f_n)$ to the quasivariational inequality (13.52). Moreover, the solution satisfies

$$\|u_n\|_X \leq \frac{1}{m_n - \gamma_n} \big(\|A_n u_{0n}\|_X + c_0 \|f_n\|_Y + \gamma_n \|u_{0n}\|_X + \beta_n\big) + \|u_{0n}\|_X, \tag{13.53}$$

where u_{0n} denotes an arbitrary element of K_n. We define the set of admissible pairs for inequality (13.52) by

$$\mathcal{V}_{ad}^n = \{\, (u_n, f_n) \in K_n \times Y \text{ such that (13.52) holds}\}. \tag{13.54}$$

Then, the optimal control problem associated with Problem $\mathcal{P}_n$ is the following.

Problem $\mathcal{Q}_n$. Find $(u_n^*, f_n^*) \in \mathcal{V}_{ad}^n$ such that

$$\mathcal{L}(u_n^*, f_n^*) = \min_{(u_n, f_n) \in \mathcal{V}_{ad}^n} \mathcal{L}(u_n, f_n). \tag{13.55}$$

Using Theorem 13.3.1 it follows that for each $n \in \mathbb{N}$ there exists at least one solution $(u_n^*, f_n^*) \in \mathcal{V}_{ad}^n$ of Problem $\mathcal{Q}_n$. We now consider the following assumptions:

$$f_n \rightharpoonup f \quad \text{in} \quad Y \quad \Longrightarrow \quad u_n(f_n) \to u(f) \quad \text{in} \quad X, \quad \text{as } n \to \infty. \tag{13.56}$$

$$\begin{cases} \text{There exists } f^0 \in Y \ \text{ such that} \\ \text{the sequence } \{u_n(f^0)\} \text{ is bounded in } X. \end{cases} \tag{13.57}$$

Concerning assumptions (13.56) and (13.57) we have the following remarks.

Remark 13.3.3 Assumptions (13.56) and (13.57) are not formulated in terms of the data K_n, A_n, and j_n. They are formulated in terms of the solutions u_n and u which are unknown and, therefore, they represent implicit assumptions. We consider these assumptions for their generality. In the next section we shall provide explicit assumptions on K_n, A_n, and j_n which guarantee that conditions (13.56) and (13.57) hold. Considering such explicit assumptions will lead us to introduce a relevant particular case in which Theorem 13.3.5 holds.

Remark 13.3.4 Condition (13.56) represents a continuous dependence condition of the solution of (13.1) with respect to the set K, the operator A, the function j, and the element $f \in Y$.

The second result in this section is a convergence result for the set of solution of Problem $\mathcal{Q}$. Its statement is as follows.

Theorem 13.3.5 *Assume that (13.9)–(13.15), (13.19), (13.20), and (13.32)–(13.34) hold and, for any $n \in \mathbb{N}$, assume that (13.47)–(13.51) hold, too. Moreover, assume that conditions (13.56)–(13.57) are satisfied and let $\{(u_n^*, f_n^*)\}$ be a sequence of solutions of Problem $\mathcal{Q}_n$. Then, there exists a subsequence of the sequence*

$\{(u_n^*, f_n^*)\}$, *again denoted* $\{(u_n^*, f_n^*)\}$, *and an element* $(u^*, f^*) \in X \times Y$ *such that*

$$f_n^* \rightharpoonup f^* \quad \text{in} \quad Y \quad \text{as} \quad n \to \infty, \tag{13.58}$$

$$u_n^* \to u^* \quad \text{in} \quad X \quad \text{as} \quad n \to \infty, \tag{13.59}$$

$$(u^*, f^*) \quad \text{is a solution of Problem } \mathcal{Q}. \tag{13.60}$$

Proof We claim that the sequence $\{f_n^*\}$ is bounded in Y. Arguing by contradiction, assume that $\{f_n^*\}$ is not bounded in Y. Then, passing to a subsequence still denoted $\{f_n^*\}$, we have

$$\|f_n^*\|_Y \to +\infty \quad \text{as} \quad n \to +\infty. \tag{13.61}$$

We use equality (13.32) and assumption (13.33)(b) to see that

$$\mathcal{L}\left(u_n^*, f_n^*\right) \geq h\left(f_n^*\right).$$

Therefore, passing to the limit as $n \to \infty$ in this inequality and using (13.61) combined with assumption (13.34)(b) we deduce that

$$\lim_{n\to\infty} \mathcal{L}\left(u_n^*, f_n^*\right) = +\infty. \tag{13.62}$$

On the other hand, since $\left(u_n^*, f_n^*\right)$ represents a solution to Problem $\mathcal{Q}_n$, for each $n \in \mathbb{N}$ we have

$$\mathcal{L}\left(u_n^*, f_n^*\right) \leq \mathcal{L}(u_n, f_n) \qquad \forall\,(u_n, f_n) \in \mathcal{V}_{ad}^n. \tag{13.63}$$

We now use assumption (13.57) and denote by u_n^0 the solution of Problem $\mathcal{P}_n$ for $f_n = f^0$, i.e., $u_n^0 = u_n\left(f^0\right)$. Then $\left(u_n^0, f^0\right) \in \mathcal{V}_{ad}^n$ and, therefore, (13.63) and (13.32) imply that

$$\mathcal{L}\left(u_n^*, f_n^*\right) \leq g\left(u_n^0\right) + h\left(f^0\right). \tag{13.64}$$

Then, since (13.57) guarantees that $\left\{u_n^0\right\}$ is a bounded sequence in X, assumption (13.33)(c) on the function g implies that there exists $D > 0$ which does not depend on n such that

$$g\left(u_n^0\right) + h\left(f^0\right) \leq D \qquad \forall\, n \in \mathbb{N}. \tag{13.65}$$

Relations (13.62), (13.64), and (13.65) lead to a contradiction, which concludes the claim.

Next, since the sequence $\{f_n^*\}$ is bounded in Y we can find a subsequence again denoted $\{f_n^*\}$ and an element $f^* \in Y$ such that (13.58) holds. Denote by u^* the solution of Problem $\mathcal{P}$ for $f = f^*$, i.e., $u^* = u(f^*)$. Then, we have

$$(u^*, f^*) \in \mathcal{V}_{ad} \tag{13.66}$$

and, moreover, assumption (13.56) implies that (13.59) holds, too.

We now prove that (u^*, f^*) is a solution to the optimal control problem $\mathcal{Q}$. To this end we use the convergences (13.58), (13.59) and the weakly lower semicontinuity of the functional $\mathcal{L}$, guaranteed by (13.32)–(13.34), to see that

$$\mathcal{L}(u^*, f^*) \le \liminf_{n\to\infty} \mathcal{L}(u_n^*, f_n^*). \tag{13.67}$$

Next, we fix a solution $\left(u_0^*, f_0^*\right)$ of Problem $\mathcal{Q}$ and, in addition, for each $n \in \mathbb{N}$ we denote by $\widetilde{u}_n^0$ the solution of Problem $\mathcal{P}_n$ for $f_n = f_0^*$. It follows from here that $\left(\widetilde{u}_n^0, f_0^*\right) \in \mathcal{V}_{ad}^n$ and, by the optimality of the pair $\left(u_n^*, f_n^*\right)$, we have that

$$\mathcal{L}\left(u_n^*, f_n^*\right) \le \mathcal{L}\left(\widetilde{u}_n^0, f_0^*\right) \qquad \forall\, n \in \mathbb{N}.$$

We pass to the upper limit in this inequality to see that

$$\limsup_{n\to\infty} \mathcal{L}\left(u_n^*, f_n^*\right) \le \limsup_{n\to\infty} \mathcal{L}\left(\widetilde{u}_n^0, f_0^*\right). \tag{13.68}$$

Now, remember that u_0^* is the solution of the inequality (13.1) for $f = f_0^*$ and $\widetilde{u}_n^0$ is the solution of the inequality (13.52) for $f_n = f_0^*$, i.e., $\widetilde{u}_n^0 = u_n\left(f_0^*\right)$ and $\widetilde{u}_n = u_n\left(f_0^*\right)$. Therefore, assumption (13.56) implies that

$$\widetilde{u}_n^0 \to u_0^* \quad \text{in} \quad X \quad \text{as} \quad n \to \infty$$

and, using the continuity of the functional $u \mapsto \mathcal{L}\left(u, f_0^*\right) : X \to \mathbb{R}$ yields

$$\lim_{n\to\infty} \mathcal{L}\left(\widetilde{u}_n^0, f_0^*\right) = \mathcal{L}\left(u_0^*, f_0^*\right). \tag{13.69}$$

We now use (13.67)–(13.69) to see that

$$\mathcal{L}\left(u^*, f^*\right) \le \mathcal{L}(u_0^*, f_0^*). \tag{13.70}$$

On the other hand, since (u_0^*, f_0^*) is a solution of Problem $\mathcal{Q}$, we have

$$\mathcal{L}\left(u_0^*, f_0^*\right) = \min_{(u,f)\in\mathcal{V}_{ad}} \mathcal{L}(u, f), \tag{13.71}$$

and, therefore, inclusion (13.66) implies that

$$\mathcal{L}\left(u_0^*, f_0^*\right) \leq \mathcal{L}\left(u^*, f^*\right). \tag{13.72}$$

We now combine the inequalities (13.70) and (13.72) to see that

$$\mathcal{L}\left(u^*, f^*\right) = \mathcal{L}\left(u_0^*, f_0^*\right). \tag{13.73}$$

Finally, relations (13.66), (13.73), and (13.71) imply that (13.60) holds, which concludes the proof. □

We end this subsection with the following remarks.

Remark 13.3.6 If Problem $\mathcal{Q}$ has a unique solution (u^*, f^*) then, under the assumption of Theorem 13.3.5 the convergences (13.58) and (13.59) are valid for the whole sequence $\left\{\left(u_n^*, f_n^*\right)\right\}$. Indeed, a careful analysis of the proof of Theorem 13.2 reveals that the sequence $\left\{f_n^*\right\}$ is bounded in Y and, moreover, each weakly convergent subsequence of $\left\{f_n^*\right\}$ converges weakly to f^*. We now use Theorem 13.2.4 to deduce that the whole sequence satisfies (13.58). Finally, using Theorem 13.2.13 it follows that (13.59) holds, too.

Remark 13.3.7 The statement of Theorem 13.3.5 still remains valid if we replace the definition (13.2) with (13.46) and the definition (13.54) with

$$\mathcal{V}_{ad}^n = \{\, (u_n, f_n) \in K_n \times U \text{ such that (13.52) holds} \,\}, \tag{13.74}$$

U being a nonempty weakly closed subset of Y. The proof of this statement is based on the property (13.45) of the set U.

13.3.3 A Relevant Particular Case

Our aim in this subsection is to present explicit conditions on the family of sets K_n, operators A_n, and functionals j_n which guarantee that assumptions (13.56) and (13.57) hold. We conclude from here that, under these conditions, the abstract result in Theorem 13.3.5 holds.

Everywhere in this subsection we assume that (13.9)–(13.15), (13.19), and (13.20) hold and, for each $f \in Y$, we denote by $u = u(f)$ the solution of inequality (13.1), guaranteed by Theorem 13.2.12. Moreover, for each $n \in \mathbb{N}$ we consider the set $K_n \subset X$, the operator $A_n : X \to X$, and the functional $j_n : X \times X \to \mathbb{R}$ such that the followings hold:

$$K_n = c_n K \quad \text{with} \quad c_n > 0. \tag{13.75}$$

$$\begin{cases} \text{(a) } A_n = A + T_n. \\ \text{(b) } A : X \to X \text{ satisfies condition (13.10) with} \\ \quad m > 0 \text{ and } M > 0. \\ \text{(c) } T_n : X \to X \text{ is a monotone Lipschitz continuous operator.} \end{cases} \tag{13.76}$$

$$\begin{cases} j_n \text{ satisfies condition (13.49) with } \alpha_n \geq 0,\ \beta_n \geq 0,\ \gamma_n \geq 0 \\ \text{such that } m > \alpha_n,\quad m > \gamma_n. \end{cases} \tag{13.77}$$

With this choice we consider Problem $\mathcal{P}_n$. It is easy to see that for each $n \in \mathbb{N}$ the set $K_n \subset X$ satisfies condition (13.47). Moreover, the operator A_n satisfies condition (13.48) with $m_n = m$ and $M_n = M + L_{T_n}$, L_{T_n} being the Lipschitz constant of the operator T_n. We now use assumption (13.77) to see that conditions (13.49)–(13.51) are also satisfied. Therefore, using Theorem 13.2.12 we deduce that for each $f_n \in Y$ there exists a unique solution $u_n = u_n(f_n)$ to the quasivariational inequality (13.52).

On the other hand, if (13.32)–(13.34) hold, then Theorem 13.3.1 guarantees the existence of at least one solution (u^*, f^*) of Problem $\mathcal{Q}$ and, for each $n \in \mathbb{N}$, the existence of at least one solution (u_n^*, f_n^*) to Problem $\mathcal{Q}_n$.

We now consider the following additional assumptions:

$$\lim_{n\to\infty} c_n = 1. \tag{13.78}$$

$$j(u, \lambda v) = \lambda j(u, v) \qquad \forall \lambda \geq 0,\ u,\ v \in X. \tag{13.79}$$

$$\begin{cases} \text{For any } n \in \mathbb{N} \text{ there exists } F_n \geq 0 \text{ and } \delta_n \geq 0 \text{ such that} \\ \text{(a) } \|T_n v\|_X \leq F_n(\|v\|_X + \delta_n) \qquad \forall v \in X. \\ \text{(b) } \lim_{n\to\infty} F_n = 0. \\ \text{(c) The sequence } \{\delta_n\} \subset \mathbb{R} \text{ is bounded.} \end{cases} \tag{13.80}$$

$$\begin{cases} \text{For any } n \in \mathbb{N} \text{ there exists } G_n \geq 0 \text{ and } H_n \geq 0 \text{ such that} \\ \text{(a) } j_n(v_1, v_2) - j_n(v_1, v_1) + j(v_2, v_1) - j(v_2, v_2) \\ \qquad \leq G_n + H_n\|v_1 - v_2\|_X + \alpha\,\|v_1 - v_2\|_X^2 \\ \qquad \forall v_1,\ v_2 \in X. \\ \text{(b) } \lim_{n\to\infty} G_n = \lim_{n\to\infty} H_n = 0. \end{cases} \tag{13.81}$$

Moreover, we reinforce assumption (13.77) by assuming that there exist two constants β_0 and γ_0 such that

$$\beta_n \leq \beta_0, \qquad \gamma_n \leq \gamma_0 < m, \qquad \forall n \in \mathbb{N}. \tag{13.82}$$

We have the following result.

Theorem 13.3.8 *Assume that (13.9)–(13.15), (13.19), (13.20), and (13.32)–(13.34) hold and, for any $n \in \mathbb{N}$, assume that (13.75)–(13.77) hold, too. Assume moreover that conditions (13.78)–(13.82) are satisfied and let $\{(u_n^*, f_n^*)\}$ be a sequence of solutions of Problem $\mathcal{Q}_n$. Then, there exists a subsequence of the sequence $\{(u_n^*, f_n^*)\}$, again denoted $\{(u_n^*, f_n^*)\}$, and an element $(u^*, f^*) \in X \times Y$ such that (13.58)–(13.60) hold.*

The proof is carried out in several steps, based on the abstract result provided by Theorem 13.3.5. The first step of the proof is the following.

Lemma 13.3.9 *Under the assumption of Theorem 13.3.8, if the sequence $\{f_n\}$ is bounded in Y, then the sequence $\{u_n(f_n)\}$ is bounded in X.*

Proof Let u_0 be a given element of K and let $n \in \mathbb{N}$. Condition (13.75) guarantees that $c_n u_0 \in K_n$ and, therefore, using inequality (13.53) with $u_{0n} = c_n u_0$ yields

$$\|u_n\|_X \leq \frac{1}{m_n - \gamma_n}\big(\|A_n(c_n u_0)\|_X + c_0\|f_n\|_Y + \gamma_n\|c_n u_0\|_X + \beta_n\big) + \|c_n u_0\|_X.$$

We now use assumption (13.82) and equality $m_n = m$ which, recall, follows from assumption (13.76). In this way we deduce that

$$\|u_n\|_X \leq \frac{1}{m - \gamma_0}\big(\|A_n(c_n u_0)\|_X + c_0\|f_n\|_Y + \gamma_0 c_n\|u_0\|_X + \beta_0\big) + c_n\|u_0\|_X. \tag{13.83}$$

Recall that $A_n(c_n u_0) = A(c_n u_0) + T_n(c_n u_0)$ and, therefore,

$$\|A_n(c_n u_0)\|_X \leq \|A(c_n u_0)\|_X + \|T_n(c_n u_0)\|_X. \tag{13.84}$$

We now write

$$\|A(c_n u_0)\|_X \leq \|A(c_n u_0) - Au_0\|_X + \|Au_0\|_X,$$

then we use assumption (13.10)(b) to deduce that

$$\|A(c_n u_0)\|_X \leq (M|c_n - 1|\,\|u_0\|_X + \|Au_0\|_X). \tag{13.85}$$

Moreover, using (13.80) we have that

$$\|T_n(c_n u_0)\|_X \leq F_n(c_n\,\|u_0\|_X + \delta_n). \tag{13.86}$$

Next, we combine inequalities (13.83)–(13.86) to find that

$$\|u_n\|_X \tag{13.87}$$

$$\leq \frac{1}{m-\gamma_0}\big(M|c_n-1|\,\|u_0\|_X + \|Au_0\|_X + F_n(c_n\,\|u_0\|_X + \delta_n)$$

$$+\frac{1}{m-\gamma_0}\big(c_0\|f_n\|_Y + \gamma_0 c_n\|u_0\|_X + \beta_0\big) + c_n\|u_0\|_X.$$

Lemma 13.3.9 is now a direct consequence of inequality (13.87) and assumptions (13.78), (13.80)(b). □

We proceed with the following result.

Lemma 13.3.10 *Under the assumption of Theorem 13.3.8, condition (13.56) holds.*

Proof Let $\{f_n\} \subset Y$, $f \in Y$ such that

$$f_n \rightharpoonup f \quad \text{in} \quad Y \quad \text{as} \quad n \to \infty. \tag{13.88}$$

Let $n \in \mathbb{N}$. Besides Problems $\mathcal{P}$ and $\mathcal{P}_n$ we consider the intermediate problems of finding two elements $\bar{u}_n$ and $\widetilde{u}_n$ such that

$$\bar{u}_n \in K, \qquad (A\bar{u}_n, v-\bar{u}_n)_X + j(\bar{u}_n, v) - j(\bar{u}_n, \bar{u}_n) \tag{13.89}$$

$$\geq (f_n, \pi v - \pi\bar{u}_n)_Y \qquad \forall\, v \in K.$$

$$\widetilde{u}_n \in K_n, \qquad (A\widetilde{u}_n, v_n-\widetilde{u}_n)_X + j(\widetilde{u}_n, v_n) - j(\widetilde{u}_n, \widetilde{u}_n) \tag{13.90}$$

$$\geq (f_n, \pi v_n - \pi\widetilde{u}_n)_Y \qquad \forall\, v_n \in K_n.$$

Note that Theorem 13.2.12 guarantees the existence of a unique solution $\bar{u}_n$ and $\widetilde{u}_n$ to the quasivariational inequalities (13.89) and (13.90), respectively. Our aim in what follows is to establish estimates for the norms $\|u_n - \widetilde{u}_n\|_X$ and $\|\widetilde{u}_n - \bar{u}_n\|_X$.

Let $n \in \mathbb{N}$. We take $v_n = u_n$ in (13.90), $v_n = \widetilde{u}_n$ in (13.52), then we add the resulting inequalities to obtain that

$$(A_n u_n - A\widetilde{u}_n, u_n - \widetilde{u}_n)_X \leq j_n\,(u_n, \widetilde{u}_n) - j_n(u_n, u_n) + j\,(\widetilde{u}_n, u_n) - j\,(\widetilde{u}_n, \widetilde{u}_n)\,.$$

We use now assumption (13.76)(a) to see that $A_n u_n = Au_n + T_n u_n$ and, therefore, we deduce that

$$(Au_n - A\widetilde{u}_n, u_n - \widetilde{u}_n)_X \leq (T_n u_n, \widetilde{u}_n - u_n)_X$$

$$+ j_n\,(u_n, \widetilde{u}_n) - j_n(u_n, u_n) + j\,(\widetilde{u}_n, u_n) - j\,(\widetilde{u}_n, \widetilde{u}_n)\,.$$

Next, we use conditions (13.10)(a), (13.80)(a), and (13.81)(a) to find that

$$m\ \|u_n - \widetilde{u}_n\|_X^2 \le F_n(\|u_n\|_X + \delta_n)\ \|u_n - \widetilde{u}_n\|_X \qquad (13.91)$$

$$+G_n + H_n\ \|u_n - \widetilde{u}_n\|_X + \alpha\ \|u_n - \widetilde{u}_n\|_X^2\ .$$

On the other hand, assumption (13.88) and Lemma 13.3.9 imply that there exists $E > 0$ which does not depend on n such that $\|u_n\|_X \le E$. Therefore, since $m > \alpha$, inequality (13.91) yields

$$\|u_n - \widetilde{u}_n\|_X^2 \le \left(\frac{H_n}{m-\alpha} + \frac{(E+\delta_n)F_n}{m-\alpha}\right)\ \|u_n - \widetilde{u}_n\|_X + \frac{G_n}{m-\alpha}.$$

Next, the elementary inequality

$$x^2 \le ax + b \implies x \le a + \sqrt{b} \qquad \forall\, x,\ a,\ b \ge 0$$

combined with assumptions (13.80)(b),(c) and (13.81)(b) implies that

$$\|u_n - \widetilde{u}_n\|_X \to 0 \qquad \text{as} \qquad n \to 0. \qquad (13.92)$$

On the other hand, condition (13.75) allows us to test in (13.90) with $v_n = c_n\bar{u}_n \in K_n$. As a result we deduce that

$$(A\widetilde{u}_n, c_n\bar{u}_n - \widetilde{u}_n)_X + j\,(\widetilde{u}_n, c_n\bar{u}_n) - j\,(\widetilde{u}_n, \widetilde{u}_n) \ge (f_n, c_n\pi\bar{u}_n - \pi\widetilde{u}_n)_Y\ . \qquad (13.93)$$

We now use condition (13.75), again, to test in (13.89) with $v = \frac{1}{c_n}\widetilde{u}_n \in K$. Then, we multiply the resulting inequality with $c_n > 0$ and use assumption (13.79) on j to find that

$$(A\bar{u}_n, \widetilde{u}_n - c_n\bar{u}_n)_X + j\,(\bar{u}_n, \widetilde{u}_n) - j\,(\bar{u}_n, c_n\bar{u}_n) \ge (f_n, \pi\widetilde{u}_n - c_n\pi\bar{u}_n)_Y\ . \qquad (13.94)$$

We now add inequalities (13.93) and (13.94) to deduce that

$$(A\widetilde{u}_n - A\bar{u}_n, \widetilde{u}_n - c_n\bar{u}_n)_X$$

$$\le j\,(\widetilde{u}_n, c_n\bar{u}_n) - j\,(\widetilde{u}_n, \widetilde{u}_n) + j\,(\bar{u}_n, \widetilde{u}_n) - j\,(\bar{u}_n, c_n\bar{u}_n)\,,$$

then we use assumption (13.11)(b) to obtain that

$$(A\widetilde{u}_n - A\bar{u}_n, \widetilde{u}_n - c_n\bar{u}_n)_X \le \alpha\ \|\widetilde{u}_n - \bar{u}_n\|_X\ \|\widetilde{u}_n - c_n\bar{u}_n\|_X\ . \qquad (13.95)$$

Next, we write

$$\widetilde{u}_n - c_n\bar{u}_n = \widetilde{u}_n - \bar{u}_n + (1 - c_n)\bar{u}_n,$$

then we substitute this equality in (13.95) and use condition (13.10)(a) to find that

$$m\|\widetilde{u}_n - \bar{u}_n\|_X^2 \leq (A\widetilde{u}_n - A\bar{u}_n, (c_n - 1)\bar{u}_n)_X$$
$$+\alpha\|\widetilde{u}_n - \bar{u}_n\|_X^2 + \alpha|1 - c_n|\,\|\widetilde{u}_n - \bar{u}_n\|_X\,\|\bar{u}_n\|_X.$$

We now use assumption (13.10)(b) and the smallness assumption (13.12) to see that

$$\|\widetilde{u}_n - \bar{u}_n\|_X \leq \frac{M + \alpha}{m - \alpha}\,|1 - c_n|\,\|\bar{u}_n\|_X. \tag{13.96}$$

Next, consider an element $u_0 \in K$. Condition (13.75) guarantees that $c_n u_0 \in K_n$ and, therefore, using inequality (13.16) for the variational inequality (13.90) yields

$$\|\bar{u}_n\|_X \leq \frac{1}{m - \gamma}\big(\|A(c_n u_0)\|_X + c_0\|f_n\|_Y + \gamma\|c_n u_0\|_X + \beta\big) + \|c_n u_0\|_X.$$

We use assumption (13.10) and convergences (13.78), (13.88) to deduce that the sequence $\{\bar{u}_n\}$ is bounded in X, i.e., there exists $\overline{E} > 0$ such that

$$\|\bar{u}_n\|_X \leq \overline{E} \qquad \forall\, n \in \mathbb{N}. \tag{13.97}$$

We now combine inequalities (13.96) and (13.97), then we use assumption (13.78) to deduce that

$$\|\widetilde{u}_n - \bar{u}_n\|_X \to 0 \qquad \text{as} \qquad n \to \infty. \tag{13.98}$$

Finally, assumption (13.88) and Theorem 13.2.13 yield

$$\|\bar{u}_n - u\|_X \to 0 \qquad \text{as} \qquad n \to \infty. \tag{13.99}$$

We now write

$$\|u_n - u\|_X \leq \|u_n - \widetilde{u}_n\|_X + \|\widetilde{u}_n - \bar{u}_n\|_X + \|\bar{u}_n - u\|_X\,,$$

then we use the convergences (13.92), (13.98), and (13.99) to see that

$$\|u_n - u\|_X \to 0 \qquad \text{as} \qquad n \to \infty.$$

It follows from here that condition (13.56) is satisfied, which concludes the proof. □

We now are in a position to provide the proof of Theorem 13.3.8.

Proof First, we use Lemma 13.3.10 to see that, under the assumptions of Theorem 13.3.8, condition (13.56) holds. On the other hand, Lemma 13.3.9 shows that the sequence $\{u_n(f^0)\}$ is bounded in X, for any $f^0 \in Y$. Therefore, condition (13.57) holds, too. Theorem 13.3.8 is now a direct consequence of Theorem 13.3.5. □

We end this section with the following remarks.

Remark 13.3.11 In contrast to conditions (13.56) and (13.57), conditions (13.75)–(13.82) are explicit conditions, since they are formulated in terms of the data K_n, A_n, and j_n. In many applications they are easy to be verified. A concrete example which illustrates this statement will be presented in Sect. 13.4.

Remark 13.3.12 If Problem $\mathcal{Q}$ has a unique solution (u^*, f^*) then, under the assumptions of Theorem 13.3.8 the convergences (13.58) and (13.59) are valid for the whole sequence $\{(u_n^*, f_n^*)\}$. This statement is a direct consequence of Remark 13.3.6.

Remark 13.3.13 The statement of Theorem 13.3.5 still remains valid if we replace the definitions (13.2) and (13.54) with definitions (13.46) and (13.74), respectively, U being a given nonempty weakly closed subset of Y.

13.4 A Frictional Contact Problem

In this section we use the abstract results presented in Sects. 13.2 and 13.3 in the study of a quasivariational inequality which models the frictional contact of an elastic body with a foundation. We start by introducing the function spaces we need, then we describe the model of contact and prove its unique weak solvability, Theorem 13.4.1. Next, we turn to the optimal control of the problem and prove existence and convergence results, Theorems 13.4.4 and 13.4.6, respectively. Finally, we exemplify our results in the study of a one-dimensional mathematical model which describes the equilibrium of an elastic rod in unilateral contact with a foundation, under the action of a body force.

13.4.1 Function Spaces

For the study of mathematical models of contact we need further notation and preliminary material that we introduce in this subsection. Everywhere below we denote by $\mathbb{S}^d$ the space of second order symmetric tensors on $\mathbb{R}^d$ or, equivalently, the space of symmetric matrices of order d. The inner product and norm on $\mathbb{R}^d$ and

$\mathbb{S}^d$ are defined by

$$\boldsymbol{u} \cdot \boldsymbol{v} = u_i v_i \ , \qquad \|\boldsymbol{v}\| = (\boldsymbol{v} \cdot \boldsymbol{v})^{\frac{1}{2}} \qquad \forall\, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d,$$
$$\boldsymbol{\sigma} \cdot \boldsymbol{\tau} = \sigma_{ij} \tau_{ij} \ , \qquad \|\boldsymbol{\tau}\| = (\boldsymbol{\tau} \cdot \boldsymbol{\tau})^{\frac{1}{2}} \qquad \forall\, \boldsymbol{\sigma}, \boldsymbol{\tau} \in \mathbb{S}^d,$$

and $\mathbf{0}$ will denote the zero element of these spaces. Let $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ be a bounded domain. We denote by Γ its boundary, assumed to be Lipschitz continuous and divided into three measurable parts Γ_1, Γ_2, and Γ_3 such that $meas\,(\Gamma_1) > 0$. We use the notation $\boldsymbol{x} = (x_i)$ for a typical point in $\Omega \cup \Gamma$ and we denote by $\boldsymbol{\nu} = (\nu_i)$ the outward unit normal at Γ. Here and below the indices i and j run between 1 and d and, unless stated otherwise, the summation convention over repeated indices is used. An index that follows a comma represents the partial derivative with respect to the corresponding component of the spatial variable $\boldsymbol{x}$, i.e., $u_{i,j} = \partial u_i / \partial x_j$. Moreover, $\boldsymbol{\varepsilon}$ represents the deformation operator, i.e.,

$$\boldsymbol{\varepsilon}(\boldsymbol{v}) = (\varepsilon_{ij}(\boldsymbol{v})), \quad \varepsilon_{ij}(\boldsymbol{v}) = \frac{1}{2}\,(v_{i,j} + v_{j,i}).$$

We use the standard notation for Sobolev and Lebesgue spaces associated with Ω and Γ and, in addition, we consider the spaces

$$V = \left\{ \boldsymbol{v} \in H^1(\Omega)^d \ : \ \boldsymbol{v} = \mathbf{0} \text{ on } \Gamma_1 \right\}, \quad Y = L^2(\Omega)^d \times L^2(\Gamma_2)^d.$$

It is well known that V is a real Hilbert space endowed with the inner product

$$(\boldsymbol{u}, \boldsymbol{v})_V = \int_\Omega \boldsymbol{\varepsilon}(\boldsymbol{u}) \cdot \boldsymbol{\varepsilon}(\boldsymbol{v})\, dx,$$

and the associated norm $\|\cdot\|_V$. Completeness of the space $(V, \|\cdot\|_V)$ follows from the assumption $meas\,(\Gamma_1) > 0$, which allows the use of Korn's inequality. For an element $\boldsymbol{v} \in V$ we still write $\boldsymbol{v}$ for the trace of $\boldsymbol{v}$ and we denote by v_ν and $\boldsymbol{v}_\tau$ the normal and tangential components of $\boldsymbol{v}$ on Γ given by $v_\nu = \boldsymbol{v} \cdot \boldsymbol{\nu}$, $\boldsymbol{v}_\tau = \boldsymbol{v} - v_\nu \boldsymbol{\nu}$. We also recall that there exists $d_0 > 0$ which depends on Ω, Γ_1, and Γ_3 such that

$$\|\boldsymbol{v}\|_{L^2(\Gamma_3)^d} \le d_0 \|\boldsymbol{v}\|_V \quad \text{for all } \boldsymbol{v} \in V. \tag{13.100}$$

Inequality (13.100) represents a consequence of the Sobolev trace theorem. The space Y will be endowed with its canonic inner product and associated norm, denoted by $(\cdot,\cdot)_Y$ and $\|\cdot\|$, respectively.

For a regular function $\boldsymbol{\sigma} : \Omega \to \mathbb{S}^d$ we denote by σ_ν and $\boldsymbol{\sigma}_\tau$ the normal and tangential stress on Γ, that is, $\sigma_\nu = (\boldsymbol{\sigma}\boldsymbol{\nu}) \cdot \boldsymbol{\nu}$ and $\boldsymbol{\sigma}_\tau = \boldsymbol{\sigma}\boldsymbol{\nu} - \sigma_\nu \boldsymbol{\nu}$, and we recall that the following Green's formula holds:

$$\int_\Omega \boldsymbol{\sigma} \cdot \boldsymbol{\varepsilon}(\boldsymbol{v})\, dx + \int_\Omega \mathrm{Div}\, \boldsymbol{\sigma} \cdot \boldsymbol{v}\, dx = \int_\Gamma \boldsymbol{\sigma}\boldsymbol{\nu} \cdot \boldsymbol{v}\, da \quad \text{for all } \boldsymbol{v} \in H^1(\Omega)^d. \tag{13.101}$$

More details on the function spaces used in contact mechanics, including their basic properties, can be found in the books [41, 42].

13.4.2 The Model

The physical setting is the following. An elastic body occupies, in its reference configuration, the domain $\Omega \subset \mathbb{R}^d$. Its boundary Γ is divided into three measurable disjoint parts $\Gamma_1, \Gamma_2, \Gamma_3$ such that $meas\,(\Gamma_1) > 0$, as already mentioned. The body is fixed on Γ_1, is acted upon by given surface tractions on Γ_2, and is in potential contact with an obstacle on Γ_3. To construct a mathematical model which corresponds to the equilibrium of the body in this physical setting above we need to prescribe specific interface boundary condition. Here, we assume that the contact is with normal compliance and finite penetration, associated with a version of Coulomb's law of dry friction. Therefore, the classical formulation of the problem is the following.

Problem P. Find a displacement field $\boldsymbol{u} : \Omega \to \mathbb{R}^d$ and a stress field $\boldsymbol{\sigma} : \Omega \to \mathbb{S}^d$ such that

$$\boldsymbol{\sigma} = \mathcal{F}\boldsymbol{\varepsilon}(\boldsymbol{u}) \quad \text{in } \Omega, \tag{13.102}$$

$$\mathrm{Div}\,\boldsymbol{\sigma} + \boldsymbol{f}_0 = \boldsymbol{0} \quad \text{in } \Omega, \tag{13.103}$$

$$\boldsymbol{u} = \boldsymbol{0} \quad \text{on } \Gamma_1, \tag{13.104}$$

$$\boldsymbol{\sigma}\boldsymbol{\nu} = \boldsymbol{f}_2 \quad \text{on } \Gamma_2, \tag{13.105}$$

$$\left.\begin{array}{l} u_\nu \le k, \quad \sigma_\nu + p(u_\nu) \le 0, \\ (u_\nu - k)(\sigma_\nu + p(u_\nu)) = 0 \end{array}\right\} \quad \text{on } \Gamma_3, \tag{13.106}$$

$$\left.\begin{array}{l} \|\boldsymbol{\sigma}_\tau\| \le \mu\, p(u_\nu), \\ -\boldsymbol{\sigma}_\tau = \mu\, p(u_\nu)\, \frac{\boldsymbol{u}_\tau}{\|\boldsymbol{u}_\tau\|} \quad if\ \boldsymbol{u}_\tau \ne \boldsymbol{0} \end{array}\right\} \quad \text{on } \Gamma_3. \tag{13.107}$$

We now provide a description of the equations and boundary conditions in Problem P. First, Eq. (13.102) represents the elastic constitutive law of the material in which $\mathcal{F}$ is assumed to be a nonlinear constitutive operator. Equation (13.103) is the equation of equilibrium. We use it here since the contact process is assumed to be static and, therefore, the inertial term in the equation of motion is neglected. Conditions (13.104) and (13.105) represent the displacement and traction boundary conditions, respectively.

Condition (13.106) represents the so-called normal compliance condition with unilateral constraint. Here, $k > 0$ is a given bound which limits the normal displacement and p is a given positive function which will be described below. This condition describes the contact with an obstacle made of a rigid body covered by a layer of thickness k made of deformable material. Condition (13.107) represents a static version of Coulomb's law of dry friction in which μ denotes the coefficient of friction and $\mu p(u_\nu)$ is the friction bound. The coupling of boundary conditions (13.106) and (13.107) was considered for the first time in [5]. Later, it was used in a number of papers, see [42] and the references therein. It describes a contact with normal compliance, as far as the normal displacement satisfies the condition $u_\nu < k$, associated with the classical Coulomb's law of dry friction. When $u_\nu = k$ the contact is with a Signorini-type condition and is associated with the Tresca friction law with the friction bound $\mu p(k)$. It follows from here that conditions (13.106), (13.107) describe a natural transition from the Coulomb law of dry friction (which is valid as far as $0 \le u_\nu < k$) to the Tresca law (which is valid when $u_\nu = k$).

In the study of the mechanical problem (13.102)–(13.107) we assume that the elasticity operator $\mathcal{F}$ and the normal compliance function p satisfy the following conditions:

$$\left\{\begin{array}{l}
\text{(a) } \mathcal{F} : \Omega \times \mathbb{S}^d \to \mathbb{S}^d. \\
\text{(b) There exists } L_{\mathcal{F}} > 0 \text{ such that} \\
\qquad \|\mathcal{F}(\boldsymbol{x}, \boldsymbol{\varepsilon}_1) - \mathcal{F}(\boldsymbol{x}, \boldsymbol{\varepsilon}_2)\| \le L_{\mathcal{F}} \|\boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_2\| \\
\qquad\quad \forall\, \boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2 \in \mathbb{S}^d, \ \text{a.e. } \boldsymbol{x} \in \Omega. \\
\text{(c) The mapping } \boldsymbol{x} \mapsto \mathcal{F}(\boldsymbol{x}, \boldsymbol{\varepsilon}) \text{ is measurable on } \Omega, \\
\qquad \text{for any } \boldsymbol{\varepsilon} \in \mathbb{S}^d. \\
\text{(d) There exists } m_{\mathcal{F}} > 0 \text{ such that} \\
\qquad (\mathcal{F}(\boldsymbol{x}, \boldsymbol{\varepsilon}_1) - \mathcal{F}(\boldsymbol{x}, \boldsymbol{\varepsilon}_2)) \cdot (\boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_2) \ge m_{\mathcal{F}} \, \|\boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_2\|^2 \\
\qquad\quad \forall\, \boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2 \in \mathbb{S}^d, \ \text{a.e. } \boldsymbol{x} \in \Omega. \\
\text{(e) } \mathcal{F}(\boldsymbol{x}, \boldsymbol{0}) = \boldsymbol{0}.
\end{array}\right. \tag{13.108}$$

$$\left\{\begin{array}{l}
\text{(a) } p : \Gamma_3 \times \mathbb{R} \to \mathbb{R}_+. \\
\text{(b) There exists } L_p > 0 \text{ such that} \\
\qquad |p(\boldsymbol{x}, r_1) - p(\boldsymbol{x}, r_2)| \le L_p |r_1 - r_2| \\
\qquad\quad \forall\, r_1, r_2 \in \mathbb{R}, \ \text{a.e. } \boldsymbol{x} \in \Gamma_3. \\
\text{(c) } (p(\boldsymbol{x}, r_1) - p(\boldsymbol{x}, r_2))\,(r_1 - r_2) \ge 0 \\
\qquad \forall\, r_1, r_2 \in \mathbb{R}, \ \text{a.e. } \boldsymbol{x} \in \Gamma_3. \\
\text{(d) The mapping } \boldsymbol{x} \mapsto p(\boldsymbol{x}, r) \text{ is measurable on } \Gamma_3, \\
\qquad \text{for any } r \in \mathbb{R}. \\
\text{(e) } p(\boldsymbol{x}, r) = 0 \quad \text{for all} \quad r \le 0, \ \text{a.e. } \boldsymbol{x} \in \Gamma_3. \\
\text{(f) There exists } p^* \in \mathbb{R} \text{ such that } p(x, r) \le p^* \\
\qquad \text{for all} \quad r \ge 0, \ \text{a.e. } \boldsymbol{x} \in \Gamma_3.
\end{array}\right. \tag{13.109}$$

The coefficient of friction is such that

$$\mu \in L^\infty(\Gamma_3), \qquad \mu(\boldsymbol{x}) \geq 0 \quad \text{a.e. } \boldsymbol{x} \in \Gamma_3. \tag{13.110}$$

Moreover, we assume that

$$d_0^2 L_p \|\mu\|_{L^\infty(\Gamma_3)} < m_{\mathcal{F}} \tag{13.111}$$

where d_0, $m_{\mathcal{F}}$, and L_p are the constants which appear in (13.100), (13.108)(d), and (13.109)(b), respectively. Note that inequality (13.111) could be interpreted as a smallness condition on the coefficient of friction. Such kind of conditions are often used in the variational analysis of frictional contact problems with elastic materials, as explained in [38] and the references therein.

Let K denote the set defined by

$$K = \{\, \boldsymbol{v} \in V \;:\; v_\nu \leq k \ \text{ a.e. on } \Gamma_3 \,\}, \tag{13.112}$$

and assume that the densities of body forces and tractions are such that $\boldsymbol{f}_0 \in L^2(\Omega)^d$, $\boldsymbol{f}_2 \in L^2(\Gamma_2)^d$. We now derive the variational formulation of Problem $\mathcal{P}$ and, to this end, we assume that $(\boldsymbol{u}, \boldsymbol{\sigma})$ are sufficiently regular functions which satisfy (13.102)–(13.107). Then, using (13.106) and (13.112) it follows that

$$\boldsymbol{u} \in K. \tag{13.113}$$

Let $\boldsymbol{v} \in K$. We use Green's formula (13.101) and equalities (13.103)–(13.105) to see that

$$\begin{aligned} \int_\Omega \boldsymbol{\sigma} \cdot (\boldsymbol{\varepsilon}(\boldsymbol{v}) - \boldsymbol{\varepsilon}(\boldsymbol{u}))\, dx = \int_\Omega \boldsymbol{f}_0 \cdot (\boldsymbol{v} - \boldsymbol{u})\, dx \\ + \int_{\Gamma_2} \boldsymbol{f}_2 \cdot (\boldsymbol{v} - \boldsymbol{u})\, da + \int_{\Gamma_3} \boldsymbol{\sigma}\boldsymbol{\nu} \cdot (\boldsymbol{v} - \boldsymbol{u})\, da. \end{aligned} \tag{13.114}$$

Moreover, using the boundary conditions (13.106) and (13.107) it is easy to see that

$$\sigma_\nu (v_\nu - u_\nu) \geq p(u_\nu)(u_\nu - v_\nu) \qquad \text{a.e. on } \Gamma_3,$$

$$\boldsymbol{\sigma}_\tau (\boldsymbol{v}_\tau - \boldsymbol{u}_\tau) \geq \mu\, p(u_\nu)(\|\boldsymbol{u}_\tau\| - \|\boldsymbol{v}_\tau\|) \qquad \text{a.e. on } \Gamma_3.$$

Therefore, since

$$\boldsymbol{\sigma}\boldsymbol{\nu} \cdot (\boldsymbol{v} - \boldsymbol{u}) = \sigma_\nu (v_\nu - u_\nu) + \boldsymbol{\sigma}_\tau (\boldsymbol{v}_\tau - \boldsymbol{u}_\tau) \qquad \text{a.e. on } \Gamma_3,$$

we deduce that

$$\int_{\Gamma_3} \boldsymbol{\sigma}\boldsymbol{\nu}\cdot(\boldsymbol{v}-\boldsymbol{u})\,da \tag{13.115}$$

$$\geq \int_{\Gamma_3} p(u_\nu)(u_\nu - v_\nu)\,da + \int_{\Gamma_3} \mu\, p(u_\nu)(\|\boldsymbol{u}_\tau\| - \|\boldsymbol{v}_\tau\|)\,da.$$

Next, we combine equality (13.114) with inequality (13.115), then we use the constitutive law (13.102) and the regularity (13.113). As a result we find the following variational formulation of Problem P.

Problem P^V. Given $\boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \in Y$, find $\boldsymbol{u}$ such that

$$\boldsymbol{u} \in K, \quad \int_\Omega \mathcal{F}\boldsymbol{\varepsilon}(\boldsymbol{u})\cdot(\boldsymbol{\varepsilon}(\boldsymbol{v})-\boldsymbol{\varepsilon}(\boldsymbol{u}))\,dx \tag{13.116}$$

$$+\int_{\Gamma_3} p(u_\nu)(v_\nu - u_\nu)\,da + \int_{\Gamma_3} \mu\, p(u_\nu)(\|\boldsymbol{v}_\tau\| - \|\boldsymbol{u}_\tau\|)\,da$$

$$\geq \int_\Omega \boldsymbol{f}_0\cdot(\boldsymbol{v}-\boldsymbol{u})\,dx + \int_{\Gamma_2} \boldsymbol{f}_2\cdot(\boldsymbol{v}-\boldsymbol{u})\,da \qquad \forall\, \boldsymbol{v}\in K.$$

Note that Problem P^V is formulated in terms of the displacement field. Once the displacement field is known, the stress field can be easily obtained by using the constitutive law (13.102). A couple $(\boldsymbol{u}, \boldsymbol{\sigma})$ which satisfies (13.102) and (13.116) is called a *weak solution* to the contact problem P.

13.4.3 Weak Solvability

Our main result in this section, which represents a continuation of our previous results in [5, 39], is the following.

Theorem 13.4.1 *Assume that (13.108)–(13.111) hold. Then, for each* $\boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \in Y$ *there exists a unique solution* $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{f})$ *to the variational inequality (13.116). Moreover, if* $\boldsymbol{f}_n = (\boldsymbol{f}_{0n}, \boldsymbol{f}_{2n}) \in Y$, $\boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \in Y$, *and* $\boldsymbol{f}_{0n} \rightharpoonup \boldsymbol{f}_0$ *in* $L^2(\Omega)^d$, $\boldsymbol{f}_{2n} \rightharpoonup \boldsymbol{f}_2$ *in* $L^2(\Gamma_2)^d$, *as* $n \to \infty$, *then* $\boldsymbol{u}_n(\boldsymbol{f}_n) \to \boldsymbol{u}(\boldsymbol{f})$ *in* X, *as* $n \to \infty$.

Note that Theorem 13.4.1 provides the existence of a unique weak solution to the frictional contact Problem P as well as its continuous dependence with respect to the density of body forces and tractions.

The proof of Theorem 13.4.1 will be carried out in several steps, based on the abstract existence and convergence results in Sect. 13.2. To present it assume in what follows that (13.108)–(13.111) hold and we consider the operator $A : V \to V$, the function $j : V \times V \to \mathbb{R}$, and the operator $\pi : V \to Y$ defined by

$$(A\boldsymbol{u}, \boldsymbol{v})_V = \int_\Omega \mathcal{F}\boldsymbol{\varepsilon}(\boldsymbol{u}) \cdot \boldsymbol{\varepsilon}(\boldsymbol{v})\,dx + \int_{\Gamma_3} p(u_\nu)v_\nu\,da \qquad \forall\,\boldsymbol{u},\ \boldsymbol{v} \in V, \quad (13.117)$$

$$j(\boldsymbol{u}, \boldsymbol{v}) = \int_{\Gamma_3} \mu\, p(u_\nu)\|\boldsymbol{v}_\tau\|\,da \qquad \forall\,\boldsymbol{u} \in V,\ \boldsymbol{v} \in V, \quad (13.118)$$

$$\pi\boldsymbol{v} = (\iota\boldsymbol{v}, \gamma_2\boldsymbol{v}) \qquad \forall\,\boldsymbol{v} \in V. \quad (13.119)$$

Here $\iota : V \to L^2(\Omega)^d$ is the canonic embedding and $\gamma_2 : V \to L^2(\Gamma_2)^d$ is the restriction to the trace map to Γ_2. The first step of the proof is the following.

Lemma 13.4.2 *Given $\boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \in Y$, an element $\boldsymbol{u} \in V$ is solution to the variational inequality (13.116) if and only if*

$$\boldsymbol{u} \in K, \qquad (A\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{u})_V + j(\boldsymbol{u}, \boldsymbol{v}) - j(\boldsymbol{u}, \boldsymbol{u}) \geq (\boldsymbol{f}, \pi\boldsymbol{v} - \pi\boldsymbol{u})_Y \qquad \forall\,\boldsymbol{v} \in K. \quad (13.120)$$

Proof The statement of Lemma 13.4.2 is a direct consequence of the notation (13.117)–(13.119). □

Lemma 13.4.3 *The function j defined by (13.118) satisfies conditions (13.11), (13.14), and (13.19) on the space $X = V$.*

Proof Condition (13.11)(a) is obviously satisfied. On the other hand, an elementary calculation based on the definition (13.118) and assumptions (13.109), (13.110) yields

$$\begin{aligned} &j(\boldsymbol{u}_1, \boldsymbol{v}_2) - j(\boldsymbol{u}_1, \boldsymbol{v}_1) + j(\boldsymbol{u}_2, \boldsymbol{v}_1) - j(\boldsymbol{u}_2, \boldsymbol{v}_2) \\ &\leq L_p\|\mu\|_{L^\infty(\Gamma_3)} \int_{\Gamma_3} \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|\,da \end{aligned}$$

for all $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{v}_1, \boldsymbol{v}_2 \in V$. Therefore, the trace inequality (13.100) shows that condition (13.11)(b) holds with $\alpha = d_0^2 L_p\|\mu\|_{L^\infty(\Gamma_3)}$. Next, using assumptions (13.109)(b),(e) and (13.110) it is easy to see that

$$\begin{aligned} j(\boldsymbol{\eta}, \boldsymbol{v}_1) - j(\boldsymbol{\eta}, \boldsymbol{v}_2) &\leq \int_{\Gamma_3} \mu p(\eta_\nu)\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|\,da \\ &\leq L_p\|\mu\|_{L^\infty(\Gamma_3)} \int_{\Gamma_3} \|\boldsymbol{\eta}\|\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|\,da \end{aligned}$$

for all $\boldsymbol{\eta}, \boldsymbol{v}_1, \boldsymbol{v}_2 \in V$. Therefore, the trace inequality (13.100) shows that condition (13.14)(b) is satisfied with $\beta = 0$ and $\gamma = d_0^2 L_p\|\mu\|_{L^\infty(\Gamma_3)}$. Finally, note that condition (13.19) holds from assumption (13.109) and the compactness of the trace operator, since $\boldsymbol{u}_k \rightharpoonup \boldsymbol{u}$ in V implies that $p(u_{k\nu}) \to p(u_\nu)$ and $\|\boldsymbol{u}_{k\tau}\| \to \|\boldsymbol{u}_\tau\|$ in $L^2(\Gamma_3)$. □

We now have all the ingredients to provide the proof of Theorem 13.4.1.

Proof The set K is obviously a convex nonempty subset of V. Moreover, using the properties of the trace map we deduce that K is closed and, therefore, (13.9) holds. Next, we use assumptions (13.108) and (13.109) and the trace inequality (13.100) to see that

$$(A\boldsymbol{u} - A\boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v})_V \geq m_{\mathcal{F}} \|\boldsymbol{u} - \boldsymbol{v}\|_V^2,$$

$$\|A\boldsymbol{u} - A\boldsymbol{v}\|_V \leq \left(L_{\mathcal{F}} + d_0^2 L_p\right) \|\boldsymbol{u} - \boldsymbol{v}\|_V$$

for all $\boldsymbol{u}, \boldsymbol{v} \in V$. Therefore, condition (13.10) holds with $X = V$ and $m = m_{\mathcal{F}}$. On the other hand, Lemma 13.4.3 guarantees that the functional (13.118) satisfies conditions (13.11), (13.14), and (13.19) on the space $X = V$, with $\alpha = d_0^2 L_p \|\mu\|_{L^\infty(\Gamma_3)}$, $\beta = 0$, and $\gamma = d_0^2 L_p \|\mu\|_{L^\infty(\Gamma_3)}$. Therefore, using (13.111) it follows that the smallness assumption (13.12) is satisfied and, moreover, (13.15) holds, too. Finally, we note that conditions (13.13) and (13.20) are a direct consequence of definition (13.119) combined with the properties of the operators ι and γ_2.

It follows from above that we are in a position to apply Theorem 13.2.12 on the space $X = V$. As a result we deduce the unique solvability of the variational inequality (13.120), for each $\boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \in Y$. This result combined with Lemma 13.4.2 proves the existence of a unique solution to the variational inequality (13.116), for each $\boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \in Y$.

Assume now that $\boldsymbol{f}_n = (\boldsymbol{f}_{0n}, \boldsymbol{f}_{2n}) \in Y$, $\boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \in Y$, and $\boldsymbol{f}_{0n} \rightharpoonup \boldsymbol{f}_0$ in $L^2(\Omega)^d$, $\boldsymbol{f}_{2n} \rightharpoonup \boldsymbol{f}_2$ in $L^2(\Gamma_2)^d$, as $n \to \infty$. Then $\boldsymbol{f}_n \rightharpoonup \boldsymbol{f}$ in Y, as $n \to \infty$. Therefore, using Theorem 13.2.13 and Lemma 13.4.2 we deduce that $\boldsymbol{u}(\boldsymbol{f}_n) \to \boldsymbol{u}(\boldsymbol{f})$ in V, as $n \to \infty$, which concludes the proof. □

13.4.4 Optimal Control

We now associate to Problem P^V the set of admissible pairs $\mathcal{V}_{ad}$ and the cost function $\mathcal{L}$ given by

$$\mathcal{V}_{ad} = \{\, (\boldsymbol{u}, \boldsymbol{f}) \in K \times Y \text{ such that } \boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \text{ and (13.116) holds} \},\quad (13.121)$$

$$\mathcal{L}(\boldsymbol{u}, \boldsymbol{f}) = a_0 \int_\Omega \|\boldsymbol{f}_0\|^2\, dx + a_2 \int_{\Gamma_2} \|\boldsymbol{f}_2\|^2\, da + a_3 \int_{\Gamma_3} |u_\nu - \theta|^2\, da \quad (13.122)$$

for all $\boldsymbol{u} \in V$, $\boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \in Y$. Here θ is a given element in $L^2(\Gamma_3)$ and a_0, a_2, a_3 are strictly positive constants. Moreover, we consider the following optimal control problem.

Problem Q^V. Find $(\boldsymbol{u}^*, \boldsymbol{f}^*) \in \mathcal{V}_{ad}$ such that

$$\mathcal{L}(\boldsymbol{u}^*, \boldsymbol{f}^*) = \min_{(\boldsymbol{u},\boldsymbol{f})\in\mathcal{V}_{ad}} \mathcal{L}(\boldsymbol{u}, \boldsymbol{f}). \tag{13.123}$$

Our first result in this subsection is the following.

Theorem 13.4.4 *Assume that (13.108)–(13.111) hold. Then, the optimal control problem Q^V has at least one solution $(\boldsymbol{u}^*, \boldsymbol{f}^*)$.*

Proof It is easy to see that the function $\mathcal{L}$ defined by (13.122) satisfy conditions (13.32)–(13.34) on the spaces $X = V$, $Y = L^2(\Omega)^d \times L^2(\Gamma_3)^d$ with

$$g(\boldsymbol{v}) = a_3 \int_{\Gamma_3} |v_\nu - \theta|^2\, da, \quad h(\boldsymbol{f}) = a_0 \int_\Omega \|\boldsymbol{f}_0\|^2\, dx + a_2 \int_{\Gamma_2} \|\boldsymbol{f}_2\|^2\, da$$

for all $\boldsymbol{v} \in V$, $\boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \in Y$. Therefore, the solvability of the optimal control problem Q^V is a direct consequence of Lemma 13.4.2 and Theorem 13.3.1. □

Next, besides conditions (13.108)–(13.111), we assume that

$$B \text{ is a closed convex set of } \mathbb{S}^d \text{ such that } \boldsymbol{0} \in B \tag{13.124}$$

and we denote by $P_B : \mathbb{S}^d \to B$ the projection operator. We also consider the sequences $\{\omega_n\}$, $\{k_n\}$, and $\{\varepsilon_n\}$ such that

$$\omega_n \geq 0, \qquad k_n > 0, \qquad \varepsilon_n \geq 0 \qquad \forall\, n \in \mathbb{N} \tag{13.125}$$

and, for each $n \in \mathbb{N}$, we define the set K_n by

$$K_n = \{\, \boldsymbol{v} \in V \ :\ v_\nu \leq k_n \ \text{ a.e. on } \Gamma_3 \,\}. \tag{13.126}$$

With these data, we consider the following perturbation of Problem P^V.

Problem P_n^V. Given $\boldsymbol{f}_n = (\boldsymbol{f}_{0n}, \boldsymbol{f}_{2n}) \in Y$, find $\boldsymbol{u}_n \in V$ such that

$$\boldsymbol{u}_n \in K_n, \quad \int_\Omega \mathcal{F}\boldsymbol{\varepsilon}(\boldsymbol{u}_n) \cdot (\boldsymbol{\varepsilon}(\boldsymbol{v}) - \boldsymbol{\varepsilon}(\boldsymbol{u}_n))\, dx \tag{13.127}$$

$$+\omega_n \int_\Omega (\boldsymbol{\varepsilon}(\boldsymbol{u}_n) - P_B\boldsymbol{\varepsilon}(\boldsymbol{u}_n)) \cdot (\boldsymbol{\varepsilon}(\boldsymbol{v}) - \boldsymbol{\varepsilon}(\boldsymbol{u}_n))\, dx + \int_{\Gamma_3} p(u_{n\nu})(v_\nu - u_{n\nu})\, da$$

$$+ \int_{\Gamma_3} \mu\, p(u_{n\nu})\big(\sqrt{\|\boldsymbol{v}_\tau\|^2 + \varepsilon_n^2} - \sqrt{\|\boldsymbol{u}_{n\tau}\|^2 + \varepsilon_n^2}\big)\, da$$

$$\geq \int_\Omega \boldsymbol{f}_{0n} \cdot (\boldsymbol{v} - \boldsymbol{u}_n)\, dx + \int_{\Gamma_2} \boldsymbol{f}_{2n} \cdot (\boldsymbol{v} - \boldsymbol{u}_n)\, da \qquad \forall\, \boldsymbol{v} \in K_n.$$

Note that Problem P_n^V represents the variational formulation of an elastic contact problem of the form (13.102)–(13.107) in which the following changes have been operated in the model:

- The elastic constitutive law (13.102) was replaced with the constitutive law $\boldsymbol{\sigma} = \mathcal{F}\boldsymbol{\varepsilon}(\boldsymbol{u})) + \omega_n(\boldsymbol{\varepsilon}(\boldsymbol{u}) - P_B\boldsymbol{\varepsilon}(\boldsymbol{u}))$. Such kind of constitutive law has been used by many authors, see [38, 41] and the references therein.
- The bound k in (13.106) was replaced by a perturbation, denoted k_n.
- The Coulomb law of dry friction (13.107) was replaced with its regularization

$$-\boldsymbol{\sigma}_\tau = \mu\, p(u_\nu)\, \frac{\boldsymbol{u}_\tau}{\sqrt{\|\boldsymbol{u}_\tau\|^2 + \varepsilon_n^2}}\,.$$

- The densities $\boldsymbol{f}_0$ and $\boldsymbol{f}_2$ of body forces and tractions, respectively, were replaced by their perturbations $\boldsymbol{f}_{0n}$ and $\boldsymbol{f}_{2n}$, respectively.

For Problem P_n^V we have the following existence and uniqueness result.

Theorem 13.4.5 *Assume (13.108)–(13.111) and, moreover, assume that (13.124)–(13.125) hold. Then, for each* $\boldsymbol{f}_n = (\boldsymbol{f}_{0n}, \boldsymbol{f}_{2n}) \in Y$ *there exists a unique solution* $\boldsymbol{u}_n = \boldsymbol{u}_n(\boldsymbol{f}_n)$ *to the variational inequality (13.127).*

Proof The proof of Theorem 13.4.5 is based on arguments similar to those used in the proof of Theorem 13.4.1 and, for this reason, we skip the details. The steps of the proof are the following.

i) *The quasivariational inequality.* Besides the operator π defined in (13.119), for each $n \in \mathbb{N}$ we consider the operator $A_n : V \to V$ and the function $j_n : V \times V \to \mathbb{R}$ given by

$$(A_n\boldsymbol{u}, \boldsymbol{v})_V = \int_\Omega \mathcal{F}\boldsymbol{\varepsilon}(\boldsymbol{u}) \cdot \boldsymbol{\varepsilon}(\boldsymbol{v})\, dx + \int_{\Gamma_3} p(u_\nu)v_\nu\, da \tag{13.128}$$

$$+ \omega_n \int_\Omega (\boldsymbol{\varepsilon}(\boldsymbol{u}) - P_B\boldsymbol{\varepsilon}(\boldsymbol{u})) \cdot \boldsymbol{\varepsilon}(\boldsymbol{v})\, dx \qquad \forall\, \boldsymbol{u},\ \boldsymbol{v} \in V.$$

$$j_n(\boldsymbol{u}, \boldsymbol{v}) = \int_{\Gamma_3} \mu\, p(u_\nu)\left(\sqrt{\|\boldsymbol{v}_\tau\|^2 + \varepsilon_n^2} - \varepsilon_n\right) da \quad \forall\, \boldsymbol{u},\ \boldsymbol{v} \in V. \tag{13.129}$$

Then, it is easy to see that, given $\boldsymbol{f}_n = (\boldsymbol{f}_{0n}, \boldsymbol{f}_{2n}) \in Y$, an element $\boldsymbol{u}_n \in V$ is a solution to inequality (13.127) if and only if

$$\boldsymbol{u}_n \in K_n, \qquad (A_n\boldsymbol{u}_n, \boldsymbol{v} - \boldsymbol{u}_n)_V + j_n(\boldsymbol{u}_n, \boldsymbol{v}) - j_n(\boldsymbol{u}_n, \boldsymbol{u}_n) \qquad (13.130)$$
$$\geq (\boldsymbol{f}_n, \pi\boldsymbol{v} - \pi\boldsymbol{u}_n)_Y \qquad \forall\, \boldsymbol{v} \in K_n.$$

ii) *The operator* A_n. First, we recall that the projector operator $P_B : \mathbb{S}^d \to B$ is nonexpansive, i.e.,

$$\|P_B\boldsymbol{\tau}_1 - P_B\boldsymbol{\tau}_2\| \leq \|\boldsymbol{\tau}_1 - \boldsymbol{\tau}_2\| \qquad (13.131)$$

for all $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2 \in \mathbb{S}^d$. This inequality implies that

$$\big((\boldsymbol{\tau}_1 - P_B\boldsymbol{\tau}_1) - (\boldsymbol{\tau}_2 - P_B\boldsymbol{\tau}_2)\big) \cdot (\boldsymbol{\tau}_1 - \boldsymbol{\tau}_2) \geq 0, \qquad (13.132)$$

for all $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2 \in \mathbb{S}^d$. Therefore, using assumptions (13.108) and (13.109), the trace inequality (13.100), and estimates (13.131), (13.132) we deduce that

$$(A_n\boldsymbol{u} - A_n\boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v})_V \geq m_{\mathcal{F}}\|\boldsymbol{u} - \boldsymbol{v}\|_V^2,$$

$$\|A_n\boldsymbol{u} - A_n\boldsymbol{v}\|_V \leq (L_{\mathcal{F}} + d_0^2 L_p + 2\omega_n)\,\|\boldsymbol{u} - \boldsymbol{v}\|_V$$

for all $\boldsymbol{u}, \boldsymbol{v} \in X$. It follows from here that condition (13.48) holds with $X = V$, $m_n = m_{\mathcal{F}}$, and $M_n = L_{\mathcal{F}} + d_0^2 L_p + 2\omega_n$.

iii) *The function* j_n. We claim that the function j_n defined by (13.129) satisfies conditions (13.11), (13.14), and (13.19) on the space $X = V$.

First, condition (13.11)(a) is obviously satisfied. On the other hand, an elementary calculation based on the definition (13.129) and assumptions (13.109), (13.110), combined with inequality

$$\Big|\sqrt{a^2 + \varepsilon^2} - \sqrt{b^2 + \varepsilon^2}\Big| \leq |a - b| \qquad \forall\, a,\, b,\, \varepsilon > 0,$$

implies that

$$j_n(\boldsymbol{u}_1, \boldsymbol{v}_2) - j_n(\boldsymbol{u}_1, \boldsymbol{v}_1) + j_n(\boldsymbol{u}_2, \boldsymbol{v}_1) - j_n(\boldsymbol{u}_2, \boldsymbol{v}_2)$$
$$\leq L_p\|\mu\|_{L^\infty(\Gamma_3)} \int_{\Gamma_3} \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|\, da$$

for all $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{v}_1, \boldsymbol{v}_2 \in V$. Therefore, the trace inequality (13.100) shows that condition (13.11)(b) holds with $\alpha = d_0^2 L_p\|\mu\|_{L^\infty(\Gamma_3)}$. A similar argument shows that condition (13.14) is satisfied with $\beta = 0$ and $\gamma = d_0^2 L_p\|\mu\|_{L^\infty(\Gamma_3)}$. Finally, note that condition (13.19) holds from assumption (13.109) and the compactness of the

trace operator, since $\boldsymbol{u}_k \rightharpoonup \boldsymbol{u}$ in V implies that $p(u_{k\nu}) \to p(u_\nu)$ and $\|\boldsymbol{u}_{k\tau}\| \to \|\boldsymbol{u}_\tau\|$ in $L^2(\Gamma_3)$.
iv) *End of proof.* The set K_n is obviously a convex nonempty subset of V. Moreover, recall that step ii) shows that condition (13.48) holds with $X = V$ and $m = m_{\mathcal{F}}$. On the other hand, step iii) guarantees that the functional (13.118) satisfies conditions (13.11), (13.14), and (13.19) on the space $X = V$, with $\alpha = \gamma = d_0^2 L_p \|\mu\|_{L^\infty(\Gamma_3)}$. Therefore, using (13.111) it follows that the smallness assumption (13.12) is satisfied and, moreover, (13.15) holds, too. Finally, we note that conditions (13.13) and (13.20) are a direct consequence of definition (13.119) combined with the properties of the operators ι and γ_2.

It follows from above that we are in a position to apply Theorem 13.2.12 on the space $X = V$. In this way we deduce the unique solvability of the quasivariational inequality (13.130), for each $\boldsymbol{f}_n = (\boldsymbol{f}_{0n}, \boldsymbol{f}_{2n}) \in Y$. This result, combined with step i), leads to the existence of a unique solution to the quasivariational inequality (13.127), for each $\boldsymbol{f}_n = (\boldsymbol{f}_{0n}, \boldsymbol{f}_{2n}) \in Y$. □

We now move to the control of Problem P_n^V. The set of admissible pairs for this problem is given by

$$\mathcal{V}_{ad}^n = \{\, (\boldsymbol{u}_n, \boldsymbol{f}_n) \in K_n \times Y \text{ s. t. } \boldsymbol{f}_n = (\boldsymbol{f}_{0n}, \boldsymbol{f}_{2n}) \text{ and (13.127) holds}\,\}. \tag{13.133}$$

Moreover, the corresponding optimal control problem is the following.

Problem Q_n^V. Find $(\boldsymbol{u}_n^*, \boldsymbol{f}_n^*) \in \mathcal{V}_{ad}^n$ such that

$$\mathcal{L}(\boldsymbol{u}_n^*, \boldsymbol{f}_n^*) = \min_{(\boldsymbol{u}_n, \boldsymbol{f}_n) \in \mathcal{V}_{ad}^n} \mathcal{L}(\boldsymbol{u}_n, \boldsymbol{f}_n). \tag{13.134}$$

Our main result in this section is the following.

Theorem 13.4.6 *Assume (13.108)–(13.111), (13.124), and (13.125). Then the following statements hold.*

i) *For each $n \in \mathbb{N}$ the optimal control problem Q_n^V has at least one solution $(\boldsymbol{u}_n^*, \boldsymbol{f}_n^*)$.*
ii) *If $\omega_n \to 0$, $k_n \to k$, and $\varepsilon_n \to \varepsilon$ as $n \to \infty$, then for any sequence $\{(\boldsymbol{u}_n^*, \boldsymbol{f}_n^*)\}$ of solutions of Problem Q_n^V there exists a subsequence, again denoted $\{(\boldsymbol{u}_n^*, \boldsymbol{f}_n^*)\}$, and an element $(\boldsymbol{u}^*, \boldsymbol{f}^*) \in X \times Y$, such that*

$$\boldsymbol{f}_n^* \rightharpoonup \boldsymbol{f}^* \quad \text{in} \quad Y, \quad \text{as} \quad n \to \infty, \tag{13.135}$$

$$\boldsymbol{u}_n^* \to \boldsymbol{u}^* \quad \text{in} \quad X \quad \text{as} \quad n \to \infty, \tag{13.136}$$

$$(\boldsymbol{u}^*, \boldsymbol{f}^*) \quad \text{is a solution of Problem } Q^V. \tag{13.137}$$

Proof i) First, it is easy to see that condition (13.75) holds with $c_n = \frac{k_n}{k} > 0$. Next, we use (13.131), (13.132) to see that condition (13.76) is satisfied with

$$(T_n \boldsymbol{u}, \boldsymbol{v})_V = \omega_n \int_\Omega (\boldsymbol{\varepsilon}(\boldsymbol{u}) - P_B \boldsymbol{\varepsilon}(\boldsymbol{u})) \cdot \boldsymbol{\varepsilon}(\boldsymbol{v})\, dx \qquad \forall\, \boldsymbol{u},\ \boldsymbol{v} \in V. \tag{13.138}$$

On the other hand, it follows from step iii) in the proof of Theorem 13.4.5 that the function j_n satisfies condition (13.77) with $\alpha_n = d_0^2 L_p \|\mu\|_{L^\infty(\Gamma_3)}$, $\beta_n = 0$, and $\gamma_n = d_0^2 L_p \|\mu\|_{L^\infty(\Gamma_3)}$. In addition, recall that the function $\mathcal{L}$ defined by (13.122) satisfies conditions (13.32)–(13.34) on the spaces $X = V$, $Y = L^2(\Omega)^d \times L^2(\Gamma_3)^d$ with

$$g(\boldsymbol{v}) = a_3 \int_{\Gamma_3} |v_\nu - \theta|^2\, da, \quad h(\boldsymbol{f}) = a_0 \int_\Omega \|\boldsymbol{f}_0\|^2\, dx + a_2 \int_{\Gamma_2} \|\boldsymbol{f}_2\|^2\, da.$$

Therefore, as already explained in page 463, we are in a position to apply Theorem 13.3.5 to deduce that the optimal control problem Q_n^V has at least one solution $(\boldsymbol{u}_n^*, \boldsymbol{f}_n^*)$, for each $n \in \mathbb{N}$.

ii) Assume now that $\omega_n \to 0$, $k_n \to k$, and $\varepsilon_n \to 0$, as $n \to \infty$. In order to prove (13.135)–(13.137) we use Theorem 13.3.8 and, to this end, we start by checking that conditions (13.78)–(13.82) are satisfied.

First, since $c_n = \frac{k_n}{k}$ we deduce that condition (13.78) holds. Moreover, it is easy to see that condition (13.79) holds, too. On the other hand, assumption $\boldsymbol{0} \in B$ combined with inequality (13.131) shows that $\|\boldsymbol{\tau} - P_B \boldsymbol{\tau}\| \le 2\,\|\boldsymbol{\tau}\|$ for all $\boldsymbol{\tau} \in \mathbb{S}^d$. Therefore, definition (13.138) implies that

$$(T_n \boldsymbol{v}, \boldsymbol{w})_V = \omega_n \int_\Omega \|\boldsymbol{\varepsilon}(\boldsymbol{v}) - P_B \boldsymbol{\varepsilon}(\boldsymbol{v})\|\, \|\boldsymbol{\varepsilon}(\boldsymbol{w})\|\, dx \le 2\omega_n \|\boldsymbol{v}\|_V \|\boldsymbol{w}\|_V$$

for all $\boldsymbol{u},\ \boldsymbol{v} \in V$, $n \in \mathbb{N}$, which shows that

$$\|T_n \boldsymbol{v}\|_V \le 2\omega_n \|\boldsymbol{v}\|_V \qquad \forall\, \boldsymbol{v} \in V,\ n \in \mathbb{N}.$$

We conclude from here that condition (13.80) holds with $F_n = 2\omega_n$ and $\delta_n = 0$.

Assume now that $n \in \mathbb{N}$ is fixed and $\boldsymbol{v}_1,\ \boldsymbol{v}_2 \in V$. We use definitions (13.129) and (13.118) to see that

$$j_n(\boldsymbol{v}_1, \boldsymbol{v}_2) - j_n(\boldsymbol{v}_1, \boldsymbol{v}_1) + j(\boldsymbol{v}_2, \boldsymbol{v}_1) - j(\boldsymbol{v}_2, \boldsymbol{v}_2)$$

$$= \int_{\Gamma_3} \mu\, p(v_{1\nu}) \Big(\sqrt{\|\boldsymbol{v}_{2\tau}\|^2 + \varepsilon_n^2} - \varepsilon_n - \|\boldsymbol{v}_{2\tau}\| \Big)\, da$$

$$+\int_{\Gamma_3} \mu\, p(v_{1\nu})\big(\|\boldsymbol{v}_{1\tau}\| - \sqrt{\|\boldsymbol{v}_{1\tau}\|^2 + \varepsilon_n^2} + \varepsilon_n\big)\, da$$

$$+\int_{\Gamma_3} \mu\, \big(p(v_{1\nu}) - p(v_{2\nu})\big)\big(\|\boldsymbol{v}_{2\tau}\| - \|\boldsymbol{v}_{1\tau}\|\big)\, da.$$

Therefore, using assumptions (13.109)(b) and (13.110) combined with the inequality

$$\Big|\sqrt{a^2+\varepsilon^2} - a - \varepsilon\Big| \le \varepsilon \qquad \forall\, a,\ \varepsilon > 0,$$

we find that

$$j_n(\boldsymbol{v}_1, \boldsymbol{v}_2) - j_n(\boldsymbol{v}_1, \boldsymbol{v}_1) + j(\boldsymbol{v}_2, \boldsymbol{v}_1) - j(\boldsymbol{v}_2, \boldsymbol{v}_2)$$

$$\le 2\varepsilon_n \|\mu\|_{L^\infty(\Gamma_3)} \int_{\Gamma_3} p(v_{1\nu})\, da + L_p \|\mu\|_{L^\infty(\Gamma_3)} \int_{\Gamma_3} \|\boldsymbol{v}_{1\tau} - \boldsymbol{v}_{2\tau}\|^2\, da.$$

Next, we use assumption (13.109)(f) and the trace inequality (13.100) to deduce that

$$j_n(\boldsymbol{v}_1, \boldsymbol{v}_2) - j_n(\boldsymbol{v}_1, \boldsymbol{v}_1) + j(\boldsymbol{v}_2, \boldsymbol{v}_1) - j(\boldsymbol{v}_2, \boldsymbol{v}_2)$$

$$\le 2\varepsilon_n p^* \|\mu\|_{L^\infty(\Gamma_3)} meas(\Gamma_3) + d_0^2 L_p \|\mu\|_{L^\infty(\Gamma_3)} \|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_V^2.$$

It follows from here that condition (13.81) holds with $G_n = 2\varepsilon_n p^* \|\mu\|_{L^\infty(\Gamma_3)}\, meas(\Gamma_3)$ and $H_n = 0$.

Recall now that $m_n = m_{\mathcal{F}}$, $\alpha = \gamma_n = d_0^2 L_p \|\mu\|_{L^\infty(\Gamma_3)}$, and $\beta_n = 0$, for each $n \in \mathbb{N}$. Moreover, $d_0^2 L_p \|\mu\|_{L^\infty(\Gamma_3)} < m_{\mathcal{F}}$, as assumed in (13.111). We conclude from here that condition (13.82) is satisfied.

Finally, note that the rest of the conditions in Theorem 13.3.8 are satisfied, as it follows from the previous results proved in this section. Theorem 13.4.6 is now a direct consequence of Theorem 13.3.8. □

Remark 13.4.7 As a consequence of Remarks 13.3.2 and 13.3.7 we deduce that the statements of Theorems 13.4.4 and 13.4.6 still remain valid if we replace the definition (13.121) and (13.133) with the following ones:

$$\mathcal{V}_{ad} = \{\, (\boldsymbol{u}, \boldsymbol{f}) \in K \times U \text{ s.t. } \boldsymbol{f} = (\boldsymbol{f}_0, \boldsymbol{f}_2) \text{ and (13.116) holds}\,\}, \tag{13.139}$$

$$\mathcal{V}_{ad}^n = \{\, (\boldsymbol{u}_n, \boldsymbol{f}_n) \in K_n \times U \text{ s.t. } \boldsymbol{f} = (\boldsymbol{f}_{0n}, \boldsymbol{f}_{2n}) \text{ and (13.127) holds}\,\}, \tag{13.140}$$

U being a given nonempty weakly closed subset of Y. The proof of this statement is based on the property (13.45) of the set U.

13.4.5 A One-Dimensional Example

In this subsection we illustrate our results in the study of a one-dimensional example. Thus, we consider Problem P in the particular case when $\Omega = (0, 1)$, $\Gamma_1 = \{0\}$, $\Gamma_2 = \emptyset$, $\Gamma_3 = \{1\}$. Note that in this case the linearized strain field is given by $\varepsilon = u'$, where, here and below, the prime denotes the derivative with respect to the spatial variable $x \in [0, 1]$. Moreover, we assume that the material is homogeneous and behaves linearly elastic. Therefore, the elasticity operator is $\mathcal{F}\varepsilon = E\varepsilon$ where $E > 0$ is the Young modulus of the material. In addition, we assume that the density of the body force does not depend on the spatial variable and we denote it by $f \in \mathbb{R}$. Then, the statement of the problem is the following.

Problem P^{1d}. Find a displacement field $u\colon [0, 1] \to \mathbb{R}$ and a stress field $\sigma\colon [0, 1] \to \mathbb{R}$ such that

$$\sigma(x) = E\,u'(x) \quad \text{for } x \in (0, 1), \tag{13.141}$$

$$\sigma'(x) + f = 0 \quad \text{for } x \in (0, 1), \tag{13.142}$$

$$u(0) = 0, \tag{13.143}$$

$$\left.\begin{array}{ll} u(1) \le k, & \sigma(1) + p(u(1)) \le 0, \\ (u(1) - k)(\sigma(1) + p(u(1))) = 0 & \end{array}\right\}. \tag{13.144}$$

Note that Problem P^{1d} models the contact of an elastic rod of length $l = 1$. The rod occupies the domain $[0, 1]$ on the Ox axis, is fixed at its end $x = 0$, as acted by a body force, and its extremity $x = 1$ is in contact with a foundation made of a deformable material of thickness $k > 0$, which covers a rigid body. The reaction of the deformable material is described with the function $p : \mathbb{R} \to \mathbb{R}$ which is positive, monotone, and vanishes for a negative argument. This physical setting is depicted in Fig. 13.1.

For the analysis of Problem P^{1d} we use the space

$$V = \{\, v \in H^1(0, 1) \; : \; v(0) = 0 \,\}$$

and the set of admissible displacement field defined by

$$K = \{\, u \in V \mid u(1) \le k \,\}.$$

The variational formulation of Problem P^{1d}, obtained using integration by parts, is the following.

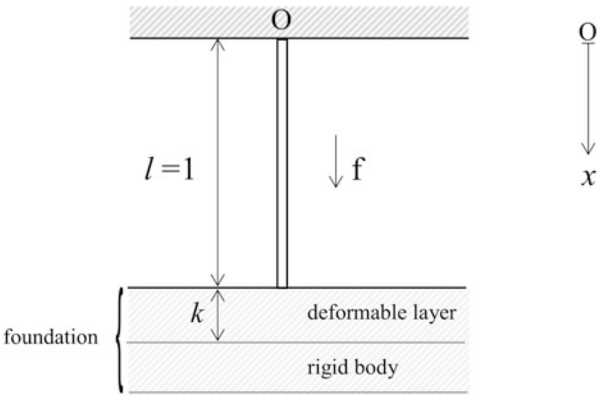

Fig. 13.1 Physical setting

Problem P_V^{1d}. Find a displacement field $u \in K$ such that

$$\int_0^1 Eu'(v' - u')\,dx + p(u(1))(v(1) - u(1)) \geq \int_0^1 f(v - u)\,dx \quad \forall\, v \in K. \tag{13.145}$$

The existence of a unique solution to Problem P_V^{1d} follows from Theorem 13.4.1. Consider now the case when

$$p(r) = r^+ = \begin{cases} 0 & \text{if} \quad r < 0, \\ r & \text{if} \quad r \geq 0. \end{cases} \tag{13.146}$$

Then, a simple calculation allows us to solve Problem P^{1d}. Three cases are possible, described below, together with the corresponding mechanical interpretations.

a) **The case** $f < 0$. In this case the body force acts in the opposite direction of the foundation and the solution of Problem P^{1d} is given by

$$\begin{cases} \sigma(x) = -fx + f, \\ u(x) = -\frac{f}{2E}\,x^2 + \frac{f}{E}\,x \end{cases} \qquad \forall\, x \in [0, 1]. \tag{13.147}$$

We have $u(1) < 0$ and $\sigma(1) = 0$ which shows that there is separation between the rod and the foundation and, therefore, there is no reaction on the point $x = 1$. This case corresponds to Fig. 13.2a.

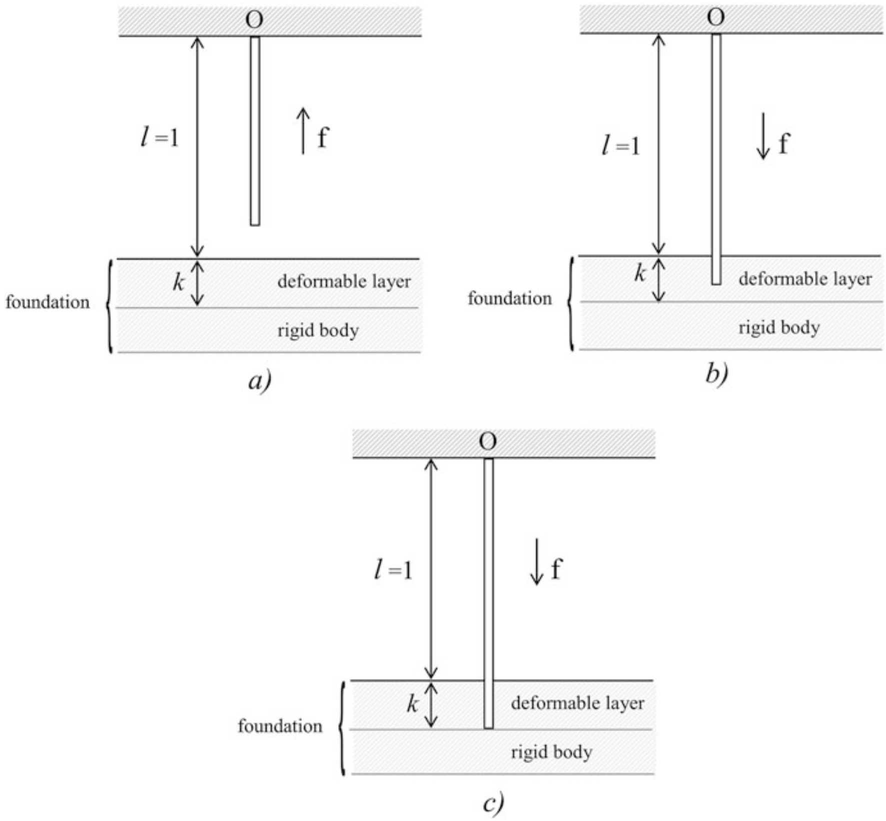

Fig. 13.2 The rod in contact with a foundation. (**a**) The case $f < 0$; (**b**) The case $0 \le f < 2k(E+1)$; (**c**) The case $f \ge 2k(E+1)$

b) **The case** $0 \le f < 2k(E+1)$. In this case the body force pushes the rod towards the foundation and the solution of Problem P^{1d} is given by

$$\begin{cases} \sigma(x) = -fx + \frac{f(2E+1)}{2(E+1)}, \\ u(x) = -\frac{f}{2E}x^2 + \frac{f(2E+1)}{2E(E+1)}x \end{cases} \quad \forall x \in [0,1]. \tag{13.148}$$

We have $0 \le u(1) < k$ and $\sigma(1) \le 0$ which shows that there is penetration into the deformable layer and the reaction of the foundation is towards the rod. Nevertheless, the penetration is partial, since $u(1) < k$. This case corresponds to Fig. 13.2b.

c) **The case** $f \ge 2k(E+1)$. In this case the solution of Problem P^{1d} is given by

$$\begin{cases} \sigma(x) = -fx + \frac{f}{2} + kE, \\ u(x) = -\frac{f}{2E}x^2 + \left(\frac{f}{2E} + k\right)x \end{cases} \quad \forall x \in [0,1]. \tag{13.149}$$

We have $u(1) = k$ which shows that the rigid-plastic layer is completely penetrated and the point $x = 1$ reaches the rigid body. This case corresponds to Fig. 13.2c).

We now formulate the optimal control problem Q^V in the one-dimensional case of Problem P^{1d}. In this particular setting $Y = L^2(0, 1)$ and we choose

$$U = \{ f \in Y \ : \ f \text{ is a constant} \}.$$

We use (13.139) to see that in this case

$$\mathcal{V}_{ad} = \{ (u, f) \in K \times U \ : (13.145) \text{ holds} \} \tag{13.150}$$

and

$$\mathcal{L}(u, f) = a_0 \, |f|^2 + a_3 \, |u(1) - \theta|^2, \tag{13.151}$$

where $\theta \in \mathbb{R}$, $a_0 > 0$, $a_3 > 0$. Then, using (13.123) we see that the problem can be formulated as follows.

Problem Q^{1d}. Find $(u^*, f^*) \in \mathcal{V}_{ad}$ such that

$$\mathcal{L}(u^*, f^*) = \min_{(u,g) \in \mathcal{V}_{ad}} \mathcal{L}(u, f). \tag{13.152}$$

We now take $E = 1$. Then, it is easy to see that if $0 \leq f < 4k$, then $0 \leq f < 2k(E+1)$ and, if $f \geq 4k$, then $f \geq 2k(E+1)$. Therefore, using (13.147)–(13.149) we have

$$u(x) = \begin{cases} -\frac{f}{2} x^2 + f x & \text{if } f < 0, \\ -\frac{f}{2} x^2 + \frac{3f}{4} x & \text{if } 0 \leq f < 4k \\ -\frac{f}{2} x^2 + (\frac{f}{2} + k) \, x & \text{if } f \geq 4k \end{cases} \qquad \forall \, x \in [0, 1]. \tag{13.153}$$

So,

$$u(1) = \begin{cases} \frac{f}{2} & \text{if } f < 0, \\ \frac{f}{4} & \text{if } 0 \leq f < 4k, \\ k & \text{if } f \geq 4k \end{cases}$$

and, using (13.151) with $\theta = 1$, $a_0 = 1$, $a_3 = 16$, we find that

$$\mathcal{L}(u, f) = \begin{cases} 5f^2 - 16f + 16 & \text{if } f < 0, \\ 2f^2 - 8f + 16 & \text{if } 0 \leq f < 4k, \\ f^2 + 16 & \text{if } f \geq 4k. \end{cases} \tag{13.154}$$

To conclude, problem (13.152) consists to minimize the function (13.154) when $f \in \mathbb{R}$, for a given $k > 0$. For this reason, in what follows we denote by J_k the function defined by (13.154), i.e.,

$$J_k(f) = \begin{cases} 5f^2 - 16f + 16 & \text{if } f < 0, \\ 2f^2 - 8f + 16 & \text{if } 0 \le f < 4k, \\ f^2 + 16(k-1)^2 & \text{if } f \ge 4k. \end{cases} \tag{13.155}$$

It is easy to see that this function is not a convex function. Nevertheless, it has a unique point of minimum given by

$$f^*(k) = \begin{cases} 4k & \text{if } 0 < k \le \frac{1}{2}, \\ 2 & \text{if } k > \frac{1}{2}. \end{cases} \tag{13.156}$$

Then, using (13.153) we find that the optimal control problem Q^{1d} has a unique solution $(u^*(k), f^*(k))$, given by

$$u^*(k) = \begin{cases} -\frac{f^*(k)}{2}\,x^2 + \left(\frac{f^*(k)}{2} + k\right)x & \forall\, x \in [0,1], \quad \text{if } 0 \le k \le \frac{1}{2}, \\ -\frac{f^*(k)}{2}\,x^2 + \frac{3f^*(k)}{4}\,x & \forall\, x \in [0,1], \quad \text{if } k > \frac{1}{2} \end{cases}$$

where, recall, $f^*(k)$ is given by (13.156). It is easy to see that when $k_n \to k$, then $f^*(k_n) \to f^*(k)$ and, therefore, $u^*(k_n) \to u^*(k)$. This represents a validation of the abstract convergence result in Theorem 13.3.8.

13.5 Conclusion

In this chapter we studied an optimal control problem for elliptic quasivariational inequalities in Hilbert spaces. We provided the existence of optimal pairs and proved a convergence result. The proofs were based on arguments of monotonicity and lower semicontinuity. Then, we applied these abstract results in the study of a mathematical model which describes the equilibrium of an elastic body in frictional contact with an obstacle, the so-called foundation. We presented various mechanical interpretations of these results and we exemplified them in the particular case of an elastic rod in contact with a rigid body covered by a layer of soft material.

The study presented in this chapter gives rise to several open problems that we describe in what follows. Any progress in these directions will complete our work and will open the way for new advances and ideas.

First, it would be interesting to derive necessary optimality conditions in the study of Problem $\mathcal{Q}$ introduced on page 447. Due to the nonsmooth and nonconvex feature of the functional $\mathcal{L}$, the treatment of this problem requires the use of its approximation by smooth optimization problems. And, in this matter, the abstract convergence results for the optimal pairs in this chapter could be a crucial tool. Next, it would be useful to establish an optimality condition for the optimal control problem Q^V stated on page 476. We are convinced that such conditions could be established for a regularization of this problem, by using arguments similar to those used in [27]. There, a boundary optimal control problem for a frictional contact problem with normal compliance has been considered.

Another interesting continuation of the results presented in this chapter would be their extension to evolutionary variational inequalities. For such inequalities both the data and the unknown depend on time variable and, moreover, the time derivative of the unknown appears in the statement of the problem. In addition, an initial condition is needed. Such kind of inequalities model quasistatic process of contact for elastic, viscoelastic, and viscoplastic materials. The optimal control of a quasistatic model of contact with linearly elastic materials was studied in [3]. There, besides the existence of the optimal pairs, necessary optimality conditions for a regularization problem have been established.

An interesting continuation of the results presented in this chapter would be their extension to variational–hemivariational inequalities. These inequalities represent a generalization of variational inequalities, in which both convex and nonconvex functions are involved. Besides arguments of convexity and monotonicity, the theory of variational–hemivariational inequalities was built based on the properties of Clarke subdifferential, defined for locally Lipschitz function. The details can be found in the books [32, 42] and the edited volume [18]. Some preliminary results in the study of optimal control for variational–hemivariational inequalities can be found in [37].

We end this section by recalling that the control of mathematical models of contact, as well as their optimal shape design, deserves to make the object of important studies in the future. These topics are of considerable theoretical and applied interest. Indeed, in most applications this is the main interest of the design engineer and any result in this direction will illustrate the cross fertilization between models and applications, in one hand, and the nonlinear functional analysis, on the other hand. The related issues are the observability properties of the models and parameter identification. Using reliable parameter identification procedures will help in establishing the validity of various mathematical models of contact with deformable bodies. This, in turn, will help in the construction of effective and efficient numerical algorithms for the problems with established convergence. As better models for specific applications are obtained, improved mathematical models and numerical simulations will be possible.

References

1. F. Abergel, E. Casas, Some optimal control problems of multistate equations appearing in fluid mechanics. Math. Model. Numer. Anal. **27**, 223–247 (1993)
2. F. Abergel, R. Temam, On some control in fluid mechanics. Theor. Comput. Fluid Dyn. **1**, 303–325 (1990)
3. A. Amassad, D. Chenais, C. Fabre, Optimal control of an elastic contact problem involving Tresca friction law. Nonlinear Anal. **48**, 1107–1135 (2002)
4. C. Baiocchi, A. Capelo, *Variational and Quasivariational Inequalities: Applications to Free-Boundary Problems* (John Wiley, Chichester, 1984)
5. M. Barboteu, X. Cheng, M. Sofonea, Analysis of a contact problem with unilateral constraint and slip-dependent friction. Math. Mech. Solids **21**, 791–811 (2016)
6. M. Barboteu, M. Sofonea, D. Tiba, The control variational method for beams in contact with deformable obstacles. Z. Angew. Math. Mech. **92**, 25–40 (2012)
7. V. Barbu, *Optimal Control of Variational Inequalities*. Research Notes in Mathematics, vol. 100 (Pitman, Boston, 1984)
8. A. Bermudez, C. Saguez, Optimal control of a Signorini problem, SIAM J. Control Optim. **25**, 576–582 (1987)
9. J.F. Bonnans, D. Tiba, Pontryagin's principle in the control of semilinear elliptic variational inequalities. Appl. Math. Optim. **23**, 299–312 (1991)
10. H. Brézis, Equations et inéquations non linéaires dans les espaces vectoriels en dualité. Ann. Inst. Fourier (Grenoble) **18**, 115–175 (1968)
11. A. Capatina, Optimal control of Signorini problem. Numer. Funct. Anal. Optim. **21**, 817–828 (2000)
12. A. Capatina, *Variational Inequalities Frictional Contact Problems*. Advances in Mechanics and Mathematics, vol. 31 (Springer, New York, 2014)
13. A. Capatina, R. Stavre, Optimal control of a non-isothermal Navier-Stokes flow. Int. J. Engn. Sci. **34**, 59–66 (1996)
14. G. Duvaut, J.-L. Lions, *Inequalities in Mechanics and Physics* (Springer-Verlag, Berlin, 1976)
15. C. Eck, J. Jarušek, M. Krbec, *Unilateral Contact Problems: Variational Methods and Existence Theorems*. Pure and Applied Mathematics, vol. 270 (Chapman/CRC Press, New York, 2005)
16. A. Freidman, Optimal control for variational inequalities. SIAM J. Control. Optim. **24**, 439–451 (1986)
17. R. Glowinski, *Numerical Methods for Nonlinear Variational Problems* (Springer-Verlag, New York, 1984)
18. W. Han, S. Migórski, M. Sofonea (eds.), *Advances in Variational and Hemivariational Inequalities*. Advances in Mechanics and Mathematics, vol. 33 (Springer, New York, 2015)
19. W. Han, M. Sofonea, *Quasistatic Contact Problems in Viscoelasticity and Viscoplasticity*. Studies in Advanced Mathematics, vol. 30 (American Mathematical Society, Providence; International Press, Somerville, 2002)
20. I. Hlaváček, J. Haslinger, J. Necǎs, J. Lovíšek, *Solution of Variational Inequalities in Mechanics* (Springer-Verlag, New York, 1988)
21. N. Kikuchi, J.T. Oden, *Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods* (SIAM, Philadelphia, 1988)
22. D. Kinderlehrer, G. Stampacchia, *An Introduction to Variational Inequalities and their Applications*. Classics in Applied Mathematics, vol. 31 (SIAM, Philadelphia, 2000)
23. T.A. Laursen, *Computational Contact and Impact Mechanics* (Springer, Berlin, 2002)
24. J.-L. Lions, *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles* (Dunod, Paris, 1968)
25. J.-L. Lions, *Quelques méthodes de résolution des problèmes aux limites non linéaires* (Gauthiers-Villars, Paris, 1969)
26. A. Matei, S. Micu, Boundary optimal control for nonlinear antiplane problems. Nonlinear Anal. Theory Methods Appl. **74**, 1641–1652 (2011)

27. A. Matei, S. Micu, Boundary optimal control for a frictional contact problem with normal compliance. Appl. Math. Optim. **78**, 379–401 (2018)
28. A. Matei, S. Micu, C. Niţă, Optimal control for antiplane frictional contact problems involving nonlinearly elastic materials of Hencky type. Math. Mech. Solids **23**, 308–328 (2018)
29. F. Mignot, Contrôle dans les inéquations variationnelles elliptiques. J. Funct. Anal. **22**, 130–185 (1976)
30. F. Mignot, J.-P. Puel, Optimal control in some variational inequalities. SIAM J. Control Optim. **22**, 466–476 (1984)
31. S. Migórski, A note on optimal control problem for a hemivariational inequality modeling fluid flow. Dyn. Syst., 545–554 (2013)
32. S. Migórski, A. Ochal, M. Sofonea, *Nonlinear Inclusions and Hemivariational Inequalities. Models and Analysis of Contact Problems*. Advances in Mechanics and Mathematics, vol.26 (Springer, New York, 2013)
33. D. Motreanu, M. Sofonea, Quasivariational inequalities and applications in frictional contact problems with normal compliance. Adv. Math. Sci. Appl. **10**, 103–118 (2000)
34. P. Neitaanmaki, J. Sprekels, D. Tiba, *Optimization of Elliptic Systems: Theory and Applications*. Springer Monographs in Mathematics (Springer, New York, 2006)
35. P. Neitamaki, D. Tiba, *Optimal Control of Nonlinear Parabolic Systems: Theory, Algorithms and Applications*, vol. 179 (Marcel Dekker, New York, 1994)
36. P.D. Panagiotopoulos, *Inequality Problems in Mechanics and Applications* (Birkhäuser, Boston, 1985)
37. Z. Peng, K. Kunisch, Optimal control of elliptic variational-hemivariational inequalities. J. Optim. Theory Appl. **178**, 1–25 (2018)
38. M. Shillor, M. Sofonea, J.J. Telega, *Models and Analysis of Quasistatic Contact*. Lecture Notes in Physics, vol. 655 (Springer, Berlin, 2004)
39. M. Sofonea, D. Danan, C. Zheng, Primal and dual variational formulation of a frictional contact problem. Mediterr. J. Math. **13**, 857–872 (2016)
40. M. Sofonea, A. Matei, *Variational Inequalities with Applications. A Study of Antiplane Frictional Contact Problems*. Advances in Mechanics and Mathematics, vol.18 (Springer, New York, 2009)
41. M. Sofonea, A. Matei, *Mathematical Models in Contact Mechanics*. London Mathematical Society Lecture Note Series, vol.398 (Cambridge University Press, Cambridge, 2012)
42. M. Sofonea, S. Migórski, *Variational-Hemivariational Inequalities with Applications*. Pure and Applied Mathematics (Chapman & Hall/CRC Press, Boca Raton-London, 2018)
43. D. Tiba, *Lectures on the Optimal Control of Elliptic Equations*. Lecture Notes, vol. 32 (University of Jyväskylä, Jyväskylä 1995)
44. A. Touzaline, Optimal control of a frictional contact problem. Acta Math. Appl. Sin. Engl. Ser. **31**, 991–1000 (2015)

Chapter 14
On Generalized Derivative Sampling Series Expansion

Zurab A. Piranashvili and Tibor K. Pogány

Abstract Master generalized sampling series expansion is presented for entire functions (signals) coming from a class whose members satisfy an extended exponential boundedness condition. Firstly, estimates are given for the remainder of Maclaurin series of those functions and consequent derivative sampling results are obtained and discussed.

The established results are employed in evaluating the related remainder term of signals which occur in sampling series expansion of stochastic processes and random fields (not necessarily stationary or homogeneous) whose spectral kernel satisfies the relaxed exponential boundedness. The derived truncation error upper bounds enable to obtain mean-square master generalized derivative sampling series expansion formulae either for harmonizable Piranashvili-type stochastic processes or for random fields.

Finally, being the sampling series convergence rate exponential, almost sure P sampling series expansion formulae are presented.

Keywords Whittaker–Kotel'nikov–Shannon sampling theorem · Derivative sampling · Exponentially bounded signals · Entire functions · Truncation error upper bound · Harmonizable stochastic process · Piranashvili process · Karhunen process · Loève process · Weak sense stationary process · Mean-square convergence · Almost sure P convergence

Z. A. Piranashvili
Vladimir Chavchanidze Institute of Cybernetics, Georgian Technical University, Tbilisi, Georgia

T. K. Pogány (✉)
Faculty of Maritime Studies, University of Rijeka, Rijeka, Croatia

Institute of Applied Mathematics, Óbuda University, Budapest, Hungary
e-mail: poganj@pfri.hr

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_14

14.1 Introduction and Motivation

Last year was the 50th anniversary of publishing first named author's article [30] which became one of the most prolific and in the same time unavoidable cornerstone works for the mean-square and also almost sure sampling reconstruction of a general class of stochastic processes which are not necessarily of weak sense (or Khintchine sense) stationary or homogeneous. Moreover the process class considered in [30] covers the harmonizable (in the Rozanov's sense [51]) processes, Karhunen processes, Loève processes, and the weak sense stationary processes as well. To achieve these results the complex analytical background had been established, employing among others the Cauchy residue theorem, by deriving a set of different types of sampling series truncation error bounds of exponentially vanishing convergence rate of the kernel functions. These general kernels occur in the integrands of spectral representations related to the correlation functions of studied input processes. Finally, thanking to the aid of the celebrated Karhunen–Cramér theorem one concludes about the similar fashion sampling series reconstruction results in both mean-square and almost sure sense for the so-called Piranashvili processes.

Thereafter the issued 1967 article [30], following the traces installed therein the first named author gradually enlarged and generalized the kernel functions class to be considered in his works taking not only sampling but the more essential derivative sampling restoration series whose initial form was modified by a cosine hyperbolic denominator term in getting substantially improved convergence rate, thereby achieving exponential vanishing behavior. Moreover, not only scalar argument signals but vectorial ones (e.g., random fields) came into the focus of his interest, compare [31–38]. Certain aspects of his ancestor results reported in [30] were explored and developed in parallel or out of his mainstream (for instance, the results by Lee in 1970s [18–21], Higgins [6–8], Houdré in 1990s [9–13], and recently by Olenko and/or Pogány [23–29, 39–43, 45, 47, 48]). Here should be mentioned Jerri's survey article [15] and the exhaustive work by Butzer et al. [3].

In this memoir our goal is to unify, extend, and generalize the first named author's earlier efforts giving a master sampling reconstruction formula in uniform sampling setting. In all occurring situations regarding the input signal which should be restored in view of "digital-to-analog" procedure the established master theorems final specialized form is the "holy grail of the digital signal processing," the celebrated Whittaker–Kotel'nikov–Shannon (WKS) sampling reconstruction formula is

$$f(z) = \sum_{n\in\mathbb{Z}} f\left(n\frac{\pi}{\alpha}\right) \frac{\sin(\alpha z - n\pi)}{\alpha z - n\pi}, \qquad \alpha > 0 \tag{14.1}$$

either for deterministic or stochastic signal f (function, stochastic process, or random field) coming from a special class for scalar or higher dimensional argument z which belongs to certain compact in $\mathbb{C}^d$, $d \in \mathbb{N}$.

14.2 Master Sampling Theorem for Deterministic Signals

Let $f(z)$ be the entire function satisfying the condition

$$|f(z)| \le L_f \left(1 + |z|^m\right) e^{\sigma |y|}, \qquad z = x + iy \in \mathbb{C}, \tag{14.2}$$

where $m \in \mathbb{N}_0$, and $L_f, \sigma > 0$ are absolute constants. Consider the integral

$$I_n(z) = \frac{1}{2\pi\, i} \int_{\gamma_n} \frac{g(\zeta, z)}{(\zeta - z)^{p+1}}\, d\zeta\,, \tag{14.3}$$

where the integrand is of the form:

$$g(\zeta, z) = \frac{f(z)\, \mathrm{sinc}^q \beta(\zeta - z)}{(\zeta - c)^{N_0+1}(a e^{\delta\zeta} + b e^{-\delta\zeta}) \sin^{N+1}(\alpha\zeta)},$$

where

$$z \ne \zeta_k := k\, \frac{\pi}{\alpha};\ k \in \mathbb{Z}.$$

Here the familiar "sinus cardinalis" (sinc) terminology has been used:

$$\mathrm{sinc}\, w = \begin{cases} \dfrac{\sin w}{w}, & w \ne 0 \\ 1, & w = 0 \end{cases},$$

while the circular integration path

$$\gamma_n = \{\zeta : |\zeta| = (n + \tfrac{1}{2})\pi/\alpha\}, \tag{14.4}$$

for $N, N_0 + 1, p, q \in \mathbb{N}_0$; $a, b, \alpha, \beta > 0$, $\delta \in \mathbb{R}$ and a certain $c \ne 0$ is coming from the punctured complex plane $\mathbb{C} \setminus \{0\}$.

Applying Cauchy's residue theorem to $I_n(z)$, mentioning that the structure of $g(\zeta, z)$ dictates that there occur poles ζ_k, $k = \overline{-n, n}$ of the order $N + 1$, $\zeta = c$ of the order $N_0 + 1$, and $\zeta = z \ne \zeta_k$ of the order $p + 1$, we conclude

$$I_n(z) = \mathrm{Res}\left[\frac{g(\zeta, z)}{(\zeta - z)^{p+1}}; c, N_0 + 1\right] + \sum_{|k| \le n} \mathrm{Res}\left[\frac{g(\zeta, z)}{(\zeta - z)^{p+1}}; \zeta_k, N + 1\right]$$
$$+ \frac{1}{p!} \lim_{\zeta \to z} \left(\frac{d}{d\zeta}\right)^p g(\zeta, z) =: R_c + \sum_{|k| \le n} R_k + \frac{1}{p!} \lim_{\zeta \to z} \left(\frac{d}{d\zeta}\right)^p g(\zeta, z). \tag{14.5}$$

By the residue calculation via the general Leibniz chain-rule for the higher order derivatives of the product function we conclude

$$\begin{aligned} R_c &= \frac{1}{N_0!} \lim_{\zeta\to c} \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^{N_0} \frac{f(z)\,\mathrm{sinc}^q \beta(\zeta-z)}{(\zeta-z)^{p+1}\,(a\mathrm{e}^{\delta\zeta}+b\mathrm{e}^{-\delta\zeta})\sin^{N+1}(\alpha\zeta)} \\ &= \frac{1}{N_0!} \lim_{\zeta\to c} \sum_{r_1=0}^{N_0} \binom{N_0}{r_1} \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^{r_1} \frac{f(\zeta)}{(\zeta-z)^{p+1}} \\ &\quad \cdot \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^{N_0-r_1} \frac{\mathrm{sinc}^q \beta(\zeta-z)}{(a\mathrm{e}^{\delta\zeta}+b\mathrm{e}^{-\delta\zeta})\sin^{N+1}(\alpha\zeta)} \\ &=: \frac{1}{N_0!} \lim_{\zeta\to c} \sum_{r_1=0}^{N_0} \binom{N_0}{r_1} U_{r_1}(\zeta)\, V_{N_0-r_1}(\zeta)\,. \end{aligned}$$

Accordingly,

$$\lim_{\zeta\to c} U_{r_1}(\zeta) = \frac{(-1)^p}{p!\,(c-z)^{p+1}} \sum_{r_2=0}^{r_1} \binom{r_1}{r_2} \frac{(p+r_1-r_2)!\, f^{(r_2)}(c)}{(c-z)^{r_1-r_2}}\,. \tag{14.6}$$

Let us introduce the Kummer's confluent hypergeometric function [1, p. 504]

$${}_1F_1\left(\begin{matrix} a \\ b \end{matrix}\,\middle|\, z\right) = \sum_{n\ge 0} \frac{(a)_n}{(b)_n}\,\frac{z^n}{n!}\,,$$

where $(x)_m = x(x+1)\cdots(x+m-1)$ stands for the familiar Pochhammer symbol (or, in other words *shifted factorial*), taking by convention $(0)_0 \equiv 1$. Obviously, for a to be equal to some negative integer, $-\kappa$ say, this series reduces to a polynomial in z of the degree $\deg\big({}_1F_1\big) = \kappa$. So, with the aid of Kummer's function we define the hypergeometric polynomial type differential operator

$${}_1F_1\left(\begin{matrix} -\kappa \\ b \end{matrix}\,\middle|\, \mu\,\frac{\mathrm{d}}{\mathrm{d}w}\right) = \sum_{n=0}^{\kappa} \frac{(-\kappa)_n}{(b)_n}\,\frac{\mu^n}{n!}\,\left(\frac{\mathrm{d}}{\mathrm{d}w}\right)^n\,.$$

Now, putting $\mu = c-z$, $\kappa = r_1$, $b = -p-r_1$, it is a routine exercise to show that the right-hand side sum one transforms (14.6) into

$$\lim_{\zeta\to c} U_{r_1}(\zeta) = \frac{(-1)^p\,(p+1)_{r_1}}{(c-z)^{p+r_1+1}}\,{}_1F_1\left(\begin{matrix} -r_1 \\ -p-r_1 \end{matrix}\,\middle|\,(c-z)\,\frac{\mathrm{d}}{\mathrm{d}w}\right) \circ f(w)\Big|_{w=c}\,;$$

where $h_1 \circ h_2(x)$ stands for the composite function $h_1(h_2(x))$. Denote

$$\lim_{\zeta\to c} V_s(\zeta) =: A_s(z;\theta_{N_0})\,,$$

where the parameter vector $\theta_M = (M, q, \alpha, \beta, \delta, a, b, c)$. Therefore,

$$R_c = \frac{(-1)^p}{N_0!} \sum_{r_1=0}^{N_0} \binom{N_0}{r_1} \frac{(p+1)_{r_1}\ A_{N_0-r_1}(z;\theta_{N_0})}{(c-z)^{p+r_1+1}} \cdot {}_1F_1\left(\begin{matrix} -r_1 \\ -p-r_1 \end{matrix} \middle| (c-z)\,\frac{\mathrm{d}}{\mathrm{d}w} \right) \circ f(w)\Big|_{w=c} . \tag{14.7}$$

Next, applying again the Leibniz rule, we can write

$$R_k = \frac{1}{N!} \sum_{j=0}^{N} \binom{N}{j} \lim_{\zeta\to\zeta_k} \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^j \frac{f(\zeta)(\zeta-\zeta_k)^{N+1}}{(\zeta-c)^{N_0+1}(\zeta-z)^{p+1}\sin^{N+1}(\alpha\zeta)} \cdot \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^{N-j} \frac{\mathrm{sinc}^q \beta(\zeta-z)}{a\mathrm{e}^{\delta\zeta}+b\mathrm{e}^{-\delta\zeta}} . \tag{14.8}$$

The first derivative can be solved by another use of the Leibniz rule, separating the indeterminate form, which results in

$$\left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^j \frac{f(\zeta)}{(\zeta-c)^{N_0+1}(\zeta-z)^{p+1}} \left(\frac{\zeta-\zeta_k}{\sin(\alpha\zeta)}\right)^{N+1} = \sum_{m_1=0}^{j} \binom{j}{m_1} \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^{m_1} \times \frac{f(\zeta)}{(\zeta-c)^{N_0+1}(\zeta-z)^{p+1}} \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^{j-m_1} \left(\frac{\zeta-\zeta_k}{\sin(\alpha\zeta)}\right)^{N+1} =: \sum_{m_1=0}^{j} \binom{j}{m_1} S_{m_1}(\zeta;z)\,T_{j-m_1}(\zeta) .$$

Now, we have

$$\begin{aligned}\lim_{\zeta\to\zeta_k} S_{m_1}(\zeta;z) &= \sum_{m_2=0}^{m_1} \binom{m_1}{m_2} f^{(m_2)}(\zeta_k) \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^{m_1-m_2} \frac{1}{(\zeta-c)^{N_0+1}(\zeta-z)^{p+1}}\Big|_{\zeta=\zeta_k} \\ &= \frac{1}{p!\,N_0!\,(\zeta_k-c)^{N_0+1}(\zeta_k-z)^{p+1}} \sum_{m_2=0}^{m_1} \sum_{m_3=0}^{m_1-m_2} \binom{m_1}{m_2}\binom{m_1-m_2}{m_3} \\ &\quad \cdot \frac{(-1)^{m_1-m_2} f^{(m_2)}(\zeta_k)\,(p+m_3)!\,(N_0+m_1-m_2-m_3)!}{(\zeta_k-c)^{m_1-m_2-m_3}(\zeta-z)^{m_3}}\end{aligned}$$

$$= \frac{1}{N_0!\,(\zeta_k - c)^{N_0+1}(\zeta_k - z)^{p+1}} \sum_{m_2=0}^{m_1} \binom{m_1}{m_2} \frac{f^{(m_2)}(\zeta_k)}{(c - \zeta_k)^{m_1-m_2}}$$

$$\cdot\, (N_0 + m_1 - m_2)!\; {}_2F_1\left(\begin{matrix} m_2 - m_1,\; p+1 \\ m_2 - m_1 - N_0 \end{matrix} \,\Big|\, \frac{\zeta_k - c}{\zeta_k - z}\right), \qquad (14.9)$$

bearing in mind that for all $p, P, M \in \mathbb{N}_0;\; P \geq M$

$$\sum_{s=0}^{M} \binom{M}{s} (p+s)!\,(P-s)!\, w^s = p!\, P!\; {}_2F_1\left(\begin{matrix} -M,\; p+1 \\ -P \end{matrix} \,\Big|\, w\right), \qquad (14.10)$$

where

$${}_2F_1\left(\begin{matrix} a,\; b \\ c \end{matrix} \,\Big|\, z\right) = \sum_{n\geq 0} \frac{(a)_n\,(b)_n}{(c)_n} \frac{z^n}{n!}$$

denotes the familiar Gaussian hypergeometric function (the case $P < M$ is redundant). Moreover, the right-hand side expression in the display (14.10) turns out to be a polynomial of $\deg\big({}_2F_1\big) = M$. Indeed, we have

$$\begin{aligned}\sum_{s=0}^{M} \binom{M}{s} \frac{(p+s)!\,(P-s)!}{p!\,P!} w^s &= \sum_{s=0}^{M} \frac{(-1)^s (-M)_s\, \Gamma(p+s+1)\,\Gamma(P-s+1)}{\Gamma(p+1)\,\Gamma(P+1)} \frac{w^s}{s!} \\ &= \sum_{s=0}^{M} (-1)^s (-M)_s\, (p+1)_s\, (P+1)_{-s} \frac{w^s}{s!} \\ &= \sum_{s=0}^{M} \frac{(-M)_s\,(p+1)_s}{(-P)_s} \frac{w^s}{s!}\,,\end{aligned}$$

which confirms (14.10) by virtue of the Pochhammer symbol transformation formula

$$(a)_n\,(1-a)_{-n} = (-1)^n, \qquad n \in \mathbb{Z}\,.$$

Unfortunately, further closed form summation in (14.9) cannot be inferred due to the unknown input function f.

Writing $x := \alpha(\zeta - \zeta_k)$ we conclude

$$\lim_{\zeta\to\zeta_k} T_r(\zeta) = \frac{(-1)^{k(N+1)}}{\alpha^{N-r+1}} \lim_{x\to 0} \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^r \left(\frac{x}{\sin x}\right)^{N+1} =: \frac{(-1)^{k(N+1)}}{\alpha^{N-r+1}}\, B_r(N)\,.$$

Denote the latter derivative term in (14.8)

$$C_{j,k}(z;\theta_N') := \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^{N-j} \frac{\operatorname{sinc}^q \beta(\zeta - z)}{a\mathrm{e}^{\delta\zeta} + b\mathrm{e}^{-\delta\zeta}}\bigg|_{\zeta=\zeta_k},$$

where $\theta_M' = \theta_M\big|_{c=0} = (M, q, \alpha, \beta, \delta, a, b, 0)$. Collecting the established formulae we can write:

$$R_k = \frac{(-1)^{k(N+1)}\,(\zeta - c)^{-N_0-1}}{N!\,N_0!\,\alpha^{N+1}\,(\zeta - z)^{p+1}} \sum_{j=0}^{N} \binom{N}{j} C_{j,k}(z;\theta_N') \sum_{m_1=0}^{j} \binom{j}{m_1} B_{j-m_1}(N)$$
$$\sum_{m_2=0}^{m_1} \frac{\binom{m_1}{m_2}(N_0 + m_1 - m_2)!\,f^{(m_2)}(\zeta_k)}{(c - \zeta_k)^{m_1-m_2}}\; {}_2F_1\left(\begin{matrix} m_2 - m_1,\, p+1 \\ m_2 - m_1 - N_0 \end{matrix}\,\Big|\, \frac{\zeta_k - c}{\zeta_k - z}\right). \tag{14.11}$$

Having in mind (14.3), (14.5), (14.7), and (14.11) we conclude

$$I_n(z) = \frac{(-1)^p}{N_0!} \sum_{r_1=0}^{N_0} \binom{N_0}{r_1} \frac{(p+1)_{r_1} A_{N_0-r_1}(z;\theta_{N_0})}{(c-z)^{p+r_1+1}}$$
$$\cdot\, {}_1F_1\left(\begin{matrix} -r_1 \\ -p - r_1 \end{matrix}\,\Big|\,(c-z)\,\frac{\mathrm{d}}{\mathrm{d}w}\right) \circ f(w)\Big|_{w=c}$$
$$+ \frac{\alpha^{-N-1}}{N!\,N_0!} \sum_{|k|\le n} \sum_{j=0}^{N} \sum_{m_1=0}^{j} \frac{(-1)^{k(N+1)}\binom{N}{j} C_{j,k}(z;\theta_N') \binom{j}{m_1} B_{j-m_1}(N)}{(\zeta_k - c)^{N_0+1}(\zeta_k - z)^{p+1}}$$
$$\sum_{m_2=0}^{m_1} \binom{m_1}{m_2} \frac{f^{(m_2)}(\zeta_k)}{(c - \zeta_k)^{m_1-m_2}} (N_0 + m_1 - m_2)!$$
$$\cdot\, {}_2F_1\left(\begin{matrix} m_2 - m_1,\, p+1 \\ m_2 - m_1 - N_0 \end{matrix}\,\Big|\, \frac{\zeta_k - c}{\zeta_k - z}\right) + \frac{1}{p!} \lim_{\zeta\to z} \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^p g(\zeta, z),$$

where the meanings of the coefficients A, B, C remain the same as above.

Our following step is to evaluate the modulus of the integral $I_n(z)$ by substituting $\zeta = r_n \mathrm{e}^{\mathrm{i}\varphi}$, where $r_n = (n + \frac{1}{2})\pi/\alpha$. Also, we apply (1) the estimate $|\sin z| \le \mathrm{e}^{|\Im(z)|}$, $z \in \mathbb{C}$, (2) the constraint (14.2) concerning the behavior of the input signal f, and (3) the lower bound

$$\left|\sin\left(\alpha r_n \mathrm{e}^{\mathrm{i}\varphi}\right)\right| \ge \frac{1}{2}\left(1 - \mathrm{e}^{-\pi}\right) \mathrm{e}^{\alpha r_n |\sin\varphi|}\,.$$

These three tools imply

$$|I_n(z)| \le \frac{2^N L_f\, \mathrm{e}^{\beta\, q|y|}\big(1+r_n^m\big)\, r_n^{-(p+q+N_0+1)}}{\pi\beta^q\big(1-\mathrm{e}^{-\pi}\big)^{N+1}\left|1-\dfrac{|z|}{r_n}\right|^{p+q+1}\left|1-\dfrac{|c|}{r_n}\right|^{N_0+1}} \cdot \int_0^{2\pi} \frac{\mathrm{e}^{-[(N+1)\alpha-\sigma-q\beta]r_n\,|\sin\varphi|}\,\mathrm{d}\varphi}{F_{a,b,\delta}(\varphi)}, \tag{14.12}$$

where the shorthand

$$F_{a,b,\delta}(\varphi) := \mathrm{e}^{\delta\, r_n\, \cos\varphi}\big(a^2 + b^2\mathrm{e}^{-4\delta\, r_n\, \cos\varphi} + 2ab\cos(2\delta\, r_n\, \sin\varphi)\, \mathrm{e}^{-2\delta\, r_n\, \cos\varphi}\big)^{\frac{1}{2}}$$

is used, $\varphi \in [0, 2\pi]$. The integrand in (14.12) is symmetric in φ with respect to π so

$$|I_n(z)| \le \frac{2^{N+1} L_f\, \mathrm{e}^{\beta q|y|}\big(1+r_n^m\big)\, r_n^{-(p+q+N_0+1)}}{\pi\beta^q\big(1-\mathrm{e}^{-\pi}\big)^{N+1}\left|1-\dfrac{|z|}{r_n}\right|^{p+q+1}\left|1-\dfrac{|c|}{r_n}\right|^{N_0+1}} \cdot \int_0^{\pi} \frac{\mathrm{e}^{-[(N+1)\alpha-\sigma-q\beta]r_n\,|\sin\varphi|}\,\mathrm{d}\varphi}{F_{a,b,\delta}(\varphi)}. \tag{14.13}$$

Because $F_{a,b,\delta}(\pi-\varphi) = F_{b,a,\delta}(\varphi),\ \varphi \in [0, \frac{\pi}{2}]$, we halve the integration domain. Pointing out that

$$\begin{aligned}\min_{0\le\varphi\le\frac{\pi}{2}} \mathrm{e}^{-\delta\, r_n\, \cos\varphi} F_{a,b,\delta}(\varphi) &= F_{a,b,\delta}\left(\frac{\pi}{2}\right)\\ &= \min\left\{a + b\mathrm{e}^{-2\delta\, r_n}, \sqrt{a^2+b^2+2ab\cos(2\delta\, r_n)}\right\}\\ &\ge \min\{a, b, |a-b|\} =: D_0(a,b)\,,\end{aligned} \tag{14.14}$$

making use of the Jordan's sine inequality [22, p. 33]

$$\sin\varphi \ge \frac{2}{\pi}\,\varphi, \qquad \varphi \in [0, \tfrac{\pi}{2}]$$

and its straightforward counterpart

$$\cos\varphi \ge 1 - \frac{2}{\pi}\,\varphi, \qquad \varphi \in [0, \tfrac{\pi}{2}],$$

from (14.13) via (14.14) we derive the estimate

$$|I_n(z)| \le \frac{2^{N+2} L_f\, \mathrm{e}^{\beta q|y|-\delta r_n}\big(1+r_n^m\big)\, r_n^{-(p+q+N_0+2)}\, \beta^{-q}}{D_0(a,b)\big(1-\mathrm{e}^{-\pi}\big)^{N+1}\left|1-\dfrac{|z|}{r_n}\right|^{p+q+1}\left|1-\dfrac{|c|}{r_n}\right|^{N_0+1}} \cdot \int_0^{\frac{\pi}{2}} \mathrm{e}^{-[(N+1)\alpha-\delta-\sigma-q\beta]r_n\,\frac{2\varphi}{\pi}}\,\mathrm{d}\varphi$$

$$= \frac{\pi\, 2^{N+1} L_f\, \mathrm{e}^{\beta q|y|-\delta r_n}\big(1+r_n^m\big)\, r_n^{-(p+q+N_0+2)}\, \beta^{-q}}{D_0(a,b)\big(1-\mathrm{e}^{-\pi}\big)^{N+1}\left|1-\dfrac{|z|}{r_n}\right|^{p+q+1}\left|1-\dfrac{|c|}{r_n}\right|^{N_0+1}} \cdot \frac{1-\mathrm{e}^{-[(N+1)\alpha-\sigma-\delta-q\beta]r_n}}{(N+1)\alpha-\sigma-\delta-q\beta}\,.$$

For any fixed z, c, and enough large n there hold the bounds

$$2\,\min\{|1-|z|/r_n|\,,\,|1-|c|/r_n|\} \ge 1; \qquad 1-\mathrm{e}^{-[(N+1)\alpha-\sigma-\delta-q\beta]r_n} \le 1.$$

Hence, for any $(N+1)\alpha-\delta-\sigma-q\beta>0$ we conclude

$$|I_n(z)| \le \pi\, L_f\, \frac{2^{p+q+N+N_0+2}}{(1-\mathrm{e}^{-\pi})^{N+1}}\, \frac{\mathrm{e}^{\beta q|y|-\delta\, r_n}\big(1+r_n^m\big)\, r_n^{-(p+q+N_0+2)}}{D_0(a,b)\,\beta^q\,[(N+1)\alpha-\sigma-\delta-q\beta]}\,; \tag{14.15}$$

quote that $\mathrm{e}^{\beta q|y|}$ remains finite for any bounded $z\in\mathbb{C}$.

Signify

$$\mathscr{T}_n^{(p)}(f;z) := \lim_{\zeta\to z}\left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^p \frac{f(\zeta)\,\mathrm{sinc}^q\beta(\zeta-z)}{(\zeta-c)^{N_0+1}(a\mathrm{e}^{\delta\zeta}+b\mathrm{e}^{-\delta\zeta})\,\sin^{N+1}(\alpha\zeta)}$$
$$-\frac{p!}{N!\,N_0!\,\alpha^{N+1}}$$
$$\times\sum_{|k|\le n}\sum_{j=0}^{N}\sum_{m_1=0}^{j}\frac{(-1)^{k(N+1)}\binom{N}{j}\binom{j}{m_1}C_{j,k}(z;\theta'_N)B_{j-m_1}(N)}{(\zeta_k-c)^{N_0+1}(\zeta_k-z)^{p+1}}$$
$$\times\sum_{m_2=0}^{m_1}\binom{m_1}{m_2}\frac{(N_0+m_1-m_2)!\,f^{(m_2)}(\zeta_k)}{(c-\zeta_k)^{m_1-m_2}}$$
$$\times\,{}_2F_1\left(\left.\begin{matrix}m_2-m_1,\,p+1\\ m_2-m_1-N_0\end{matrix}\right|\frac{\zeta_k-c}{\zeta_k-z}\right)$$

$$- \frac{(-1)^p\ p!}{N_0!} \sum_{r_1=0}^{N_0} \frac{\binom{N_0}{r_1}(p+1)_{r_1} A_{N_0-r_1}(z;\theta_{N_0})}{(c-z)^{p+r_1+1}}$$

$$\times\ {}_1F_1\left(\begin{matrix}-r_1\\ -p-r_1\end{matrix}\Big| (c-z)\frac{\mathrm{d}}{\mathrm{d}c}\right) \circ f(c) \tag{14.16}$$

the truncation error of the generalized pth order derivative sampling expansion of the size $2n+1$ of a suitable input function (signal) f with respect to the circular integration path γ_n defined by (14.4).

Here and in what follows, we write by convention

$${}_1F_1\left(\begin{matrix}-r_1\\ -p-r_1\end{matrix}\Big| (c-z)\frac{\mathrm{d}}{\mathrm{d}c}\right) \circ f(c) := {}_1F_1\left(\begin{matrix}-r_1\\ -p-r_1\end{matrix}\Big| (c-z)\frac{\mathrm{d}}{\mathrm{d}w}\right) \circ f(w)\Big|_{w=c}.$$

Thus, we deduce the following truncation error bound result.

Theorem 1 *Let $f(z)$ be entire satisfying (14.2) for a non-negative integer m. Then for all $N, N_0, p, q \in \mathbb{N}_0$, $a, b, \alpha, \beta > 0$, $\delta \in \mathbb{R}$ for which $(N+1)\alpha - \delta - \sigma - q\beta > 0$, and $c \in \mathbb{C} \setminus \{0\}$, we have*

$$\left|\mathscr{T}_n^{(p)}(f;z)\right| \le \pi\, L_f \left(\frac{2}{1-\mathrm{e}^{-\pi}}\right)^{N+1} \frac{p!\, 2^{p+q+N_0+1}\, \mathrm{e}^{\beta q|y|}\, \mathrm{e}^{-\delta\left(n+\frac12\right)\frac{\pi}{\alpha}}}{\beta^q\, D_0(a,b)\,[(N+1)\alpha - \sigma - \delta - q\beta]}$$

$$\cdot \left(\frac{\alpha}{\pi\left(n+\frac12\right)}\right)^{p+q+N_0-m+2} \left\{1 + \left(\frac{\alpha}{\pi\left(n+\frac12\right)}\right)^m\right\}, \tag{14.17}$$

provided n is enough large positive integer, $z \ne c$ is coming from a bounded subregion of the complex plane, and $z \ne \zeta_k = k\frac{\pi}{\alpha}$, $k = \overline{-n,n}$. The truncation error $\mathscr{T}_n^{(p)}(f;z)$ contains

$$A_s(z;\theta_{N_0}) = \lim_{\zeta\to c}\left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^s \frac{\operatorname{sinc}^q \beta(\zeta - z)}{(a\mathrm{e}^{\delta\zeta} + b\mathrm{e}^{-\delta\zeta})\sin^{N+1}(\alpha\zeta)}, \qquad s \in \mathbb{N}_0$$

$$B_r(N) = \lim_{x\to 0}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^r \left(\frac{x}{\sin x}\right)^{N+1}, \qquad r \in \mathbb{N}_0$$

$$C_{j,k}(z;\theta'_N) = \lim_{\zeta\to\zeta_k}\left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^{N-j} \frac{\operatorname{sinc}^q \beta(\zeta - z)}{a\mathrm{e}^{\delta\zeta} + b\mathrm{e}^{-\delta\zeta}},$$

while $\theta_M = (M, q, \alpha, \beta, \delta, a, b, c)$ *and* $\theta'_M = (M, q, \alpha, \beta, \delta, a, b, 0)$, M *non-negative integer, while*

$$D_0(a, b) = \min\{a, b, |a - b|\}\,.$$

Remark 1 The truncation error upper bound (14.17) is useful for all positive values of β, but it is senseless in the case $\beta \to 0$. Therefore we have changed the estimation procedure in evaluating the integrand by the bound $|\sin z| \le \mathrm{e}^{|y|}$ to achieve (14.12), which finally lead to the upper bound (14.15). ■

Next, making use of the less sensitive bound $|\mathrm{sinc}\, z| \le \mathrm{e}^{|\Im(z)|}$ in estimating the integrand of $I_n(z)$ we avoid the term β^{-q} in (14.17). The resulting upper bound is applicable for all $\beta \ge 0$.

Theorem 2 *Let the parameter space be the same as in Theorem 1 and* $\beta \ge 0$. *Then we have the truncation error upper bound*

$$\left|\mathscr{T}_n^{(p)}(f; z)\right| \le \pi\, L_f \left(\frac{2}{1 - \mathrm{e}^{-\pi}}\right)^{N+1} \frac{p!\, 2^{p+N_0+1}\, \mathrm{e}^{\beta q |y|}}{D_0(a, b)\,[(N+1)\alpha - \sigma - \delta - q\beta]}\, \mathrm{e}^{-\delta\left(n+\frac{1}{2}\right)\frac{\pi}{\alpha}} \cdot \left(\frac{\alpha}{\pi\left(n + \frac{1}{2}\right)}\right)^{p+N_0-m+2} \left\{1 + \left(\frac{\alpha}{\pi\left(n+\frac{1}{2}\right)}\right)^m\right\}, \tag{14.18}$$

which holds for all bounded $z \in \mathbb{C}$.

By letting $n \to \infty$ in both truncation error upper bounds (14.17) and (14.18) for certain fixed complex z satisfying the constraints of Theorem 1 with the weaker $\beta \ge 0$, we arrive at the master generalized derivative sampling expansion formula.

Theorem 3 *Let* $f(z)$ *be entire function satisfying* (14.2) *for certain non-negative integer* m. *Then for all*

$$\begin{aligned} \alpha(N+1) - \sigma &> 0, \\ [(N+1)\alpha - \sigma]q^{-1} &> \beta \ge 0, \\ (N+1)\alpha - \sigma - q\beta &> \delta > 0, \end{aligned}$$

the following representation holds true pointwise in any bounded z*-region of* $\mathbb{C}$ *which satisfies the conditions of Theorem 1*

$$\lim_{\zeta \to z} \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^p \frac{f(\zeta)\, \mathrm{sinc}^q \beta(\zeta - z)}{(\zeta - c)^{N_0+1}(a\mathrm{e}^{\delta\zeta} + b\mathrm{e}^{-\delta\zeta})\sin^{N+1}(\alpha\zeta)} = \frac{p!\,(-1)^{N_0+p}}{N!\, N_0!\, \alpha^{N-N_0-p-1}} \sum_{k\in\mathbb{Z}} \sum_{j=0}^{N} \sum_{m_1=0}^{j} \frac{(-1)^{k(N+1)}\binom{N}{j} C_{j,k}(z; \theta'_N)\binom{j}{m_1} B_{j-m_1}(N)}{(\alpha c - k\pi)^{N_0+1}\,(\alpha z - k\pi)^{p+1}}$$

$$\sum_{m_2=0}^{m_1} \binom{m_1}{m_2} \frac{(N_0+m_1-m_2)!\, f^{(m_2)}\left(k\frac{\pi}{\alpha}\right)}{\left(c-k\frac{\pi}{\alpha}\right)^{m_1-m_2}} \; {}_2F_1\left(\begin{matrix} m_2-m_1,\, p+1 \\ m_2-m_1-N_0 \end{matrix} \,\Big|\, \frac{\alpha c-k\pi}{\alpha z-k\pi}\right)$$

$$+ \frac{(-1)^p\; p!}{N_0!} \sum_{r_1=0}^{N_0} \frac{\binom{N_0}{r_1}(p+1)_{r_1} A_{N_0-r_1}(z;\theta_{N_0})}{(c-z)^{p+r_1+1}} \; {}_1F_1\left(\begin{matrix} -r_1 \\ -p-r_1 \end{matrix} \,\Big|\, (c-z)\frac{\mathrm{d}}{\mathrm{d}c}\right) \circ f(c) \tag{14.19}$$

uniformly for all $z \neq \zeta_n = n\frac{\pi}{\alpha}$, $n \in \mathbb{Z}$.

14.3 Discussion of Certain Special Cases

The detailed presentation and evolution of certain corollaries of Theorem 3 (14.19) we illustrate by few groups of results.

A. We begin with the case $N_0 = -1$. Then the function $g(\zeta, z)$ is without poles; therefore, the residue $R_c = 0$, accordingly the second sum in (14.19) vanishes.

The resulting formula reads

$$\lim_{\zeta\to z} \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^p \frac{f(\zeta)\,\mathrm{sinc}^q \beta(\zeta-z)}{(a\mathrm{e}^{\delta\zeta}+b\mathrm{e}^{-\delta\zeta})\sin^{N+1}(\alpha\zeta)}$$

$$= \frac{p!\,(-1)^{p+1}}{N!\,\alpha^{N-p}} \sum_{k\in\mathbb{Z}} \frac{(-1)^{k(N+1)}}{(\alpha z-k\pi)^{p+1}} \sum_{j=0}^{N} \frac{C_{j,k}(z;\theta'_N)}{(N-j)!} \sum_{m_1=0}^{j} \frac{B_{j-m_1}(N)}{(j-m_1)!}$$

$$\sum_{m_2=0}^{m_1} \frac{f^{(m_2)}\left(k\frac{\pi}{\alpha}\right)}{m_2!} \; {}_2F_1\left(\begin{matrix} m_2-m_1,\, p+1 \\ m_2-m_1+1 \end{matrix} \,\Big|\, \frac{\alpha}{\alpha z-k\pi}\right), \tag{14.20}$$

which coincides with the formula [34, Eq. (1)] and with [35, Eq. (2)] in the limiting case when $n \to \infty$.

Next, specifying $\delta = 0$ in (14.20), having in mind that $B_r(N)$ does not depend on δ and

$$C_{j,k}(z;\theta''_N)\big|_{\delta=0} = \frac{1}{a+b} \lim_{\zeta\to\zeta_k} \left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^{N-j} \mathrm{sinc}^q \beta(\zeta-z),$$

where $\theta''_N = \theta'_N\big|_{\delta=0} = (N, q, \alpha, \beta, 0, a, b, 0)$, we conclude

$$\begin{aligned}\lim_{\zeta\to z}\left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^p \frac{f(\zeta)\,\mathrm{sinc}^q\beta(\zeta-z)}{\sin^{N+1}(\alpha\zeta)} &= \frac{p!\,(-1)^{p+1}(a+b)}{N!\,\alpha^{N-p}}\sum_{k\in\mathbb{Z}}\frac{(-1)^{k(N+1)}}{(\alpha z-k\pi)^{p+1}}\sum_{j=0}^{N}\frac{C_{j,k}(z;\theta''_N)}{(N-j)!}\sum_{m_1=0}^{j}\frac{B_{j-m_1}(N)}{(j-m_1)!}\\ &\quad \sum_{m_2=0}^{m_1}\frac{f^{(m_2)}\left(k\frac{\pi}{\alpha}\right)}{m_2!}\,{}_2F_1\left(\begin{matrix}m_2-m_1,\,p+1\\ m_2-m_1+1\end{matrix}\,\Big|\,\frac{\alpha}{\alpha z-k\pi}\right),\end{aligned} \tag{14.21}$$

provided $p+q+1-m>0$. The condition for which holds true the general truncation error upper bound result (14.17) in Theorem 1 takes the reduced form being $N_0=-1, \delta=0$. The resulting sampling series expansion is in fact [33, Eq. (1)].

The associated special case $N=0$ implies that all summation indices $j=m_1=m_2=0$, that is,

$$B_0(0)=1 \qquad \text{and} \qquad C_{00}(z;\theta''_N)=\frac{1}{a+b}\,\mathrm{sinc}^q\beta\left(z-\frac{k\pi}{\alpha}\right),$$

and since the hypergeometric term becomes unity, we have

$$\begin{aligned}\lim_{\zeta\to z}\left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^p \frac{f(\zeta)\,\mathrm{sinc}^q\beta(\zeta-z)}{\sin(\alpha\zeta)} &= p!\,\alpha^p\sum_{k\in\mathbb{Z}}\frac{(-1)^{k+p+1}f\left(k\frac{\pi}{\alpha}\right)}{(\alpha z-k\pi)^{p+1}}\\ &\quad\times\mathrm{sinc}^q\beta\left(z-\frac{k\pi}{\alpha}\right).\end{aligned} \tag{14.22}$$

From the last expansion letting $q=0$, for all $\alpha>\sigma$ and $p\le m$ we deduce the derivative sampling series expansion

$$\lim_{\zeta\to z}\left(\frac{\mathrm{d}}{\mathrm{d}\zeta}\right)^p \frac{f(\zeta)}{\sin(\alpha\zeta)} = p!\,\alpha^p\sum_{k\in\mathbb{Z}}\frac{(-1)^{k+p+1}f\left(k\frac{\pi}{\alpha}\right)}{(\alpha z-k\pi)^{p+1}}. \tag{14.23}$$

Finally, $f\in B^2_\sigma$ indicate[1] $m=0$ and $p\in\mathbb{N}_0$. For these kind of input functions f for $p=0$ (14.23) becomes the classical WKS formula. We notice that for $p=1$ this formula appears in [14, pp. 115–120].

[1]The Bernstein class B^p_σ consists of entire functions (in the complex plane) of exponential type at most σ, whose restriction to $\mathbb{R}$ belongs to $L^p(\mathbb{R})$. We are interested here in B^2_σ-functions since our study belongs to the L^2-correlation theory area.

Consider now (14.22) for $p = 0$. Rewriting $(-1)^k \sin(\alpha z) = \sin(\alpha z - k\pi)$, we obviously recover the extended WKS interpolation formula [30, Eq. (19)]

$$f(z) = \sum_{k\in\mathbb{Z}} f\left(k\frac{\pi}{\alpha}\right) \operatorname{sinc}(\alpha z - k\pi) \operatorname{sinc}^q \left(\beta z - \frac{\beta k\pi}{\alpha}\right), \quad \frac{\alpha - \sigma}{\beta} > q \geq m. \tag{14.24}$$

This extension significantly improves the convergence rate of the WKS sampling series and recovers the classical WKS sampling theorem (14.1) from (14.24) when $q = 0$:

$$f(z) = \sum_{k\in\mathbb{Z}} f\left(k\frac{\pi}{\alpha}\right) \operatorname{sinc}(\alpha z - k\pi), \qquad \alpha > \sigma.$$

B. Different kinds of parameter specifications in (14.21) lead to another class of sampling series. Firstly, letting $p = 0$ we obtain by reusing the hypergeometric polynomial's property (14.10) the relation

$$f(z) = \frac{(a+b)\sin^{N+1}(\alpha z)}{N!\,\alpha^N} \sum_{k\in\mathbb{Z}} \sum_{j=0}^{N} \sum_{m_1=0}^{j} \frac{(-1)^{k(N+1)+1}}{\alpha z - k\pi} \frac{C_{j,k}(z;\theta_N'')}{(N-j)!} \frac{B_{j-m_1}(N)}{(j-m_1)!}$$
$$\cdot \sum_{m_2=0}^{m_1} \frac{f^{(m_2)}\left(k\frac{\pi}{\alpha}\right)}{m_2!(m_1-m_2-1)!} \sum_{s=0}^{m_1-m_2} \binom{m_1-m_2}{s} \frac{s!(m_1-m_2-1-s)!}{(z-k\pi/\alpha)^s}.$$

Now, for $N = 1$ this expansion one reduces to the first order derivative sampling series [31, Eq. (5)]:

$$f(z) = \sum_{k\in\mathbb{Z}} \Bigg\{\bigg[(1+q)\operatorname{sinc}\beta\Big(z - \frac{k\pi}{\alpha}\Big) - q\cos\beta\Big(z - \frac{k\pi}{\alpha}\Big)\bigg] f\left(\frac{k\pi}{\alpha}\right)$$
$$+ \Big(z - \frac{k\pi}{\alpha}\Big)\operatorname{sinc}\beta\Big(z - \frac{k\pi}{\alpha}\Big) f'\left(\frac{k\pi}{\alpha}\right)\Bigg\}$$
$$\cdot \operatorname{sinc}^2(\alpha z - k\pi)\operatorname{sinc}^{q-1}\beta\left(z - \frac{k\pi}{\alpha}\right),$$

where

$$\alpha > \frac{\sigma}{2}, \qquad 0 < \beta < \frac{2\alpha - \sigma}{q}.$$

Next, putting $q = 0$, this formula takes the form

$$f(z) = \sum_{k \in \mathbb{Z}} \left[f\left(\frac{k\pi}{\alpha}\right) + \left(z - \frac{k\pi}{\alpha}\right) f'\left(\frac{k\pi}{\alpha}\right) \right] \operatorname{sinc}^2(\alpha z - k\pi),$$

under assumption $2\alpha > \sigma$.

14.4 Brief Invitation to Piranashvili Processes

In this section we give some key results concerning harmonizable processes and their spectral representations with respect to bimeasures; for harmonizability, we refer to [17, 50, 52, 53]. Let $\{\xi(t), t \in \mathbb{R}\}$ be a centered (which means $\mathbb{E}\xi(t) = 0$) finite second order random process defined on certain standard fixed probability space $(\Omega, \mathfrak{F}, \mathsf{P})$. The process $\xi(t)$ has a covariance function (associated with a domain $\Lambda \subseteq \mathbb{R}$ with the sigma-algebra $\sigma(\Lambda)$) in the form:

$$B(t, s) = \int_\Lambda \int_\Lambda f(t, \lambda) f^*(s, \mu) F_\xi(\mathrm{d}\lambda, \mathrm{d}\mu), \tag{14.25}$$

where, for (usually taken) each $\lambda \in \Lambda$, $f(\cdot, \lambda)$ can be extended to the complex plane as a complex analytic exponentially bounded kernel function, that is, for some $M > 0$, $\alpha \in \mathbb{R}$

$$|f(t, \lambda)| \le M\mathrm{e}^{\alpha|t|},$$

while $F_\xi(\cdot, \cdot)$ is a positive definite measure on Λ^2. The total variation $\|F_\xi\|(\Lambda, \Lambda)$ of the spectral distribution function F_ξ satisfies

$$\|F_\xi\|(\Lambda, \Lambda) = \int_\Lambda \int_\Lambda \left|F_\xi(\mathrm{d}\lambda, \mathrm{d}\mu)\right| = V_{F_\xi} < \infty;$$

the constant V_{F_ξ} is also called the *Vitali variation* [53, p. 153]. Notice that the sample function $\xi(t) \equiv \xi(t, \omega_0)$ and $f(t, \lambda)$ possess the same exponential types [2, Theorem 4] and [30, Theorem 3]. Then, by the Karhunen–Cramér theorem the process $\xi(t)$ has the spectral representation

$$\xi(t) = \int_\Lambda f(t, \lambda) Z_\xi(\mathrm{d}\lambda), \tag{14.26}$$

where $Z_\xi(\cdot)$ is the spectral process (associated with ξ) which is a spectral measure and

$$\mathbb{E} Z_\xi(S) Z_\xi^*(S') = F_\xi(S, S'), \qquad S, S' \in \sigma(\Lambda).$$

In turn, the first named author's approach was more general in [30] in which he considered instead of exponentially bounded kernel function $f(t,\lambda)$, appearing in the spectral representation (14.26), the kernel function satisfying similarly to (14.2) the condition

$$|f(t,\lambda)| \le L_f(\lambda)\,\left(1+|t|^m\right)\,\mathrm{e}^{c(\lambda)\,|\Im(t)|}, \qquad t\in\mathbb{C}; \quad m\in\mathbb{N}_0\,, \tag{14.27}$$

where

$$\sup_\Lambda L_f(\lambda) < \infty; \qquad \sup_\Lambda c(\lambda) = \sigma < \infty\,.$$

The background for this kind extension is the following. Being $f(t,\lambda)$ entire, it possesses the Maclaurin expansion

$$f(t,\lambda) = \sum_{n=0}^\infty \frac{f^{(n)}(0,\lambda)}{n!}\,t^n\,.$$

As

$$\sigma = \sup_\Lambda c(\lambda) = \sup_\Lambda \limsup_n \sqrt[n]{|f^{(n)}(0,\lambda)|} < \infty\,,$$

the exponential type of $f(t,\lambda)$ is equal to σ. The process possessing this type of kernel has been called *Piranashvili process* [44]. Consequently, for all $w>\sigma$ there holds

$$\xi(t) = \sum_{n\in\mathbb{Z}} \xi\left(\frac{n\pi}{w}\right)\frac{\sin(wt-n\pi)}{wt-n\pi}, \tag{14.28}$$

and the series converges uniformly in the mean-square and almost surely [30, Theorem 1]. This result we call as Whittaker–Kotel'nikov–Shannon (WKS) stochastic sampling theorem, also consult Remark 2 in [44, §2. The Piranashvili–Lee theory].

The class of Piranashvili processes includes various well-known subclasses of stochastic processes. Some of their particular cases are listed below. Specifying $F_\xi(x,y) = \delta_{xy}F_\xi(x)$ in (14.25) one easily concludes the Karhunen-representation of the covariance

$$B(t,s) = \int_\Lambda f(t,\lambda)f^*(s,\lambda)F_\xi(\mathrm{d}\lambda).$$

Next, putting $f(t,\lambda) = e^{it\lambda}$ in (14.25) one gets the Loève representation:

$$B(t,s) = \int_\Lambda\int_\Lambda e^{i(t\lambda - s\mu)}F_\xi(\mathrm{d}\lambda,\mathrm{d}\mu).$$

Here is $c(\lambda) = |\lambda|$; therefore, the WKS formula (14.28) holds for all $w > \sigma = \sup|\Lambda|$.

Note that the Karhunen process with the Fourier kernel $f(t,\lambda) = e^{it\lambda}$ is the weakly stationary stochastic process having the covariance

$$B(\tau) = \int_\Lambda e^{i\tau\lambda} F_\xi(\mathrm{d}\lambda), \qquad \tau = t - s.$$

A deeper insight into different kinds of harmonizability is presented in [17, 49–51, 53] and the related references therein.

Finally, using $\Lambda = [-w, w]$ for some finite $w > 0$ (called *bandwidth*) for a Karhunen process we arrive at

$$B(\tau) = \int_{-w}^{w} e^{i\tau\lambda} F_\xi(\mathrm{d}\lambda),$$

getting the band-limited process.

For further reading—including historical background, exhaustive results overview and references list until 1987—we refer to the monumental two-tom monograph by Yaglom [52, 53].

14.5 Master Sampling Theorem for Stochastic Signals

Here and in what follows we will concentrate to transferring the results of deterministic results from Sect. 14.2 to stochastic framework. Firstly, we introduce the concept of mean-square derivative of a stochastic process. Denote $L^2(\Omega)$ the space of finite second order complex random variables X defined on a standard probability space $(\Omega, \mathfrak{F}, \mathsf{P})$; $L^2(\Omega)$ is a Hilbert space equipped with the scalar product $\langle X, Y\rangle = \mathbb{E}XY^*$, that is, with the norm $\|\cdot\|_2 = (\mathbb{E}|\cdot|^2)^{1/2}$ endowed. The linear mean-square sense closure $\mathscr{H}(\xi) = \overline{L^2\{\xi(t)\colon t \in \mathbb{R}\}}$ of the process $\xi(t)$ is a subspace of $L^2(\Omega)$, and possesses an H-space structure by itself too.

If there exists a random variable $\xi'(t)$ such that

$$\lim_{h\to 0} \left\| \frac{\xi(t+h) - \xi(t)}{h} - \xi'(t) \right\|_2 = 0,$$

it is called the first mean-square (or *in medio*) sense derivative of the stochastic process $\xi(t) \in L^2(\Omega)$, see, e.g., [46, p. 388, Definition 2]. Conventionally, we will denote by l.i.m. the shorthand of *limes-in-medio* (or mean-square limit) in the sequel. The higher order derivatives we introduce by induction.

Let $L^2(F_\xi; \Lambda)$ be the class of all complex valued functions integrable with respect to the measure $F_\xi(\cdot, \cdot)$, i.e.,

$$L^2(F_\xi; \Lambda) = \left\{ \phi(z) \colon \int_\Lambda \int_\Lambda \phi(z)\, \phi^*(w)\, F_\xi(\mathrm{d}z, \mathrm{d}w) < \infty \right\} .$$

Moreover, equipped with the inner product

$$\langle \phi, \psi \rangle_{L^2(F_\xi; \Lambda)} = \int_\Lambda \int_\Lambda \phi(z)\, \psi^*(w)\, F_\xi(\mathrm{d}z, \mathrm{d}w) \,,$$

$L^2(F_\xi; \Lambda)$ possesses also Hilbert space structure [52, 53, p. 155]. Accordingly, the spectral representation of the rth mean-square derivative $\xi^{(r)}(t)$ is of the form [46, p. 388, Lemma 3, Eq. (17)]

$$\xi^{(r)}(t) = \int_\Lambda \frac{\partial^r}{\partial t^r} f(t, \lambda)\, Z_\xi(\mathrm{d}\lambda), \qquad r \in \mathbb{N}_0 \tag{14.29}$$

assuming $\frac{\partial^r}{\partial t^r} f(t, \lambda) \in L^2(F_\xi; \Lambda)$; for the weakly stationary processes case consult [5, p. 281].

Finally, introduce the mean-square truncation error

$$\tau_n^{(p)}(\xi; t) := \| \mathscr{T}_n^{(p)}(\xi; t) \|_2^2 \tag{14.30}$$

via the truncation error of a deterministic signal f treated in the previous sections.

Lemma 1 *If $f(t, \lambda)$ appears as the kernel function in the spectral representation* (14.26)*, then we have the spectral representation*

$$\mathscr{T}_n^{(p)}(\xi; t) = \int_\Lambda \mathscr{T}_n^{(p)}(f; t)\, Z_\xi(\mathrm{d}\lambda) \tag{14.31}$$

in the mean-square sense.

Lemma 1 is the immediate consequence of the definition (14.16) applied to the process $\xi(t)$ and the spectral representation (14.26), bearing in mind the isometric isomorphism between $\mathscr{H}(\xi)$ and $L^2(F_\xi; \Lambda)$.

Now we are ready to formulate the stochastic setting counterparts of Theorems 1 and 2.

Theorem 4 *Suppose that $\{\xi(t) \colon t \in \Lambda \subseteq \mathbb{R}\}$ is a stochastic process with covariance function of the form* (14.25)*:*

$$B(t, s) = \int_\Lambda \int_\Lambda f(t, \lambda) f^*(s, \mu) F_\xi(\mathrm{d}\lambda, \mathrm{d}\mu) \,,$$

where $F(S_1, S_2)$ is a complex additive set function with respect to both variables, positive definite with bounded total variation $\|F_\xi\|(\Lambda, \Lambda) < \infty$ and the kernel function $f(t, \lambda)$, which satisfies the condition (14.27) for certain $m \in \mathbb{N}_0$, has analytic continuation to the whole $\mathbb{C}$ with respect to the argument t. Then for all parameters $N, N_0, p, q \in \mathbb{N}_0$, $a, b, \alpha, \beta, \sigma > 0$, $c \in \mathbb{C} \setminus \{0\}$, and $\delta \in \mathbb{R}$ satisfying

$$\alpha(N+1) - \sigma > 0,$$
$$[(N+1)\alpha - \sigma]q^{-1} > \beta \geq 0,$$
$$(N+1)\alpha - \sigma - q\beta > \delta > 0,$$

the mean-square truncation error upper bound in generalized pth order derivative sampling series expansion reads as follows:

$$\tau_n^{(p)}(\xi; t) \leq L_f^2 \left(\frac{2}{1 - \mathrm{e}^{-\pi}}\right)^{2(N+1)} \frac{\pi^2 (p!)^2 4^{p+q+N_0+1} \|F_\xi\|(\Lambda, \Lambda)\, \mathrm{e}^{-\delta(2n+1)\frac{\pi}{\alpha}}}{\beta^{2q} D_0^2(a, b)\, [(N+1)\alpha - \sigma - \delta - q\beta]^2} \cdot \left(\frac{\alpha}{\pi\left(n + \frac{1}{2}\right)}\right)^{2(p+q+N_0-m+2)} \left\{1 + \left(\frac{\alpha}{\pi\left(n + \frac{1}{2}\right)}\right)^m\right\}^2 =: U_n(t), \tag{14.32}$$

provided n is enough large positive integer, and finite Vitali total variation

$$V_{F_\xi} = \|F_\xi\|(\Lambda, \Lambda) = \int_\Lambda \int_\Lambda |F_\xi(\mathrm{d}\lambda, \mathrm{d}\mu)|,$$

while t is any bounded real $t \notin \{c, x_k = k\frac{\pi}{\alpha} : k = \overline{-n, n}\}$.

Proof Direct calculation via (14.30) and Lemma 1 (14.31) gives

$$\tau_n^{(p)}(\xi; t) = \|\mathscr{T}_n^{(p)}(\xi; t)\|_2^2 = \left\|\int_\Lambda \mathscr{T}_n^{(p)}(f; t)\, Z_\xi(\mathrm{d}\lambda)\right\|_2^2$$
$$\leq \sup_{t \in \mathbb{R}} |\mathscr{T}_n^{(p)}(f; t)|^2 \int_\Lambda \big|F_\xi(\mathrm{d}\lambda, \mathrm{d}\mu)\big| \leq \|F_\xi\|(\Lambda, \Lambda) \sup_{t \in \mathbb{R}} |\mathscr{T}_n^{(p)}(f; t)|^2.$$

By virtue of the estimate (14.17) (remarking that $\Im(t) = 0$) applied to the truncation error $\mathscr{T}_n^{(p)}(f; t)$ of the kernel function, we conclude the asserted bound (14.32). □

Take now into account the facts upon the parameter $\beta \to 0$ by which we obtained Theorem 2, that is, the bound (14.18). By similar argumentation like in the previous proof assuming only $\beta \geq 0$ we deduce the

Theorem 5 *Under the same hypotheses then in Theorem 4 and using $\beta \geq 0$ we have*

$$\tau_n^{(p)}(\xi; t) \leq \pi^2 L_f^2 \left(\frac{2}{1-\mathrm{e}^{-\pi}}\right)^{2(N+1)} \frac{(p!)^2\, 4^{p+N_0+1}\, \|F_\xi\|(\Lambda, \Lambda)\, \mathrm{e}^{-\delta\,(2n+1)\frac{\pi}{\alpha}}}{D_0^2(a,b)\,[(N+1)\alpha - \sigma - \delta - q\beta]^2}$$
$$\cdot \left(\frac{\alpha}{\pi\left(n+\frac{1}{2}\right)}\right)^{2(p+N_0-m+2)} \left\{1 + \left(\frac{\alpha}{\pi\left(n+\frac{1}{2}\right)}\right)^m\right\}^2 =: V_n(t), \tag{14.33}$$

for all bounded real $t \notin \{c, x_k = k\frac{\pi}{\alpha} : k = \overline{-n, n}\}$.

Remark 2 The decay rate of both upper bounds for fixed t is exponential; (14.32) and (14.33) have magnitudes of

$$U_n(t) = \mathscr{O}\big(n^{-2(p+q+N_0-m+3)}\, \mathrm{e}^{-2(\delta\pi/\alpha)\,n}\big)$$
$$V_n(t) = \mathscr{O}\big(n^{-2(p+N_0-m+3)}\, \mathrm{e}^{-2(\delta\pi/\alpha)\,n}\big)$$

as $n \to \infty$, respectively. ■

So the final generalized derivative sampling series results.

Theorem 6 *Suppose that $\{\xi(t) : t \in \Lambda \subseteq \mathbb{R}\}$ is a Piranashvili process with covariance function of the form (14.25):*

$$B(t,s) = \int_\Lambda \int_\Lambda f(t,\lambda) f^*(s,\mu) F_\xi(\mathrm{d}\lambda, \mathrm{d}\mu)\,,$$

where $F(S_1, S_2)$ is a complex additive set function with respect to both variables, positive definite with finite Vitali variation $V_{F_\xi} = \|F_\xi\|(\Lambda, \Lambda) < \infty$ and the kernel function $f(t,\lambda)$, which satisfies the condition (14.27) for certain $m \in \mathbb{N}_0$, has analytic continuation to the whole $\mathbb{C}$ with respect to the argument t. Then for all parameters $N, N_0, p, q \in \mathbb{N}_0$, $a, b, \alpha, \beta, \sigma > 0$, $\delta \in \mathbb{R}$ satisfying the constraint set

$$\alpha(N+1) - \sigma > 0, \quad [(N+1)\alpha - \sigma]q^{-1} > \beta \geq 0, \quad (N+1)\alpha - \sigma - q\beta > \delta > 0,$$

there holds true the master generalized pth order mean-square derivative sampling formula

$$\underset{x\to t}{\mathrm{l.i.m.}} \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^p \frac{\xi(x)\,\mathrm{sinc}^q \beta(x-t)}{(x-c)^{N_0+1}(a\mathrm{e}^{\delta x} + b\mathrm{e}^{-\delta x})\sin^{N+1}(\alpha x)}$$
$$= \frac{p!\,(-\alpha)^{N_0+p}}{N!\,N_0!\,\alpha^{N-1}} \sum_{k\in\mathbb{Z}} \sum_{j=0}^{N} \sum_{m_1=0}^{j} \frac{(-1)^{k(N+1)} \binom{N}{j} \binom{j}{m_1} C_{j,k}(t; \theta_N') B_{j-m_1}(N)}{(\alpha c - k\pi)^{N_0+1}\,(\alpha t - k\pi)^{p+1}}$$

$$\sum_{m_2=0}^{m_1} \binom{m_1}{m_2} \frac{(N_0+m_1-m_2)!\,{}_2F_1\left(\begin{matrix} m_2-m_1,\, p+1 \\ m_2-m_1-N_0 \end{matrix}\,\bigg|\, \frac{\alpha c-k\pi}{\alpha t-k\pi}\right)}{\left(c-k\frac{\pi}{\alpha}\right)^{m_1-m_2}}\, \xi^{(m_2)}\left(k\frac{\pi}{\alpha}\right)$$

$$+\frac{(-1)^p\, p!}{N_0!} \sum_{r_1=0}^{N_0} \frac{\binom{N_0}{r_1}(p+1)_{r_1} A_{N_0-r_1}(t;\theta_{N_0})}{(c-t)^{p+r_1+1}}\, {}_1F_1\left(\begin{matrix} -r_1 \\ -p-r_1 \end{matrix}\,\bigg|\,(c-t)\frac{\mathrm{d}}{\mathrm{d}c}\right)\circ\xi(c), \tag{14.34}$$

for all bounded real $t \neq c, n\frac{\pi}{\alpha}, n \in \mathbb{Z}$. The notations A, B, C have the same meaning as above.

Proof As to the generalized derivative mean-square sense sampling series result for Piranashvili processes, it is enough letting $n \to \infty$ in both (14.32) and (14.33). Indeed,

$$0 \leq \lim_{n\to\infty} \tau_n^{(p)}(\xi;t) \leq \lim_{n\to\infty} \max\{U_n(t), V_n(t)\} = 0$$

implies the assertion. □

Remark 3 It is worth to mention another approach which leads to the same result of considerable interest since it will be exploited in obtaining the sampling series expansion for random fields.

Consider the pth mean-square derivative $\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^p \xi(x)\, h(x,t)$ where

$$h(x,t) := \frac{\operatorname{sinc}^q \beta(x-t)}{(x-c)^{N_0+1}(a\mathrm{e}^{\delta x}+b\mathrm{e}^{-\delta x})\sin^{N+1}(\alpha x)}.$$

Thus, by virtue of the spectral representations (14.26) and (14.29) of the input process and its higher order derivatives, we get

$$\begin{aligned}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^p \xi(x)\, h(x,t) &= \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^p \int_\Lambda f(x,\lambda) Z_\xi(\mathrm{d}\lambda)\, h(x,t)\\ &= \int_\Lambda \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^p \big\{f(x,\lambda)\, h(x,t)\big\}\, Z_\xi(\mathrm{d}\lambda).\end{aligned}$$

Applying to the last expression the derivative series expansion (14.19) of Theorem 3 to the kernel function $\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^p f(x,\lambda)$, by several legitimate exchange of the order of integration and summation, we conclude

$$\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^p \xi(x)\, h(x,t) = C_1 \sum_{k\in\mathbb{Z}} \sum_{j=0}^{N} \sum_{m_1=0}^{j} \frac{(-1)^{k(N+1)}\binom{N}{j} C_{j,k}(z;\theta_N')\binom{j}{m_1} B_{j-m_1}(N)}{(\alpha c - k\pi)^{N_0+1}\,(\alpha z - k\pi)^{p+1}}$$
$$\times \sum_{m_2=0}^{m_1} \binom{m_1}{m_2} \frac{(N_0+m_1-m_2)!}{\left(c - k\dfrac{\pi}{\alpha}\right)^{m_1-m_2}}$$
$$\times\ {}_2F_1\left(\begin{matrix} m_2-m_1,\ p+1 \\ m_2-m_1-N_0\end{matrix}\,\Big|\,\frac{\alpha c - k\pi}{\alpha z - k\pi}\right)$$
$$\cdot \int_\Lambda \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^{m_2} f(x,\lambda)\Big|_{x=x_k} Z_\xi(\mathrm{d}\lambda)$$
$$+ C_2 \sum_{r_1=0}^{N_0} \frac{\binom{N_0}{r_1}(p+1)_{r_1} A_{N_0-r_1}(z;\theta_{N_0})}{(c-z)^{p+r_1+1}}$$
$$\times\ {}_1F_1\left(\begin{matrix} -r_1 \\ -p-r_1\end{matrix}\,\Big|\,(c-z)\frac{\mathrm{d}}{\mathrm{d}c}\right) \circ f(c), \tag{14.35}$$

where

$$C_1 = \frac{p!\,(-\alpha)^{N_0+p}}{N!\,N_0!\,\alpha^{N-1}}; \quad x_k = k\frac{\pi}{\alpha}; \quad C_2 = \frac{(-1)^p\, p!}{N_0!}\,.$$

Being

$$\int_\Lambda \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^{m_2} f(x,\lambda)\Big|_{x=x_k} Z_\xi(\mathrm{d}\lambda) = \xi^{(m_2)}(x_k),$$

it is sufficient to apply $\text{l.i.m.}_{x\to t}$ in (14.35) to obtain the series expansion statement (14.34) of Theorem 6. ■

The last topic of this section is to derive the *almost sure sense* (called also a.s. P, or *with probability* 1) master generalized derivative sampling series expansion formula. This result we will deduce by the upper bounds either (14.32) or (14.33) and the celebrated Borel–Cantelli Lemma.

Theorem 7 *Assume the same range of parameters applied in Theorem 6. Then we have*

$$\mathsf{P}\Bigg\{(-1)^p \frac{N_0!}{p!} \lim_{x\to t} \frac{\mathrm{d}^p}{\mathrm{d}x^p} \frac{\xi(x)\,\mathrm{sinc}^q \beta(x-t)}{(x-c)^{N_0+1}(a\mathrm{e}^{\delta x}+b\mathrm{e}^{-\delta x})\sin^{N+1}(\alpha x)}$$
$$= \frac{(-1)^{N_0}\alpha^{N_0+p}}{N!\,\alpha^{N-1}} \sum_{k\in\mathbb{Z}} \sum_{j=0}^{N} \sum_{m_1=0}^{j} \frac{(-1)^{k(N+1)}\binom{N}{j}\binom{j}{m_1} C_{j,k}(t;\theta_N') B_{j-m_1}(N)}{(\alpha c - k\pi)^{N_0+1}\,(\alpha t - k\pi)^{p+1}}$$

$$\sum_{m_2=0}^{m_1} \frac{\binom{m_1}{m_2}(N_0+m_1-m_2)!\,{}_2F_1\left(\begin{matrix} m_2-m_1,\, p+1 \\ m_2-m_1-N_0 \end{matrix}\,\middle|\,\frac{\alpha c-k\pi}{\alpha t-k\pi}\right)}{\alpha^{m_2-m_1}\,(c\alpha-k\pi)^{m_1-m_2}}\,\xi^{(m_2)}(x_k)$$

$$+\sum_{r_1=0}^{N_0}\frac{\binom{N_0}{r_1}(p+1)_{r_1}A_{N_0-r_1}(t;\theta_{N_0})}{(c-t)^{p+r_1+1}}\,{}_1F_1\left(\begin{matrix}-r_1\\ -p-r_1\end{matrix}\,\middle|\,(c-t)\frac{\mathrm{d}}{\mathrm{d}c}\right)\circ\xi(c)\Bigg\}=1 \tag{14.36}$$

for all bounded real $t \notin \{c, x_k = k\frac{\pi}{\alpha} : k \in \mathbb{Z}\}$.

Proof To prove the almost sure convergence of the partial sum of the right-hand side series (14.34) it is suitable to evaluate, using the Markov inequality, the probability

$$\mathsf{P}_n = \mathsf{P}\left\{|\mathscr{T}_n^{(p)}(\xi;t)| \geq \epsilon;\ \text{for at least one}\ n \geq n_0\right\} \leq \epsilon^{-2}\,\tau_n^{(p)}(\xi;t).$$

Accordingly, since the asymptotic described in Remark 2 we have

$$\mathsf{P}_n = \mathscr{O}\left(n^{-2(p+q+N_0-m+3)}\,\mathrm{e}^{-2(\delta\pi/\alpha)\,n}\right).$$

The series

$$\sum_{n\geq n_0}\mathsf{P}_n < C\,\mathrm{Li}_{2(p+q+N_0-m+3)}\left(\mathrm{e}^{-2(\delta\pi/\alpha)}\right); \qquad C>0$$

obviously converges. (Here $\mathrm{Li}_a(w) = \sum_{n\geq 1} n^{-a}w^n$; $|w| < 1$ stands for the so-called polylogarithm or Jonquière's function, see [4, pp. 30–31], [16].)

Thus, by the Borel–Cantelli Lemma we deduce that the convergence in (14.36) holds with probability 1 uniformly in any bounded t. □

Now, we make use of a Belyaev's idea [2, p. 443] to evaluate the a.s. P convergence rate $\rho(n)$ in (14.34).

Theorem 8 *There exists a positive integer* $N(\omega)$; $\omega \in \Omega$ *for which*

$$\mathsf{P}\left\{|\mathscr{T}_n^{(p)}(\xi;t)| < \rho(n);\ \text{for all}\ n \geq N(\omega)\right\} = 1,$$

where the series

$$\sum_{n\geq 1}\frac{\mathrm{e}^{-2(\delta\pi/\alpha)n}}{n^{2(p+N_0-m+3)}}\frac{1}{\rho^2(n)}$$

converges.

Proof For fixed parameters $p, q, N_0, m, \alpha, \delta$ and bounded t we have

$$\mathsf{P}\left\{\exists n \geq N(\omega) \colon |\mathscr{T}_n^{(p)}(\xi; t)| > \rho(n)\right\} \leq \sum_{n \geq N(\omega)} \frac{\tau_n^{(p)}(\xi; t)}{\rho^2(n)} \leq \sum_{n \geq N(\omega)} \frac{\mathrm{e}^{-2(\delta\pi/\alpha)n}}{n^{2(p+N_0-m+3)}} \frac{1}{\rho^2(n)}.$$

The existence of such non-negative integer $N(\omega)$ that for all $n \geq N(\omega)$ there holds $|\mathscr{T}_n^{(p)}(\xi; t)| < \rho(n)$ with probability 1 is ensured by the Borel–Cantelli Lemma as the right-hand side series converges. The rest is obvious. □

Remark 4 The truncation error upper bound $U_n(t)$ (14.32) is of greater magnitude than $V_n(t)$, which is the applied one in Theorem 8, consult Remark 2 as well. Therefore Theorem 8 covers both approaches to the matter discussed in detail in Theorems 1 and 2 for deterministic signals. ■

Finally, a plethora of illustrative examples can be constructed by Theorem 7 in obtaining the almost sure P rate of convergence. So, following the traces of Belyaev's classical example [2, p. 443]

$$|\mathscr{T}_n^{(p)}(\xi; t)| < \frac{\mathrm{e}^{-2(\delta\pi/\alpha)n}}{n^{2(p+N_0-m+5/2)}} (\log n)^{\frac{1+\epsilon}{2}}, \qquad \epsilon > 0$$

can be considered for instance.

14.6 Generalized Sampling Series for Random Fields

Let $\{\Xi(\boldsymbol{t}) \colon \boldsymbol{t} = (t_1, \cdots, t_d) \in \mathbb{R}^d\}$ be a scalar complex valued random field with d-dimensional argument, separable simultaneously with all its mean-square derivatives if any. For the sake of simplicity let us assume that $\Xi(\boldsymbol{t})$ is centered, that is, $\mathbb{E}\Xi(\boldsymbol{t}) = 0, \boldsymbol{t} \in \mathbb{R}^d$ and the covariance

$$B(\boldsymbol{t}, \boldsymbol{s}) = \int_{\boldsymbol{\Lambda}} \int_{\boldsymbol{\Lambda}} \prod_{j=1}^{d} f_j(t_j, \lambda_j)\, f_j^*(s_j, \mu_j)\, F_\Xi(\mathrm{d}\boldsymbol{\lambda}, \mathrm{d}\boldsymbol{\mu})$$

, where Λ_j is a parameter space $\boldsymbol{\Lambda} := \prod_{j=1}^d \Lambda_j$ and $\boldsymbol{u} := (u_1, \cdots, u_d)$ denotes a d-dimensional vector throughout, while $F(\boldsymbol{S}, \boldsymbol{S}')$ is a complex positive definite function of sets $S_1, \cdots, S_d; S_1', \cdots, S_d'$, additive with respect to all arguments having finite total variation, provided

$$\int_{\boldsymbol{\Lambda}} \int_{\boldsymbol{\Lambda}} |F_\Xi(\mathrm{d}\boldsymbol{\lambda}, \mathrm{d}\boldsymbol{\mu})| < \infty.$$

Similarly to the spectral representation of the Piranashvili process, the random field $\Xi(\boldsymbol{t})$ possesses the mean-square stochastic integral representation[2]

$$\Xi(\boldsymbol{t}) = \int_{\boldsymbol{\Lambda}} \prod_{j=1}^{d} f_j(t_j, \lambda_j)\, Z_\Xi(\boldsymbol{\lambda})\,.$$

Here the so-called *spectral field* $Z_\Xi(\boldsymbol{\lambda})$ is orthogonally scattered, that is,

$$\mathbb{E} Z_\Xi(\boldsymbol{S}) Z_\Xi^*(\boldsymbol{S}') = F_\Xi(\boldsymbol{S}, \boldsymbol{S}'), \qquad \boldsymbol{S}, \boldsymbol{S}' \in \sigma(\boldsymbol{\Lambda})\,.$$

Moreover, we assume that all $f_j(t_j, \lambda_j);\ j = \overline{1, d}$ are kernel functions satisfying (14.27), that is,

$$|f_j(t_j, \lambda_j)| \le L_{f_j}(\lambda_j)\, \big(1 + |t_j|^{m_j}\big)\, \mathrm{e}^{c_j(\lambda_j)\,|\Im(t_j)|}, \qquad t_j \in \mathbb{C};\ \ m_j \in \mathbb{N}_0,$$

where

$$\sup_{\Lambda_j} L_j(\lambda_j) < \infty; \quad \sup_{\Lambda_j} c_j(\lambda_j) = \sigma_j < \infty; \qquad j = \overline{1, d}.$$

At this point we are ready to formulate the appropriate generalized derivative mean-square sampling series result for random fields. For the sake of simplicity we take the shorthand notations

$$\frac{\partial^{|p|}}{\partial \boldsymbol{x}^{|p|}} = \frac{\partial^{p_1+\cdots+p_d}}{\partial x_1^{p_1} \cdots \partial x_d^{p_d}}; \quad \boldsymbol{u}^{|v|} = \prod_{j=1}^{d} u_j^{v_j}; \quad \boldsymbol{p}! = \prod_{j=1}^{d} p_j!\,.$$

Theorem 9 *Assume that the two-dimensional random field* $\{\Xi(t_1, t_2) \colon t_1, t_2 \in \mathbb{R}\}$ *satisfies the above exposed conditions. Also let all parameters* N_1, N_2, N_{01}, N_{02}, $p_1, p_2, q_1, q_2 \in \mathbb{N}_0$*;* $a_1, a_2, b_1, b_2, \alpha_1, \alpha_2, \beta_1, \beta_2, \sigma_1, \sigma_2 > 0$*;* $c_1, c_2 \in \mathbb{C} \setminus \{0\}$*, and* $\delta_1, \delta_2 \in \mathbb{R}$ *satisfy*

$$\begin{aligned} \alpha_j(N_j + 1) - \sigma_j &> 0, \\ [(N_j + 1)\alpha_j - \sigma_j] q_j^{-1} &> \beta_j \ge 0, \\ (N_j + 1)\alpha_j - \sigma_j - q_j \beta_j &> \delta_j, > 0, \qquad j = 1, 2. \end{aligned}$$

Denote

$$h_j(x_j, t_j) = \frac{\mathrm{sinc}^{q_j} \beta_j(x_j - t_j)}{(x_j - c_j)^{N_{0j}+1} (a_j \mathrm{e}^{\delta_j x_j} + b_j \mathrm{e}^{-\delta_j x_j}) \sin^{N_j+1}(\alpha_j x_j)}, \qquad j = 1, 2\,.$$

[2]The existence of such representation is ensured by the Karhunen–Cramèr theorem.

Then for the almost all sample paths there holds true

$$\underset{\boldsymbol{x}\to\boldsymbol{t}}{\text{l.i.m.}}\frac{\partial^{|p|}}{\partial x^{p_1}\partial x_2^{p_2}}\left\{\Xi(x_1,x_2)h_1(x_1,t_1)\,h_2(x_2,t_2)\right\} = \frac{\boldsymbol{p}!\,(-\boldsymbol{\alpha})^{|N_0+p|}}{\boldsymbol{N}!\,\boldsymbol{N}_0!\,\boldsymbol{\alpha}^{|N-1|}}$$

$$\cdot\sum_{k_1,k_2\in\mathbb{Z}}\sum_{j_1=0}^{N_1}\sum_{m_{11}=0}^{j_1}\sum_{j_2=0}^{N_2}\sum_{m_{21}=0}^{j_2}\frac{(-1)^{k_1(N_1+1)}\binom{N_1}{j_1}\binom{j_1}{m_{11}}C_{j_1,k_1}(t_1;\theta_{N_1'})B_{j_1-m_{11}}(N_1)}{(\alpha_1c_1-k_1\pi)^{N_{01}+1}\,(\alpha_1t_1-k_1\pi)^{p_1+1}}$$

$$\cdot\frac{(-1)^{k_2(N_2+1)}\binom{N_2}{j_2}\binom{j_2}{m_{21}}C_{j_2,k_2}(t_2;\theta_{N_2'})B_{j_2-m_{21}}(N_2)}{(\alpha_2c_2-k_2\pi)^{N_{02}+1}\,(\alpha_2t_2-k_2\pi)^{p_2+1}}$$

$$\sum_{m_{12}=0}^{m_{11}}\frac{\binom{m_{11}}{m_{12}}(N_{01}+m_{11}-m_{12})!\,{}_2F_1\left(\begin{matrix}m_{12}-m_{11},\,p_1+1\\ m_{12}-m_{11}-N_{01}\end{matrix}\,\Big|\,\frac{\alpha_1c_1-k_1\pi}{\alpha_1t_1-k_1\pi}\right)}{\alpha_1^{m_{12}-m_{11}}\,(c_1\alpha_1-k_1\pi)^{m_{11}-m_{12}}}$$

$$\sum_{m_{22}=0}^{m_{21}}\frac{\binom{m_{21}}{m_{22}}(N_{02}+m_{21}-m_{22})!\,{}_2F_1\left(\begin{matrix}m_{22}-m_{21},\,p_2+1\\ m_{22}-m_{21}-N_{02}\end{matrix}\,\Big|\,\frac{\alpha_2c_2-k_2\pi}{\alpha_2t_2-k_2\pi}\right)}{\alpha_2^{m_{22}-m_{21}}\,(c_2\alpha_2-k_2\pi)^{m_{21}-m_{22}}}$$

$$\cdot\frac{\partial^{|m|}}{\partial x_1^{m_{12}}\partial x_2^{m_{21}}}\,\Xi\left(k_1\frac{\pi}{\alpha_1},k_2\frac{\pi}{\alpha_2}\right)+\frac{(-1)^{p_2}\boldsymbol{p}!\,(-\alpha_1)^{N_{01}+p_1}}{N_1!N_{01}!N_{02}!\alpha_1^{N_1-1}}$$

$$\cdot\sum_{k_1\in\mathbb{Z}}\sum_{r_2=0}^{N_{02}}\sum_{j_1=0}^{N_1}\sum_{m_{11}=0}^{j_1}\frac{(-1)^{k_1(N_1+1)}\binom{N_1}{j_1}\binom{j_1}{m_{11}}C_{j_1,k_1}(t_1;\theta_{N_1'})B_{j_1-m_{11}}(N_1)}{(\alpha_1c_1-k_1\pi)^{N_{01}+1}\,(\alpha_1t_1-k_1\pi)^{p_1+1}}$$

$$\cdot\frac{\binom{N_{02}}{r_2}(p_2+1)_{r_2}A_{N_{02}-r_2}(t_2;\theta_{N_{02}})}{(c_2-t_2)^{p_2+r_2+1}}$$

$$\sum_{m_{12}=0}^{m_{11}}\frac{\binom{m_{11}}{m_{12}}(N_{01}+m_{11}-m_{12})!\,{}_2F_1\left(\begin{matrix}m_{12}-m_{11},\,p_1+1\\ m_{12}-m_{11}-N_{01}\end{matrix}\,\Big|\,\frac{\alpha_1c_1-k_1\pi}{\alpha_1t_1-k_1\pi}\right)}{\alpha_1^{m_{12}-m_{11}}\,(c_1\alpha_1-k_1\pi)^{m_{11}-m_{12}}}$$

$$\cdot\frac{\partial^{m_{12}}}{\partial x_1^{m_{12}}}\,{}_1F_1\left(\begin{matrix}-r_2\\ -p_2-r_2\end{matrix}\,\Big|\,(c_2-t_2)\frac{\partial}{\partial c_2}\right)\circ\,\Xi\left(k_1\frac{\pi}{\alpha_1},c_2\right)$$

$$+\frac{(-1)^{p_1}\boldsymbol{p}!\,(-\alpha_2)^{N_{02}+p_2}}{N_2!N_{02}!N_{01}!\alpha_2^{N_2-1}}$$

$$\cdot\sum_{k_2\in\mathbb{Z}}\sum_{r_1=0}^{N_{01}}\sum_{j_2=0}^{N_2}\sum_{m_{21}=0}^{j_2}\frac{(-1)^{k_2(N_2+1)}\binom{N_2}{j_2}\binom{j_2}{m_{21}}C_{j_2,k_2}(t_2;\theta_{N_2'})B_{j_2-m_{21}}(N_2)}{(\alpha_2c_2-k_2\pi)^{N_{02}+1}\,(\alpha_2t_2-k_2\pi)^{p_2+1}}$$

$$\cdot\frac{\binom{N_{01}}{r_1}(p_1+1)_{r_1}A_{N_{01}-r_1}(t_1;\theta_{N_{01}})}{(c_1-t_1)^{p_1+r_1+1}}$$

$$\sum_{m_{22}=0}^{m_{21}} \frac{\binom{m_{21}}{m_{22}}(N_{02}+m_{21}-m_{22})!\,{}_2F_1\left(\begin{matrix} m_{22}-m_{21},\, p_2+1 \\ m_{22}-m_{21}-N_{02}\end{matrix}\,\middle|\,\frac{\alpha_2 c_2 - k_2\pi}{\alpha_2 t_2 - k_2\pi}\right)}{\alpha_2^{m_{22}-m_{21}}\;(c_2\alpha_2 - k_2\pi)^{m_{21}-m_{22}}}$$

$$\cdot \frac{\partial^{m_{21}}}{\partial x_2^{m_{21}}}\,{}_1F_1\left(\begin{matrix} -r_1 \\ -p_1-r_1\end{matrix}\,\middle|\,(c_1-t_1)\frac{\partial}{\partial c_1}\right)\circ \Xi\left(c_1, k_2\frac{\pi}{\alpha_2}\right)$$

$$+ \frac{(-1)^{|p|}\boldsymbol{p}!}{\boldsymbol{N}_0!}\sum_{r_1=0}^{N_{01}} \frac{\binom{N_{01}}{r_1}(p_1+1)_{r_1} A_{N_{01}-r_1}(t_1;\theta_{N_{01}})}{(c_1-t_1)^{p_1+r_1+1}}$$

$$\sum_{r_2=0}^{N_{02}} \frac{\binom{N_{02}}{r_2}(p_2+1)_{r_2} A_{N_{02}-r_2}(t_2;\theta_{N_{02}})}{(c_2-t_2)^{p_2+r_2+1}}$$

$${}_1F_1\left(\begin{matrix} -r_1 \\ -p_1-r_1\end{matrix}\,\middle|\,(c_1-t_1)\frac{\partial}{\partial c_1}\right)\,{}_1F_1\left(\begin{matrix} -r_2 \\ -p_2-r_2\end{matrix}\,\middle|\,(c_2-t_2)\frac{\partial}{\partial c_2}\right)\circ \Xi(c_1,c_2). \tag{14.37}$$

The convergence is uniform with respect to bounded $t_j \notin \{c_j, x_{jn} = n_j\frac{\pi}{\alpha_j}\}$, $j = 1, 2$.

Remark 5 We point out that certain formulae listed in [32–34] are corollaries of the master two-dimensional generalized sampling series (14.37). Moreover, in the case $p_j = q_j = N_j = 0$; $j = 1, 2$ (14.37) becomes the formula [37, p. 17, Eq. (7)], while specifying $p_j = N_j = N_{0j} = 0$; $j = 1, 2$ implies [37, p. 18, Eq. (8)]. Finally, by $p_j = N_j = 0$; $N_{0j} = -1$; $j = 1, 2$ we infer the formula [37, p. 18, Eq. (9)]. ■

References

1. M. Abramowitz, I.A. Stegun (eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Applied Mathematics Series, vol. 55 (National Bureau of Standards, Washington, D. C., 1964); 9th Reprinted edition, Dover Publications, New York, 1972
2. Yu.K. Belyaev, Analytical random processes. Teor. Veroyat. Primenen. **III/4**, 437–444 (1959). (in Russian)
3. P.L. Butzer, W. Splettstösser, R.L. Stens, The sampling theorem and linear prediction in signal analysis. Jahresber. Deutsch. Math.-Verein. **90**(1), 1–70 (1988)
4. A. Erdélyi, W. Magnus, F. Oberhettinger, F.G. Tricomi, *Higher Transcendental Functions*, vol. 1 (McGraw–Hill, New York, 1953)
5. I.I. Gikhman, A.V. Skorokhod, *Introduction to the Theory of Stochastic Processes*. 2nd edn. (Nauka, Moscow, 1971). (in Russian)
6. J.R. Higgins, Five short stories about the cardinal series. Bull. Am. Math. Soc. (N.S.) **12**(1), 45–89 (1985)
7. J.R. Higgins, Sampling theorem and the contour integral method. Appl. Anal. **41**, 155–171 (1991)

8. J.R. Higgins, Derivative sampling – a paradigm example of multichannel methods, in *Sampling Theory in Fourier and Signal Analysis: Advanced Topics*, ed. by J.R. Higgins, R.L. Stens (Oxford University Press, Oxford, 1999), pp. 67–95
9. C. Houdré, Harmonizability, V-boundedness, (2, p)-boundedness of stochastic processes. Probab. Theory Relat. Fields **84**, 39–54 (1990)
10. C. Houdré, On the linear prediction of multivariate (2, p)-bounded processes. Ann. Probab. **19**(2), 843–867 (1991)
11. C. Houdré, On the spectral SLLN and the pointwise ergodic theorem in L^α. Ann. Probab. **20**(4), 1731–1753 (1992)
12. C. Houdré, Wavelets, probability, and statistics: some bridges, in *Wavelets: Mathematics and Applications*, ed. by J.J. Benedetto, M.W. Frazier (CRC Press, Boca Raton, 1994), pp. 365–399
13. C. Houdré, Reconstruction of band limited processes form irregular samples. Ann. Probab. **23**(2), 674–696 (1995)
14. A. Hurwitz, R. Courant, *Allgemeine Funktionentheorie und elliptische Funktionen Geometrische Funktionentheorie* (Springer, Berlin, 1922)
15. J.A. Jerri, The Shannon sampling theorem – its various extensions and applications: a tutorial review. Proc. IEEE **65**(11), 1565–1596 (1977)
16. A. Jonquière, Note sur la série $\sum_{n=1}^{\infty} \frac{x^n}{n^s}$. Bull. Soc. Math. France **17**, 142–152 (1889)
17. Y. Kakihara, *Multidimensional Second Order Stochastic Processes* (World Scientific, Singapore, 1997)
18. A.J. Lee, On band limited stochastic processes. SIAM J. Appl. Math. **30**, 267–277 (1976)
19. A.J. Lee, Characterization of bandlimited functions and processes. Inform. Control **31**, 258–271 (1976)
20. A.J. Lee, Approximate interpolation and the sampling theorem. SIAM J. Appl. Math. **32**, 730–743 (1977)
21. A.J. Lee, Sampling theorems for nonstationary processes. Trans Am. Math. Soc. **242**, 225–241 (1978)
22. D.S. Mitrinović, *Analytic Inequalities*. In cooperation with Vasić P. M. Die Grundlehren der mathematischen Wissenschaften, Band 165 (Springer, New York, 1970)
23. A.Ya. Olenko, T.K. Pogány, Direct Lagrange–Yen type interpolation of random fields. Theory Stoch. Process. **9**(25)(3–4), 242–254 (2003)
24. A.Ya. Olenko, T.K. Pogány, A precise upper bound for the error of interpolation of stochastic processes. Theory Probab. Math. Stat. **71**, 151–163 (2005)
25. A.Ya. Olenko, T.K. Pogány, Time shifted aliasing error upper bounds for truncated sampling cardinal series. J. Math. Anal. Appl. **324**(1), 262–280 (2006)
26. A.Ya. Olenko, T.K. Pogány, On sharp bounds for remainders in multidimensional sampling theorem. Sampl. Theory Signal Image Process. **6**(3), 249–272 (2007)
27. A.Ya. Olenko, T.K. Pogány, Universal truncation error upper bounds in sampling restoration. Georgian Math. J. **17**, 765–786 (2010)
28. A.Ya. Olenko, T.K. Pogány, Universal truncation error upper bounds in irregular sampling restoration. Appl. Anal. **90**(3–4), 595–608 (2010)
29. A.Ya. Olenko, T.K. Pogány, Average sampling reconstruction of harmonizable processes. Commun. Stat. Theory Methods **40**(19–20), 3587–3598 (2011)
30. Z.A. Piranašvili, The problem of interpolation of random processes. Teor. Verojatnost. i Primenen. **12**(4), 708–717 (1967). (in Russian)
31. Z. Piranashvili, On the extension of Kotelnikoff–Shannon formula. Bull. Georgian Acad. Sci. **154**(1), 52–54 (1996)
32. Z. Piranashvili, Kotelnikov-Shannon's generalized formula for the random fields. Bull. Georgian Acad. Sci. **163**(3), 448–451 (2001)
33. Z. Piranashvili, On the generalization of one interpolation formula. Bull. Georgian Acad. Sci. **166**(2), 251–254 (2002)
34. Z. Piranashvili, On the generalized formula of exponentially convergent sampling. Bull. Georgian Acad. Sci. **170**(1), 50–53 (2004)

35. Z. Piranashvili, On estimation of the remainder term of the generalized sampling series. Bull. Georgian Natl. Acad. Sci. (N.S.) **3**(2), 30–32 (2009)
36. Z.A. Piranashvili, On the generalized fast convergent sampling series. Bull. Georgian Natl. Acad. Sci. (N.S.) **5**(2), 19–24 (2011)
37. Z. Piranashvili, Some remarks on certain sampling formulas. Bull. Georgian Natl. Acad. Sci. (N.S.) **7**(3), 15–19 (2013)
38. Z.A. Piranashvili, T.K. Pogány, Some new generalizations of the Kotel'nikov–Shannon formula for stochastic signals, in *Probability Theory and Mathematical Statistics Conference Dedicated to* 100*th Birth Anniversary of A. N. Kolmogorov*, held September 21–27, Tbilisi, Georgia (2003). (unpublished conference talk)
39. T. Pogány, An approach to the sampling theorem for continuous time processes. Austral. J. Stat. **31**(3), 427–432 (1989)
40. T. Pogány, On the irregular-derivative sampling with uniformly dense sample points for bandlimited stochastic processes, in *Exploring Stochastic Laws. Festschrift in Honour of the 70th Birthday of Academician Vladimir Semenovich Korolyuk*, ed. by A.V. Skorokhod, Yu.A. Borovskikh (VSP Utrecht, The Netherlands, 1995), pp. 395–408
41. T. Pogány, On certain correlation properties of the sampling cardinal series expansion of stationary stochastic processes. Glasnik Mat. Ser. III **31**(51), 353–362 (1996)
42. T. Pogány, Some irregular derivative sampling theorems, in *Functional Analysis*, ed. by D. Butković et al., vol. IV. Proceedings of Postgraduate School and Conference. (held in Dubrovnik, Croatia, November 10–17, 1993), Matematisk Institut, Aarhus Universitet (1997), pp. 205–219
43. T. Pogány, Aliasing in the sampling restoration of non-band-limited homogeneous random fields. Math. Pannon. **8**(1), 155–164 (1997)
44. T. Pogány, Almost sure sampling restoration of bandlimited stochastic signals, in *Sampling Theory in Fourier and Signal Analysis: Advanced Topics*, ed. by J.R. Higgins, R.L. Stens (Oxford University Press, Oxford, 1999), 203–232
45. T. Pogány, Multidimensional Lagrange-Yen type interpolation via Kotel'nikov–Shannon sampling formulae. Ukrainian Math. J. **55**(11), 1810–1827 (2003)
46. T.K. Pogány, Whittaker-type derivative sampling reconstruction of stochastic $L^{\alpha}(\Omega)$-processes. Appl. Math. Comput. **187**(1), 384–394 (2007)
47. T. Pogány, P. Peruničić, On the sampling theorem for homogeneous random fields. Theory Probab. Math. Stat. **53**, 153–159 (1996)
48. T. Pogány, P. Peruničić, On the multidimensional sampling theorem. Glas. Mat. Ser. III **36**(56), 155–167 (2001)
49. M. Priestley, *Non-linear and Non-stationary Time Series* (Academic Press, New York, 1988)
50. M.M. Rao, Harmonizable processes: structure theory. Enseign. Math. (2) **28**(3–4), 295–351 (1982)
51. Ju.A. Rozanov, Spectral analysis of abstract functions. Theor. Probab. Appl. **4**(3), 271–287 (1959)
52. A.M. Yaglom, in *Correlation Theory of Stationary and Related Random Functions I. Basic Results*. Springer Series in Statistics (Springer, New York, 1987)
53. A.M. Yaglom, in *Correlation Theory of Stationary and Related Random Functions II. Supplementary Notes and References*. Springer Series in Statistics (Springer, New York, 1987)

Chapter 15
Voronoi Polygonal Hybrid Finite Elements and Their Applications

Hui Wang and Qing-Hua Qin

This chapter describes the polygonal hybrid finite element formulation with fundamental solution kernels for two-dimensional elasticity in isotropic and homogeneous solids. The n-sided polygonal discretization is implemented by the Voronoi diagram in a given domain. Then the element formulation is established by introducing two independent displacements, respectively, defined within the element domain and over the element boundary. The element interior fields approximated by the fundamental solutions of problem can naturally satisfy the governing equations and the element frame fields approximated by one-dimensional shape functions are used to guarantee the conformity of elements. As a result, only element boundary integrals caused in the modified hybrid functional are needed for practical computation. Finally, the present method is verified by three examples involving the usage of general and special n-sided polygonal hybrid finite elements.

15.1 Introduction

The finite element method (FEM) is the most popular tool for finding numerical solutions of a wide range of engineering problems subjected to loadings, because an engineering problem defined in a complex continuum can be modeled by a finite number of small elements with simple geometric shapes. In the conventional two-

H. Wang
College of Civil Engineering and Architecture, Henan University of Technology, Zhengzhou, China

Q.-H. Qin (✉)
Research School of Engineering, Australian National University, Canberra, ACT, Australia
e-mail: qinghua.qin@anu.edu.au

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_15

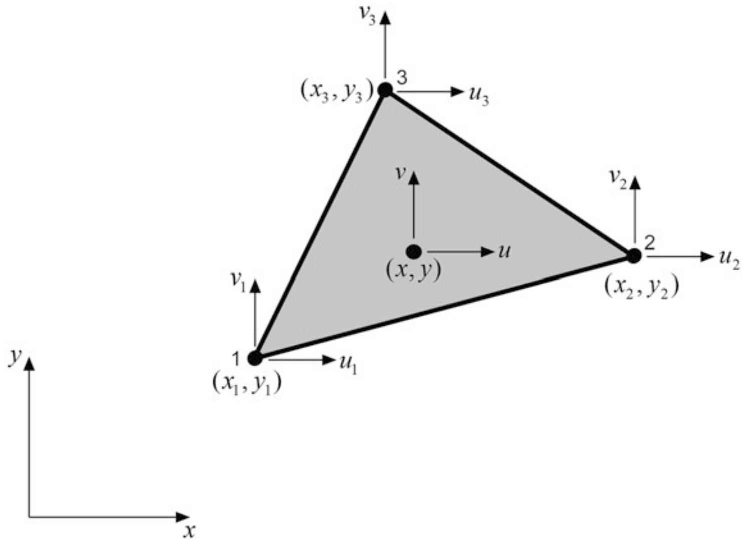

Fig. 15.1 Schematic of triangular linear finite element

dimensional finite element formulation, triangular or quadrilateral finite elements are well developed for various engineering applications [1, 2]. Such element formulation is generally based on the use of first- or second-order polynomial interpolation functions for the dependent field variable approximation, i.e., displacement and temperature. This requires that the values of the dependent field variable at element nodes can uniquely determine the coefficients of its interpolating polynomial. For example, for the two-dimensional triangular linear finite element consisting of three nodes, displayed in Fig. 15.1, the displacement component at any point (x, y) or (x_1, x_2) in the whole element can be approximated by the following first-order polynomial interpolation as:

$$u(x, y) = a + bx + cy \tag{15.1}$$

where a, b, and c are the coefficients to be determined.

We notice that the displacement field variable $u(x, y)$ should be equal to the nodal displacement when the coordinate (x, y) moves to that nodal point. Thus, we have

$$\begin{aligned} a + bx_1 + cy_1 &= u(x_1, y_1) = u_1 \\ a + bx_2 + cy_2 &= u(x_2, y_2) = u_2 \\ a + bx_3 + cy_3 &= u(x_3, y_3) = u_3 \end{aligned} \tag{15.2}$$

Equation (15.2) can be rewritten in matrix form as

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{Bmatrix} a \\ b \\ c \end{Bmatrix} = \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} \tag{15.3}$$

from which we obtain

$$\begin{Bmatrix} a \\ b \\ c \end{Bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix}^{-1} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} \tag{15.4}$$

Hence, substituting these coefficients into Eq. (15.1) finally gives

$$u(x, y) = N_1(x, y)u_1 + N_2(x, y)u_2 + N_3(x, y)u_3 \tag{15.5}$$

in which

$$\begin{aligned} N_1(x, y) &= \frac{1}{2A_e}[(x_2 y_3 - x_3 y_2) + (y_2 - y_3)x + (x_3 - x_2)y] \\ N_2(x, y) &= \frac{1}{2A_e}[(x_3 y_1 - x_1 y_3) + (y_3 - y_1)x + (x_1 - x_3)y] \\ N_3(x, y) &= \frac{1}{2A_e}[(x_1 y_2 - x_2 y_1) + (y_1 - y_2)x + (x_2 - x_1)y] \end{aligned} \tag{15.6}$$

are shape functions and A_e is the area of the triangular element that can be calculated using the determinant of the coefficient matrix in Eq. (15.3)

$$A_e = \frac{1}{2} \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix} \tag{15.7}$$

As an alternative to conventional finite elements, there has been growing interest in developing nontraditional finite elements with arbitrary polygonal and polyhedral meshes over the past decade [3–5]. In convex polygonal or polyhedral finite elements, the number of element sides n is not restricted to three (triangles) or four (quadrilaterals) in the two-dimensional cases, so that they are capable of possessing higher degrees of geometric isotropy and thus the meshing effort can be significantly simplified for modeling complex geometries without introducing numerical instability and the quality of mesh can be improved. However, it may not be always possible to uniquely determine the coefficients of the polynomials when $n \geq 4$, typically for polygonal elements of arbitrary shapes. For such case, although n linear equations for the n polynomial coefficients can still be defined, the equations may not be independent. Moreover, when $n \geq 4$, it is impossible to ensure inter-element displacement compatibility with n-term polynomial representations. To overcome these obstacles, the Laplace/Wachspress interpolants based on barycentric coordinates [3–6] are usually employed as shape functions for approximated displacement fields in the polygonal finite element [7–9]. However, the construction of Laplace/Wachspress shape functions requires complicated mathematical transformations, especially for polygonal elements with curved edges. Moreover, different polygonal finite elements require different interpolants, as like conventional finite elements. Besides, the numerical domain integration associated with arbitrary polygonal finite element is a non-trivial task and usually needs special integration rule such as homogeneous integration, Green-Gauss quadrature, and strain smoothing [10–12].

Apart from the polygonal finite element technique with Laplace/Wachspress interpolants, the hybrid Trefftz finite element method (HT-FEM) using T-complete solutions of problem as approximation kernels can be utilized for developing polygonal finite elements, because of the distinctive characteristic of element boundary integral in the HT-FEM [13]. Different to the shape-function-based finite elements, the hybrid Trefftz finite element introduces the two different displacement fields, which are, respectively, defined inside the element and on its boundary. The interior displacement field is approximated by the linear combination of truncated T-complete functions of the problem, which exactly satisfy the governing partial differential equations of the problem, such that the interior displacement field can also satisfy the governing partial differential equations of the problem inside the element, but not the conformity conditions between adjacent elements, which can be guaranteed by introducing the frame displacement field defined on the element boundary. To provide a linkage between these two independent displacement fields, the hybrid integral functional should be elaborately developed. Typically, the natural feature of the interior displacement field can be used to remove the domain integral in it, and only element boundary integrals are encountered for computation, which can be easily evaluated by standard Gaussian numerical quadrature. This means the polygonal hybrid Trefftz finite elements with any number of sides can be flexibly constructed for modeling the computational domain. Figure 15.2 shows a general triangular hybrid Trefftz finite element with three nodes, for which the interior displacement field (gray region) is only defined inside the element, and the independent linear frame displacement field (color lines) is just defined over the element boundary. Although the same three nodes are included in this element, the interior displacement field based on T-complete interpolant can give non-uniform stress and strain fields inside the element, rather than the constant stress and strain fields in the conventional triangular linear finite element shown in Fig. 15.1. However, the expressions of T-complete solutions of some problems are either complex or difficult to be derived. Moreover, one needs to properly arrange truncated terms for hybrid polygonal Trefftz finite elements with large numbers of sides to prevent spurious energy modes and keep the solving matrix be of full rank.

In order to improve the hybrid Trefftz finite element formulation, a novel hybrid finite element formulation is formed with the help of fundamental solutions of problem which usually have unified expressions in practice [13–19], as is called

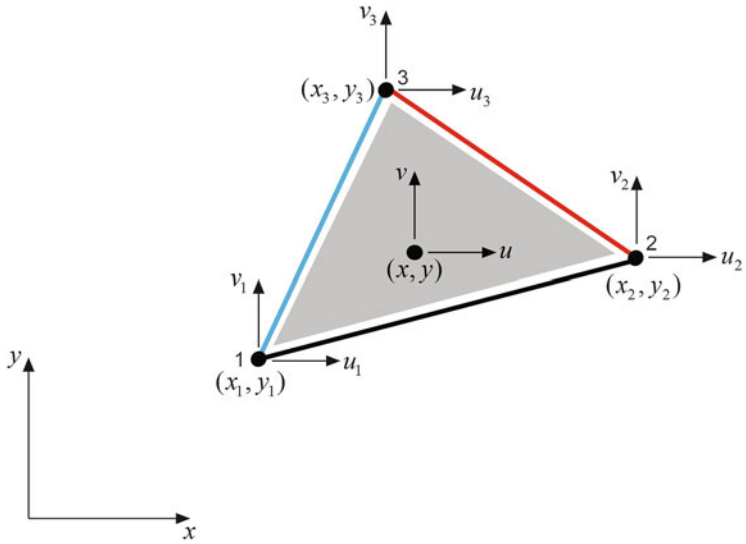

Fig. 15.2 Schematic of triangular hybrid Trefftz finite element

HFS-FEM for short. In this chapter, Voronoi convex polygons with any number of sides (n-sided convex polygons), and its performances on convergence and accuracy are numerically studied via a few benchmark problems in the context of two-dimensional linear isotropic elasticity. To generate convex polygons in various shaped geometric domains, the PolyMesher written in Matlab codes based on Voronoi diagram theory is employed by introducing related distance functions [20–37]. For each Voronoi n-sided polygonal hybrid finite element, two types of independent fields are introduced into the double-variable hybrid variational functional. One is the intra-element displacement and stress fields, which are approximated by the linear combination of fundamental solutions associated with several fictitious source points so that they can naturally satisfy the elastic equilibrium equations. Another is the auxiliary conforming element displacement field, which is defined along the element boundary and interpolated by the conventional shape functions which is the same as that in the conventional FEM [38] and boundary element method (BEM) [39–43] to enforce the conformity requirement of displacement field on the common interface of adjacent elements. The independence of the intra-element fields and the inter-element field makes us conveniently construct arbitrary n-sided polygonal elements. Moreover, the mathematical definition of the intra-element fields allows the domain integral in the hybrid functional be converted into integrals on element boundary wireframe, which are suitable for n-sided polygonal finite elements and can be easily evaluated by summing Gaussian numerical quadrature values on each segment of the element wireframe. This means that multiple types of polygonal elements with different number of sides can be used together to model a specific domain with same kernel functions, i.e., fundamental solutions in a unified form.

This is the main advantage of the present hybrid polygonal finite element over the conventional polygonal finite element with Laplace/Wachspress interpolants or Trefftz polygonal finite element with T-complete functions.

The outline of this chapter is as follows: The basic theory of Voronoi diagram is reviewed in Sect. 15.2, and then the governing equations for plane linear elasticity and the basic procedure of the Voronoi polygonal hybrid finite elements with fundamental solution kernels are described in Sect. 15.3. In Sect. 15.4, three numerical examples are presented to validate the present element formulation and assess its accuracy and convergence. Finally, some concluding remarks are drawn in Sect. 15.5.

15.2 Basics of Voronoi Polygons

Voronoi cells provide rather convenience to develop unconventional polygonal elements to discretize the computational domain due to the fact that different polygonal cells are permitted to have a different number of sides. In mathematics and computational geometry, a Voronoi polygonal cell is defined as a partition of a plane based on distance to points in a specific subset of the plane. That set of points is usually called as seeds, and for each seed there is a corresponding region consisting of all points closer to that seed than to any other.

Given an arbitrary plane domain $\Omega \subseteq \mathbb{R}^2$. Let $|\mathbf{x} - \mathbf{z}|$ denote the Euclidean distance between the point $\mathbf{x}$ and the subset $\mathbf{z}$. Given a set of seeds $\{\mathbf{z}_i\}_{i=1}^{k}$ belonging to Ω, the Voronoi polygonal cell V_i corresponding to the point $\mathbf{z}_i$ is defined by

$$V_i = \left\{\mathbf{x} \in \Omega : |\mathbf{x} - \mathbf{z}_i| \leq \left|\mathbf{x} - \mathbf{z}_j\right|, \quad \text{for all} \quad i \neq j,\ j = 1, \ldots, k\right\} \tag{15.8}$$

from which it is found that each such cell is obtained from the intersection of $(k-1)$ half-planes, and every common edge of it is defined as bisector to the line connecting two neighboring seeds. Hence it is a convex polygon and has up to $(k - 1)$ edges in the boundary.

In order to understand how the Voronoi partition works, we consider two simple cases including two and seven seeds, respectively, as indicated in Fig. 15.3. For the case of two seeds, the definition in Eq. (15.8) gives the bisector of the line connecting the two seeds and the shaded region is the Voronoi region corresponding to the seed $\mathbf{z}_1$, while, for the case of seven seeds, the bisectors of lines connecting the seed $\mathbf{z}_1$ and any other seeds $\mathbf{z}_j$ $(j \neq 1)$ meet at circumcenters of related Delaunay triangles. The connection of these centers of the circumcircles produces the Voronoi region corresponding to the seed $\mathbf{z}_1$.

As a typical application of Voronoi diagram, Fig. 15.4 shows the Voronoi polygonal cells corresponding to 200 randomly selected seeds in a square. It is clearly seen that the Voronoi diagram discretizes the square with 200 polygonal cells, which can be used to model topology change of material phase in material science.

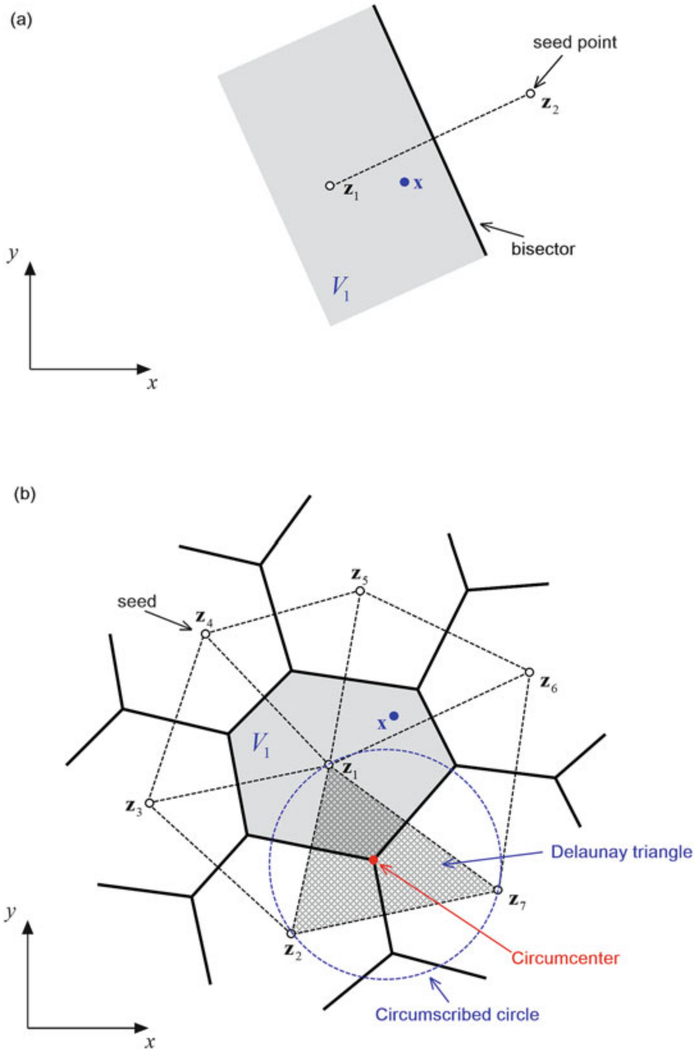

Fig. 15.3 Schematic of producing Voronoi diagram. (**a**) Voronoi region generated by two seeds. (**b**) Voronoi region generated by seven seeds

However, it is seen from Fig. 15.4 that there are some very small cells in the generated Voronoi diagram, due to the randomness of seeds in the square domain, and they are not beneficial to produce high-quality convex polygonal discretization in the computing domain, so that the solution with higher accuracy can be achieved. To improve mesh quality, the centroid Voronoi technique can be employed by iteratively setting the seed of each cell to coincide with the cell centroid [44]. Figure 15.5 shows a centroidal Voronoi tessellation in a square with

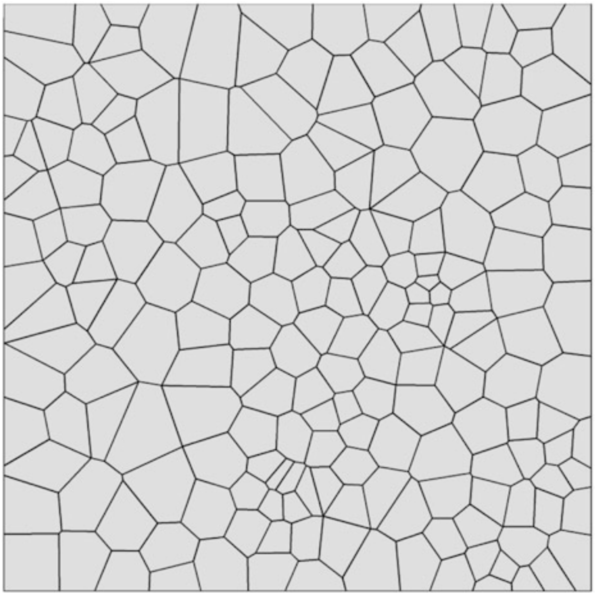

Fig. 15.4 Voronoi diagram corresponding to 200 randomly selected seeds in a square

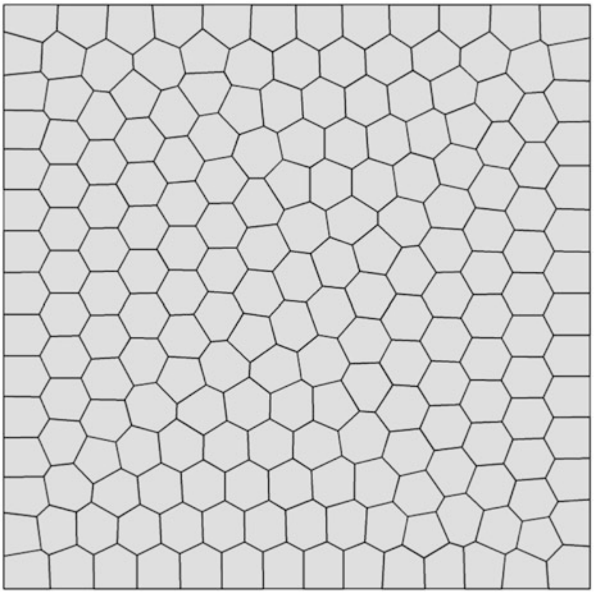

Fig. 15.5 200-cell centroidal Voronoi diagram corresponding to 200 seeds in a square

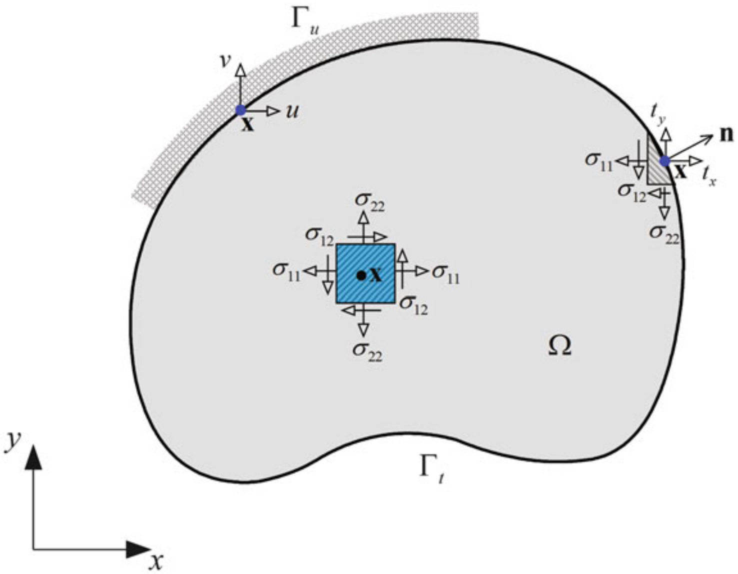

Fig. 15.6 Schematic of two-dimensional elasticity

200 seeds, which are simultaneously the generators for the Voronoi tessellation and the centroids of the Voronoi regions. It is clearly seen that the sizes of Voronoi polygonal cells are relatively uniform, which are important to obtain high-accurate finite element solutions.

15.3 Formulations of Polygonal Hybrid Finite Element

15.3.1 Governing Equations

In order to derive the element formulation, the governing equations of two-dimensional linear elasticity in isotropic and homogeneous solids are firstly reviewed to keep the content of this chapter intact.

Let us consider a two-dimensional (2D) static elasticity problem defined in the isotropic and homogeneous domain Ω bounded by its boundary $\Gamma = \Gamma_u \cup \Gamma_t$, $\Gamma_u \cap \Gamma_t = \emptyset$, where Γ_u and Γ_t are the displacement and traction boundaries, respectively. In the absence of body forces, the 2D static governing partial differential equations describing elastic small element equilibrium at a point $\mathbf{x} \in \Omega$, as indicated in Fig. 15.6, are given by [45]

$$\mathbf{L}^{\mathrm{T}}\sigma(\mathbf{x}) = 0, \qquad \mathbf{x} \in \Omega \tag{15.9}$$

where $\sigma(\mathbf{x}) = \{\sigma_x(\mathbf{x}), \sigma_y(\mathbf{x}), \tau_{xy}(\mathbf{x})\}^{\mathrm{T}}$ or $\{\sigma_{11}(\mathbf{x}), \sigma_{22}(\mathbf{x}), \sigma_{12}(\mathbf{x})\}^{\mathrm{T}}$ is the stress vector, and $\mathbf{L}$ is the strain–displacement operator matrix

$$\mathbf{L} = \begin{bmatrix} \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial y} & \frac{\partial}{\partial x} \end{bmatrix} \tag{15.10}$$

It's assumed that the rigid motion of an elastic body is constrained fully so that no displacements of particles of the body are possible without a deformation of it. Therefore, for small deformation case, the strain vector $\varepsilon(\mathbf{x}) = \{\varepsilon_x(\mathbf{x}), \varepsilon_y(\mathbf{x}), \gamma_{xy}(\mathbf{x})\}^{\mathrm{T}}$ or $\{\varepsilon_{11}(\mathbf{x}), \varepsilon_{22}\ (\mathbf{x}), \gamma_{12}(\mathbf{x})\}^{\mathrm{T}}$ at point $\mathbf{x}$ can be defined through considering the deformation of small element at point $\mathbf{x}$ as

$$\varepsilon(\mathbf{x}) = \mathbf{L}\mathbf{u}(\mathbf{x}) \tag{15.11}$$

where $\mathbf{u}(\mathbf{x}) = \{u(\mathbf{x}), v(\mathbf{x})\}^{\mathrm{T}}$ or $\{u_1(\mathbf{x}), u_2(\mathbf{x})\}^{\mathrm{T}}$ is the displacement vector.

The relations between the stress and the strain components can be described by the Hooke's law in matrix form as

$$\sigma(\mathbf{x}) = \mathbf{D}\varepsilon(\mathbf{x}) \tag{15.12}$$

where $\mathbf{D}$ stands by the constant constitutive matrix

$$\mathbf{D} = \begin{bmatrix} \lambda + 2\mu & \lambda & 0 \\ \lambda & \lambda + 2\mu & 0 \\ 0 & 0 & \mu \end{bmatrix} \tag{15.13}$$

with

$$\mu = \frac{E}{2(1+\nu)}, \qquad \lambda = \begin{cases} \dfrac{\nu E}{(1+\nu)(1-\nu)} & \text{for plane stress} \\ \dfrac{\nu E}{(1+\nu)(1-2\nu)} & \text{for plane strain} \end{cases} \tag{15.14}$$

In Eq. (15.14), ν is Poisson's ratio and E is Young's modulus of isotropic and homogeneous material.

Besides, the following displacement and traction boundary conditions prescribed on the displacement boundary Γ_u and the traction boundary Γ_t

$$\begin{aligned} \mathbf{u}(\mathbf{x}) &= \bar{\mathbf{u}}, \qquad \mathbf{x} \in \Gamma_u \\ \mathbf{s}(\mathbf{x}) &= \bar{\mathbf{s}}, \qquad \mathbf{x} \in \Gamma_s \end{aligned} \tag{15.15}$$

and the continuity condition between adjacent material constituents R_i and R_j

$$\begin{aligned} \mathbf{u}_i &= \mathbf{u}_j \\ & \qquad\qquad \text{on } R_i \cap R_j \\ \mathbf{s}_i + \mathbf{s}_j &= \mathbf{0} \end{aligned} \tag{15.16}$$

should be augmented to form a complete solving system. $\bar{\mathbf{u}}$ and $\bar{\mathbf{s}}$ are, respectively, the specified displacement and traction constraints, and the traction vector $\mathbf{s} = \{s_x(\mathbf{x}), s_y(\mathbf{x})\}^{\mathrm{T}}$ or $\{s_1(\mathbf{x}), s_2(\mathbf{x})\}^{\mathrm{T}}$ is given by

$$\mathbf{s} = \mathbf{A}\sigma \tag{15.17}$$

with

$$\mathbf{A} = \begin{bmatrix} n_1 & 0 & n_2 \\ 0 & n_2 & n_1 \end{bmatrix} \tag{15.18}$$

and $n_i \ (i = 1, 2)$ are components of the unit outward normal vector to the boundary.

Alternatively, the governing partial differential equations and boundary conditions listed above can be written in extremely concise form as

$$\sigma_{ij,j} = 0 \tag{15.19}$$

$$\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}) \tag{15.20}$$

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl} \tag{15.21}$$

and

$$\begin{aligned} lu_i(\mathbf{x}) &= \bar{u}_i, & \mathbf{x} \in \Gamma_u \\ s_i(\mathbf{x}) &= \sigma_{ij}(\mathbf{x})n_j = \bar{s}_i, & \mathbf{x} \in \Gamma_s \end{aligned} \tag{15.22}$$

where the repeated indices imply summation and the comma represents the differential operation, i.e., $u_{i,j} = \partial u_i / \partial x_j$. C_{ijkl} is the general material tensor.

15.3.2 Mesh Discretization

In the computation, the solid body is firstly divided into N Voronoi polygonal elements, which can be achieved by the so-called pre-processors, i.e., PolyMesher [46]. This is especially true for problems with complex geometries. The meshing procedure in a continuum solid generates unique numbers for all the polygonal

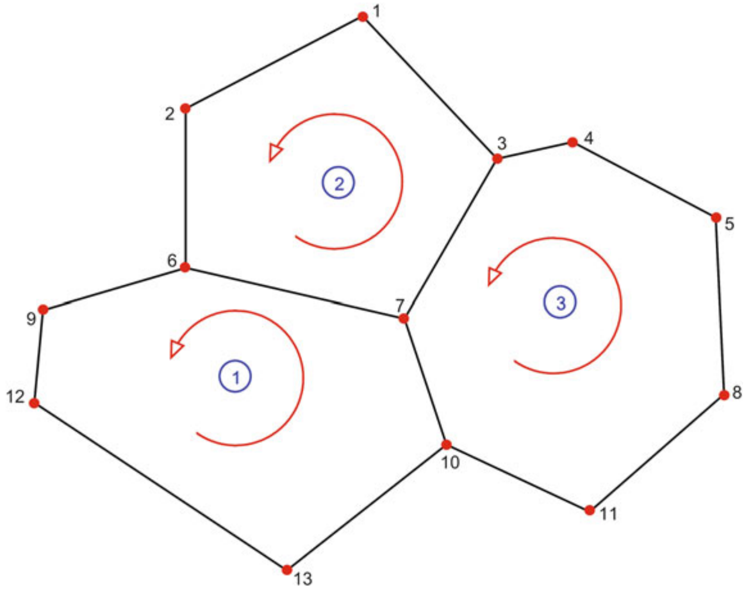

Fig. 15.7 Example of a mesh with three polygonal elements and nodes properly numbered

Table 15.1 Example of element connectivity

Elements	Nodes
①	6, 9, 12, 13, 10, 7
②	7, 3, 1, 2, 6
③	10, 11, 8, 5, 4, 3, 7

elements and nodes in a proper numbering manner. A polygonal element is formed by connecting its nodes in a pre-defined consistent fashion, i.e., anti-clockwise direction, to create the connectivity of the element. Figure 15.7 shows an example of a polygonal mesh for a two-dimensional solid, and Table 15.1 gives the corresponding nodal connectivity data. It is evident that all the three polygonal elements together form the entire computing domain of the problem without any gap or overlapping. Moreover, it is seen from Fig. 15.7 that it is possible for the domain to consist of different types of polygonal elements with different numbers of nodes, as long as they are geometrical compatible (no gaps and overlapping) on the boundaries between adjacent elements. Besides, you can freely control the density of mesh according to your requirements of accuracy and computational time. Generally, a finer mesh will yield more accurate results, but will increase the computational cost.

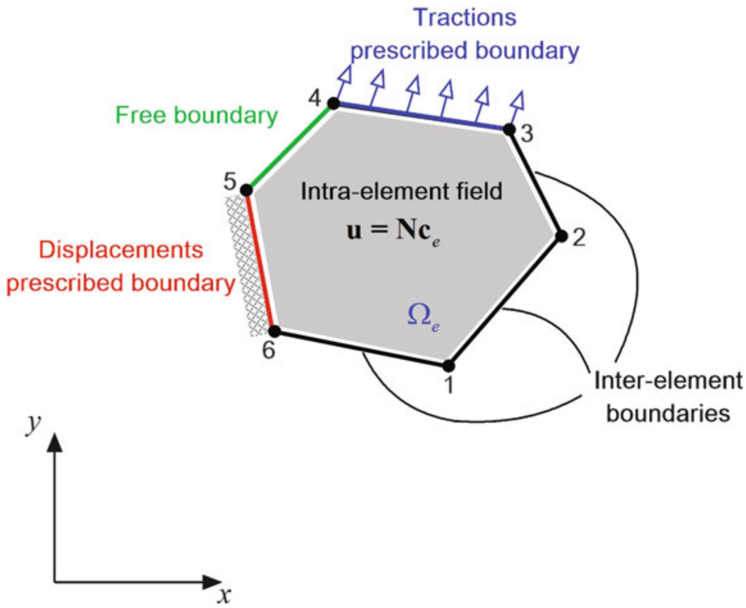

Fig. 15.8 Illustration of polygonal hybrid finite element with linear boundary

15.3.3 Displacement Interpolations

Once the domain discretization is finished, we can introduce the assumed pattern of displacement field in each polygonal element. As an example, for a typical polygonal hybrid finite element shown in Fig. 15.8, it will be assumed that the whole element domain consists of the interior domain Ω_e and the boundary Γ_e. Such partition will make it easier in introducing different patterns of the displacement field on the boundary and in the interior domain.

Within the interior of the element domain, the linear combination of displacement fundamental solutions of the problem at a set of sources is used as the approximation function to model the intra-element displacement field at any point $\mathbf{x} \in \Omega_e$, that is,

$$u_k(\mathbf{x}) = \sum_{i=1}^{m} [u_{1k}^*(\mathbf{x}, \mathbf{x}_i^s)c_1^i + u_{2k}^*(\mathbf{x}, \mathbf{x}_i^s)c_2^i] = \sum_{i=1}^{m}\sum_{l=1}^{2} u_{lk}^*(\mathbf{x}, \mathbf{x}_i^s)c_l^i, \qquad \mathbf{x} \in \Omega_e \tag{15.23}$$

where m is the number of source points distributed outside the element domain, $u_{lk}^*(\mathbf{x}, \mathbf{x}_i^s)$ represents the displacement response along the kth direction at the point $\mathbf{x}$ due to the unit force along the lth direction at the source point $\mathbf{x}_i^s$, and c_l^i is the source intensity along the lth direction at the source point $\mathbf{x}_i^s$. Besides, u_1 and u_2 stand by the displacement component u and v in the study, respectively.

For plane strain problems, the induced displacement solutions, the so-called Kelvin solutions, at $\mathbf{x}$ caused by the l-direction unit force at $\mathbf{x}_i^s$, can be written as [40, 47]

$$u_{lk}^*(\mathbf{x}, \mathbf{x}_i^s) = \frac{1}{8\pi G(1-\nu)}\left[(3-4\nu)\,\delta_{lk}\ln\frac{1}{r} + r_{,l}r_{,k}\right] \tag{15.24}$$

where r stands for the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_i^s$, and

$$r_{,l} = \frac{\partial r}{\partial x_l} \tag{15.25}$$

Besides, as indicated in Eq. (15.23), the source points placed outside the element are needed for the intra-element displacement approximation. This can be done by distributing these source points on the pseudo-boundary geometrically similar to the element physical boundary Γ_e, as well done in the classic method of fundamental solution [48–55], and their locations on the pseudo-boundary can be generated by a simple geometrical expression with a dimensionless parameter $\gamma > 0$ [56, 57]

$$\mathbf{x}^s = \mathbf{x}_b + \gamma(\mathbf{x}_b - \mathbf{x}_c) \tag{15.26}$$

where $\mathbf{x}^s,\ \mathbf{x}_b,\ \mathbf{x}_c$ represent the coordinates of source point, node, and center of the element, respectively.

For convenience, we rewrite Eq. (15.23) in the following matrix form:

$$\mathbf{u}(\mathbf{x}) = \sum_{i=1}^{m}\mathbf{N}_i(\mathbf{x})\mathbf{c}_i = \mathbf{N}_e(\mathbf{x})\mathbf{c}_e, \qquad \mathbf{x}\in\Omega_e \tag{15.27}$$

where

$$\mathbf{c}_e = \{\,\mathbf{c}_1{}^{\mathrm{T}}\ \mathbf{c}_2{}^{\mathrm{T}}\ \cdots\ \mathbf{c}_m{}^{\mathrm{T}}\,\}^{\mathrm{T}} \tag{15.28}$$

is a column vector of $2\times m$ undetermined coefficients,

$$\mathbf{N}_e(\mathbf{x}) = [\,\mathbf{N}_1(\mathbf{x})\ \mathbf{N}_2(\mathbf{x})\ \cdots\ \mathbf{N}_m(\mathbf{x})\,] \tag{15.29}$$

denotes a matrix containing displacement fundamental solutions of coordinates $\mathbf{x}$, and

$$\mathbf{N}_i(\mathbf{x}) = \begin{bmatrix} u_{11}^*(\mathbf{x}, \mathbf{x}_i^s) & u_{21}^*(\mathbf{x}, \mathbf{x}_i^s) \\ u_{12}^*(\mathbf{x}, \mathbf{x}_i^s) & u_{22}^*(\mathbf{x}, \mathbf{x}_i^s) \end{bmatrix}, \qquad \mathbf{c}_i = \{\,c_1^i,\ c_2^i\,\}^{\mathrm{T}} \tag{15.30}$$

Furthermore, the corresponding strain and stress fields can be derived by considering the strain–displacement relation in (15.11) and the stress–strain relation in (15.12)

$$\varepsilon(\mathbf{x}) = \mathbf{S}_e(\mathbf{x})\mathbf{c}_e \tag{15.31}$$

$$\sigma(\mathbf{x}) = \mathbf{T}(\mathbf{x})\mathbf{c}_e \tag{15.32}$$

where

$$\mathbf{S}_e(\mathbf{x}) = \mathbf{L}\mathbf{N}_e(\mathbf{x}) \tag{15.33}$$

and

$$\mathbf{T}_e(\mathbf{x}) = \mathbf{D}\mathbf{S}_e(\mathbf{x}) \tag{15.34}$$

respectively, denote the induced strain and stress matrices that consist of fundamental solutions of the strain and stress fields.

Besides, from Eqs. (15.17) and (15.32), the column vector of boundary traction can be expressed as

$$\mathbf{t} = \mathbf{A}\sigma = \mathbf{A}\mathbf{T}_e\mathbf{c}_e = \mathbf{Q}_e\mathbf{c}_e \tag{15.35}$$

It's evident that the intra-element stress field (15.32) can naturally satisfy the linear elastic governing partial differential equations (15.9) for the sake of the physical definition of fundamental solutions. This attractive feature is beneficial to simplify the hybrid functional below.

However, the intra-element displacement field defined within the interior of the element given by Eq. (15.27) is non-conforming across the inter-element boundary. To deal with such problem, the hybrid technique popularly used in the hybrid finite element method pioneered by Pian [58–62] is employed to introduce an auxiliary conforming displacement field along the element boundaries. Here, the independent boundary displacement field is interpolated using the displacement vector $\mathbf{d}_e$ at the nodes of the element as

$$\tilde{\mathbf{u}}(\mathbf{x}) = \tilde{\mathbf{N}}_e(\mathbf{x})\mathbf{d}_e, \qquad \mathbf{x} \in \Gamma_e \tag{15.36}$$

where the elements of the interpolation matrix $\tilde{\mathbf{N}}_e$ are functions of the element boundary coordinates. For the two-dimensional case under consideration, the boundary displacement interpolation matrix $\tilde{\mathbf{N}}_e$ consists of the standard one-dimensional shape functions as used in the conventional one-dimensional bar element. For example, for the six-sided polygonal hybrid finite element including 1, 2, 3, 4, 5, and 6 nodes, as indicated in Fig. 15.8, when $\mathbf{x}$ locates at its second side connecting the global nodes 2 and 3, the matrix $\tilde{\mathbf{N}}_e$ and the vector $\mathbf{d}_e$ can be written as

$$\tilde{\mathbf{N}}_e = \begin{bmatrix} \mathbf{0} & \tilde{\mathbf{N}}_1 & \tilde{\mathbf{N}}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{15.37}$$

$$\mathbf{d}_e = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & \mathbf{u}_5 & \mathbf{u}_6 \end{bmatrix}^{\mathrm{T}} \tag{15.38}$$

where the elements in the interpolation matrix $\tilde{\mathbf{N}}_e$ are 2×2 submatrix, i.e.,

$$\tilde{\mathbf{N}}_1 = \begin{bmatrix} \tilde{N}_1(\xi) & 0 \\ 0 & \tilde{N}_1(\xi) \end{bmatrix}, \qquad \tilde{\mathbf{N}}_2 = \begin{bmatrix} \tilde{N}_2(\xi) & 0 \\ 0 & \tilde{N}_2(\xi) \end{bmatrix} \tag{15.39}$$

and

$$\tilde{N}_1(\xi) = \frac{1-\xi}{2}, \qquad \tilde{N}_2(\xi) = \frac{1+\xi}{2} \qquad (-1 \le \xi \le 1) \tag{15.40}$$

represent the standard one-dimensional linear shape functions. In Eq. (15.40), ξ is a natural coordinate system defined along the side connecting the global nodes 2 and 3. Besides, in Eq. (15.38), $\mathbf{u}_i{}^{\mathrm{T}} = \{u_i, v_i\}$ is the nodal displacement vector.

For the isoparametric element formulation, the linear shape functions $\tilde{N}_1(\xi)$ and $\tilde{N}_2(\xi)$ in Eq. (15.40) are also used for the element geometry interpolation through the following relation:

$$\mathbf{x}(\xi) = \tilde{N}_1(\xi)\mathbf{x}_2 + \tilde{N}_2(\xi)\mathbf{x}_3 \tag{15.41}$$

which maps the natural ξ coordinate of any point in the mapped element to the actual $\mathbf{x}$ coordinate of the point in the physical element, as indicated in Fig. 15.9.

Additionally, as illustrated above, the linear shape functions are employed for the approximation of inter-element displacement distribution on each element side. Actually, the quadratic shape functions can also be employed for inter-element

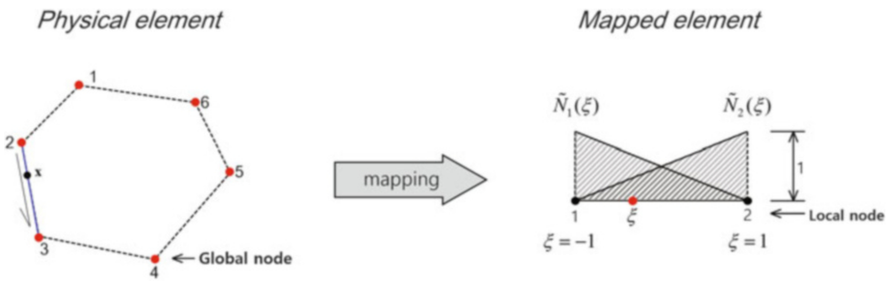

Fig. 15.9 Geometry interpolation from one side of the physical element to the mapped line element

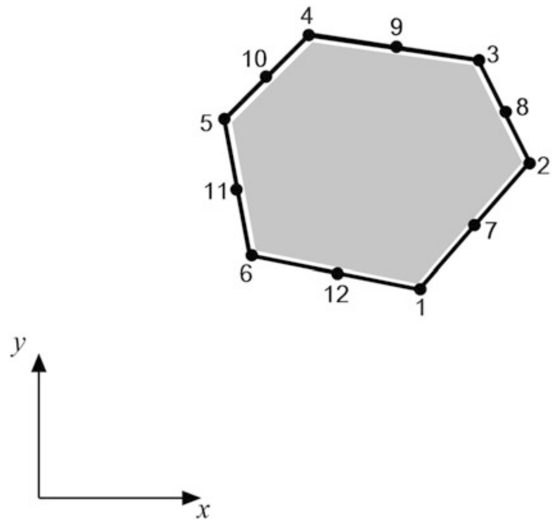

Fig. 15.10 Illustration of polygonal hybrid finite element with quadratic boundary

displacement interpolation. This means that there are three nodes for each element side. As an example, for the element shown in Fig. 15.10 containing 6 quadratic sides and 12 nodes, the shape function matrix $\tilde{\mathbf{N}}_e$ and the nodal vector $\mathbf{d}_e$ over the second side consisting of nodes 2, 8, and 3 can be written as

$$\tilde{\mathbf{N}}_e = \left[\mathbf{0}\ \tilde{\mathbf{N}}_1\ \tilde{\mathbf{N}}_3\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\ \tilde{\mathbf{N}}_2\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\ \mathbf{0} \right] \tag{15.42}$$

$$\mathbf{d}_e = \left[\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3\ \mathbf{u}_4\ \mathbf{u}_5\ \mathbf{u}_6\ \mathbf{u}_7\ \mathbf{u}_8\ \mathbf{u}_9\ \mathbf{u}_{10}\ \mathbf{u}_{11}\ \mathbf{u}_{12} \right]^{\mathrm{T}} \tag{15.43}$$

where each element in the interpolation matrix $\tilde{\mathbf{N}}_e$ is 2×2 submatrix, i.e.,

$$\tilde{\mathbf{N}}_1 = \begin{bmatrix} \tilde{N}_1(\xi) & 0 \\ 0 & \tilde{N}_1(\xi) \end{bmatrix}, \quad \tilde{\mathbf{N}}_2 = \begin{bmatrix} \tilde{N}_2(\xi) & 0 \\ 0 & \tilde{N}_2(\xi) \end{bmatrix}, \quad \tilde{\mathbf{N}}_3 = \begin{bmatrix} \tilde{N}_3(\xi) & 0 \\ 0 & \tilde{N}_3(\xi) \end{bmatrix} \tag{15.44}$$

with

$$\tilde{N}_1(\xi) = -\frac{\xi(1-\xi)}{2}, \qquad \tilde{N}_2(\xi) = 1 - \xi^2, \qquad \tilde{N}_3(\xi) = \frac{\xi(1+\xi)}{2} \tag{15.45}$$

Correspondingly, the geometry interpolation over this element side can be given through the following relation:

$$\mathbf{x}(\xi) = \tilde{N}_1(\xi)\mathbf{x}_2 + \tilde{N}_2(\xi)\mathbf{x}_8 + \tilde{N}_3(\xi)\mathbf{x}_3 \tag{15.46}$$

15.3.4 Double-Variable Hybrid Functional

In mechanics of solids, our problem is to determine the displacement distribution of the body shown in Fig. 15.6, which should satisfy the governing equations (15.9), (15.11), (15.12) and the boundary conditions (15.15). Then strains and stresses can be subsequently determined because they are related to displacements. This leads to requiring solutions of the second-order partial differential equations. However, the exact solutions to such problem are only available for simple geometries and loading conditions [63–69]. For problems of complex geometries and general boundary and loading conditions, obtaining exact solutions is an almost impossible task. For such case, finding approximate solutions is popularly performed and can usually be achieved through energy (variational) methods or weighted residual methods, which require a weaker continuity on the field variables and generally are written in an integral form. The typical application of weak-form method is the finite element method. However, the stationary conditions of the traditional potential or complementary variational functional cannot guarantee satisfaction of the inter-element continuity conditions. To meet the requirement of inter-element continuity, a modified variational functional based on two independent field variables is defined in integral form

$$\Pi_m = \sum_e \Pi_{me} \tag{15.47}$$

where

$$\Pi_{me} = \int_{\Omega_e} U_\varepsilon \mathrm{d}\Omega - \int_{\Gamma_e^s} \bar{s}_i \tilde{u}_i \mathrm{d}\Gamma + \int_{\Gamma_e} s_i \left(\tilde{u}_i - u_i\right) \mathrm{d}\Gamma \tag{15.48}$$

is the elementary variational functional and U_ε is the strain energy per unit volume defined as

$$U_\varepsilon = \frac{1}{2}\sigma_{ij}\varepsilon_{ij} \tag{15.49}$$

In Eq. (15.47), the governing equations (15.9), (15.11), (15.12) hold true, a priori, within the element domain due to the use of the fundamental solutions as intra-element trial functions, and the boundary displacement satisfies the essential boundary conditions. The similar principle to construct variational functional with two independent variables can be found for potential problems [13, 20] and elastic problems [22, 23]. As indicated in Fig. 15.8, the boundary Γ_e of a particular polygonal element e consists of the following parts:

$$\Gamma_e = \Gamma_e^u \cup \Gamma_e^s \cup \Gamma_e^I \tag{15.50}$$

where $\Gamma_e^u = \Gamma_e \cap \Gamma_u$ is the prescribed displacement boundary, $\Gamma_e^s = \Gamma_e \cap \Gamma_s$ is the prescribed traction boundary, and Γ_e^I stands for the inter-element boundary of the element "*e*."

15.3.4.1 Stationary Condition of the Proposed Variational Functional

Next we will show that the stationary condition of the elementary functional (15.48) leads to the equilibrium equation (15.9), the boundary conditions (15.15), and the continuity conditions (15.16) between elements.

For this purpose, Hamilton's principle can be used as a simple yet powerful tool to derive discretized system equations. It states simply that of all the admissible displacement the most accurate solution makes the static Lagrangian functional a minimum. Here, an admissible displacement refers to that satisfying the displacement compatible (continuous) requirement in the problem domain and the prescribed displacement constraints.

Therefore, in mathematics, Hamilton's principle states

$$\delta \Pi_{me} = 0 \tag{15.51}$$

The first-order variational of Eq. (15.51) yields

$$\delta \Pi_{me} = \int_{\Omega_e} \delta U_\varepsilon \mathrm{d}\Omega - \int_{\Gamma_e^s} \bar{s}_i \delta \tilde{u}_i \mathrm{d}\Gamma + \int_{\Gamma_e} \delta s_i \left(\tilde{u}_i - u_i\right) \mathrm{d}\Gamma + \int_{\Gamma_e} s_i \left(\delta \tilde{u}_i - \delta u_i\right) \mathrm{d}\Gamma \tag{15.52}$$

in which the first term is given as

$$\begin{aligned} \int_{\Omega_e} \delta U_\varepsilon \mathrm{d}\Omega &= \int_{\Omega_e} \sigma_{ij} \delta \varepsilon_{ij} \mathrm{d}\Omega = \int_{\Omega_e} \sigma_{ij} \delta u_{i,j} \mathrm{d}\Omega \\ &= \int_{\Omega_e} \left(\sigma_{ij} \delta u_i\right)_{,j} \mathrm{d}\Gamma - \int_{\Omega_e} \sigma_{ij,j} \delta u_i \mathrm{d}\Omega \end{aligned} \tag{15.53}$$

Using the Gaussian theorem

$$\int_{\Omega} f_{,i} \mathrm{d}\Omega = \int_{\Gamma} f n_i \mathrm{d}\Gamma \tag{15.54}$$

we have

$$\int_{\Omega_e} \left(\sigma_{ij} \delta u_i\right)_{,j} \mathrm{d}\Gamma = \int_{\Gamma_e} s_i \delta u_i \mathrm{d}\Gamma \tag{15.55}$$

Thus, Eq. (15.53) can be written as

$$\int_{\Omega_e} \delta U_\varepsilon \mathrm{d}\Omega = \int_{\Gamma_e} s_i \delta u_i \mathrm{d}\Gamma - \int_{\Omega_e} \sigma_{ij,j} \delta u_i \mathrm{d}\Omega \tag{15.56}$$

Substituting Eq. (15.56) into Eq. (15.52) gives

$$\delta\Pi_{me} = -\int_{\Omega_e} \sigma_{ij,j} \delta u_i \mathrm{d}\Omega - \int_{\Gamma_e^s} \bar{s}_i \delta\tilde{u}_i \mathrm{d}\Gamma + \int_{\Gamma_e} \delta s_i \left(\tilde{u}_i - u_i\right) \mathrm{d}\Gamma + \int_{\Gamma_e} s_i \delta\tilde{u}_i \mathrm{d}\Gamma \tag{15.57}$$

For the proposed method, the admissible boundary displacement $\delta\tilde{u}_i$ satisfies the displacement conformity in advance, that is,

$$\begin{aligned} \delta\tilde{u}_i &= 0 \quad &\text{on } \Gamma_e^u \quad &(\tilde{u}_i = \bar{u}_i) \\ \delta\tilde{u}_i^e &= \delta\tilde{u}_i^f \quad &\text{on } \Gamma_{ef}^I \quad &(\tilde{u}_i^e = \tilde{u}_i^f) \end{aligned} \tag{15.58}$$

then, Eq. (15.57) can be rewritten as

$$\begin{aligned} \delta\Pi_{me} = &-\int_{\Omega_e} \sigma_{ij,j} \delta u_i \mathrm{d}\Omega + \int_{\Gamma_e^s} \left(s_i - \bar{s}_i\right) \delta\tilde{u}_i \mathrm{d}\Gamma + \int_{\Gamma_e^I} s_i \delta\tilde{u}_i \mathrm{d}\Gamma \\ &+ \int_{\Gamma_e} \delta s_i \left(\tilde{u}_i - u_i\right) \mathrm{d}\Gamma \end{aligned} \tag{15.59}$$

from which the governing equation in the domain Ω_e and boundary conditions on Γ_e^s can be obtained

$$\begin{aligned} \sigma_{ij,j} &= 0 \quad &\text{in } \Omega_e \\ s_i &= \sigma_{ij} n_j = \bar{s}_i \quad &\text{on } \Gamma_e^s \\ \tilde{u}_i &= u_i \quad &\text{on } \Gamma_e^u \end{aligned} \tag{15.60}$$

by using the stationary condition $\delta\Pi_{me} = 0$ and the arbitrariness of quantities δu_i, $\delta\tilde{u}_i$, and δs_i.

As to the continuity requirement between the two adjacent elements "*e*" and "*f*" given in Eq. (15.16), we can obtain it in the following way. When assembling elements "*e*" and "*f*," we have

$$\begin{aligned} \delta\Pi_{m(e+f)} = &-\int_{\Omega_e+\Omega_f} \sigma_{ij,j} \delta u_i \mathrm{d}\Omega + \int_{\Gamma_e^s+\Gamma_f^s} \left(s_i - \bar{s}_i\right) \delta\tilde{u}_i \mathrm{d}\Gamma \\ &+ \int_{\Gamma_{ef}^I} \left(s_{ie} + s_{if}\right) \delta\tilde{u}_i \mathrm{d}\Gamma + \int_{\Gamma_e+\Gamma_f} \delta s_i \left(\tilde{u}_i - u_i\right) \mathrm{d}\Gamma \end{aligned} \tag{15.61}$$

from which the vanishing variation of $\Pi_{m(e+f)}$ leads to the continuity condition $s_{ie} + s_{if} = 0$ on the inter-element boundary Γ^I_{ef}.

15.3.4.2 Theorem on the Existence of Extremum

If the expression

$$\int_{\Omega} \delta^2 U_\varepsilon \mathrm{d}\Omega - \sum_e \left[\int_{\Gamma_e} \delta s_{ie}(\delta\tilde{u}_{ie} - \delta u_{ie})\mathrm{d}\Gamma + \int_{\Gamma^I_e} \delta s_{ie}\delta\tilde{u}_{ie}\mathrm{d}\Gamma \right] \tag{15.62}$$

is uniformly positive (or negative) in the neighborhood of u_{i0}, where the displacement u_{i0} has such a value that $\Pi_m(u_{i0}) = (\Pi_m)_0$, and $(\Pi_m)_0$ stands for the stationary value of Π_m, we have

$$\Pi_m \geq (\Pi_m)_0 \qquad [\text{or } \Pi_m \leq (\Pi_m)_0] \tag{15.63}$$

in which the relation that $\tilde{u}_{ie} = \tilde{u}_{if}$ is identical on $\Gamma_e \cap \Gamma_f$ has been used. This is due to the definition of continuity condition in Eq. (15.16).

Proof The proof of the theorem on the existence of extremum may be completed by way of the so-called second variational approach [23, 70]. In doing this, performing variation of $\delta\Pi_m$ and using the constrained conditions, we find

$$\begin{aligned} \delta^2\Pi_m &= \int_{\Omega} \delta^2 U_\varepsilon \mathrm{d}\Omega - \sum_e \left[\int_{\Gamma_e} \delta s_{ie}(\delta\tilde{u}_{ie} - \delta u_{ie})\mathrm{d}\Gamma + \int_{\Gamma^I_e} \delta s_{ie}\delta\tilde{u}_{ie}\mathrm{d}\Gamma \right] \\ &= \text{expression (15.62)} \end{aligned} \tag{15.64}$$

Therefore, the theorem has been proved from the sufficient condition of the existence of a local extreme of a functional. This completes the proof.

15.3.5 Formation of Resulting Linear Equations

15.3.5.1 Element Equations

Once the two independent displacement fields, which are, respectively, defined within the interior and along the boundary of the element domain, are assumed, the hybrid finite element equation for a polygonal element can be formulated using the following process.

The element double-variable hybrid functional in Eq. (15.48) can be written in concise form as

$$\Pi_{me} = \frac{1}{2}\int_{\Omega_e} \sigma_{ij}\varepsilon_{ij}\mathrm{d}\Omega - \int_{\Gamma_e^s} \bar{s}_i\tilde{u}_i\mathrm{d}\Gamma + \int_{\Gamma_e} s_i\left(\tilde{u}_i - u_i\right)\mathrm{d}\Gamma \tag{15.65}$$

By substituting for the strain component from the element strain–displacement relationship in Eq. (15.20) into the first integral in the right-hand side of Eq. (15.65), we have

$$\Pi_{me} = \frac{1}{4}\int_{\Omega_e} \sigma_{ij}(u_{i,j} + u_{j,i})\mathrm{d}\Omega - \int_{\Gamma_e^s} \bar{s}_i\tilde{u}_i\mathrm{d}\Gamma + \int_{\Gamma_e} s_i\left(\tilde{u}_i - u_i\right)\mathrm{d}\Gamma \tag{15.66}$$

Considering the symmetry of stress components, that is, $\sigma_{ij} = \sigma_{ji}$, we have

$$\sigma_{ij}u_{i,j} = \sigma_{ij}u_{j,i} \tag{15.67}$$

Then Eq. (15.66) is further expressed as

$$\Pi_{me} = \frac{1}{2}\int_{\Omega_e} \sigma_{ij}u_{i,j}\mathrm{d}\Omega - \int_{\Gamma_e^s} \bar{s}_i\tilde{u}_i\mathrm{d}\Gamma + \int_{\Gamma_e} s_i\left(\tilde{u}_i - u_i\right)\mathrm{d}\Gamma \tag{15.68}$$

Using the Gaussian theorem given in Eq. (15.54), we get

$$\begin{aligned}\frac{1}{2}\int_{\Omega_e} \sigma_{ij}u_{i,j}\mathrm{d}\Omega &= \frac{1}{2}\int_{\Omega_e} \left(\sigma_{ij}u_i\right)_{,j}\mathrm{d}\Omega - \frac{1}{2}\int_{\Omega_e} \sigma_{ij,j}u_i\mathrm{d}\Omega \\ &= \frac{1}{2}\int_{\Gamma_e} \sigma_{ij}u_i n_j\mathrm{d}\Omega - \frac{1}{2}\int_{\Omega_e} \sigma_{ij,j}u_i\mathrm{d}\Omega\end{aligned} \tag{15.69}$$

Remembering the relationship in Eq. (15.22), we further have

$$\frac{1}{2}\int_{\Omega_e} \sigma_{ij}u_{i,j}\mathrm{d}\Omega = \frac{1}{2}\int_{\Gamma_e} s_i u_i\mathrm{d}\Omega - \frac{1}{2}\int_{\Omega_e} \sigma_{ij,j}u_i\mathrm{d}\Omega \tag{15.70}$$

Because of the natural satisfaction of the equilibrium equations by the assumed intra-element displacement field inside the element domain, the domain integral in Eq. (15.70) can be really removed. Hence, we finally have

$$\frac{1}{2}\int_{\Omega_e} \sigma_{ij}u_{i,j}\mathrm{d}\Omega = \frac{1}{2}\int_{\Gamma_e} s_i u_i\mathrm{d}\Omega \tag{15.71}$$

Substituting Eq. (15.71) into Eq. (15.68) can produce

$$\begin{aligned}\Pi_{me} &= \frac{1}{2}\int_{\Gamma_e} s_i u_i \mathrm{d}\Omega - \int_{\Gamma_e^s} \bar{s}_i \tilde{u}_i \mathrm{d}\Gamma + \int_{\Gamma_e} s_i \left(\tilde{u}_i - u_i\right) \mathrm{d}\Gamma \\ &= -\frac{1}{2}\int_{\Gamma_e} s_i u_i \mathrm{d}\Omega - \int_{\Gamma_e^s} \bar{s}_i \tilde{u}_i \mathrm{d}\Gamma + \int_{\Gamma_e} s_i \tilde{u}_i \mathrm{d}\Gamma\end{aligned} \tag{15.72}$$

which can be rewritten in matrix form as

$$\Pi_{me} = -\frac{1}{2}\int_{\Gamma_e} \mathbf{s}^{\mathrm{T}}\mathbf{u}\mathrm{d}\Gamma - \int_{\Gamma_e^t} \bar{\mathbf{s}}^{\mathrm{T}}\tilde{\mathbf{u}}\mathrm{d}\Gamma + \int_{\Gamma_e} \mathbf{s}^{\mathrm{T}}\tilde{\mathbf{u}}\mathrm{d}\Gamma \tag{15.73}$$

Subsequently, the substitution of the intra-element displacement field (15.27), the induced traction field (15.35), and the inter-element displacement field (15.36) into the functional (15.73) yields

$$\Pi_{me} = -\frac{1}{2}\mathbf{c}_e{}^{\mathrm{T}}\mathbf{H}_e\mathbf{c}_e - \mathbf{d}_e{}^{\mathrm{T}}\mathbf{g}_e + \mathbf{c}_e{}^{\mathrm{T}}\mathbf{G}_e\mathbf{d}_e \tag{15.74}$$

where

$$\begin{aligned}\mathbf{H}_e &= \int_{\Gamma_e} \mathbf{Q}_e{}^{\mathrm{T}}\mathbf{N}_e\mathrm{d}\Gamma \\ \mathbf{G}_e &= \int_{\Gamma_e} \mathbf{Q}_e{}^{\mathrm{T}}\tilde{\mathbf{N}}_e\mathrm{d}\Gamma \\ \mathbf{g}_e &= \int_{\Gamma_e^s} \tilde{\mathbf{N}}_e^{\mathrm{T}}\bar{\mathbf{s}}\mathrm{d}\Gamma\end{aligned} \tag{15.75}$$

In the practical numerical implementation, the matrices $\mathbf{H}_e$ and $\mathbf{G}_e$ can be evaluated by means of side-by-side Gaussian quadrature over the entire element boundary, while the element equivalent nodal force vector $\mathbf{g}_e$ representing the contribution of prescribed external tractions on the element side can be evaluated by the same numerical quadrature scheme just along this side, that is,

$$\begin{aligned}\mathbf{H}_e &= \sum_{s=1}^{ns}\left[\int_{-1}^{1} \mathbf{Q}^{\mathrm{T}}(\mathbf{x}(\xi))\mathbf{N}(\mathbf{x}(\xi))J(\xi)\mathrm{d}\xi\right] = \sum_{s=1}^{ns}\sum_{k=1}^{ng} w_k\mathbf{Q}^{\mathrm{T}}(\mathbf{x}(\xi_k))\mathbf{N}(\mathbf{x}(\xi_k))J(\xi_k) \\ \mathbf{G}_e &= \sum_{s=1}^{ns}\left[\int_{-1}^{1} \mathbf{Q}^{\mathrm{T}}(\mathbf{x}(\xi))\tilde{\mathbf{N}}(\mathbf{x}(\xi))J(\xi)\mathrm{d}\xi\right] = \sum_{s=1}^{ns}\sum_{k=1}^{ng} w_k\mathbf{Q}^{\mathrm{T}}(\mathbf{x}(\xi_k))\tilde{\mathbf{N}}(\mathbf{x}(\xi_k))J(\xi_k) \\ \mathbf{g}_e &= \int_{-1}^{1} \tilde{\mathbf{N}}^{\mathrm{T}}(\mathbf{x}(\xi))\bar{\mathbf{t}}J(\xi)\mathrm{d}\xi = \sum_{k=1}^{ng} w_k\tilde{\mathbf{N}}^{\mathrm{T}}(\mathbf{x}(\xi_k))\bar{\mathbf{t}}J(\xi_k)\end{aligned} \tag{15.76}$$

where ns is the number of sides of polygonal element, ng is the number of Gaussian quadrature points on each element side, w_k is the weighting factors for the kth Gaussian points ξ_k, and $J(\xi_k)$ is the corresponding Jacobian coefficient which can be evaluated by the geometrical mapping equation given in Eqs. (15.41) or (15.46)

$$J(\xi) = \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}\xi}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}\xi}\right)^2} = \sqrt{\left(\sum_{i=1}^{nn} \frac{\mathrm{d}\tilde{N}_i}{\mathrm{d}\xi} x_i\right)^2 + \left(\sum_{i=1}^{nn} \frac{\mathrm{d}\tilde{N}_i}{\mathrm{d}\xi} x_i\right)^2} \tag{15.77}$$

where nn is the number of nodes on each side of the element.

The minimization of the functional Π_{me} in Eq. (15.74) with respect to $\mathbf{c}_e$ and $\mathbf{d}_e$ yields

$$\begin{aligned} \frac{\partial \Pi_{me}}{\partial \mathbf{c}_e{}^{\mathrm{T}}} &= -\mathbf{H}_e \mathbf{c}_e + \mathbf{G}_e \mathbf{d}_e = \mathbf{0} \\ \frac{\partial \Pi_{me}}{\partial \mathbf{d}_e{}^{\mathrm{T}}} &= \mathbf{G}_e{}^{\mathrm{T}} \mathbf{c}_e - \mathbf{g}_e = \mathbf{0} \end{aligned} \tag{15.78}$$

from which we can obtain the element stiffness equation

$$\mathbf{k}_e \mathbf{d}_e = \mathbf{g}_e \tag{15.79}$$

and the optional relationship of $\mathbf{c}_e$ and $\mathbf{d}_e$

$$\mathbf{c}_e = \mathbf{H}_e{}^{-1} \mathbf{G}_e \mathbf{d}_e \tag{15.80}$$

where

$$\mathbf{k}_e = \mathbf{G}_e{}^{\mathrm{T}} \mathbf{H}_e{}^{-1} \mathbf{G}_e \tag{15.81}$$

It is necessary to note that in the element equation (15.79), the computation of the right term is same as that in the conventional finite element formulation. Besides, since the element matrix $\mathbf{H}_e$ is symmetric, the caused element stiffness matrix $\mathbf{k}_e$ in Eq. (15.81) keeps symmetric too.

More importantly, from the procedure described above, the evaluations of both $\mathbf{H}_e$ and $\mathbf{G}_e$ involve the element boundary integrals only. The employment of fundamental solutions of the problem in the intra-element displacement field can directly convert the domain integral in the hybrid functional into the boundary line integrals, which obviously decreases the complexity of the integration by one dimension and can be more easily evaluated in practice than the domain integrals, especially in the natural polygonal finite elements [36, 71]. Furthermore, the limitation of number of element sides is entirely removed. Such feature allows for greater flexibility in constructing polygonal elements of arbitrary shapes with the same kernel functions, the fundamental solutions of problem here, for discretizing complex geometry domains. This means that we can establish a family of n-sided

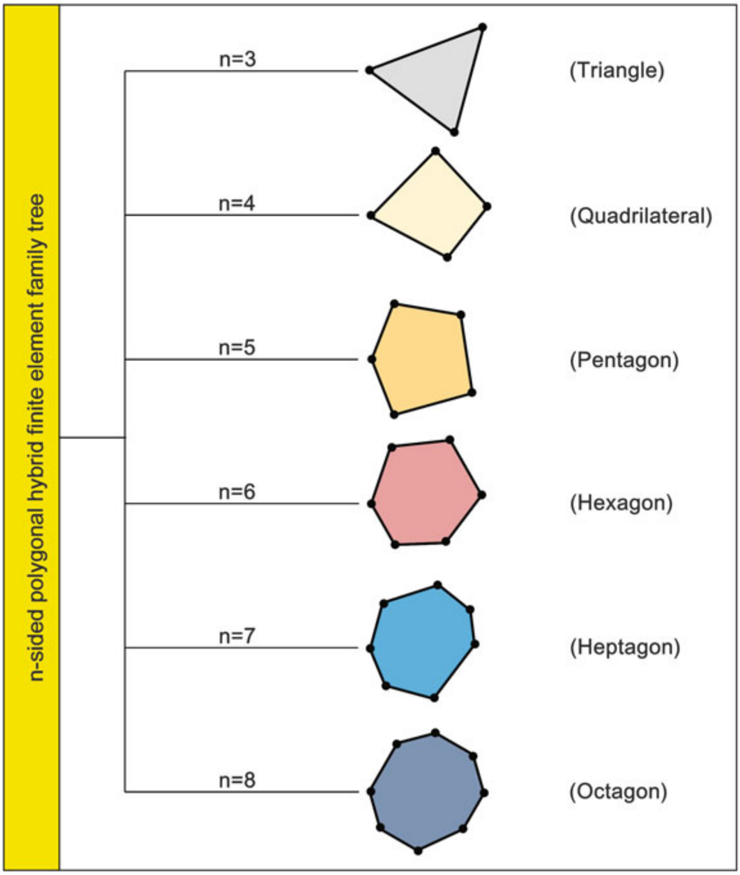

Fig. 15.11 Schematic of *n*-sided polygonal hybrid finite element family tree

polygonal hybrid finite element ($n \geq 3$) in a unified form for computational application, as shown in Fig. 15.11. But we need to note that the element in Fig. 15.11 can be modified to that with quadratic side consisting of three nodes, according to your needs.

15.3.5.2 Assembly of Global Equation

The element stiffness equations (15.79) for all the individual elements can be assembled together to form the global stiffness equation:

$$\mathbf{KD} = \mathbf{F} \tag{15.82}$$

where **K** is the global stiffness matrix, which is sparse and symmetric in practice. **D** is a vector of all the displacements at all the nodes in the entire problem domain, and **F** is a vector of the total equivalent nodal force vector.

This process of assembly is one of simply adding up the contributions from all the elements connected at a node, as done in the conventional finite element procedure. For example, for the system consisting of three polygonal elements given in Fig. 15.7, the node 6 corresponding to the global degrees of freedom 11 and 12 occupies the rows 11 and 12 and the columns 11 and 12 in the global stiffness matrix **K**, while it is the starting node of the six-sided polygonal element 1 so it has the local degrees of freedom 1, 2. Similarly, for the five-sided polygonal element 2, it has the local degrees of freedom 9, 10, as indicated in element connectivity Table 15.1, so we can simply add the corresponding elements in the element matrices to produce the counterpart in the global stiffness matrix

$$\begin{bmatrix} k_{1,1}^1 & k_{1,2}^1 \\ k_{1,2}^1 & k_{2,2}^1 \end{bmatrix} + \begin{bmatrix} k_{9,9}^2 & k_{9,10}^2 \\ k_{9,10}^2 & k_{10,10}^2 \end{bmatrix} = \begin{bmatrix} K_{11,11} & K_{11,12} \\ K_{11,12} & K_{12,12} \end{bmatrix} \tag{15.83}$$

where $k_{i,j}^e$ and $K_{i,j}$ denote the matrix component of the local element matrix and the global matrix, respectively.

15.3.5.3 Imposition of Displacement Constraints

The global stiffness matrix **K** in Eq. (15.82) does not usually have a full rank, because the prescribed displacement constraints are not yet imposed. Physically, an unconstrained solid or structure is capable of performing rigid movement; thus, the displacement solutions may be not unique. To obtain unique displacement solutions, the prescribed displacement conditions must be introduced into the global stiffness equations.

For constrained solids and structures, the most direct way to impose the displacement constraints is simply removing the rows and columns corresponding to the constrained nodal degree of freedoms. But such treatment may lead to the disorder of the global matrix and the size of it is changed too. Besides, the penalty approach can be employed, which is achieved by adding a large number (penalty term), i.e., 10^{20}, to the leading diagonal of the global matrix in the row in which the prescribed constraint value is required, and simultaneously, the term in the same row of the nodal force vector is replaced by the multiplication of the prescribed constraint value and the augmented stiffness diagonal element. For example, we may assume that the condition at the degree of freedom D_5 is known to be $\bar{D}_5$, then following the rule described above, the fifth row of the unconstrained set of stiffness equations (15.82) with n unknowns totally would be modified as

$$K_{5,1}D_1 + \cdots + \left(K_{5,5} + 10^{20}\right) D_5 + \cdots + K_{5,n}D_n = \bar{D}_5 \left(K_{5,5} + 10^{20}\right) \tag{15.84}$$

which would have the approximated effect of making $D_5 = \bar{D}_5$, since the terms $\sum_{i=1, i\neq 5}^{n} K_{5,i} D_i$ are very small relative to the larger diagonal term $(K_{5,5} + 10^{20}) D_5$ and can be neglected in the practical computation.

Additionally, the so-called large number 10^{20} can be replaced by a smaller number. In our application, the penalty parameter is chosen by

$$\max_{i,j=1\to n}(|\mathbf{K}_{i,j}|) \times 10^6 \tag{15.85}$$

After the treatment of constraints, the modified stiffness matrix $\mathbf{K}$ in Eq. (15.82) will be of full rank, and will be positive definite. Next, the modified stiffness equations can be solved by the standard solver of linear system of equations such as Gaussian elimination method.

15.3.6 Recovery of Rigid-Body Motion

By checking the above procedure, we know that the solution fails if any of the functions u_{li}^* is in a rigid-body motion mode. As a consequence, the matrix $\mathbf{H}_e$ is not in full rank and becomes singular for inversion. Therefore, special care should be taken to discard all rigid-body motion terms from $\mathbf{u}_e$ to prevent the element deformability matrix $\mathbf{H}_e$ from being singular.

However, it is necessary to reintroduce the discarded rigid-body modes in the internal field $\mathbf{u}_e$ of a particular element and then to calculate the corresponding rigid-body amplitude by requiring local or average fitting, for example, the least squares adjustment of $\mathbf{u}_e$ and $\tilde{\mathbf{u}}_e$ at all nodes of the polygonal element. In this case, these missing terms can easily be recovered by setting for the augmented intra-element displacement field of the polygonal element e

$$\hat{\mathbf{u}}_e = \mathbf{u}_e + \begin{bmatrix} 1 & 0 & x_2 \\ 0 & 1 & -x_1 \end{bmatrix} \mathbf{c}_0 \tag{15.86}$$

where $\mathbf{c}_0$ is the undetermined rigid-body amplitude vector, which can be calculated using the least square matching of $\mathbf{u}_e$ and $\tilde{\mathbf{u}}_e$ at all nodes of the polygonal element, that is,

$$\sum_{i=1}^{n} \left[(u_{1i} - \tilde{u}_{1i})^2 + (u_{2i} - \tilde{u}_{2i})^2 \right] = \min \tag{15.87}$$

which finally yields

$$\mathbf{R}_e \mathbf{c}_0 = \mathbf{r}_e \tag{15.88}$$

with

$$\mathbf{R}_e = \sum_{i=1}^{n} \begin{bmatrix} 1 & 0 & x_{2i} \\ 0 & 1 & -x_{1i} \\ x_{2i} & -x_{1i} & x_{1i}^2 + x_{2i}^2 \end{bmatrix} \tag{15.89}$$

$$\mathbf{r}_e = \sum_{i=1}^{n} \left\{ \begin{matrix} d_{e1}^i - u_{e1}^i \\ d_{e2}^i - u_{e2}^i \\ \left(d_{e1}^i - u_{e1}^i\right) x_{2i} - \left(d_{e2}^i - u_{e2}^i\right) x_{1i} \end{matrix} \right\} \tag{15.90}$$

As a result, once the nodal field $\mathbf{d}_e$ and the interpolation coefficient $\mathbf{c}_e$ are determined by solving Eq. (15.80), then $\mathbf{c}_0$ can be evaluated from Eq. (15.88). Finally, the complete displacement field $\hat{\mathbf{u}}_e$ at any interior point in a particular element can be obtained by means of Eq. (15.86).

15.3.7 Algorithm for Implementing the Solution Procedure

A step-to-step algorithm is presented here to implement the element formulations and the solution procedures for the analysis of linear elastic problems using the Voronoi polygonal hybrid finite elements, as follows:

Step 1: The computing domain is discretized using the Voronoi polygonal cells, and correspondingly, the element connectivity is established for use
Step 2: Evaluate the matrices $\mathbf{H}_e$ and $\mathbf{G}_e$ for each polygonal hybrid finite element by Eq. (15.75)
Step 3: Evaluate the element stiffness matrix k_e for each polygonal hybrid finite element by Eq. (15.81)
Step 4: Assemble for producing the global stiffness matrix $\mathbf{K}$
Step 5: Evaluate the equivalent nodal force vector $\mathbf{F}$
Step 6: Introduce the prescribed displacement constraints in the global equations
Step 7: Solve the modified global equations for the nodal displacement solutions $\mathbf{D}$, and then determine the element nodal displacement $\mathbf{d}_e$
Step 8: Evaluate the element interpolation coefficient $\mathbf{c}_e$ by Eq. (15.80)
Step 9: Perform recovery of rigid-body motion for each element by solving Eq. (15.88)
Step 10: Evaluate the element interior displacement and stress fields by Eqs. (15.86) and (15.32)

15.4 Applications

In this section, three numerical examples including the Cook's problem, the thick cylinder under internal pressure, and the elliptical hole problem in infinite plate are tested to validate the present polygonal hybrid finite element and demonstrate its accuracy and convergence.

15.4.1 Cook's Problem

To validate the present polygonal hybrid finite element in a simple way, the Cook's problem in plane stress state [12, 72] is taken into consideration. It is assumed that the left edge of the beam is fixed without rotation constraint and the right edge is subjected to uniformly distributed shearing loads, as shown in Fig. 15.12. For this problem, the deflection at the midpoint of the loaded edge is evaluated by several meshes with the conventional quadrilateral finite elements implemented by ABAQUS and the present *n*-sided polygonal hybrid finite elements to assess the performance of the present elements, as indicated in Fig. 15.13. The results are compared with the available reference value 23.96 [73, 74]. Besides, the results by the present elements are also compared to those in Reference [74] by the 16 and 64 natural polygonal finite elements with Laplace interpolants.

The results in Table 15.2 indicate that the three types of elements can produce better results with the increase of number of elements. However, the present polygonal hybrid finite elements are not as stiff as the quadrilateral elements and the natural polygonal finite elements and behave the best accuracy, especially with coarser meshes. Similar conclusions can be found in literature [73].

Besides, it is observed that both the present polygonal hybrid finite elements and the natural polygonal finite elements have more nodes than the conventional quadrilateral finite elements with the same number of elements. For example, 19, 65, and 148 polygonal elements have 39, 132, and 296 nodes, respectively, but the

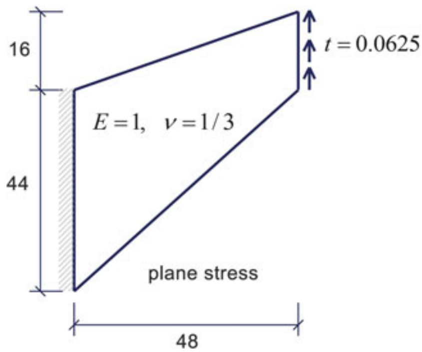

Fig. 15.12 Cook's problem

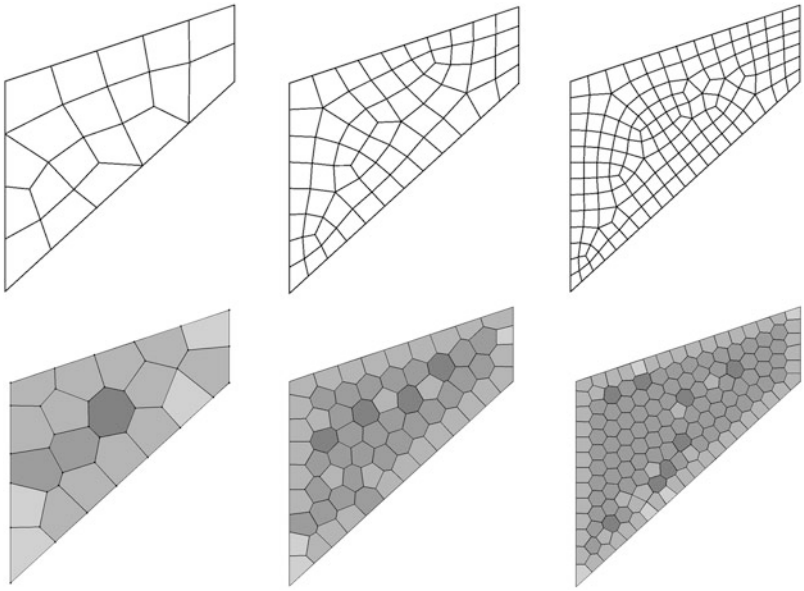

Fig. 15.13 Quadrilateral and polygonal discretization with 19, 65, and 148 elements

Table 15.2 Comparison of deflection results at the midpoint of the loaded edge for Cook's problem

Number of elements	Polygonal hybrid FE	Quadrilateral FE	Polygonal FE	Reference value
19	23.116	25.183	21.9240	23.96
65	23.801	24.195	23.4488	
148	23.880	24.087	/	

same number of quadrilateral elements just has 28, 83, and 175 nodes, respectively. This can be attributed to the multi-node connection of polygonal mesh. Thus, the polygonal elements inevitably lead to the bigger size of the solving system than that for the quadrilateral elements.

Finally, to demonstrate the reasonable choice of the parameter γ in Eq. (15.26), we investigate the variation of the deflection at the midpoint of the loaded edge when the value of the parameter γ changes from 0.1 to 20. The results in Fig. 15.14 indicate that there is a large range to produce stable results for the three polygonal meshes given in Fig. 15.13. However, it is found that too small values of γ, that is, the source points are too close to the physical boundary of the hybrid element, lead to vibrating results due to the near singularity of fundamental solutions. However, when the source points are too remote from the element physical boundary, the numerical accuracy may also decrease due to the round-off error caused by large magnitude difference between the unknown interpolating coefficient $\mathbf{c}_e$ and nodal

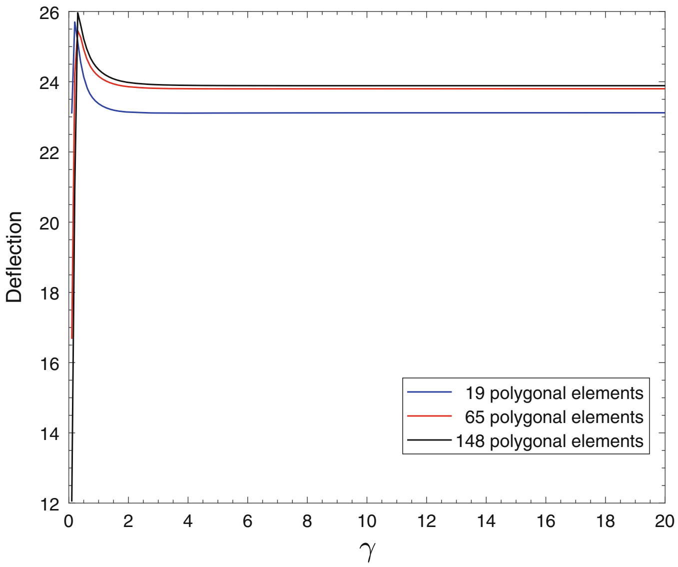

Fig. 15.14 Illustration of the effect of the parameter γ on the deflection at the midpoint of the loaded edge

displacement $\mathbf{d}_e$. Here, we take $\gamma = 10$ for the following computation, unless otherwise stated.

15.4.2 Thick Cylinder Under Internal Pressure

To demonstrate the applicability of the present Voronoi polygonal element for dealing with curved boundaries, a long thick circular cylinder under internal pressure p is accounted for, as indicated in Fig. 15.15. This problem has been studied by many researchers to demonstrate the efficiency of the developed numerical methods such as radial basis collocation methods [75, 76], in which the strong RBF interpolation can produce exponential convergence rate. Due to axisymmetric feature of the cylinder model, only one quarter of it, the shaded region in Fig. 15.15, is chosen for computation and the corresponding boundary conditions are also displayed in the figure. For this particular problem, the theoretical solutions of displacements and stresses in the polar coordinate system (r, θ) are expressed as [45, 77]

$$u_r = \frac{1+\nu}{E}\left[-\frac{A}{r} + 2B\,(1-2\nu)\,r\right], \qquad u_\theta = 0 \tag{15.91}$$

$$\sigma_r = \frac{A}{r^2} + 2B, \qquad \sigma_\theta = -\frac{A}{r^2} + 2B, \qquad \sigma_{r\theta} = 0 \tag{15.92}$$

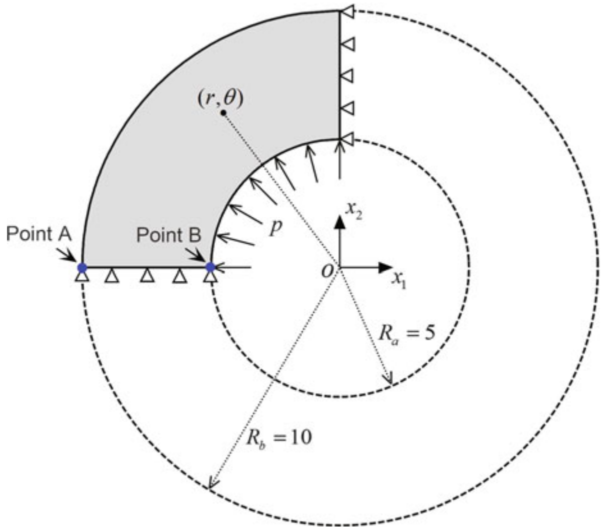

Fig. 15.15 Schematic diagram of thick cylinder under internal pressure

where

$$A = -\frac{{R_a}^2 {R_b}^2}{{R_b}^2 - {R_a}^2} p, \qquad B = \frac{{R_a}^2}{2({R_b}^2 - {R_a}^2)} p \tag{15.93}$$

In the practical computation, the inner and outer radii are $R_a = 5$ and $R_b = 10$, respectively. The applied internal uniform pressure is chosen as $p = 10$. In Fig. 15.16, total three polygonal mesh configurations are used to model the computing domain: (a) 150 Voronoi polygonal elements including 3 four-sided elements, 46 five-sided elements, 96 six-sided elements, and 5 seven-sided elements; (b) 400 Voronoi polygonal elements including 5 four-sided elements, 97 five-sided elements, 273 six-sided elements, and 25 seven-sided elements; (c) 560 Voronoi polygonal elements including 9 four-sided elements, 119 five-sided elements, 398 six-sided elements, and 34 seven-sided elements. For comparison, in Fig. 15.16, the mesh divisions using general 4-node quadrilateral finite elements (CPE4R) in ABAQUS are also provided. It's noticed that the general finite element mesh is produced by setting the same number of segments as that in Voronoi mesh along the boundary of the computing domain.

Firstly, the numerical convergence of the relative error in the stress norm is shown in Fig. 15.17. It is seen from Fig. 15.17 that both the present Voronoi polygonal elements and the general 4-node quadrilateral elements yield optimal convergence with mesh refinement. From Fig. 15.17, it can be observed that the present hybrid

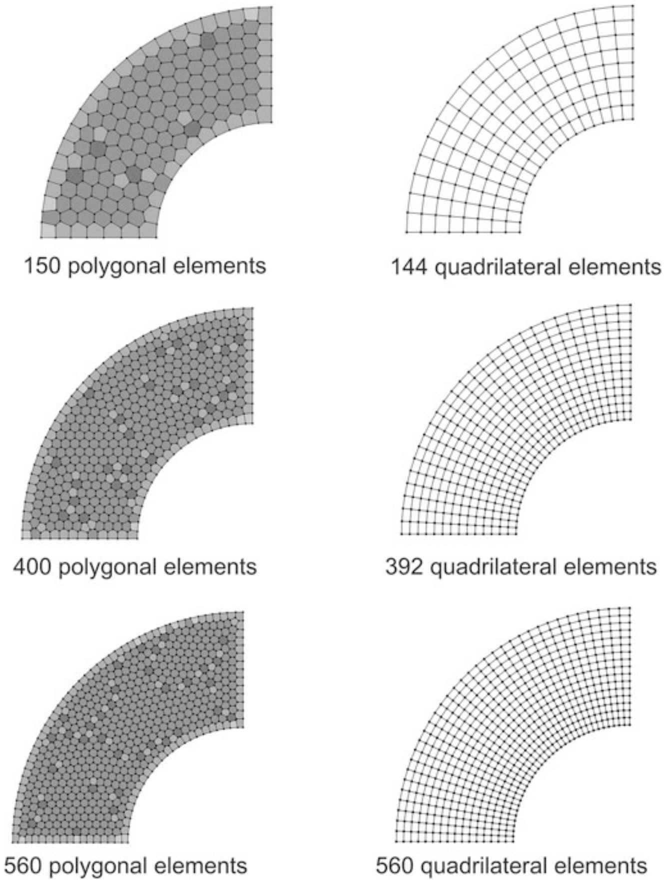

Fig. 15.16 Various mesh configurations of the thick cylinder with hybrid polygonal elements (left) and general 4-node quadrilateral element CPE4R in ABAQUS (right)

polygonal elements yield more accurate results than general 4-node quadrilateral elements. Also, similar convergence ratio for the two types of elements is presented in Fig. 15.17. Next, the variations of radial displacement and radial and hoop stresses along the bottom edge of the computing domain using 150 polygonal elements are displayed in Fig. 15.18, from which it's found that the numerical results from the present polygonal elements agree well with the available exact results, except for the radial stress σ_r at $r = 5$. The main reason is that the linear approximation of equivalent nodal loads brings large error along the curved edge. Same problem can be found when we solve this example using the commercial finite element software ABAQUS with linear general element. To improve the numerical accuracy at $r = 5$,

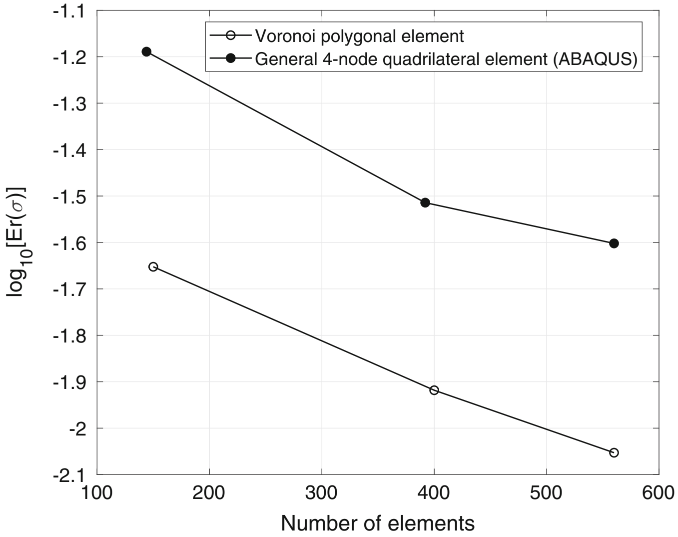

Fig. 15.17 Convergent results of stress for the thick cylinder under internal pressure

we can use more elements in the computing domain. For clarifying this, comparison of exact solutions and numerical results from the present method and ABAQUS at two key positions (point A and B in Fig. 15.15) is performed in Table 15.3, from which it is illustrated that with mesh refinement, both the two methods converge to the exact solution. Again, one observes that the present hybrid Voronoi polygonal element can produce better accuracy than the general 4-node quadrilateral finite element.

15.4.3 Infinite Plate with a Centered Elliptical Hole Under Tension

In this example, for which the exact solution is available, an infinite plate with a centered elliptical hole is considered to demonstrate the accuracy of the constructed special n-sided elliptical-hole hybrid finite element. It is assumed that the plate is subjected to unidirectional tension. Under the plane stress state, the exact complex potential solutions for such case can be determined by using Cauchy integral methods, as found in Muskhelishvili [77]. The related stress concentration factor

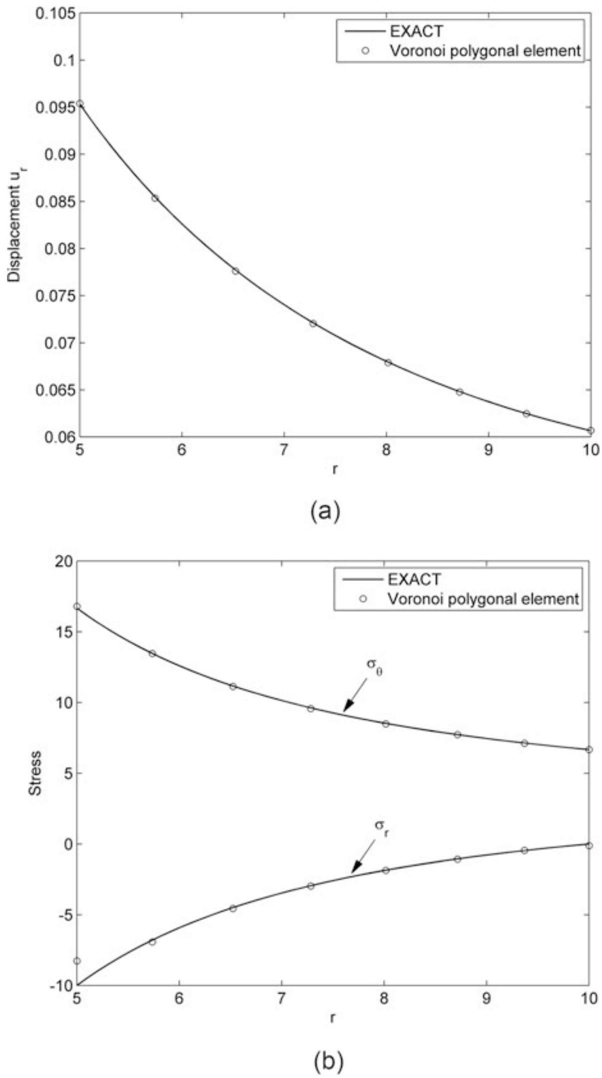

Fig. 15.18 Variations of radial displacement (**a**) and stresses (**b**) along the bottom edge

(SCF) on the hoop stress σ_θ at the top point $(a, 0)$ of the elliptical boundary is given by

$$\frac{(\sigma_\theta)_{\max}}{p} = 1 + 2\frac{a}{b} \tag{15.94}$$

In the computation, the infinite domain is idealized by a large square domain taken from the work of Piltner [78] with the side length 100 and the major axis of

Table 15.3 Comparison of exact solutions and different numerical results

		Point A	Point B
EXACT	σ_r	−10.000	0.0000
	σ_θ	16.667	6.6667
Present	σ_r	−8.3379 (150 elements)	−0.1284 (150 elements)
		−9.6818 (560 elements)	−0.0107 (560 elements)
	σ_θ	16.8756 (150 elements)	6.6746 (150 elements)
		16.6193 (560 elements)	6.6591 (560 elements)
ABAQUS	σ_r	−8.4733 (144 elements)	−0.2087 (144 elements)
		−9.2030 (560 elements)	−0.1041 (560 elements)
	σ_θ	15.1400 (144 elements)	6.8753 (144 elements)
		15.8697 (560 elements)	6.7707 (560 elements)

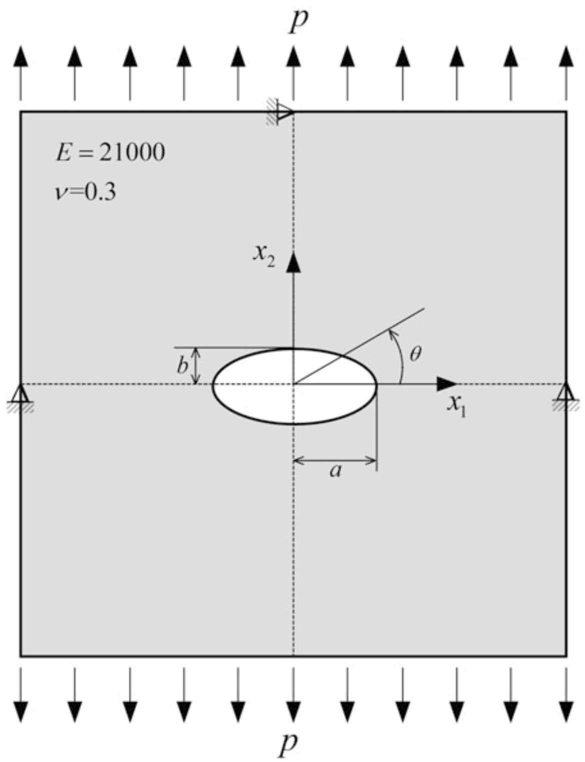

Fig. 15.19 Infinite square plate with a centered elliptical hole under tension

the elliptical hole is taken to be a=2, as shown in Fig. 15.19. The applied uniform tension load p is assumed to be 1. The mesh configuration associated with one 8-node special hybrid element and 48 8-node general hybrid elements is displayed in Fig. 15.20. The intra-element displacement field in the special hybrid element can be formulated by employing the special fundamental solutions corresponding to

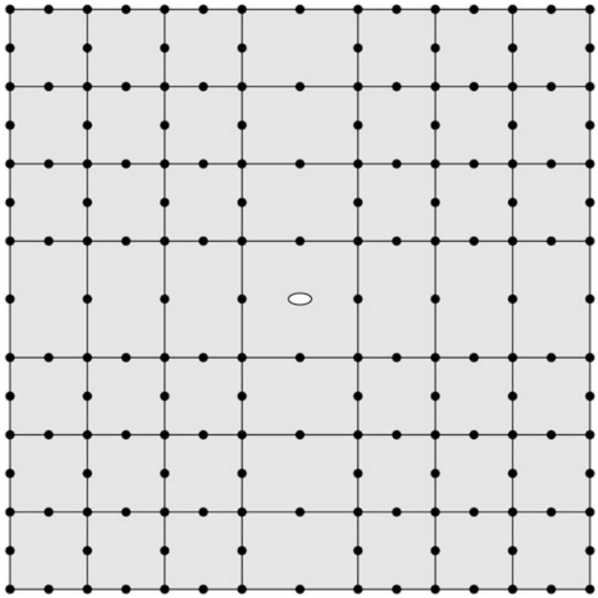

Fig. 15.20 Mesh configuration of special element

the elliptical hole [78] to replace the general fundamental solutions in Eq. (15.24), such that we can simply use a special hybrid polygonal element to enclose the elliptical hole to avoid the mesh division around the hole boundary. Such strategy can significantly decrease the computational cost when there are many holes or inclusions in the computational domains [20]. The total number of nodes is 176. Four symmetric displacement constraints at points (±50, 0), (0, ±50) are imposed during the computation.

The corresponding variation of hoop stress along the elliptical hole boundary is displayed in Fig. 15.21, and it is seen that there is good agreement between the numerical results using the present hybrid finite element model and the analytical solutions. As well, the decays of stress components σ_{11} and σ_{22} away from the edge of the elliptical hole are shown in Figs. 15.22 and 15.23, from which it can be seen that the magnitude of the stress components decays rapidly to the state without the elliptical hole, so the correctness of the truncated size of the infinite plate is illustrated.

Finally, the stress concentration factor for various ratios of a/b is calculated, and numerical and analytical solutions are tabulated in Table 15.4, from which it is found that the maximum relative error is just 0.27%, indicating that the present hybrid model can accurately capture the dramatic variation of stress on the elliptical hole boundary and the constructed special element is verified.

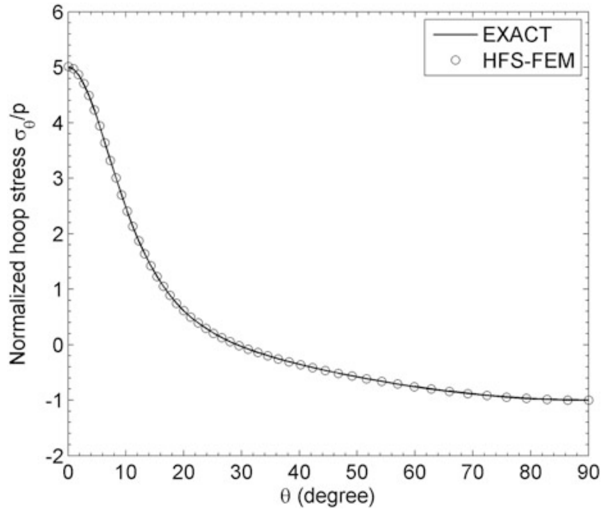

Fig. 15.21 Variation of hoop stress along the boundary of elliptical hole

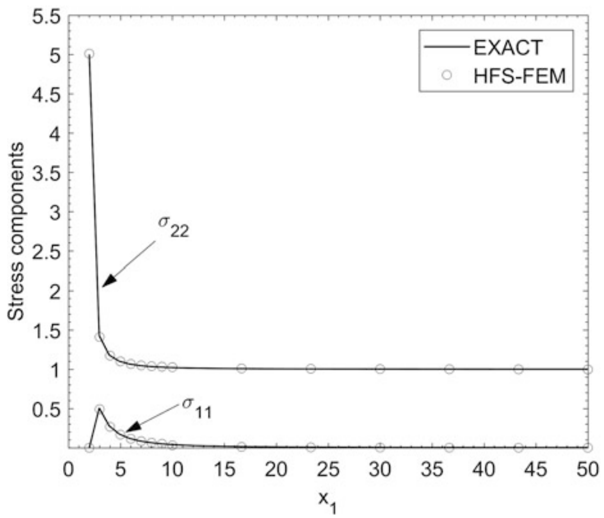

Fig. 15.22 Stress decay along the horizontal coordinate axis

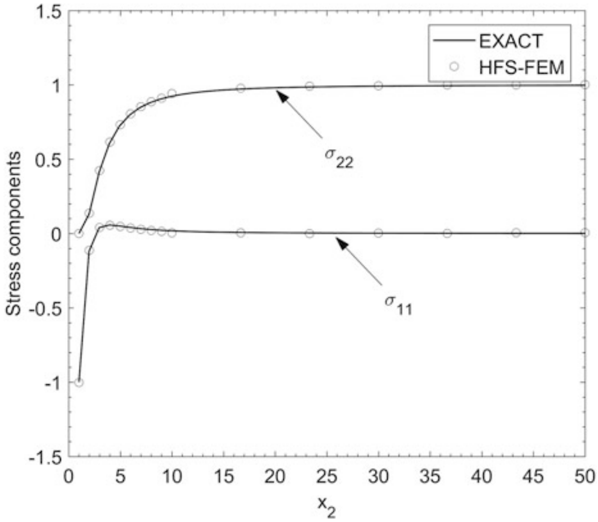

Fig. 15.23 Stress decay along the vertical coordinate axis

Table 15.4 Stress concentration factor for various ratios of *a*/*b*

a/b	1	2	3	4	5	6	7	8
EXACT	3	5	7	9	11	13	15	17
This study	3.008	5.010	7.012	9.015	11.018	13.021	15.024	17.026
Percentage relative error (%)	0.27	0.20	0.17	0.17	0.16	0.16	0.16	0.15

15.5 Conclusions

Voronoi cells can easily possess more connected neighbors and thus are suitable for generating unstructured polygonal mesh with high level of geometric isotropy. In this chapter, the formulation and implementation of *n*-sided polygonal hybrid finite element based on Voronoi partition are presented for two-dimensional linear elastic problems in isotropic and homogeneous materials. Different to the conventional conforming finite element and the natural polygonal finite element which is based on shape function interpolation at the whole element level, the present Voronoi polygonal hybrid finite element is formulated by introducing two independent displacement fields, respectively, within the interior and along the boundary of the element domain. The attractive property of element boundary integrals is achieved and permits versatile construction of convex polygons of arbitrary order to model the computing domain. It is demonstrated from three numerical experiments that the present Voronoi polygonal hybrid finite element has good convergence and accuracy for handling two-dimensional linear elastic analysis and hence significantly extends the potential applications of finite elements to convex *n*-sided polygons. Moreover, it's straightforward to integrate the present technique with conventional finite elements when necessary. It should also be mentioned that the present polygonal element formulation can be extended to the solutions of three-dimensional problems, nonlinear problem, and coupled problems if the related fundamental solutions are available.

References

1. T.R. Chandrupatla, A.D. Belegundu, *Introduction to Finite Elements in Engineering* (Pearson, London, 2012)
2. Q.H. Qin, Trefftz finite element method and its applications. Appl. Mech. Rev. **58**, 316–337 (2005)
3. N. Sukumar, E.A. Malsch, Recent advances in the construction of polygonal finite element interpolants. Arch. Comput. Method Eng. **13**, 129–163 (2006)
4. G. Manzini, A. Russo, N. Sukumar, New perspectives on polygonal and polyhedral finite element methods. Math. Models Methods Appl. Sci. **24**, 1665–1699 (2014)
5. A. Francis, A. Ortiz-Bernardin, S. Bordas, S. Natarajan, Linear smoothed polygonal and polyhedral finite elements. Int. J. Numer. Method Eng. **109**, 1263–1288 (2016)
6. K. Hormann, N. Sukumar, *Generalized Barycentric Coordinates in Computer Graphics and Computational Mechanics* (CRC Press, Boca Raton, 2017)
7. S. Natarajan, S.P. Bordas, E.T. Ooi, Virtual and smoothed finite elements: a connection and its application to polygonal/polyhedral finite element methods. Int. J. Numer. Methods Eng. **104**, 1173–1199 (2015)
8. S. Natarajan, E.T. Ooi, I. Chiong, C. Song, Convergence and accuracy of displacement based finite element formulations over arbitrary polygons: Laplace interpolants, strain smoothing and scaled boundary polygon formulation. Finite Elem. Anal. Des. **85**, 101–122 (2014)
9. L.B.D. Veiga, F. Brezzi, A. Cangiani, G. Manzini, L.D. Marini, A. Russo, Basic principles of virtual element methods. Math. Models Methods Appl. Sci. **23**, 199–214 (2013)
10. E.L. Wachspress, *A Rational Finite Element Basis* (Academic Press, New York, 1975)
11. M.S. Floater, K. Hormann, G. Kós, A general construction of barycentric coordinates over convex polygons. Adv. Comput. Math. **24**, 311–331 (2006)
12. E.B. Chin, J.B. Lasserre, N. Sukumar, Numerical integration of homogeneous functions on convex and nonconvex polygons and polyhedra. Comput. Mech. **56**, 967–981 (2015)
13. Q.H. Qin, *The Trefftz Finite and Boundary Element Method* (WIT Press, Boston, 2000)
14. J. Jirousek, A. Zieliński, Survey of Trefftz-type element formulations. Comput. Struct. **63**, 225–242 (1997)
15. J.A.T. de Freitas, Formulation of elastostatic hybrid-Trefftz stress elements. Comput. Methods Appl. Mech. Eng. **153**, 127–151 (1998)
16. Q.H. Qin, H. Wang, *Matlab and C Programming for Trefftz Finite Element Methods* (CRC Press, Boca Raton, 2009)
17. Q.H. Qin, Hybrid-Trefftz finite element method for Reissner plates on an elastic foundation. Comput Methods Appl. Mech. Eng. **122**, 379–392 (1995)
18. H. Wang, Q.H. Qin, D. Arounsavat, Application of hybrid Trefftz finite element method to non-linear problems of minimal surface. Int. J. Numer. Methods Eng. **69**, 1262–1277 (2007)
19. Q.H. Qin, Fundamental solution based finite element method. J. Appl. Mech. Eng. **2**, e118 (2013)
20. H. Wang, Q.H. Qin, Hybrid FEM with fundamental solutions as trial functions for heat conduction simulation. Acta Mech. Solida Sin. **22**, 487–498 (2009)
21. H. Wang, L.L. Cao, Q.H. Qin, Hybrid graded element model for nonlinear functionally graded materials. Mech. Adv. Mater. Struct. **19**, 590–602 (2012)

22. H. Wang, Q.H. Qin, Fundamental-solution-based finite element model for plane orthotropic elastic bodies. Eur. J. Mech. Solid **29**, 801–809 (2010)
23. H. Wang, Q.H. Qin, Fundamental-solution-based hybrid FEM for plane elasticity with special elements. Comput. Mech. **48**, 515–528 (2011)
24. H. Wang, Q.H. Qin, Special fiber elements for thermal analysis of fiber-reinforced composites. Eng. Comput. **28**, 1079–1097 (2011)
25. H. Wang, Q.H. Qin, A new special element for stress concentration analysis of a plate with elliptical holes. Acta Mech. **223**, 1323–1340 (2012)
26. H. Wang, Q.H. Qin, Numerical implementation of local effects due to two-dimensional discontinuous loads using special elements based on boundary integrals. Eng. Anal. Bound. Elem. **36**, 1733–1745 (2012)
27. H. Wang, Q.H. Qin, Boundary integral based graded element for elastic analysis of 2D functionally graded plates. Eur. J. Mech. Solid **33**, 12–23 (2012)
28. H. Wang, Q.H. Qin, X.P. Liang, Solving the nonlinear Poisson-type problems with F-Trefftz hybrid finite element model. Eng. Anal. Bound. Elem. **36**, 39–46 (2012)
29. H. Wang, Q.H. Qin, FE approach with Green's function as internal trial function for simulating bioheat transfer in the human eye. Arch. Mech. **62**, 493–510 (2010)
30. H. Wang, Q.H. Qin, A fundamental solution-based finite element model for analyzing multi-layer skin burn injury. J. Mech. Med. Biol. **12**, 1250027 (2012)
31. Q.H. Qin, H. Wang, Special circular hole elements for thermal analysis in cellular solids with multiple circular holes. Int. J. Comput. Methods **10**, 1350008 (2013)
32. H. Wang, X.J. Zhao, J.S. Wang, Interaction analysis of multiple coated fibers in cement composites by special n-sided interphase/fiber elements. Compos. Sci. Technol. **118**, 117–126 (2015)
33. H. Wang, Q.H. Qin, Y. Xiao, Special n-sided Voronoi fiber/matrix elements for clustering thermal effect in natural-hemp-fiber-filled cement composites. Int. J. Heat Mass Transf. **92**, 228–235 (2016)
34. H. Wang, Q.H. Qin, A new special coating/fiber element for analyzing effect of interface on thermal conductivity of composites. Appl. Math. Comput. **268**, 311–321 (2015)
35. H. Wang, Y. Gao, Q.H. Qin, Green's function based finite element formulations for isotropic seepage analysis with free surface. Latin Am. J. Solids Struct. **12**, 1991–2005 (2015)
36. H. Wang, Q.H. Qin, Voronoi polygonal hybrid finite elements with boundary integrals for plane isotropic elastic problems. Int. J. Appl. Mech. **09**, 1750031 (2017)
37. H. Wang, Q.H. Qin, Y.P. Lei, Green's-function-based-finite element analysis of fully plane anisotropic elastic bodies. J. Mech. Sci. Technol. **31**, 1305–1313 (2017)
38. O.C. Zienkiewicz, R.L. Taylor, *The Finite Element Method for Solid and Structural Mechanics* (Butterworth-Heinemann, Oxford, 2005)
39. H. Wang, B. Liu, Y. Kang, Q.H. Qin, Analysing effective thermal conductivity of 2D closed-cell foam based on shrunk Voronoi tessellations. Arch. Mech. **69**, 451–470 (2017)
40. C.A. Brebbia, J.C.F. Telles, L.C. Wrobel, *Boundary Element Techniques: Theory and Applications in Engineering* (Springer, Berlin, 1984)
41. Q.H. Qin, *Green's Function and Boundary Elements of Multifield Materials* (Elsevier, Oxford, 2007)
42. C.Y. Lee, H. Wang, Q.H. Qin, Dual reciprocity boundary element method using compactly supported radial basis functions for 3D linear elasticity with body forces. Int. J. Mech. Mater. Des. **12**, 463–476 (2016)
43. C.Y. Lee, H. Wang, Q.H. Qin, Efficient hypersingular line and surface integrals direct evaluation by complex variable differentiation method. Appl. Math. Comput. **316**, 256–281 (2018)
44. Q. Du, V. Faber, M. Gunzburger, Centroidal Voronoi tessellations: applications and algorithms. SIAM Rev. **41**, 637–676 (1999)
45. S.P. Timoshenko, J.N. Goodier, *Theory of Elasticity* (McGraw-Hill, New York, 1987)

46. C. Talischi, G.H. Paulino, A. Pereira, I.F. Menezes, PolyMesher: a general-purpose mesh generator for polygonal elements written in Matlab. Struct. Multidiscip. Optim. **45**, 309–328 (2012)
47. Q.H. Qin, Y. Huang, BEM of postbuckling analysis of thin plates. Appl. Math. Model. **14**, 544–548 (1990)
48. G. Fairweather, A. Karageorghis, The method of fundamental solutions for elliptic boundary value problems. Adv. Comput. Math. **9**, 69–95 (1998)
49. H. Wang, Q.H. Qin, Meshless approach for thermo-mechanical analysis of functionally graded materials. Eng. Anal. Bound. Elem. **32**, 704–712 (2008)
50. H. Wang, Q.H. Qin, Y. Kang, A new meshless method for steady-state heat conduction problems in anisotropic and inhomogeneous media. Arch. Appl. Mech. **74**, 563–579 (2005)
51. Z. Zhang, H. Wang, Q.H. Qin, Meshless method with operator splitting technique for transient nonlinear bioheat transfer in two-dimensional skin tissues. Int. J. Mol. Sci. **16**, 2001–2019 (2015)
52. Z.W. Zhang, H. Wang, Q.H. Qin, Method of fundamental solutions for nonlinear skin bioheat model. J. Mech. Med. Biol. **14**, 1450060 (2014)
53. H. Wang, Q.H. Qin, Some problems on the method of fundamental solution with radial basis functions. Acta. Mech. Solida Sin. **20**, 21–29 (2007)
54. C.S. Chen, A. Karageorghis, Y.S. Smyrlis, *The Method of Fundamental Solutions: A Meshless Method* (Dynamic Publishers, Atlanta, 2008)
55. C.Y. Lee, H. Wang, Q.H. Qin, Method of fundamental solutions for 3D elasticity with body forces by coupling compactly supported radial basis functions. Eng. Anal. Bound. Elem. **60**, 123–136 (2015)
56. H. Wang, Q.H. Qin, A meshless method for generalized linear or nonlinear Poisson-type problems. Eng. Anal. Bound. Elem. **30**, 515–521 (2006)
57. W.A. Yao, H. Wang, Virtual boundary element integral method for 2-D piezoelectric media. Finite Elem. Anal. Des. **41**, 875–891 (2005)
58. T.H.H. Pian, C.C. Wu, *Hybrid and Incompatible Finite Element Methods* (CRC Press, Boca Raton, 2005)
59. J. Jirousek, Q.H. Qin, Application of hybrid-Trefftz element approach to transient heat conduction analysis. Comput. Struct. **58**, 195–201 (1996)
60. Q.H. Qin, Hybrid Trefftz finite-element approach for plate bending on an elastic foundation. Appl. Math. Model. **18**, 334–339 (1994)
61. Q.H. Qin, Variational formulations for TFEM of piezoelectricity. Int. J. Solids Struct. **40**, 6335–6346 (2003)
62. Q.H. Qin, Solving anti-plane problems of piezoelectric materials by the Trefftz finite element approach. Comput. Mech. **31**, 461–468 (2003)
63. Q.H. Qin, *Fracture Mechanics of Piezoelectric Materials* (WIT Press, Southampton, 2001)
64. Q.H. Qin, S.W. Yu, An arbitrarily-oriented plane crack terminating at the interface between dissimilar piezoelectric materials. Int. J. Solids Struct. **34**, 581–590 (1997)
65. Q.H. Qin, Y.W. Mai, A closed crack tip model for interface cracks in thermopiezoelectric materials. Int. J. Solids Struct. **36**, 2463–2479 (1999)
66. S.W. Yu, Q.H. Qin, Damage analysis of thermopiezoelectric properties: part I–crack tip singularities. Theor. Appl. Fract. Mech. **25**, 263–277 (1996)
67. Q.H. Qin, Y.W. Mai, Crack growth prediction of an inclined crack in a half-plane thermopiezoelectric solid. Theor. Appl. Fract. Mech. **26**, 185–191 (1997)
68. Q.H. Qin, Y.W. Mai, S.W. Yu, Some problems in plane thermopiezoelectric materials with holes. Int. J. Solids Struct. **36**, 427–439 (1999)
69. Q.H. Qin, Y.W. Mai, S.W. Yu, Effective moduli for thermopiezoelectric materials with microcracks. Int. J. Fract. **91**, 359–371 (1998)
70. H.C. Simpson, S.J. Spector, On the positivity of the second variation in finite elasticity. Arch. Ration. Mech. Anal. **98**, 1–30 (1987)
71. Q.H. Qin, H. Wang, Special elements for composites containing hexagonal and circular fibers. Int. J. Comput. Methods **12**, 1540012 (17 pages) (2015)

72. H. Wang, Q.H. Qin, C.Y. Lee, n-sided polygonal hybrid finite elements with unified fundamental solution kernels for topology optimization. Appl. Math. Model. **66**, 97–117 (2019)
73. C. Talischi, G.H. Paulino, A. Pereira, I.F. Menezes, Polygonal finite elements for topology optimization: a unifying paradigm. Int. J. Numer. Methods Eng. **82**, 671–698 (2010)
74. Y.Q. Long, Y. Xu, Generalized conforming triangular membrane element with vertex rigid rotational freedoms. Finite Elem. Anal. Des. **17**, 259–271 (1994)
75. H.Y. Hu, J.S. Chen, W. Hu, Weighted radial basis collocation method for boundary value problems. Int. J. Numer. Methods Eng. **69**, 2736–2757 (2007)
76. L. Wang, Z. Zhong, Radial basis collocation method for nearly incompressible elasticity. J. Eng. Mech. **139**, 439–451 (2013)
77. N.I. Muskhelishvili, *Some Basic Problems of the Mathematical Theory of Elasticity* (Noordhoff Ltd, Groningen, 1953)
78. R. Piltner, Some remarks on finite elements with an elliptic hole. Finite Elem. Anal. Des. **44**, 767–772 (2008)

Chapter 16
Variational Methods for Schrödinger Type Equations

Giovany Malcher Figueiredo, Edwin Gonzalo Murcia, and Gaetano Siciliano

16.1 Introduction

It is well known that the Schrödinger equation is one of the most important equations in physics. It was formulated by E. Schrödinger in 1925 (which later in 1933 received the Nobel Prize in Physics) and introduced by taking into account the de Broglie hypothesis according to which matter particles possess a wave packet delocalized in space. According to the Copenhagen interpretation the square modulus of the wave function $\psi : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{C}$ encloses the physical information on the particle; in particular, $|\psi|^2$ is related to the probability of finding the particle in a specific space region. Since its formulation the Schrödinger equation is the object of many research from a physical and mathematical point of view. Mathematically the Schrödinger equation is a partial differential equation of the form

$$i\hbar\partial_t \psi = -\frac{\hbar}{2m}\Delta\psi + V\psi + f(|\psi|)\psi$$

where ψ is the unknown function, m the mass particle, $\hbar$ the normalized Plank constant, $V : \mathbb{R}^3 \to \mathbb{R}$ is a potential, and f a suitable nonlinearity. The symbol Δ is

G. M. Figueiredo
Departamento de Matemática, Universidade de Brasilia, Brasilia, Brazil
e-mail: giovany@unb.br

E. G. Murcia
Departamento de Matemáticas, Pontificia Universidad Javeriana, Bogotá, Colombia
e-mail: murciae@javeriana.edu.co

G. Siciliano (✉)
Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, SP, Brazil
e-mail: sicilian@ime.usp.br

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_16

the Laplacian with respect to the spatial variables. We remark that, depending on the physical model described, the potential V can be given a priori, which is the case studied in Sect. 16.3, as well as it can be an unknown of the problem, for example it can depend on the same ψ, which is the case studied in Sect. 16.4.

Actually here we will concern with fractional versions of the Schrödinger equation driven by the Fractional Laplacian $(-\Delta)^s, s \in (0, 1)$. In fact, since the paper of Laskin [45–47], where he gave the bases for a Fractional Quantum Mechanics, it has been recognized that the fractional version of the Schrödinger equation is more appropriate to describe some physical and real models. We will not enter in details here referring the reader to the appropriate sections below.

The search of standing wave solutions of the above fractional Schrödinger equation, that is solution of the particular form

$$\psi(x,t) = u(x)e^{i\omega t}, \quad u : \mathbb{R}^3 \to \mathbb{R}, \quad \omega > 0,$$

(which are even interesting from a physical point of view) led to consider a semilinear elliptic equation of type

$$(-\Delta)^s u + \omega u + Vu = f(u).$$

Roughly speaking, we will treat the case in which the potential V is given, and the case in which it is unknown, depending on the same wave function ψ. We will employ variational methods in order to find multiple solutions. The choice of working with variational methods is mainly due to the fact that, as it is well known, the Schrödinger equation has a Lagrangian formulation. Indeed our assumptions on f and V will be "compatible" with the use of variational tools.

The organization of this chapter is as follows.

Section 16.2 is devoted to recall basic mathematical facts useful to study our equations. In particular the Sobolev spaces we will use, as well as few notions of Differential Calculus in Hilbert spaces and Critical Point Theory. Indeed the solutions of our fractional equations will be found as critical points of suitable functionals defined in infinite dimensional Hilbert spaces, or Hilbert manifolds. This part can be found in [52].

Section 16.3 is devoted to study the problem under a given external potential V and a multiplicity result of solutions is presented. The results of this section are given in [39].

Finally in Sect. 16.4 the problem under an unknown potential depending on the same wave function (in other words, a system of partial differential equations) is studied. Again a multiplicity result is obtained. The results presented here are taken from [53].

Few Basic Notations As a matter of notations, we alert the reader that the Lebesgue measure $dx, dy, \ldots$ in the integration will be omitted, unless strictly necessary.

We use the symbols C, C_i, for $i = 1, 2, \ldots$ for positive constants which may also change from line to line.

We denote with $B_r(x)$ the closed ball in $\mathbb{R}^3$ centered in x with radius $r > 0$, with $B_r^c(x)$ its complementary; if $x = 0$, we simply write B_r.

Other notations will be introduced whenever we need.

16.2 Background Material

In this section we recall few mathematical facts that will be used in the next two sections where Schrödinger type equations are studied. Besides [52], we give also other references where the interested reader may find all the details.

For the reader convenience we have divided this part into the following subsections.

- **Recalling Sobolev Spaces**. Here the function spaces where we will work are introduced.
- **Basic notions of differential calculus in Hilbert spaces**. Here few notions of differential calculus in Hilbert spaces are given: the notion of critical point of functionals as well as few calculation rules and important examples. This is a quite important topic since the solutions of our equations will be found as critical points of suitable functionals restricted to Hilbert manifolds.
- **The Ljusternick-Schnirelmann category**. Here the main theorem of the Ljusternick-Schnirelmann theory is proved. This will be fundamental in order to prove existence of many critical points of the functionals, hence solutions of our equations.
- **Schrödinger type equations**. Here we apply the abstract results given in the previous subsections to a Schrödinger equation involving the "genuine" Laplace operator. Even though this is not exactly the equation studied in the subsequent sections (indeed we will consider the case of fractional Laplacian and with a more general nonlinearity) we believe this example is important in order to understand the basic ideas and then generalize to the fractional case.

16.2.1 Recalling Sobolev Spaces

We begin by recalling basic facts in measure theory, see, e.g., [56, pg. 86 and pg. 91] for more details. Here Ω is a domain in $\mathbb{R}^3$.

We will concern just with the three-dimensional case since this is the case of our applications, but of course the same analysis can be done for $\mathbb{R}^N$. If $\Omega \subset \mathbb{R}^3$ is a domain, $L^p(\Omega)$ is the usual Lebesgue space endowed with norm $|\cdot|_{p,\Omega}$ or simply $|\cdot|_p$ if no confusion arises.

For our purpose, important results in Measure Theory are the following.

Theorem 16.2.1 (Fatou's Lemma) *Let $\{f_n\}$ be a sequence of nonnegative, measurable functions on Ω, such that $f_n(x) \to f(x)$ a.e. in Ω. Then*

$$\int_\Omega f \le \liminf_{n\to\infty} \int_\Omega f_n.$$

Theorem 16.2.2 (Lebesgue Convergence Theorem) *Let g be integrable over Ω and let $\{f_n\}$ be a sequence of measurable functions such that $|f_n| \le g$ on Ω and for almost all x in Ω we have $f_n(x) \to f(x)$ when $n \to \infty$. Then*

$$\int_\Omega f_n \to \int_\Omega f.$$

Lemma 16.2.3 (Brezis-Lieb Lemma) *Given $1 < p < \infty$, if $\{f_n\} \subset L^p(\Omega)$ is a bounded sequence of functions such that $f_n(x) \to f(x)$ a.e. $x \in \Omega$, then $f \in L^p(\Omega)$ and*

$$\int_\Omega (|f_n|^p - |f_n - f|^p) \to \int_\Omega |f|^p.$$

In virtue of the Brezis-Lieb lemma, if $\{f_n\} \subset L^p(\Omega)$ is a sequence of functions such that

$$f_n(x) \to f(x) \text{ a.e. } x \in \Omega \quad \text{and} \quad |f_n|_p \to |f|_p,$$

then

$$f_n \to f \text{ in } L^p(\Omega).$$

The natural setting for the equations we are going to study involves Sobolev spaces. $H^1(\Omega)$ is the Hilbert space endowed with the inner product and norm

$$\langle u, v\rangle_\Omega := \int_\Omega \nabla u \nabla v + \int_\Omega uv, \quad \|u\|_\Omega^2 := \langle u, u\rangle_\Omega.$$

The space $H_0^1(\Omega)$ is the subspace of $H^1(\Omega)$ defined as the completion of the test functions $C_c^\infty(\mathbb{R}^3)$, the space of infinitely differentiable functions with compact support, with respect to $\|\cdot\|_\Omega$. When $\Omega \subset \mathbb{R}^3$ is a bounded domain, $H_0^1(\Omega)$ is a Hilbert space endowed with the equivalent inner product and norm given by

$$\langle u, v\rangle_{H_0^1(\Omega)} := \int_\Omega \nabla u \nabla v, \quad \|u\|_{H_0^1(\Omega)}^2 := \langle u, u\rangle_{H_0^1(\Omega)}$$

If $\Omega = \mathbb{R}^3$ we will suppress the symbol Ω in scalar products and Sobolev norms, as well as in the L^p−norm. We will make use also of the Hilbert space

$$D^{1,2}(\mathbb{R}^3) = \{u \in L^{2^*}(\mathbb{R}^3) : \|\nabla u\| \in L^2(\mathbb{R}^3)\}, \quad 2^* = 6$$

which coincides with the completion of the test functions with respect to the (square) norm

$$\|u\|_D^2 := \int_{\mathbb{R}^3} \|\nabla u\|^2.$$

The dual spaces of $H_0^1(\Omega)$ and $D^{1,2}(\mathbb{R}^3)$ are denoted by $H^{-1}(\Omega)$ and D', respectively.

Recall that given X, Y Banach spaces, the notation $X \hookrightarrow Y$ means that X is *continuously embedded into* Y, i.e. $X \subset Y$ and the inclusion map $\iota : X \to Y$ is continuous. In this case there exists a constant $C > 0$ such that

$$\|u\|_Y \leq C\|u\|_X, \quad \forall u \in X.$$

When $X \hookrightarrow Y$ and the inclusion map ι is also a compact map we say that X *is compactly embedded into* Y and we use the notation $X \hookrightarrow\hookrightarrow Y$.

For the next result see, e.g., [14, 15].

Theorem 16.2.4 *If $\Omega \subset \mathbb{R}^3$ is a bounded domain, we have*

$$H^1(\Omega) \hookrightarrow L^p(\Omega), \quad \textit{for } 1 \leq p \leq 2^*$$

and

$$H^1(\Omega) \hookrightarrow\hookrightarrow L^p(\Omega), \quad \textit{for } 1 \leq p < 2^*.$$

If $\Omega = \mathbb{R}^3$, then

$$H^1(\mathbb{R}^3) \hookrightarrow L^p(\mathbb{R}^3), \quad \textit{for } 2 \leq p \leq 2^*$$

and

$$D^{1,2}(\mathbb{R}^3) \hookrightarrow L^{2^*}(\mathbb{R}^3).$$

The embeddings in the whole space are not compact.

Let us introduce now the fractional Sobolev spaces. Given $\beta \in (0, 1)$, we recall that the fractional Laplacian $(-\Delta)^\beta$ is the pseudodifferential operator which can be defined via the Fourier transform

$$\mathcal{F}((-\Delta)^\beta u) = |\cdot|^{2\beta}\mathcal{F}u.$$

If u has sufficient regularity, it can be given also by

$$(-\Delta)^{\beta}u(z) = -\frac{C_{\beta}}{2}\int_{\mathbb{R}^3}\frac{u(z+y)+u(z-y)-2u(z)}{|y|^{3+2\beta}}dy, \quad z\in\mathbb{R}^3,$$

where C_β is a suitable normalization constant depending also on the dimension. Let

$$H^{\beta}(\mathbb{R}^3) = \left\{u\in L^2(\mathbb{R}^3) : (-\Delta)^{\beta/2}u \in L^2(\mathbb{R}^3)\right\}$$

be the Hilbert space with scalar product and (squared) norm given by

$$(u,v) = \int_{\mathbb{R}^3}(-\Delta)^{\beta/2}u(-\Delta)^{s/2}v + \int_{\mathbb{R}^3}uv, \qquad \|u\|^2 = |(-\Delta)^{\beta/2}u|_2^2 + |u|_2^2.$$

It is known that $H^{\beta}(\mathbb{R}^3)\hookrightarrow L^p(\mathbb{R}^3)$, $p\in[2,2^*_\beta]$ with $2^*_\beta := 6/(3-2\beta)$.

We will consider also the homogeneous Sobolev spaces $\dot{H}^{\alpha/2}(\mathbb{R}^3)$, $\alpha\in(0,3)$, defined as the completion of $C_c^\infty(\mathbb{R}^3)$ with respect to the norm $|(-\Delta)^{\alpha/4}u|_2$. This is a Hilbert space with scalar product and (squared) norm

$$(u,v)_{\dot{H}^{\alpha/2}} = \int_{\mathbb{R}^3}(-\Delta)^{\alpha/4}u(-\Delta)^{\alpha/4}v, \qquad \|u\|^2_{\dot{H}^{\alpha/2}} = |(-\Delta)^{\alpha/4}u|_2^2.$$

It is well known that $\dot{H}^{\alpha/2}(\mathbb{R}^3)\hookrightarrow L^{2^*_{\alpha/2}}(\mathbb{R}^3)$, $2^*_\alpha = 6/(3-\alpha)$. For more general facts about the fractional Laplacian, we refer the reader to the beautiful paper [34].

We remark that, recently, especially after the formulation of the Fractional Quantum Mechanics, the derivation of the Fractional Schrödinger equation given by Laskin in [45–47], and the notion of fractional harmonic extension of a function studied in the pioneering paper [23], equations involving fractional operators are receiving a great attention. Fractional Schrödinger type equations have been studied, e.g., in [10, 30, 40]. Actually pseudodifferential operators appear in many problems in Physics and Chemistry, see, e.g., [48, 49]; but also in obstacle problems [50, 59], optimization and finance [29, 35], conformal geometry and minimal surfaces [22, 24, 25], phase transition [1, 60], material science [16], anomalous diffusion [33, 48, 49]. But also to crystal dislocation, soft thin films, multiple scattering, quasi-geostrophic flows, water waves, and so on. It is really difficult to give an extensive list of references; the interested reader is invited to see also the references in the above cited papers.

16.2.2 Basic Notions of Differential Calculus in Hilbert Spaces

We recall here some basic fact of calculus in the infinite dimensional case. This will be fundamental in order to find solutions of our Schrödinger like equations as critical point of suitable functionals. For more details on the subject, the reader may consult, e.g., [14]. We remark that all that we say here also holds, with suitable changes, for Banach spaces and manifolds.

Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ be Hilbert spaces. We denote by $\mathcal{L}(X; Y)$ the space of linear and continuous maps from X to Y with the operator norm

$$\|A\|_\star := \sup_{\|u\|_X \le 1} \|A[u]\|_Y$$

for which it is also complete. Here is the key definition which gives rise to differential calculus in Banach spaces.

Definition 16.2.5 Let $U \subset X$ be an open set, and $I : U \to Y$ a map. We say that I is (Fréchét) differentiable at $u_0 \in U$ if there is $A_{u_0} \in \mathcal{L}(X; Y)$ such that

$$I(u_0 + h) - I(u_0) - A_{u_0}[h] = o(\|h\|) \quad \text{as} \quad h \to 0.$$

In other words,

$$\lim_{h\to 0} \frac{I(u_0 + h) - I(u_0) - A_{u_0}[h]}{\|h\|} = 0.$$

The function I is said to be (Fréchét) differentiable in U if it is differentiable in every $u \in U$.

If I is differentiable in U and the map $I' : U \to \mathcal{L}(X, Y)$ is continuous, we say that I is C^1 and write $I \in C^1(U; Y)$.

If $I \in C^1(U; \mathbb{R})$, we say that $u_0 \in U$ is a critical point of I if $A_{u_0} = 0$.

It is easy to see that whenever exists, A_{u_0} is unique, and will be denoted with $I'(u_0)$. Observe that if I is differentiable at u_0, then I is continuous at u_0.

Of course the definition can be "iterated" in the following sense. If $I' : U \subset X \to \mathcal{L}(X, Y)$ is differentiable at u_0, we say that I is twice differentiable at u_0, and we denote $I''(u_0) := (I')'(u_0)$. Note that $I''(u_0) \in \mathcal{L}(X; \mathcal{L}(X; Y)) \simeq \mathcal{L}(X \times X; Y)$, where $\mathcal{L}(X \times X; Y)$ is the set of bilinear, continuous maps from $X \times X$ to Y. Thus, if I' is differentiable at u_0 and $v, w \in X$, $I''(u_0)[v, w] \in Y$.

If I is twice differentiable at every point $u \in U$, we say that I is twice differentiable in U. In this case, if the map $I'' : U \subset X \to \mathcal{L}(X \times X; Y)$ is continuous in U, we write $I \in C^2(U; Y)$.

Example We give here two basic but very important examples of differentiable functionals.

1. If $I \in \mathcal{L}(X; Y)$, then I is differentiable at every point u, and $I'(u) = I$. Moreover $I'' \equiv 0$.
2. Let X, Y, and Z be Hilbert spaces, and $\mathfrak{b} : X \times Y \to Z$ a bilinear continuous map. For any pair $(u, v) \in X \times Y$

$$\mathfrak{b}(u + h, v + k) - \mathfrak{b}(u, v) = \mathfrak{b}(h, v) + \mathfrak{b}(u, k) + \mathfrak{b}(h, k).$$

Since $\|\mathfrak{b}(h,k)\| \le C\|h\|\|k\|$, it follows that $\|\mathfrak{b}(h,k)\| = o(\|(h,k)\|)$ and thus

$$\mathfrak{b}'(u,v)[h,k] = \mathfrak{b}(h,v) + \mathfrak{b}(u,k).$$

Now assume that $Y = X$ and that $\mathfrak{b}$ is symmetric. Denoting

$$I(u) := \mathfrak{b}(u,u),$$

we have, for every $u \in X$,

$$I'(u)[h] = 2\mathfrak{b}(u,h).$$

In particular this applies to $I(u) = \|u\|^2$. In this case $I'(u)[v] = 2(u,v)_X$.

As in the finite dimensional case we have the important

Proposition 16.2.6 (Chain Rule) *Let*

(i) X, Y, Z be Hilbert spaces,
(ii) U, V be open sets of X, Y, respectively,
(iii) $I : U \to Y$, $J : V \to Z$ be two maps with $f(U) \subset V$.

If I is differentiable at $u_0 \in U$ and J is differentiable at $v_0 := I(u_0)$, then $J \circ I$ is differentiable at u_0 and we have

$$(J \circ I)'(u_0)[h] = J'(v_0)\left[I'(u_0)[h]\right], \quad \textit{for every } h \in X.$$

A useful tool in order to find the differential of a map is the Gâteaux differential, the analogous of the directional derivative.

Definition 16.2.7 Let $U \subset X$ be an open set, and $I : U \to Y$ a map. The function I is Gâteaux differentiable at $u_0 \in U$ in the direction of $h \in X \setminus \{0\}$, if

$$\exists \lim_{t\to 0} \frac{I(u_0+th) - I(u_0)}{t} \in Y.$$

In that case we write

$$I'_G(u_0)[h] := \lim_{t\to 0} \frac{I(u_0+th) - I(u_0)}{t} = \frac{d}{dt} f(u_0+th)\Big|_{t=0}.$$

Proposition 16.2.8 *Let $I : U \subset X \to Y$ be a map such that*

(i) I is Gâteaux differentiable in U in any direction $h \in X$.
(ii) $I'_G(u) \in \mathcal{L}(X,Y)$, for every $u \in U$.
(iii) $I'_G : U \to \mathcal{L}(X,Y)$ is continuous at u.

Then, I is differentiable at u, and $I'(u) = I'_G(u)$.

By the previous proposition, to calculate the differential of I at $u \in U$, we can find first $I'_G(u)$. Then, if $I'_G(u) \in \mathcal{L}(X, Y)$ and I'_G maps continuously U in $\mathcal{L}(X, Y)$, we have that I is differentiable at u and $I'(u) = I'_G(u)$. This is a standard procedure.

Now we want to extent the notion of partial derivatives of functions defined in $\mathbb{R}^N$.

Definition 16.2.9 Let X, Y, and Z be Hilbert spaces, and consider the map $F : X \times Y \to Z$. We define the maps

$$\varphi_v : X \to Z \quad \text{by} \quad \varphi_v(u) := F(u, v)$$
$$\varphi_u : Y \to Z \quad \text{by} \quad \varphi_u(v) := F(u, v)$$

The partial derivative with respect to u of F at (u_0, v_0) is defined by $\partial_u F(u_0, v_0) := (\varphi_{v_0})'(u_0)$; the partial derivative with respect to v of F at (u_0, v_0) is defined by $\partial_v F(u_0, v_0) := (\varphi_{u_0})'(v_0)$.

Of course higher partial derivatives are defined in a similar way.

Observe that, if F is differentiable in the point (u_0, v_0) in the Hilbert space $X \times Y$ in the sense of Definition 16.2.5, then its partial derivatives exist at (u_0, v_0) and

$$\partial_u F(u_0, v_0)[h] = (\varphi_{v_0})'(u_0)[h] = F'(u_0, v_0)[h, 0] \quad \forall h \in X,$$
$$\partial_v F(u_0, v_0)[k] = (\varphi_{u_0})'(v_0)[k] = F'(u_0, v_0)[0, k] \quad \forall k \in Y.$$

Moreover a sort of "total differential theorem" holds.

Proposition 16.2.10 *Assume that $F : X \times Y \to Z$ has the partial derivatives with respect to u and v in a neighborhood $Q \subset X \times Y$ of (u_0, v_0), and that $\partial_u F$ maps continuously Q in $\mathcal{L}(X, Z)$ and $\partial_v F$ maps continuously Q in $\mathcal{L}(Y, Z)$. Then F is differentiable at (u_0, v_0) and*

$$F'(u_0, v_0)[h, k] = \partial_u F(u_0, v_0)[h] + \partial_v F(u_0, v_0)[k].$$

We state the generalization to Hilbert spaces of the implicit function theorem of calculus in $\mathbb{R}^n$.

Theorem 16.2.11 (Implicit Function Theorem) *Let T, X, and Y be Hilbert spaces; $\Lambda \subset T$, $U \subset X$ open subsets. Assume $F \in C^1(\Lambda \times U, Y)$ and let $(\lambda_0, u_0) \in \Lambda \times U$ be such that $F(\lambda_0, u_0) = 0$. If $F'_u(\lambda_0, u_0) \in \mathcal{L}(X, Y)$ is invertible, then there are neighborhoods Λ_0 of λ_0, U_0 of u_0 and a map $g \in C^1(\Lambda_0, X)$ that satisfy*

$$F(\lambda, u) = 0, \ (\lambda, u) \in \Lambda_0 \times U_0 \quad \Longleftrightarrow \quad u = g(\lambda), \ \textit{for every } \lambda \in \Lambda_0.$$

The next result is about locally invertible functions. If $I \in C(U, Y)$, I is *locally invertible* at $u_0 \in U$ if there are neighborhoods U_0 of u_0, V_0 of $v_0 := I(u_0)$ and a map $J \in C(V_0, U_0)$ such that

$$J(I(u)) = u, \ \forall u \in U_0, \quad \text{and} \quad I(J(v)) = v, \ \forall v \in V_0.$$

The common used notation is $I^{-1} := J$.

Theorem 16.2.12 (Local Inversion Theorem) *Let $I \in C^1(U, Y)$ be a map such that $I'(u_0) \in \mathcal{L}(X, Y)$ is invertible. Then I is locally invertible at u_0, with I^{-1} in $C^1(V_0, U_0)$. Also*

$$(I^{-1})'(v) = (I'(u))^{-1}, \quad \forall v \in V_0, \quad v := I(u).$$

As for the finite dimensional case, we give here some basic facts about functionals on infinite dimensional Hilbert manifolds. We restrict here to the Hilbertian case because this will be the case treated in the next sections. The same theory can be developed on Banach manifold (see, e.g., [54]).

Let us begin by introducing some basic concepts about infinite dimensional manifolds. For more details on this section, the reader may consult [14, Chapter 6, pg. 89].

Definition 16.2.13 Let X, Y be Hilbert spaces, $U \subset X$, $V \subset Y$ open subsets in X, Y, respectively. A map $I : U \to V$ is a C^1 diffeomorphism if I is C^1, bijective, and its inverse I^{-1} is also C^1.

To be coherent with some notations used in Differential Geometry, we will write the differential of I at u also as $dI(u) := I'(u) \in \mathcal{L}(X, Y)$.

Definition 16.2.14 Let E be a Hilbert space and $\mathcal{M}$ a topological space. We say that $\mathcal{M}$ is a C^1 Hilbert manifold modelled on the Hilbert subspace $X \subset E$ if it can be covered with a family of open sets in $\mathcal{M}$, $\{U_i\}_{i \in \mathcal{I}}$ such that

(i) for each $i \in \mathcal{I}$, there is a map $\psi_i : U_i \to X$ such that $V_i := \psi_i(U_i)$ is an open set in X; ψ_i is a homeomorphism onto its image V_i,
(ii) if there are two pairs (U_i, ψ_i), (U_j, ψ_j) with $U_i \cap U_j \neq \emptyset$, then the map $\psi_j \circ \psi_i^{-1} : \psi_i(U_i \cap U_j) \to \psi_j(U_i \cap U_j)$ is C^1 differentiable.

We say that the pair (U_i, ψ_i) is a *chart of the manifold*, the maps $\psi_j \circ \psi_i^{-1}$ are *changes of charts*, and the pair (V_i, ψ_i^{-1}) is a *local parametrization* of $\mathcal{M}$.

Henceforth, we assume that for every point $p \in \mathcal{M}$, there are $\widetilde{U}$, $\widetilde{V}$ open subsets of E and $\widetilde{\varphi} : \widetilde{V} \to \widetilde{U}$ a C^1 diffeomorphism such that

- $p \in \widetilde{U}$,
- defining $U := \widetilde{U} \cap \mathcal{M}$, $V := \widetilde{V} \cap X$ and $\varphi := \widetilde{\varphi}|_V$, we have $x := \varphi^{-1}(p) \in V$ and $U = \varphi(V)$.

Thus the pair (V, φ) is a local parametrization of $\mathcal{M}$.

Definition 16.2.15 Given $p \in \mathcal{M}$, the tangent space to $\mathcal{M}$ at p is defined by

$$T_p\mathcal{M} := d\widetilde{\varphi}(x)[X].$$

The following facts can be proved (see the details in [14, pg. 90–93]):

(i) $T_p\mathcal{M}$ is independent of the local parametrization.
(ii) $T_p\mathcal{M}$ is a Hilbert space homeomorphic to X.
(iii) $T_p\mathcal{M}$ is the set of tangent vectors at p to C^1 curves on $\mathcal{M}$.

Each tangent space $T_p\mathcal{M}$ inherits an inner product from X, that allows us to measure length of smooth curves on $\mathcal{M}$. Moreover if $\mathcal{M}$ is arcwise connected it can be considered as a metric space. From now on we will always assume this.

Let now $\mathcal{M}_i \subset E_i$ be manifolds modelled on the Hilbert spaces $X_i \subset E_i$, respectively, $i = 1, 2$. Suppose that there exist an open subset $U_1 \subset E_1$ and a C^1 map $\widetilde{I} : U_1 \to E_2$ such that

$$\mathcal{M}_1 \subset U_1 \quad \text{and} \quad I = \widetilde{I}|_{\mathcal{M}_1}.$$

Definition 16.2.16 The differential of $I : \mathcal{M}_1 \to \mathcal{M}_2$ at p is defined by

$$d_{\mathcal{M}_1} I(p) := d\widetilde{I}(p)|_{T_p\mathcal{M}_1}.$$

Note that $d\widetilde{I}(p) \in \mathcal{L}(E_1, E_2)$.

The definition does not depend on the map $\widetilde{I}$.

Example Let E be a Hilbert space and let $I : U \subset E \to \mathbb{R}$ be a C^1 functional, with U an open subset of E that contains a manifold $\mathcal{M}$ modelled on the Hilbert subspace X of E. Then given $p \in \mathcal{M}$,

$$d_{\mathcal{M}} I(p) = dI(p)|_{T_p\mathcal{M}}. \tag{16.1}$$

Let us denote by $(\cdot, \cdot)$ the inner product on E. Since $dI(p) \in E'$, the Riesz representation theorem gives the existence of a unique $\nabla I(p) \in E$ such that

$$(\nabla I(p), v) = dI(p)[v], \quad \forall v \in E.$$

$\nabla I(p)$ is called the *gradient* of I at p.

Analogously is defined the *constrained gradient* of I at p:

$$(\nabla_{\mathcal{M}} I(p), v) = d_{\mathcal{M}} I(p)[v], \quad \forall v \in T_p\mathcal{M}. \tag{16.2}$$

Thus, by using (16.1)

$$(\nabla_{\mathcal{M}} I(p), v) = (\nabla I(p), v) \quad \forall v \in T_p\mathcal{M}. \tag{16.3}$$

Therefore, the constraint gradient is just the projection of the (free) gradient on the tangent space $T_p\mathcal{M}$.

Definition 16.2.17 Suppose that $I : E \to \mathbb{R}$ is a C^1 functional and $\mathcal{M}$ is a C^1 manifold modelled on $X \subset E$. The point $z \in \mathcal{M}$ is a *constraint critical point* of I on $\mathcal{M}$ if

$$d_{\mathcal{M}}I(z) = 0, \quad \text{i.e.,} \quad d_{\mathcal{M}}I(z)[v] = 0,\ \forall v \in T_z\mathcal{M}.$$

Thus from (16.2) and (16.3),

$$z \text{ constrained critical point of } I \text{ on } \mathcal{M} \quad \Longrightarrow \quad (\nabla I(z), v) = 0,\ \forall v \in T_z\mathcal{M},$$

in other words, if z is constrained critical point of I on $\mathcal{M}$, then $\nabla I(z)$ is orthogonal to $T_z\mathcal{M}$.

Example Assume that $E, \mathcal{M}, X$, and $I \in C^1(E;\mathbb{R})$ are as in the previous definition. An element $z \in \mathcal{M}$ is a *local constrained minimum* of I on $\mathcal{M}$ if there is a neighborhood Θ of z in E such that

$$I(z) \le I(u), \quad \forall u \in \Theta \cap \mathcal{M}.$$

If (V, φ) is a local parametrization of $\mathcal{M}$ such that $0 \in V$ and $z = \varphi(0)$, this means that for some $W \subset X$ neighborhood of 0,

$$I(\varphi(0)) \le I(\varphi(w)) \quad \forall w \in W.$$

Therefore, z is a local constrained minimum of I on $\mathcal{M}$ if and only if $0 \in X$ is a local minimum of $I \circ \varphi$. Hence

$$0 \text{ is a critical point of the functional } I \circ \varphi : W \subset X \to \mathbb{R}.$$

Using the chain rule, the last statement can be written as

$$dI(z)[d\varphi(0)[w]] = d(I \circ \varphi)(0)[w] = 0,\ \forall w \in X.$$

From $\varphi = \widetilde{\varphi}|_V$ we have

$$d\varphi(0)[X] = d\widetilde{\varphi}(0)[X] = T_z\mathcal{M},$$

so that

$$dI(z)[d\varphi(0)[w]] = 0,\ \forall w \in X \quad \Leftrightarrow \quad dI(z)[v] = 0, \forall v \in T_z\mathcal{M}.$$

Summing up, if z is a local constrained minimum of I on $\mathcal{M}$, then z is a constrained critical point of I on $\mathcal{M}$.

If $\mathcal{M} \subset E$ is modelled on a subspace $X \subset E$ of codimension one, $\mathcal{M}$ is called a manifold of *codimension* one. This happens, for example, when

$$\mathcal{M} := J^{-1}(0), \quad \text{where } J \in C^1(E, \mathbb{R}), \quad \text{and } \nabla J(u) \neq 0,\ \forall u \in \mathcal{M}.$$

Indeed, for $p \in \mathcal{M}$, we define a linear subspace of E,

$$X_p := \{v \in E : (\nabla J(p), v) = 0\}.$$

We have the decomposition $E = X_p \oplus \langle w \rangle$, where

$$w := \frac{\nabla J(p)}{\|\nabla J(p)\|^2}.$$

Thus X_p is a subspace of codimension one. Let us show that $\mathcal{M}$ is a manifold modelled on X_p. Let $\psi : E \to E$ be the map given by

$$\psi(u) := u - p - (\nabla J(p), u - p)w + J(u)w.$$

Taking the inner product with $\nabla J(p)$ and using the fact that

$$(\nabla J(p), w) = \frac{(\nabla J(p), \nabla J(p))}{\|\nabla J(p)\|^2} = 1,$$

we have

$$\begin{aligned}
(\nabla J(p), \psi(u)) &= (\nabla J(p), u - p) - (\nabla J(p), u - p)(\nabla J(p), w) \\
&\quad + J(u)(\nabla J(p), w) \\
&= (\nabla J(p), u - p) - (\nabla J(p), u - p) + J(u) \\
&= J(u),
\end{aligned}$$

and then

$$\psi(u) \in X_p \iff (\nabla J(p), \psi(u)) = 0 \iff J(u) = 0 \iff u \in \mathcal{M}.$$

Note that $\psi(p) = 0$ and that ψ is clearly differentiable. Then,

$$\begin{aligned}
d\psi(u)[v] &= v - (\nabla J(p), v)w + (dJ(u)[v])\, w \\
&= v - (\nabla J(p), v)w + (\nabla J(u), v)w,
\end{aligned}$$

implying that

$$d\psi(p)[v] = v - (\nabla J(p), v)w + (\nabla J(p), v)w = v = id[v].$$

Therefore, by the local inversion theorem, ψ is a diffeomorphism between some $\widetilde{U} \subset E$ neighborhood of p and $\widetilde{V} \subset E$ neighborhood of 0. Let $\widetilde{\varphi} := \psi^{-1}$. Defining

$$U := \widetilde{U} \cap \mathcal{M}, \quad V := \widetilde{V} \cap X_p, \quad \varphi := \widetilde{\varphi}|_V,$$

it holds

$$0 = \varphi^{-1}(p), \quad \text{and} \quad \varphi(V) = U.$$

Then $\mathcal{M}$ is a manifold of codimension one modelled on X, with local parametrization (V, φ) at p. Furthermore

$$d\varphi(0) = d(\psi^{-1})(0) = (d\psi(p))^{-1} = id,$$

and then

$$T_p\mathcal{M} = d\varphi(0)[X_p] = id[X_p] = X_p,$$

or equivalently

$$T_p\mathcal{M} = \{v \in E : (\nabla J(p), v) = 0\}. \tag{16.4}$$

Now let $I \in C^1(E; \mathbb{R})$ be a functional. We know that if $z \in \mathcal{M}$ is a constrained critical point of I on $\mathcal{M}$, then

$$(\nabla I(z), v) = 0, \quad \forall v \in T_z\mathcal{M},$$

and hence by (16.4), there exists $\lambda \in \mathbb{R}$ that satisfies

$$\nabla I(z) = \lambda \nabla J(z). \tag{16.5}$$

Recalling that the constrained gradient is the projection onto $T_p\mathcal{M}$ of the free gradient,

$$\nabla_{\mathcal{M}} I(p) = \nabla I(p) - (\nabla I(p), w)w = \nabla I(p) - \frac{(\nabla I(p), \nabla J(p))}{\|\nabla J(p)\|^2}\nabla J(p).$$

Then if z is a constrained critical point, $\nabla_{\mathcal{M}} I(z) = 0$ and (16.5) holds with

$$\lambda = \frac{(\nabla I(p), \nabla J(p))}{\|\nabla J(p)\|^2}.$$

Definition 16.2.18 Given a functional $I \in C^1(E; \mathbb{R})$, we say that a manifold $\mathcal{M}$ is a *natural constraint* for I if there exists a functional $\widetilde{I} \in C^1(E; \mathbb{R})$ such that all the critical points of $\widetilde{I}$ constrained on $\mathcal{M}$ are critical points of I, i.e.

$$z \in \mathcal{M}, \ \nabla_{\mathcal{M}} \widetilde{I}(z) = 0 \quad \Longleftrightarrow \quad \nabla I(z) = 0.$$

Definition 16.2.19 Let $I \in C^1(E; \mathbb{R})$ be a functional. The Nehari set associated with I is given by

$$\mathcal{N} := \{u \in E \setminus \{0\} : I'(u)[u] = 0\}.$$

Proposition 16.2.20 *Assume that $I \in C^2(E; \mathbb{R})$ and that the Nehari set $\mathcal{N}$ is nonempty. Furthermore, suppose that*

$$I''(u)[u, u] \neq 0, \ \forall u \in \mathcal{N}. \tag{16.6}$$

Then, $\mathcal{N}$ is a manifold (called the Nehari manifold*), and it is a natural constraint for I.*

Proof Defining $J(u) := I'(u)[u]$ we see that

(i) $J \in C^1(E, \mathbb{R})$,
(ii) if $u \in \mathcal{N}$, by (16.6)

$$J'(u)[u] = I''(u)[u, u] + I'(u)[u] = I''(u)[u, u] \neq 0.$$

Hence $J'(u) \neq 0$ for $u \in \mathcal{N}$.

So $\mathcal{N}$ is a manifold of codimension one. If now $z \in \mathcal{N}$ is a constrained critical point of I on $\mathcal{N}$, then for some $\lambda \in \mathbb{R}$ we have

$$\nabla I(z) = \lambda \nabla J(z).$$

So

$$0 = I'(z)[z] = (\nabla I(z), z) = \lambda(\nabla J(z), z) = \lambda J'(z)[z], \quad \text{with } J'(z)[z] \neq 0,$$

and thus $\lambda = 0$. Thence $\nabla I(z) = 0$, i.e., z is a critical point of the (free) functional I.

If $z \neq 0$ is a critical point of I, then

$$\nabla I(z) = 0 \quad \Leftrightarrow \quad I'(z) = 0.$$

In particular, $I'(z)[z] = 0$ and clearly $z \in \mathcal{N}$. Also it is clear that $\nabla_{\mathcal{N}} I(z) = 0$, completing the proof. □

The main objective in the Theory of Critical Points is to find conditions under which a functional possesses many critical points, since in many cases they are related to weak solutions of partial differential equations. We will show in Sect. 16.2.4 an application to a Schrödinger type equation.

16.2.3 The Ljusternick-Schnirelmann Category

The Ljusternick-Schnirelmann theory is a key ingredient in order to obtain multiplicity results of critical points of functionals, and then solutions for variational equations. The theory is based on suitable topological properties. We recall here the main elements of the theory. For details, we refer the reader to [14, Chapter 9].

Definition 16.2.21 Let $\mathcal{M}$ be a topological space; $A \subset \mathcal{M}$ is contractible if the inclusion $A \hookrightarrow \mathcal{M}$ is homotopic to a constant map defined on A with value in $\mathcal{M}$, i.e., there is $H \in C([0, 1] \times A, \mathcal{M})$ such that

$$H(0, u) = u, \quad \text{and} \quad H(1, u) = p, \quad \forall u \in A; \; p \text{ is a fixed element of } \mathcal{M}.$$

In this section, we assume that E is a Hilbert space, $\mathcal{M} \subset E$ is a manifold given by

$$\mathcal{M} = J^{-1}(0), \quad \text{with } J \in C^1(E; \mathbb{R}) \text{ and } \quad J'(u) \neq 0, \; \forall u \in \mathcal{M}. \tag{16.7}$$

Definition 16.2.22 Let $\mathcal{M}$ be a topological space. The Ljusternick-Schnirelmann category (or LS-category, or simply the category) of A with respect to $\mathcal{M}$ is defined by

$$\operatorname{cat}_{\mathcal{M}}(A) := \begin{cases} 0, & \text{if } A = \emptyset, \\ k, & \text{if } k \text{ is the least natural number } l \text{ in the property } (C), \end{cases}$$

where

(C) There are $A_1, \ldots, A_l$ closed, contractible subsets in $\mathcal{M}$ with $A \subset A_1 \cup \cdots \cup A_l$.

If $A \neq \emptyset$ and A does not satisfy property (C), we set $\operatorname{cat}_{\mathcal{M}}(A) = \infty$.

We usually set $\operatorname{cat} \mathcal{M} := \operatorname{cat}_{\mathcal{M}}(\mathcal{M})$.

From the definition it holds:

1. if $A \subset B$ are subsets of $\mathcal{M}$, $\operatorname{cat}_{\mathcal{M}}(A) \leq \operatorname{cat}_{\mathcal{M}}(B)$.
   Indeed any covering of B by subsets of $\mathcal{M}$ is also a covering of A by subsets of $\mathcal{M}$.
2. $\operatorname{cat}_{\mathcal{M}}(A) = \operatorname{cat}_{\mathcal{M}}(\overline{A})$.
   Assume that $A \neq \emptyset$, $\operatorname{cat}_{\mathcal{M}}(A) = k < +\infty$, otherwise the statement is trivial. Therefore, there are $A_1, \ldots, A_k$ closed, contractible sets in $\mathcal{M}$ with

$$A \subset A_1 \cup \cdots \cup A_k$$

Since $A_1 \cup \cdots \cup A_k$ is closed in $\mathcal{M}$ and it contains A,

$$\overline{A} \subset A_1 \cup \cdots \cup A_k,$$

and thus

$$\operatorname{cat}_{\mathcal{M}}(\overline{A}) \le k = \operatorname{cat}_{\mathcal{M}}(A) \le \operatorname{cat}_{\mathcal{M}}(\overline{A}).$$

3. $A \subset \mathcal{M} \subset \mathcal{N}$ with $\mathcal{M}$ closed in $\mathcal{N}$ implies $\operatorname{cat}_{\mathcal{M}}(A) \ge \operatorname{cat}_{\mathcal{N}}(A)$.

Definition 16.2.23 Let $\mathcal{M}$ be a topological space. A deformation of $A \subset \mathcal{M}$ in $\mathcal{M}$ is a map $\eta \in C(A; \mathcal{M})$ homotopic to the inclusion $A \hookrightarrow \mathcal{M}$, i.e., there exists $H \in C([0,1] \times A; \mathcal{M})$ such that

$$H(0,u) = u, \quad H(1,u) = \eta(u), \quad \forall u \in A.$$

Through this section, the following topological fact will be used in some arguments (see [51, Lemma 18.3, pg. 106]).

Lemma 16.2.24 (Gluing Lemma) *Let X, Y be topological spaces. Assume that $X = A \cup B$, where A, B are closed. Let $f : A \to Y$ and $g : B \to Y$ be continuous. If $f(x) = g(x)$, for every $x \in A \cap B$, then we can define a continuous function $h : X \to Y$ using f and g, namely by putting $h(x) := f(x)$, for every $x \in A$, and $h(x) := g(x)$, for every $x \in B$.*

Lemma 16.2.25 *Let A, B be subsets of $\mathcal{M}$.*

1. *if $A \subset B$ then* $\operatorname{cat}_{\mathcal{M}}(A) \le \operatorname{cat}_{\mathcal{M}}(B)$
2. $\operatorname{cat}_{\mathcal{M}}(A \cup B) \le \operatorname{cat}_{\mathcal{M}}(A) + \operatorname{cat}_{\mathcal{M}}(B)$
3. *Let A be closed in $\mathcal{M}$, η a deformation of A in $\mathcal{M}$. Then* $\operatorname{cat}_{\mathcal{M}}(A) \le \operatorname{cat}_{\mathcal{M}}\left(\overline{\eta(A)}\right)$

Proof

1. The monotonicity of cat was already proved.
2. If $\operatorname{cat}_{\mathcal{M}}(A) = \infty$ or $\operatorname{cat}_{\mathcal{M}}(B) = \infty$, the result is immediate. Let us assume that $\operatorname{cat}_{\mathcal{M}}(A) = k$ and $\operatorname{cat}_{\mathcal{M}}(B) = l$. If $A \subset A_1 \cup \cdots \cup A_k$ and $B \subset B_1 \cup \cdots \cup B_l$, with A_i, B_j closed, contractible in $\mathcal{M}$, then
$$A \cup B \subset (A_1 \cup \cdots \cup A_k) \cup (B_1 \cup \cdots \cup B_l),$$
and the item follows.
3. We assume that $\operatorname{cat}_{\mathcal{M}}(\eta(A)) = k < +\infty$, so that there are $B_1, \ldots, B_k$ closed, contractible subsets in $\mathcal{M}$ such that
$$\eta(A) \subset B_1 \cup \cdots \cup B_k.$$

Taking $A_i := \eta^{-1}(B_i)$, A_i is closed in A, and since A is closed in $\mathcal{M}$, it is also closed in $\mathcal{M}$. Furthermore

$$A \subset A_1 \cup \cdots \cup A_k.$$

We claim that A_i is contractible in $\mathcal{M}$. Indeed, since η is a deformation of A in $\mathcal{M}$ and B_i contractible in $\mathcal{M}$, we have, respectively, maps $H \in C([0,1]\times A; \mathcal{M})$ with

$$H(0,u) = u, \quad H(1,u) = \eta(u), \quad \forall u \in A,$$

and $H_i \in C([0,1] \times B_i; \mathcal{M})$ such that

$$H_i(0,v) = v, \quad H_i(1,v) = p_i, \quad \forall v \in B_i,\ p_i \text{ fixed element in } \mathcal{M}.$$

Consider the map $\widetilde{H}_i : [0,1] \times A_i \to \mathcal{M}$ given by

$$\widetilde{H}_i(t,u) := \begin{cases} H(2t,u), & \text{for } 0 \le t \le 1/2 \\ H_i(2t-1, \eta(u)), & \text{for } 1/2 \le t \le 1, \end{cases}$$

for every $u \in A_i$. By the Gluing Lemma $\widetilde{H}_i \in C([0,1] \times \mathcal{M})$ and satisfies

$$\widetilde{H}_i(0,u) = u, \quad \widetilde{H}_i(1,v) = p_i, \quad \forall u \in A_i.$$

Hence A_i is contractible in $\mathcal{M}$, from which

$$\text{cat}_{\mathcal{M}}(A) \le k = \text{cat}_{\mathcal{M}}(\eta(A)) = \text{cat}_{\mathcal{M}}\left(\overline{\eta(A)}\right)$$

concluding the proof. □

Proposition 16.2.26 (Dominating Property) *Let X and Y be topological spaces, let $f : X \to Y$ and $g : Y \to X$ be continuous functions such that $g \circ f$ is homotopic to Id_X, then* $\text{cat}(X) \le \text{cat}(Y)$.

The reader may consult, e.g., [43].

Definition 16.2.27 (Extension Property) Let $\mathcal{M}$ be a metric space. $\mathcal{M}$ satisfies the extension property if for every metric space $\mathcal{Y}$, every subset S closed in $\mathcal{Y}$ and every continuous map $f : S \to \mathcal{M}$, there are N a neighborhood of S in $\mathcal{Y}$ and a map $\widetilde{f} \in C(N; \mathcal{M})$ such that $\widetilde{f}|_S = f$.

For the next result, see [51, Exercise 2, pg. 177].

Proposition 16.2.28 *Let $\mathcal{X}$ be a metric space and $A \subset \mathcal{X}$. The ε-neighborhood of A in $\mathcal{X}$ is defined as*

$$U(A, \varepsilon) := \{x \in \mathcal{X} : d(x, A) < \varepsilon\}.$$

1. $U(A, \varepsilon)$ is the union of the open balls $B_d(a, \varepsilon)$, for $a \in A$. Here

$$B_d(a, \varepsilon) := \{y \in \mathcal{X} : d(a, y) < \varepsilon\}$$

2. If A is compact and U is an open set in $\mathcal{X}$ containing A, there is $\varepsilon > 0$ such that the ε-neighborhood $U(A, \varepsilon)$ of A is contained in U.

Lemma 16.2.29 *Let $\mathcal{M}$ be a metric space with the extension property and let $A \subset \mathcal{M}$ be a compact subset. Then*

(1) $\text{cat}_{\mathcal{M}}(A) < \infty$.
(2) There exists U_A neighborhood of A in $\mathcal{M}$ such that $\text{cat}_{\mathcal{M}}(\overline{U}_A) = \text{cat}_{\mathcal{M}}(A)$.

Proof We divide the proof in steps.

Step 1 For $A \subset \mathcal{M}$ compact with $\text{cat}_{\mathcal{M}}(A) = 1$, there is U_A neighborhood of A in $\mathcal{M}$ such that $\text{cat}_{\mathcal{M}}(\overline{U}_A) = 1$.

Since $\text{cat}_{\mathcal{M}}(A) = 1$ we have $A \subset B$, with B closed, contractible in $\mathcal{M}$ and also a homotopy $H_1 : [0, 1] \times B \to \mathcal{M}$ with

$$H_1(0, u) = u, \quad \text{and} \quad H_1(1, u) = p, \quad \text{for every } u \in B;\ p \text{ fixed element in } \mathcal{M}.$$

In particular A is contractible, considering $H_2 := H_1|_{[0,1]\times A} \in C([0, 1] \times A; \mathcal{M})$ such that

$$H_2(0, u) = u, \quad \text{and} \quad H_2(1, u) = p, \quad \text{for every } u \in A.$$

Note that $\mathcal{Y} := [0, 1] \times \mathcal{M}$ is a metric space, since $\mathcal{M}$ is a metric space. Clearly the set

$$S := (\{0\} \times \mathcal{M}) \cup ([0, 1] \times A) \cup (\{1\} \times \mathcal{M})$$

is closed in $\mathcal{Y}$. By the pasting lemma the map $H : S \to \mathcal{M}$ given by

$$H(t, u) := \begin{cases} u, & \text{for } t = 0,\ u \in \mathcal{M} \\ H_2(t, u), & \text{for } t \in [0, 1],\ u \in A \\ p, & \text{for } t = 1,\ u \in \mathcal{M} \end{cases}$$

satisfies $H \in C(S; \mathcal{M})$. As a consequence of the extension property, there is N neighborhood of S in $\mathcal{Y}$ and a map $\widetilde{H} \in C(N; \mathcal{M})$ such that $\widetilde{H}|_S = H$. Note that

$[0, 1] \times A$ is compact, and $[0, 1] \times A \subset N$ open in $\mathcal{Y}$. Using the Proposition 16.2.28, there is U_A neighborhood of A in $\mathcal{M}$ with $[0, 1] \times \overline{U}_A \subset N$. Furthermore

$$\widetilde{H}(0, u) = H(0, u) = u, \quad \text{and} \quad \widetilde{H}(1, u) = H(1, u) = p, \quad \text{for every } u \in \overline{U}_A$$

Thus $\overline{U}_A$ is contractible in $\mathcal{M}$ and thence $\text{cat}_{\mathcal{M}}(\overline{U}_A) = 1$.

Step 2 Since for every $q \in \mathcal{M}$, $\text{cat}_{\mathcal{M}}(\{q\}) = 1$, as a particular case of Step 1 there exists a neighborhood U_q of q in $\mathcal{M}$ with $\overline{U}_q$ contractible in $\mathcal{M}$, so that $\text{cat}_{\mathcal{M}}(\overline{U}_q) = 1$.

Step 3 Proofs of items (1) and (2) of the lemma.

Since A is compact, we can find points $q_1, \ldots, q_k \in \mathcal{M}$ and the corresponding neighborhoods $U_{q_1}, \ldots, U_{q_k}$ with $\overline{U}_{q_1}, \ldots, \overline{U}_{q_k}$ contractibles in $\mathcal{M}$, such that

$$A \subset \overline{U}_{q_1} \cup \cdots \cup \overline{U}_{q_k},$$

from which

$$\text{cat}_{\mathcal{M}}(A) \leq \text{cat}_{\mathcal{M}} \left(\overline{U}_{q_1} \cup \cdots \cup \overline{U}_{q_k}\right) \leq \text{cat}_{\mathcal{M}}(\overline{U}_{q_1}) + \cdots \text{cat}_{\mathcal{M}}(\overline{U}_{q_k}) = k,$$

and item (1) of lemma follows.

Now assume that $\text{cat}_{\mathcal{M}}(A) = k$. Then there exist $A_1, \ldots, A_k$ closed, contractible sets in $\mathcal{M}$ with $A \subset A_1 \cup \cdots \cup A_k$. Defining $B_i := A_i \cap A$, we have $A \subset B_1 \cup \cdots \cup B_k$, with each B_i compact. Using Step 1, for each i there is U_i neighborhood of B_i in $\mathcal{M}$ with $\overline{U}_i$ contractible in $\mathcal{M}$. Defining

$$U_A := U_1 \cup \cdots \cup U_k,$$

it is clear that $\overline{U}_A = \overline{U}_1 \cup \cdots \cup \overline{U}_k$ and hence

$$\text{cat}_{\mathcal{M}}(\overline{U}_A) \leq k = \text{cat}_{\mathcal{M}}(A).$$

On the other hand, since $A \subset \overline{U}_A$, we have $\text{cat}_{\mathcal{M}}(A) \leq \text{cat}_{\mathcal{M}}(\overline{U}_A)$, and item (2) holds true, finishing the proof. □

Definition 16.2.30 Let $\mathcal{M}$ be a manifold of the type (16.7). We define

$$\text{cat}_k(\mathcal{M}) := \sup\{\text{cat}_{\mathcal{M}}(A) : A \subset \mathcal{M}, \ A \text{ compact}\}.$$

Of course if $\mathcal{M}$ is compact, $\text{cat}_k(\mathcal{M}) = \text{cat}\, \mathcal{M}$, by the monotonicity property of the category.

Let us consider natural numbers m such that

$$m \in \begin{cases} \{1, \ldots, \text{cat}_k(\mathcal{M})\}, & \text{if } \text{cat}_k(\mathcal{M}) < \infty \\ \mathbb{N}, & \text{if } \text{cat}_k(\mathcal{M}) = \infty \end{cases} \tag{16.8}$$

and the collection

$$\mathscr{C}_m := \{A \subset \mathcal{M} : A \text{ is compact and } \operatorname{cat}_{\mathcal{M}}(A) \geq m\}.$$

Observe that $\mathscr{C}_m \neq \emptyset$; indeed by the definition of cat_k, when $\operatorname{cat}_k(\mathcal{M}) < \infty$, $\operatorname{cat}_k(\mathcal{M}) = \operatorname{cat}_{\mathcal{M}}(K)$, for some compact subset K in $\mathcal{M}$, and thus $K \in \mathscr{C}_m$, for every $m = 1, \ldots, \operatorname{cat}_k(\mathcal{M})$. When $\operatorname{cat}_k(\mathcal{M}) = \infty$, for every $m \in \mathbb{N}$, there is $A_m \subset \mathcal{M}$ compact such that $m \leq \operatorname{cat}_{\mathcal{M}}(A_m)$, from which $A_m \in \mathscr{C}_m$.

Now, for $I \in C(\mathcal{M}, \mathbb{R})$ and every m satisfying (16.8), we define the minimax levels of I

$$c_m := \inf_{A \in \mathscr{C}_m} \max_{u \in A} I(u).$$

We have:

(i) $c_1 = \inf_{\mathcal{M}} I$.
In fact, for every $u \in \mathcal{M}$, $\operatorname{cat}_{\mathcal{M}}(\{u\}) = 1$, and being $\{u\}$ is compact, we infer $\{u\} \in \mathscr{C}_1$. Thus $c_1 \leq I(u)$, for every $u \in \mathcal{M}$, and so

$$c_1 \leq \inf_{u \in \mathcal{M}} I(u).$$

On the other hand, $\inf_{u \in \mathcal{M}} I(u) \leq I(v)$, for every $v \in \mathcal{M}$. In particular, for any $v \in A$, with $A \in \mathscr{C}_m$, $\inf_{u \in \mathcal{M}} I(u) \leq \max_{v \in A} I(v)$. Hence

$$\inf_{u \in \mathcal{M}} I(u) \leq c_1.$$

(ii) $c_1 \leq c_2 \leq \ldots \leq c_{m-1} \leq c_m \leq \ldots$
This follows from the fact that $\mathscr{C}_m \subset \mathscr{C}_{m-1}$.

(iii) $c_m < \infty$, for every m satisfying (16.8).
For any $A \in \mathscr{C}_m$,

$$c_m \leq \max_{u \in A} I(u) < \infty,$$

since A is compact and $I \in C(\mathcal{M}; \mathbb{R})$.

(iv) $\inf_{\mathcal{M}} I > -\infty \;\Rightarrow\; c_m \in \mathbb{R}$, i.e., $-\infty < c_1 \leq c_m < \infty$:
Immediate from the hypothesis, in virtue of items (i)–(iii).

Remark 16.2.31 Observe that the class $\mathscr{C}_m$ is invariant by deformations, in the following sense. Let η be a deformation in $\mathcal{M}$. By the item (3) of Lemma 16.2.25, for any $A \in \mathscr{C}_m$,

$$m \leq \operatorname{cat}_{\mathcal{M}}(A) \leq \operatorname{cat}_{\mathcal{M}}(\eta(A)).$$

Furthermore, since η is continuous and $\eta(A)$ is compact, we have $\eta(A) \in \mathscr{C}_m$. Summing up,

$$A \in \mathscr{C}_m,\ \eta \text{ deformation in } \mathcal{M} \quad \Rightarrow \quad \eta(A) \in \mathscr{C}_m \tag{16.9}$$

The next definition is a useful tool in Critical Point Theory.

Definition 16.2.32 Let $\mathcal{M}$ be a differentiable manifold in the Hilbert space X (eventually we may have $\mathcal{M} = X$). A sequence $\{u_n\} \subset \mathcal{M}$ is said a *Palais-Smale* sequence (PS sequence) on $\mathcal{M}$ for I if $\{I(u_n)\}$ is bounded and $\nabla_{\mathcal{M}} I(u_n) \to 0$.

It is a $(PS)_c$ sequence if $I(u_n) \to c$ and $\nabla_{\mathcal{M}} I(u_n) \to 0$.

We say that I satisfies the $(PS)_c$ condition on $\mathcal{M}$ if every $(PS)_c$ sequence has a convergent subsequence. Analogously we say that I satisfies the (PS) condition if every PS sequence has a convergent subsequence.

Now we can sketch the proof of the main theorem of the Ljusternick-Schnirelmann theory. We first establish some notation and state an auxiliary lemma concerning deformations (see [14, Lemma 7.10, pg. 108 and Lemma 9.12, pg. 151]).

- given $a \in \mathbb{R}$, $\mathcal{M}^a := \{p \in \mathcal{M} : I(p) \le a\}$, the sublevel of I on $\mathcal{M}$ under the level a.
- $Z := \{z \in \mathcal{M} : \nabla_{\mathcal{M}} I(z) = 0\}$, the set of critical points of I on $\mathcal{M}$.
- $Z_c := \{z \in Z : I(z) = c\}$. c is a critical level of I on $\mathcal{M}$ if $Z_c \neq \emptyset$. Note that Z_c is compact if $(PS)_c$ is satisfied.

Lemma 16.2.33 *Let $\mathcal{M}$ be a manifold of the type (16.7).*

(i) Suppose that $c \in \mathbb{R}$ is not a critical level of I on $\mathcal{M}$ and that $(PS)_c$ holds. Then, there exist $\delta > 0$ and a deformation η in $\mathcal{M}$ such that $\eta(\mathcal{M}^{c+\delta}) \subset \mathcal{M}^{c-\delta}$.

(ii) Assume that $(PS)_c$ holds. For every neighborhood U of Z_c, there are $\delta > 0$ and a deformation $\eta \in \mathcal{M}$ such that

$$\eta(\mathcal{M}^{c+\delta} \setminus U) \subset \mathcal{M}^{c-\delta}.$$

Theorem 16.2.34 (Ljusternick-Schnirelmann Theorem) *Let $\mathcal{M}$ be a manifold of the type (16.7) and $I \in C^1(E, \mathbb{R})$ bounded from below on $\mathcal{M}$, satisfying (PS). Then I has at least* $\text{cat}_k(\mathcal{M})$ *critical points on $\mathcal{M}$. More precisely:*

(1) any c_m is a critical level of I on $\mathcal{M}$.

(2) Suppose that $c := c_m = c_{m+1} = \cdots = c_{m+l}$, for some $l \in \mathbb{N}$. Then $\text{cat}_{\mathcal{M}}(Z_c) \ge l + 1$.

Proof Note that $c_m \in \mathbb{R}$, since I is bounded from below on $\mathcal{M}$. We argue by contradiction.

Suppose that $Z_{c_m} = \emptyset$. Since I is bounded from below on $\mathcal{M}$ and satisfies (PS), by item (i) of Lemma 16.2.33, there are $\delta > 0$ and η a deformation of $\mathcal{M}$ with

$$\eta(\mathcal{M}^{c_m+\delta}) \subset \mathcal{M}^{c_m-\delta}.$$

Also, by the definition of c_m, there exists $A_m \in \mathscr{C}_m$ such that

$$I(u) \leq \max_{v \in A_m} I(v) \leq c_m + \delta, \quad \forall u \in A_m,$$

so that $A_m \subset \mathcal{M}^{c_m+\delta}$. Therefore

$$\eta(A_m) \subset \mathcal{M}^{c_m-\delta}. \tag{16.10}$$

But by (16.9), $\eta(A_m) \in \mathscr{C}_m$, from which, for any $v \in \eta(A_m)$,

$$c_m \leq I(v) \leq \max_{v \in \eta(A_m)} I(v),$$

in contradiction with (16.10), and this proves (1).

Let us assume that $\text{cat}_{\mathcal{M}}(Z_c) \leq l$. Since Z_c is compact, by Lemma 16.2.29, there is U neighborhood of Z_c with

$$\text{cat}_{\mathcal{M}}(\overline{U}) = \text{cat}_{\mathcal{M}}(Z_c) \leq l.$$

Now by item (ii) of Lemma 16.2.33, there exist $\delta > 0$ and η a deformation in $\mathcal{M}$ with

$$\eta(\mathcal{M}^{c+\delta} \setminus U) \subset \mathcal{M}^{c-\delta} \tag{16.11}$$

Since $c = c_{m+l}$, there is $A_c \in \mathscr{C}_{m+l}$ such that

$$I(u) \leq \max_{v \in A_c} I(v) \leq c + \delta, \quad \forall u \in A_c,$$

and thus

$$A_c \subset \mathcal{M}^{c+\delta} \tag{16.12}$$

Now define $B_c := \overline{A_c \setminus U}$. Using the monotonicity property of cat,

$$\overline{A}_c \subset \overline{U} \cup B_c \quad \Rightarrow \quad \text{cat}_{\mathcal{M}}(A_c) = \text{cat}_{\mathcal{M}}(\overline{A}_c) \leq \text{cat}_{\mathcal{M}}(\overline{U}) + \text{cat}_{\mathcal{M}}(B_c),$$

and thence

$$\text{cat}_{\mathcal{M}}(B_c) \geq \text{cat}_{\mathcal{M}}(A_c) - \text{cat}_{\mathcal{M}}(\overline{U}) \geq m + l - l = m.$$

As a consequence, $B_c \in \mathscr{C}_m$, and by (16.9), $\eta(B_c) \in \mathscr{C}_m$. In particular, for any $v \in \eta(B_c)$,

$$c = c_m \leq \max_{v \in \eta(B_c)} I(v).$$

But on the other hand, it follows from (16.11) and (16.12) that

$$\eta(B_c) \subset \mathcal{M}^{c-\delta},$$

so that $I(v) \leq c - \delta$, a contradiction which proves (2). □

Remark 16.2.35 Observe the following facts:

- In Lemma 16.2.33 item (ii) includes item (i), as well as in the Theorem of Ljusternick-Schnirelmann item (2) includes item (1). Indeed, in the Lemma, if c is not a critical point of I on $\mathcal{M}$, $U = \emptyset$ is a neighborhood of $Z_c = \emptyset$. In the Theorem, item (1) is included in item (2) with $l = 0$; when $\text{cat}_{\mathcal{M}}(Z_c) \geq 1$, we have $Z_c \neq \emptyset$.
- Item (2) of the Ljusternick-Schnirelmann theorem implies that there are non-countable critical points of I on $\mathcal{M}$ at level c. Indeed, since

$$A \text{ discrete set } \implies \text{cat}_{\mathcal{M}}(A) = 1,$$

  if item (2) holds true, Z_c cannot be countable.
- The proof of the theorem can be done if we have a weaker restriction than (PS); namely, if the minimax levels satisfy

$$c_m < b, \qquad \text{for every } m,$$

  then we only need that $(PS)_c$ holds for any $c < b$.

16.2.4 Schrödinger Type Equations

As already said, in the next sections we will be interested in finding solutions for two kinds of Schrödinger type equations: (1) one with a given external potential, and (2) one with un unknown potential. We see here how to define the functional and using some of the previous abstract results to the slightly more difficult case of unknown potential, that physically speaking, can be interpreted as related to the motion of the charged particle. The case of external and fixed potential can be easily obtained by repeating the same arguments.

Without entering in physical details here (the interested reader is referred to the pioneering paper [19]), the search of standing wave solutions $\psi(x,t) = u(x)e^{i\omega t}$ leads us to consider the following system

$$\begin{cases} -\Delta u + u + \phi u = |u|^{p-2}u & \text{in } \mathbb{R}^3 \\ -\Delta \phi = u^2 & \text{in } \mathbb{R}^3, \end{cases} \tag{16.13}$$

with $2 < p < 6$. Here the nonlinearity $f(u) = |u|^{p-2}u$ simulates the interaction between many particles. However a more general nonlinearity can be (and indeed will be) considered in Sect. 16.4. Observe that the potential ϕ is related to the other unknown u by the second equation.

For a (weak) solution of the previous system we mean a pair $(u, \phi) \in H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3)$ that satisfies

$$\int_{\mathbb{R}^3} \nabla u \nabla v + \int_{\mathbb{R}^3} uv + \int_{\mathbb{R}^3} \phi uv = \int_{\mathbb{R}^3} |u|^{p-2}uv,$$
$$\int_{\mathbb{R}^3} \nabla\phi\nabla\psi = \int_{\mathbb{R}^3} u^2\psi,$$

for every pair $(v, \psi) \in H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3)$. Let us define the functional $F : H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3) \to \mathbb{R}$ by

$$F(u, \phi) := \frac{1}{2}\int_{\mathbb{R}^3} |\nabla u|^2 + \frac{1}{2}\int_{\mathbb{R}^3} (1+\phi)u^2 - \frac{1}{p}\int_{\mathbb{R}^3} |u|^p - \frac{1}{4}\int_{\mathbb{R}^3} |\nabla\phi|^2.$$

Proposition 16.2.36 *The functional F is C^1 and its critical points are the solutions of* (16.13).

Proof Consider the functional

$$L(u, \phi) := \frac{1}{2}\int_{\mathbb{R}^3} \phi u^2$$

and let us show it is C^1, by proving it has continuous partial derivatives, by using Proposition 16.2.8.

First note that the map $L(u, \cdot) : D^{1,2}(\mathbb{R}^3) \to \mathbb{R}$ is linear and continuous because

$$\int_{\mathbb{R}^3} \phi u^2 \le |\phi|_6 |u^2|_{6/5} = |\phi|_6 |u|^2_{12/5} \le C\|u\|^2\|\phi\|_D. \tag{16.14}$$

Therefore

$$\partial_\phi L(u, \phi)[\psi] = L(u, \psi) = \frac{1}{2}\int_{\mathbb{R}^3} \psi u^2, \quad \forall \psi \in D^{1,2}(\mathbb{R}^3).$$

For the partial derivative $\partial_u L$ We have

$$(L'_u)_G(u, \phi)[v] := \frac{d}{dt}L(u+tv, \phi)\Big|_{t=0} = \int_{\mathbb{R}^3} \phi uv,$$

and since

$$\int_{\mathbb{R}^3} \phi uv \leq |\phi|_6 |u|_{12/5} |v|_{12/5} \leq C\|u\|\|v\|\|\phi\|_D, \tag{16.15}$$

it follows that

$$(L'_u)_G(u,\phi) \in H^{-1} \quad \text{and} \quad (L'_u)_G(\,\cdot\,,\phi) : H^1(\mathbb{R}^3) \to H^{-1} \text{ is continuous.}$$

Hence L'_u exists at (u,ϕ) and

$$\partial_u L(u,\phi) = (L'_u)_G(u,\phi).$$

To sum up, L has partial derivatives with respect to u and ϕ at any point (u,ϕ).

Let us see that the maps

$$\partial_u L : H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3) \to H^{-1} \qquad \text{and} \qquad \partial_\phi L : H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3) \to D'$$

are continuous. Assume that $(u_n,\phi_n) \to (u,\phi)$ in $H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3)$. For the continuity of L_u, note that if $v \in H^1(\mathbb{R}^3)$ with $\|v\| \leq 1$,

$$\begin{aligned}\Big|(L_u(u_n,\phi_n) - L_u(u,\phi))[v]\Big| &\leq \int_{\mathbb{R}^3} |(\phi_n u_n - \phi u)v| \\ &\leq \int_{\mathbb{R}^3} |\phi_n - \phi||u_n||v| \\ &\quad + \int_{\mathbb{R}^3} |\phi||u_n - u||v|.\end{aligned}$$

Using the Hölder inequality and the Sobolev embeddings

$$\begin{aligned}\int_{\mathbb{R}^3} |\phi||u_n - u||v| &\leq |\phi|_6 |u_n - u|_{12/5} |v|_{12/5} \leq C\|\phi\|_D \|u_n - u\|\|v\| \\ &\leq C\|u_n - u\|.\end{aligned}$$

On the other hand, since $\{u_n\}$ is bounded in $H^1(\mathbb{R}^3)$

$$\begin{aligned}\int_{\mathbb{R}^3} |\phi_n - \phi||u_n||v| &\leq |\phi_n - \phi|_6 |u_n|_{12/5} |v|_{12/5} \leq C\|\phi_n - \phi\|_D \|u_n\|\|v\| \\ &\leq C\|\phi_n - \phi\|_D.\end{aligned}$$

Therefore

$$\|\partial_u L(u_n,\phi_n) - \partial_u L(u,\phi)\|_\star \leq C(\|\phi_n - \phi\|_D + \|u_n - u\|) = o_n(1). \tag{16.16}$$

For the continuity of $\partial_\phi L$, taking an element $\psi \in D^{1,2}(\mathbb{R}^3)$ with $\|\psi\|_D \leq 1$,

$$\begin{aligned}\left|(\partial_\phi L(u_n, \phi_n) - \partial_\phi L(u, \phi))[\psi]\right| &\leq \int_{\mathbb{R}^3} |\psi||u_n^2 - u^2| \\ &\leq |u_n^2 - u^2|_{6/5}|\psi|_6 \\ &\leq C|u_n^2 - u^2|_{6/5}\|\psi\|_D \\ &\leq C|u_n^2 - u^2|_{6/5}.\end{aligned}$$

Since $u_n \to u$ in $H^1(\mathbb{R}^3)$, up to a subsequence

$$\begin{aligned} u_n \to u \text{ in } L^{12/5}(\mathbb{R}^3) \quad &\Longrightarrow \quad u_n(x) \to u(x) \text{ a.e. } x \in \mathbb{R}^3 \\ u_n \to u \text{ in } L^{12/5}(\mathbb{R}^3) \quad &\Longrightarrow \quad |u_n^2|_{6/5} = |u_n|_{12/5}^2 \to |u|_{12/5}^2 = |u^2|_{6/5},\end{aligned}$$

and by the Brezis-Lieb lemma $u_n^2 \to u^2$ in $L^{6/5}(\mathbb{R}^3)$. Thus

$$\|\partial_\phi L(u_n, \phi_n) - \partial_\phi L(u, \phi)\|_\star = o_n(1) \tag{16.17}$$

Because of (16.16), (16.17) and Proposition 16.2.10, L is C^1. The fact that the other terms of F are C^1 is a consequence of the differentiability of the norm (see Example 16.2.2) and the Sobolev embeddings, for what concerns the $p - power$ nonlinearity; indeed, for the functional

$$K(u) := \frac{1}{p}\int_{\mathbb{R}^3} |u|^p,$$

it is easy to see that

$$K'_G(u)[v] = \int_{\mathbb{R}^3} |u|^{p-2}uv,$$

from which it follows that $K'_G(u) \in H^{-1}$. Now, if $u_n \to u$ in $H^1(\mathbb{R}^3)$, it can be shown that

$$\int_{\mathbb{R}^3} |u_n|^{p-2}u_n v \to \int_{\mathbb{R}^3} |u|^{p-2}uv, \quad \forall v \in H^1(\mathbb{R}^3),$$

arguing as in the proof of [62, Lemma 8.1].

The partial derivatives of F in (u, ϕ) are given by

$$\begin{aligned}\partial_u F(u, \phi)[v] &= \int_{\mathbb{R}^3} \nabla u \nabla v + \int_{\mathbb{R}^3} (1 + \phi)uv - |u|^{p-2}uv, \\ \partial_\phi F(u, \phi)[\psi] &= \frac{1}{2}\int_{\mathbb{R}^3} \psi u^2 - \int_{\mathbb{R}^3} \nabla\phi\nabla\psi.\end{aligned}$$

for $v \in H^1(\mathbb{R}^3)$, $\psi \in D^{1,2}(\mathbb{R}^3)$. Finally,

$$\begin{aligned}(u,\phi) \text{ is a critical point of } F &\iff \partial_u F(u,\phi) = 0,\ \partial_\phi F(u,\phi) = 0\\ &\iff (u,\phi) \text{ is a weak solution of (16.13)},\end{aligned}$$

and we conclude. □

The main difficulty concerning the functional F is that it is unbounded from below and above. Indeed when $t \to +\infty$

$$F(tu,\phi) \leq \frac{t^2}{2}\int_{\mathbb{R}^3}|\nabla u|^2 + \int_{\mathbb{R}^3}(1+\phi)u^2 - \frac{t^p}{p}\int_{\mathbb{R}^3}|u|^p \to -\infty.$$

On the other hand, considering the function $u_k \in H^1(\mathbb{R}^3)$ for $k \in \mathbb{N}$ defined by

$$u_k(x) := \begin{cases}\sin k|x|, & 0 \leq |x| \leq \pi\\ 0, & |x| > \pi,\end{cases}$$

we have

$$\begin{aligned}F(u_k,0) &= \frac{1}{2}\int_{\mathbb{R}^3}|\nabla u_k|^2 + \frac{1}{2}\int_{\mathbb{R}^3}u_k^2 - \frac{1}{p}\int_{\mathbb{R}^3}|u_k|^p\\ &\geq \frac{1}{2}\int_{\mathbb{R}^3}|\nabla u_k|^2 - \frac{1}{p}\int_{\mathbb{R}^3}|u_k|^p\\ &= \pi k^2\left[\int_0^\pi r^2 dr + \int_0^\pi r^2\cos 2kr\, dr\right] - \frac{4\pi}{p}\int_0^\pi r^2|\sin kr|^p dr\\ &\geq \pi^2 k^2\left(\frac{\pi^2}{3} + \frac{1}{2k^2}\right) - \frac{4\pi^4}{3p} \to \infty,\end{aligned}$$

when $k \to \infty$.

To avoid this unboundedness of F, we use the so-called reduction method. First observe that

Proposition 16.2.37 *Given $u \in H^1(\mathbb{R}^3)$, let $\phi_u \in D^{1,2}(\mathbb{R}^3)$ be the unique solution for the problem*

$$-\Delta\phi = u^2 \text{ in } \mathbb{R}^3. \tag{16.18}$$

Define the map $\Phi : H^1(\mathbb{R}^3) \to D^{1,2}(\mathbb{R}^3)$ by $\Phi[u] := \phi_u$ and let G_Φ be its graph. Then

$$G_\Phi = \{(u,\phi) \in H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3) : \partial_\phi F(u,\phi) = 0\}. \tag{16.19}$$

Proof Let $u \in H^1(\mathbb{R}^3)$ be fixed and consider the linear transformation $T_u : D^{1,2}(\mathbb{R}^3) \to \mathbb{R}$ given by

$$T_u[\psi] := \int_{\mathbb{R}^3} \psi u^2.$$

This map is continuous because of (16.14); thus, $T \in D^{1,2}(\mathbb{R}^3)'$. By Riesz representation theorem, there is a unique element $\phi_u \in D^{1,2}(\mathbb{R}^3)$ such that

$$\int_{\mathbb{R}^3} \nabla\phi_u \nabla\psi = T[\psi] = \int_{\mathbb{R}^3} \psi u^2, \text{ for every } \psi \in D^{1,2}(\mathbb{R}^3), \tag{16.20}$$

i.e. ϕ_u is the unique solution of (16.18). If G_Φ denotes the graph of Φ, given $(u, \phi) \in H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3)$

$$\begin{aligned} \partial_\phi F(u,\phi) = 0 &\iff \int_{\mathbb{R}^3} \nabla\phi\nabla\psi = \int_{\mathbb{R}^3} \psi u^2 \\ &\iff \phi \text{ is a solution of (16.18)} \\ &\iff \phi = \phi_u = \Phi[u], \end{aligned}$$

concluding the proof. □

Remark 16.2.38 Recall that from (16.14) we have the inequality

$$\int_{\mathbb{R}^3} \psi u^2 \le C\|u\|^2\|\psi\|_D, \text{ for every } \psi \in D^{1,2}(\mathbb{R}^3).$$

Also by (16.20), the solution ϕ_u of (16.18) satisfies

$$\int_{\mathbb{R}^3} \nabla\phi_u \nabla\psi = \int_{\mathbb{R}^3} \psi u^2, \text{ for every } \psi \in D^{1,2}(\mathbb{R}^3). \tag{16.21}$$

Therefore, taking $\psi = \phi_u$ we can write

$$\|\phi_u\|_D^2 = \int_{\mathbb{R}^3} \|\nabla\phi_u\|^2 = \int_{\mathbb{R}^3} \phi_u u^2 \le C\|u\|^2\|\phi_u\|_D,$$

from which

$$\|\phi_u\|_D \le C\|u\|^2.$$

Thus, we obtain

$$\int_{\mathbb{R}^3} \phi_u u^2 \le C\|u\|^4.$$

The last inequality, in the adapted version in the following sections, will be very useful. We also have the representation

$$\phi_u(x) = \left(\frac{1}{4\pi|\cdot|} \star u^2\right)(x) = \frac{1}{4\pi}\int_{\mathbb{R}^3} \frac{u^2(y)}{|x-y|}dy, \ \forall x \in \mathbb{R}^3.$$

Proposition 16.2.39 *We have*

$$\Phi \in C^1(H^1(\mathbb{R}^3); D^{1,2}(\mathbb{R}^3)).$$

Proof We define the map $J : H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3) \to D'$ by $J(u,\phi) := \partial_\phi F(u,\phi)$ and we claim that

$$J \in C^1(H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3); D') \tag{16.22}$$

In fact, since

$$J(u,\phi)[\cdot] = L(u,\cdot) - \frac{1}{2}\int_{\mathbb{R}^3} \nabla\phi\nabla(\cdot),$$

where L is the C^1 map used in the proof of the Proposition 16.2.36, (16.22) is immediate. Thus we can write

$$(\partial_u J(u,\phi)[v])[\psi] = \partial^2_{u,\phi}F(u,\phi)[v,\psi] = \int_{\mathbb{R}^3} uv\psi$$

$$(\partial_\phi J(u,\phi)[\nu])[\psi] = \partial^2_{\phi,\phi}F(u,\phi)[\nu,\psi] = -\frac{1}{2}\int_{\mathbb{R}^3} \nabla\nu\nabla\psi.$$

Now we show that

$$\partial_\phi J(u,\phi) \in \mathcal{L}(D^{1,2}(\mathbb{R}^3); D') \text{ is invertible.} \tag{16.23}$$

Indeed, since

$$\partial_\phi J(u,\phi)[\nu] = \partial^2_{\phi,\phi}F(u,\phi)[\nu, \cdot] = -\frac{1}{2}\int_{\mathbb{R}^3} \nabla\nu\nabla(\cdot),$$

up to a constant factor, $\partial_\phi J(u,\phi)$ is the Riesz isomorphism between $D^{1,2}(\mathbb{R}^3)$ and its dual, D'.

Finally, let $u_0 \in H^1(\mathbb{R}^3)$. Because of (16.19),

$$(u_0, \Phi[u_0]) \in G_\Phi \quad \Longrightarrow \quad J(u_0, \Phi[u_0]) = 0.$$

Since (16.22) holds true and $\partial_\phi J(u_0, \Phi[u_0]) \in \mathcal{L}(D^{1,2}(\mathbb{R}^3), D')$ is invertible by (16.23), in virtue of the implicit theorem function there exist neighborhoods U_0 of u_0, V_0 of $\Phi[u_0]$ and a map $g \in C^1(U_0, D^{1,2}(\mathbb{R}^3))$ such that

$$J(u, \phi) = 0 \text{ and } (u, \phi) \in U_0 \times V_0 \quad \Longleftrightarrow \quad \phi = g(u), \ \forall u \in U_0.$$

But then, for every $u \in U_0$,

$$J(u, g(u)) = 0 \quad \Longrightarrow \quad (u, g(u)) \in G_\Phi,$$

from which

$$g = \Phi|_{U_0}.$$

In particular $\Phi|_{U_0} \in C^1(U_0, D^{1,2}(\mathbb{R}^3))$, and by the arbitrariness of u_0 we conclude □

Let us define the functional $I : H^1(\mathbb{R}^3) \to \mathbb{R}$ by

$$\begin{aligned} I(u) &:= F(u, \phi_u) \\ &= \frac{1}{2}\int_{\mathbb{R}^3} [|\nabla u|^2 + (1 + \phi_u)u^2] - \frac{1}{p}\int_{\mathbb{R}^3} |u|^p - \frac{1}{4}\int_{\mathbb{R}^3} |\nabla \phi_u|^2 \\ &= \frac{1}{2}\int_{\mathbb{R}^3} (|\nabla u|^2 + u^2) + \frac{1}{4}\int_{\mathbb{R}^3} \phi_u u^2 - \frac{1}{p}\int_{\mathbb{R}^3} |u|^p, \end{aligned}$$

where we have used the relation (see (16.21))

$$\int_{\mathbb{R}^3} |\nabla \phi_u|^2 = \int_{\mathbb{R}^3} \phi_u u^2.$$

Since $F \in C^1(H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3); \mathbb{R})$ and $\Phi \in C^1(H^1(\mathbb{R}^3); D^{1,2}(\mathbb{R}^3))$, I is a C^1 functional in $H^1(\mathbb{R}^3)$ and by (16.19) we get

$$I'(u) = \partial_u F(u, \Phi[u]) + \partial_\phi F(u, \Phi[u]) \circ \Phi'[u] = \partial_u F(u, \Phi[u]).$$

Proposition 16.2.40 *For* $(u, \phi) \in H^1(\mathbb{R}^3) \times D^{1,2}(\mathbb{R}^3)$*,*

$$(u, \phi) \text{ is a critical point of } F \iff u \text{ is a critical point of } I \text{ and } \phi = \Phi[u]$$

Proof Indeed

$$
\begin{aligned}
(u,\phi) \text{ is a critical point of } F &\iff \partial_u F(u,\phi) = 0 \text{ and } \partial_\phi F(u,\phi) = 0 \\
&\iff \partial_u F(u,\phi) + \partial_\phi F(u,\phi)\Phi'[u] \\
&\quad = 0 \text{ and } \phi = \Phi[u] \\
&\iff I'(u) = 0 \text{ and } \phi = \Phi[u] \\
&\iff u \text{ is a critical point of } I \text{ and } \phi = \Phi[u],
\end{aligned}
$$

concluding the proof. □

Thus system (16.13) has been reduced to the equation

$$
-\Delta u + u + \Phi[u]u = |u|^{p-2}u \quad \text{in} \quad \mathbb{R}^3, \tag{16.24}
$$

whose solutions are the critical points of I.

It is easy to see that $I \in C^2(H^1(\mathbb{R}^3); \mathbb{R})$. Indeed by using the same techniques as before, after straightforward computations one sees that for every $v, w \in H^1(\mathbb{R}^3)$

$$
\begin{aligned}
I'(u)[v] &= \int_{\mathbb{R}^3} \nabla u \nabla v + \int_{\mathbb{R}^3} uv + \int_{\mathbb{R}^3} \phi_u uv - \int_{\mathbb{R}^3} |u|^{p-2}uv, \\
I''(u)[v,w] &= \int_{\mathbb{R}^3} \nabla v \nabla w + \int_{\mathbb{R}^3} vw + \int_{\mathbb{R}^3} \phi_u vw \\
&\quad + 2\int_{\mathbb{R}^3} \phi_{u,w} uv - (p-1)\int_{\mathbb{R}^3} |u|^{p-2}vw,
\end{aligned}
$$

where $\phi_{u,w}$ is the unique solution of

$$
-\Delta\phi = uw \quad \text{in } \mathbb{R}^3.
$$

which can be obtained by the Riesz Representation Theorem.

The Nehari set for I is defined by

$$
I'(u)[u] = 0 \quad \iff \quad \int_{\mathbb{R}^3} (|\nabla u|^2 + u^2) + \int_{\mathbb{R}^3} \phi_u u^2 = \int_{\mathbb{R}^3} |u|^p. \tag{16.25}
$$

Let now $u \in H^1(\mathbb{R}^3)$, $u \neq 0$. For $t > 0$ we have

$$
I(tu) = \frac{t^2}{2}\int_{\mathbb{R}^3} (|\nabla u|^2 + u^2) + \frac{t^4}{4}\int_{\mathbb{R}^3} \phi_u u^2 - \frac{t^p}{p}\int_{\mathbb{R}^3} |u|^p
$$

from which we see that

- $I(tu)$ has a strict local minimum at $t = 0$,
- there is $T > 0$ large enough such that $I(Tu) < 0$.

Therefore, for some $t_u > 0$, $t_u u \in \mathcal{N}$ showing that $\mathcal{N}$ is not empty. Moreover, for every $u \in \mathcal{N} \setminus \{0\}$, using (16.25) we have

$$\begin{aligned} I''(u)[u,u] &= \int_{\mathbb{R}^3} (|\nabla u|^2 + u^2) + 3\int_{\mathbb{R}^3} \phi_u u^2 - (p-1)\int_{\mathbb{R}^3} |u|^p \\ &< 3\left[\int_{\mathbb{R}^3} (|\nabla u|^2 + u^2) + \int_{\mathbb{R}^3} \phi_u u^2\right] - (p-1)\int_{\mathbb{R}^3} |u|^p \\ &= 3\int_{\mathbb{R}^3} |u|^p - (p-1)\int_{\mathbb{R}^3} |u|^p \\ &= -(p-4)\int_{\mathbb{R}^3} |u|^p < 0, \end{aligned}$$

so that (16.6) is satisfied. By Proposition 16.2.20, $\mathcal{N}$ is a natural constraint for I. Moreover from (16.25) and the Sobolev embedding

$$\|u\|^2 \le |u|_p^p \le C\|u\|^p$$

we deduce that

$$0 < \frac{1}{C} \le \|u\|^{p-2}$$

showing that $\mathcal{N}$ is bounded away from zero. Then any critical point of I constrained to $\mathcal{N}$ is a solution of (16.24).

Analogously it can be seen that the solutions of

$$-\Delta u + V(x)u = |u|^{p-2}u \quad \text{in } \mathbb{R}^3 \tag{16.26}$$

are critical points of the C^2 functional

$$I(u) = \frac{1}{2}\int_{\mathbb{R}^3} |\nabla u|^2 + \frac{1}{2}\int V(x)u^2 - \frac{1}{p}\int_{\mathbb{R}^3} |u|^p$$

on a suitable Hilbert space on which it is well defined. A similar equation to (16.26) will be studied in Sect. 16.3 below.

Actually the equations studied in the next two sections are generalizations of that we have seen here. Indeed we will consider fractional equations; moreover, the nonlinearity will be a $f(u)$ of “power” type; finally, a small parameter ε appears in the equations. Physically the parameter is related to the Planck constant. From a mathematically point of view it is important since, whenever it tends to zero, that is whenever it is sufficiently small, we are in a position to apply all the previous machinery in order to prove the (PS) condition and implementing the barycenter technique to find multiple critical points, hence multiple solutions of the problem. The case when ε tends to zero is known in the physical literature as "the semiclassical limit."

16.3 The Case of Given Potential: The Fractional Schrödinger Equation

The results shown in this section are taken from [39].

This kind of Schrödinger equation was first derived and studied by Laskin [47]. Then many papers appeared studying existence, multiplicity, and behavior of solutions to such equations. In these last years problems involving fractional operators are receiving a special attention, due to the fact that fractional and nonlocal equations have important applications in many sciences, as we said before.

More specifically, the problem addressed here concerns with existence and multiplicity of positive solutions for the following equation

$$\begin{cases} \varepsilon^{2s}(-\Delta)^s u + V(z)u = f(u) \\ u \in H^s(\mathbb{R}^3) \end{cases} \tag{P_ε}$$

where $s \in (0, 1)$ and $\varepsilon > 0$.

Problem (P_ε) appears when one look for standing wave solutions, that is solution of the special form

$$\psi(z,t) = u(z)e^{-iEt/\varepsilon}, \quad u(z) \in \mathbb{R},\ E \text{ a real constant}$$

to the following Fractional Schrödinger equation

$$i\varepsilon\frac{\partial\psi}{\partial t} = \varepsilon^{2s}(-\Delta)^s\psi + W(z)\psi - f(|\psi|)$$

where $W : \mathbb{R}^3 \to \mathbb{R}$ is an external potential and f a suitable nonlinearity. The parameter ε corresponds, after some normalizations, to the Planck constant.

We introduce the basic assumptions on f and V:

(V) $V : \mathbb{R}^3 \to \mathbb{R}$ is a continuous function satisfying

$$0 < \min_{\mathbb{R}^3} V(x) =: V_0 < \liminf_{|x|\to\infty} V(x) =: V_\infty \in (0, +\infty]\,;$$

(f1) $f : \mathbb{R} \to \mathbb{R}$ is a function of class C^1 and $f(u) = 0$ for $u \le 0$;
(f2) $\exists\, q \in (2, 2_s^* - 1)$ such that $\lim_{u\to\infty} f'(u)/u^{q-1} = 0$, where $2_s^* := 6/(3-2s)$;

(f3) $\exists\, \theta > 2$ such that $0 < \theta F(u) := \theta \int_0^u f(t)dt \le uf(u)$ for all $u > 0$;
(f4) the function $u \to f(u)/u$ is strictly increasing in $(0, +\infty)$.

We say that $u \in W_\varepsilon$ (see Sect. 16.3.1 below for the definition of W_ε) is a solution of (P_ε) if for every $v \in W_\varepsilon$

$$\varepsilon^{2s} \int_{\mathbb{R}^3} (-\Delta)^{s/2} u (-\Delta)^{s/2} v + \int_{\mathbb{R}^3} V(z)uv = \int_{\mathbb{R}^3} f(u)v$$

that is, as we will see, u is a critical point of a suitable energy functional I_ε. The solution with "minimal energy" is what we call a *ground state*.

We remark that the assumptions on V and f are quite natural in this context. Assumption (V) was first introduced by Rabinowitz in [55] to take into account potentials which are possibly not coercive. Hypothesis (f1) is not restrictive since we are looking for positive solutions (see, e.g., [38, pag. 1247]) and (f2)–(f4) are useful to use variational techniques which involve the Palais-Smale condition, the Mountain Pass Theorem, and the Nehari manifold.

The existence of ground states solutions is the aim of our first result.

Theorem 16.3.1 *Suppose that f verifies (f1)–(f4) and V verifies (V). Then there exists a ground state solution $\mathrm{u}_\varepsilon \in W_\varepsilon$ of (P_ε),*

1. *for every $\varepsilon > 0$, if $V_\infty = +\infty$;*
2. *for every $\varepsilon \in (0, \bar{\varepsilon}]$, for some $\bar{\varepsilon} > 0$, if $V_\infty < +\infty$.*

We treat also the case of the multiplicity of solutions. This result involves topological properties of the set of minima of the potential V:

$$M := \Big\{ x \in \mathbb{R}^3 : V(x) = V_0 \Big\}.$$

Indeed by means of the Ljusternik-Schnirelman theory we arrive at the following result.

Theorem 16.3.2 *Suppose that f satisfies (f1)–(f4) and the function V satisfies (V). Then, there exists $\varepsilon^* > 0$ such that for every $\varepsilon \in (0, \varepsilon^*]$ problem (P_ε) has at least*

1. $\operatorname{cat} M$ *positive solutions;*
2. $\operatorname{cat} M + 1$ *positive solutions, if M is bounded and* $\operatorname{cat} M \ge 2$.

Even if we will not prove the next result, we just say that using the Morse Theory we can obtain a result relating the number of positive solutions with the homology of the domain M.

Theorem 16.3.3 *Suppose that f satisfies (f1)–(f4) and the function V satisfies (V). Then there exists $\varepsilon^* > 0$ such that for every $\varepsilon \in (0, \varepsilon^*]$ problem (P_ε) has at least* $2\mathcal{P}_1(M) - 1$ *solutions, possibly counted with their multiplicity.*

Here

$$\mathcal{P}_t(M) = \sum_k \dim H_k(M) t^k$$

where $H_*(M)$ be its singular homology with coefficients in some field $\mathbb{F}$. If, for example, M is obtained by a contractible domain cutting off k disjoint contractible sets, it is $\mathrm{cat}(M) = 2$ and $\mathcal{P}_1(M) = 1 + k$. For the proof of this last result, we refer the reader to [39].

Remark 16.3.4 As it will be evident by the proofs, Theorems 16.3.1 and 16.3.2 remain true if we replace conditions (f2) with the weaker condition

- $\exists\, q \in (2, 2_s^* - 1)$ such that $\lim_{u\to\infty} f(u)/u^q = 0$, where $2_s^* := 6/(3-2s)$.

On the other hand, for Theorem 16.3.3 we need (f2) to have the compactness of a certain operator (see [39]). We have preferred to state our theorems under the stronger conditions just for the sake of simplicity.

To prove our result will be mainly inspired by some papers of Benci, Cerami and Passaseo [17, 18, 21] and Cingolani and Lazzo [27, 28] who treated the case $s = 1$. Indeed Cingolani and Lazzo prove a multiplicity result on the existence of solutions based on the topological richness of the set of minima of the potential appearing in the equation, as in [27], or a suitable function involving the potentials in the case of competing potentials, as in [28]. These ideas and techniques have been extensively used to attack also other type of problems, and indeed similar results are obtained for other equations and operators, like the $p-$laplacian [3, 9], the biharmonic operator [4], $p\&q-$laplacian, fractional laplacian in expanding domain [41], magnetic Laplacian [5, 6], or quasilinear operators [7, 12, 13].

We point out that the existing literature on the semiclassical limit for fractional equations like in (P_ε) deals with the study of the concentration points of a single solution u^ε whenever $\varepsilon \to 0$, see, e.g., [8, 31, 32, 36]. The interest in studying the semiclassical case lies in the fact that such solutions u^ε develop some spikes around one or more different points of the space. However, to the best of our knowledge there are no results dealing with the multiplicity of solutions, for small ε involving the "topological richness" of the set of the minima of the potential V.

For the reader convenience the subdivision of this section is the following.

- **The variational setting**. Here after a change of variable, we introduce an equivalent problem (P_ε^*) and the related variational setting; actually, we will prove Theorems 16.3.1, 16.3.2 and 16.3.3 by referring to this equivalent problem. The variational setting of the problem is settled and suitable functionals I_ε and E_μ are introduced.
- **Compactness for $\mathbf{I_\varepsilon}$ and $\mathbf{E_\mu}$. Existence of a ground state solution**. Here some compactness properties for the functionals are proved and the proof of Theorem 16.3.1 is given, that is the existence of the ground state solution is proved.
- **The barycenter map**. Here barycenter maps are introduced in order to estimate the category of suitable sublevels of the energy functional whenever $\varepsilon \to 0$.
- **Proof of Theorem 16.3.2**. Here the proof of the multiplicity result is completed.

16.3.1 The Variational Setting

First of all, it is easy to see that our problem is equivalent, after a change of variable to the following one

$$\begin{cases} (-\Delta)^s u + V(\varepsilon x)u = f(u) \ \text{ in } \ \mathbb{R}^3, \\ u \in H^s(\mathbb{R}^3) \\ u(x) > 0, \ x \in \mathbb{R}^3 \end{cases} \tag{P_ε^*}$$

to which we will refer from now on. Once we find solutions u_ε for (P_ε^*), the function $w_\varepsilon(x) := u_\varepsilon(x/\varepsilon)$ will be a solution of (P_ε).

We fix now some notations involving the functionals used to get the solutions to (P_ε^*).

Let us consider first the autonomous case. For a given constant (potential) $\mu > 0$ consider the problem

$$\begin{cases} (-\Delta)^s u + \mu u = f(u) \ \text{ in } \ \mathbb{R}^3, \\ u \in H^s(\mathbb{R}^3) \\ u(x) > 0, \ x \in \mathbb{R}^3 \end{cases} \tag{A_μ}$$

and the C^1 functional in $H^s(\mathbb{R}^3)$

$$E_\mu(u) = \frac{1}{2}\int_{\mathbb{R}^3} |(-\Delta)^{s/2}u|^2 + \frac{\mu}{2}\int_{\mathbb{R}^3} u^2 - \int_{\mathbb{R}^3} F(u)$$

whose critical points are the solutions of (A_μ). In this case $H^s(\mathbb{R}^3)$ is endowed with the (squared) norm

$$\|u\|_\mu^2 = \int_{\mathbb{R}^3} |(-\Delta)^{s/2}u|^2 + \mu \int_{\mathbb{R}^3} u^2.$$

The following are well-known facts. The functional E_μ has a mountain pass geometry and, defining $\mathcal{H} = \{\gamma \in C([0, 1], H^s(\mathbb{R}^3)) : \gamma(0) = 0, E_\mu(\gamma(1)) < 0\}$, the mountain pass level

$$m(\mu) := \inf_{\gamma \in \mathcal{H}} \sup_{t \in [0,1]} E_\mu(\gamma(t)) \tag{16.27}$$

satisfies

$$m(\mu) = \inf_{u \in H^s(\mathbb{R}^3)\setminus\{0\}} \sup_{t \geq 0} E_\mu(tu) = \inf_{u \in \mathcal{M}_\mu} E_\mu(u) > 0, \tag{16.28}$$

where

$$\mathcal{M}_\mu := \Big\{u \in H^s(\mathbb{R}^3) \setminus \{0\} : \int_{\mathbb{R}^3} |(-\Delta)^{s/2}u|^2 + \mu \int_{\mathbb{R}^3} u^2 = \int_{\mathbb{R}^3} f(u)u\Big\}.$$

It is easy to see that $\mathcal{M}_\mu$ is bounded away from zero in $H^s(\mathbb{R}^3)$, and is a differentiable manifold radially diffeomorphic to the unit sphere. It is usually called the *Nehari manifold* associated with E_μ.

On the other hand, the solutions of (P_ε^*) can be characterized as critical points of the C^1 functional given by

$$I_\varepsilon(u) = \frac{1}{2}\int_{\mathbb{R}^3} |(-\Delta)^{s/2}u|^2 + \frac{1}{2}\int_{\mathbb{R}^3} V(\varepsilon x)u^2 - \int_{\mathbb{R}^3} F(u)$$

which is well defined on the Hilbert space

$$W_\varepsilon := \Big\{u \in H^s(\mathbb{R}^3) : \int_{\mathbb{R}^3} V(\varepsilon x)u^2 < \infty\Big\}$$

endowed with the (squared) norm

$$\|u\|^2_{W_\varepsilon} = \int_{\mathbb{R}^3} |(-\Delta)^{s/2}u|^2 + \int_{\mathbb{R}^3} V(\varepsilon x)u^2.$$

Note that if $V_\infty = +\infty$, W_ε has compact embedding into $L^p(\mathbb{R}^3)$ for $p \in [2, 2_s^*)$, see, e.g., [26, Lemma 3.2].

The Nehari manifold associated with I_ε is

$$\mathcal{N}_\varepsilon = \Big\{u \in W_\varepsilon \setminus \{0\} : J_\varepsilon(u) = 0\Big\}$$

where

$$J_\varepsilon(u) := \int_{\mathbb{R}^3} |(-\Delta)^{s/2}u|^2 + \int_{\mathbb{R}^3} V(\varepsilon x)u^2 - \int_{\mathbb{R}^N} f(u)u \qquad (16.29)$$

and its tangent space in u is given by

$$\mathrm{T}_u\mathcal{N}_\varepsilon = \Big\{v \in H^s(\mathbb{R}^3) : J'(u)[v] = 0\Big\}.$$

Let us introduce also

$$\mathcal{S}_\varepsilon := \Big\{u \in W_\varepsilon : \|u\|_{W_\varepsilon} = 1\Big\} \setminus \{u \in H^s(\mathbb{R}^3) : u \leq 0 \text{ a.e.}\}$$

which is a smooth manifold of codimension 1. The next result is standard; the proof follows the same lines of [18, Lemma 2.1 and Lemma 2.2].

Lemma 16.3.5 *The following propositions hold true:*

1. *for every* $u \in \mathcal{N}_\varepsilon$ *it is* $J'_\varepsilon(u)[u] < 0$;
2. $\mathcal{N}_\varepsilon$ *is a differentiable manifold radially diffeomorphic to* $\mathcal{S}_\varepsilon$ *and there exists* $k_\varepsilon > 0$ *such that*

$$\|u\|_{W_\varepsilon} \geq k_\varepsilon, \quad I_\varepsilon(u) \geq k_\varepsilon$$

As in [18, Lemma 2.1], it is easy to see that the functions in $\mathcal{N}_\varepsilon$ have to be positive on some set of nonzero measure. It is also easy to check that I_ε has the mountain pass geometry, as given in the next

Lemma 16.3.6 *Fixed* $\varepsilon > 0$, *for the functional* I_ε *the following statements hold:*

i) there exists $\alpha, \rho > 0$ *such that* $I_\varepsilon(u) \geq \alpha$ *with* $\|u\|_\varepsilon = \rho$,
ii) there exist $e \in W_\varepsilon$ *with* $\|e\|_{W_\varepsilon} > \rho$ *such that* $I_\varepsilon(e) < 0$.

Then, defining the mountain pass level of I_ε,

$$c_\varepsilon := \inf_{\gamma \in \mathcal{H}} \sup_{t \in [0,1]} I_\varepsilon(\gamma(t))$$

where $\mathcal{H} = \{\gamma \in C([0, 1], W_\varepsilon) : \gamma(0) = 0, I_\varepsilon(\gamma(1)) < 0\}$, well-known arguments imply that

$$c_\varepsilon = \inf_{u \in W_\varepsilon \setminus \{0\}} \sup_{t \geq 0} I_\varepsilon(tu) = \inf_{u \in \mathcal{N}_\varepsilon} I_\varepsilon(u) \geq m(V_0).$$

16.3.2 Compactness for I_ε and E_μ: Existence of a Ground State Solution

This section is devoted to prove compactness properties related to the functionals I_ε and E_μ.

It is standard by now to see that hypothesis (f3) is used to obtain the boundedness of the (PS) sequences for I_ε or E_μ.

We need to recall the following Lions type lemma.

Lemma 16.3.7 *If* $\{u_n\}$ *is bounded in* $H^s(\mathbb{R}^3)$ *and for some* $R > 0$ *and* $2 \leq r < 2^*_s$ *we have*

$$\sup_{x \in \mathbb{R}^N} \int_{B_R(x)} |u_n|^r \to 0 \quad as \quad n \to \infty,$$

then $u_n \to 0$ *in* $L^p(\mathbb{R}^3)$ *for* $2 < p < 2^*_s$.

For a proof see, e.g., [30, Lemma 2.3].

In order to prove compactness, some preliminary work is needed.

Lemma 16.3.8 *Let $\{u_n\} \subset W_\varepsilon$ be such that $I'_\varepsilon(u_n) \to 0$ and $u_n \rightharpoonup 0$ in W_ε. Then we have either*

a) $u_n \to 0$ in W_ε, or
b) there exist a sequence $\{y_n\} \subset \mathbb{R}^3$ and constants $R, c > 0$ such that

$$\liminf_{n\to+\infty} \int_{B_R(y_n)} u_n^2 \geq c > 0.$$

Proof Suppose that b) does not occur. Using Lemma 16.3.7 it follows

$$u_n \to 0 \ \text{ in } \ L^p(\mathbb{R}^3) \ \text{ for } \ p \in (2, 2_s^*).$$

Given $\xi > 0$, by (f1) and (f2), for some constant $C_\xi > 0$ we have

$$0 \leq \int_{\mathbb{R}^3} f(u_n)u_n \leq \xi \int_{\mathbb{R}^3} u_n^2 + C_\xi \int_{\mathbb{R}^3} |u_n|^{q+1}.$$

Using the fact that $\{u_n\}$ is bounded in $L^2(\mathbb{R}^3)$, $u_n \to 0$ in $L^{q+1}(\mathbb{R}^3)$, and that ξ is arbitrary, we can conclude that

$$\int_{\mathbb{R}^3} f(u_n)u_n \to 0.$$

Recalling that

$$\|u_n\|_{W_\varepsilon}^2 - \int_{\mathbb{R}^3} f(u_n)u_n = I'_\varepsilon(u_n)[u_n] = o_n(1),$$

it follows that $u_n \to 0$ in W_ε. □

Lemma 16.3.9 *Assume that $V_\infty < \infty$ and let $\{v_n\}$ be a $(PS)_d$ sequence for I_ε in W_ε with $v_n \rightharpoonup 0$ in W_ε. Then*

$$v_n \not\to 0 \ \text{ in } \ W_\varepsilon \implies d \geq m(V_\infty)$$

(recall that $m(V_\infty)$ is the mountain pass level of E_{V_∞}, see (16.28)).

Proof Let $\{t_n\} \subset (0, +\infty)$ be a sequence such that $\{t_n v_n\} \subset \mathcal{M}_{V_\infty}$. We start by showing the following

Claim The sequence $\{t_n\}$ satisfies $\limsup_{n\to\infty} t_n \leq 1$.

In fact, supposing by contradiction that the claim does not hold, there exists $\delta > 0$ and a subsequence still denoted by $\{t_n\}$, such that

$$t_n \geq 1 + \delta \quad \text{for all} \quad n \in \mathbb{N}. \tag{16.30}$$

Since $\{v_n\}$ is bounded in W_ε, $I'_\varepsilon(v_n)[v_n] = o_n(1)$, that is,

$$\int_{\mathbb{R}^3} \left[|(-\Delta)^{s/2} v_n|^2 + V(\varepsilon x) v_n^2 \right] = \int_{\mathbb{R}^3} f(v_n) v_n + o_n(1).$$

Moreover, since $\{t_n v_n\} \subset \mathcal{M}_{V_\infty}$, we get

$$t_n^2 \int_{\mathbb{R}^3} \left[|(-\Delta)^{s/2} v_n|^2 + V_\infty v_n^2 \right] = \int_{\mathbb{R}^3} f(t_n v_n) t_n v_n.$$

The last two equalities imply that

$$\int_{\mathbb{R}^3} \left[\frac{f(t_n v_n) v_n^2}{t_n v_n} - \frac{f(v_n) v_n^2}{v_n} \right] = \int_{\mathbb{R}^3} [V_\infty - V(\varepsilon x)] v_n^2 + o_n(1). \tag{16.31}$$

Given $\xi > 0$, by condition (16.3) there exists $R = R(\xi) > 0$ such that

$$V(\varepsilon x) \geq V_\infty - \xi \quad \text{for any} \quad |x| \geq R.$$

Let $C > 0$ be such that $\|v_n\|_{W_\varepsilon} \leq C$. Since $v_n \to 0$ in $L^2(B_R(0))$, we conclude by (16.31)

$$\int_{\mathbb{R}^3} \left[\frac{f(t_n v_n)}{t_n v_n} - \frac{f(v_n)}{v_n} \right] v_n^2 \leq \xi C V_\infty + o_n(1). \tag{16.32}$$

Since $v_n \not\to 0$ in W_ε, we may invoke Lemma 16.3.8 to obtain $\{y_n\} \subset \mathbb{R}^3$ and $\check{R}, c > 0$ such that

$$\int_{B_{\check{R}}(y_n)} v_n^2 \geq c. \tag{16.33}$$

Defining $\check{v}_n := v_n(\cdot + y_n)$, we may suppose that, up to a subsequence,

$$\check{v}_n \rightharpoonup \check{v} \ \text{ in } \ H^s(\mathbb{R}^3).$$

Moreover, in view of (16.33), there exists a subset $\Omega \subset \mathbb{R}^N$ with positive measure such that $\check{v} > 0$ in Ω. From (f4), we can use (16.30) to rewrite (16.32) as

$$0 < \int_\Omega \left[\frac{f((1+\delta)\check{v}_n)}{(1+\delta)\check{v}_n} - \frac{f(\check{v}_n)}{\check{v}_n} \right] \check{v}_n^2 \leq \xi C V_\infty + o_n(1), \quad \text{for any } \ \xi > 0.$$

Letting $n \to \infty$ in the last inequality and applying Fatou's Lemma, it follows that

$$0 < \int_{\Omega} \left[\frac{f((1+\delta)\check{v})}{(1+\delta)\check{v}} - \frac{f(\check{v})}{\check{v}} \right] \check{v}^2 \leq \xi C V_\infty, \quad \text{for any } \xi > 0.$$

which is an absurd, proving the claim.

Now, it is convenient to distinguish the following cases:

Case 1 $\limsup_{n\to\infty} t_n = 1$.

In this case there exists a subsequence, still denoted by $\{t_n\}$, such that $t_n \to 1$. Thus,

$$d + o_n(1) = I_\varepsilon(v_n) \geq m(V_\infty) + I_\varepsilon(v_n) - E_{V_\infty}(t_n v_n). \tag{16.34}$$

Recalling that

$$I_\varepsilon(v_n) - E_{V_\infty}(t_n v_n) = \frac{(1-t_n^2)}{2} \int_{\mathbb{R}^3} |(-\Delta)^{s/2} v_n|^2 + \frac{1}{2} \int_{\mathbb{R}^3} V(\varepsilon x) v_n^2$$
$$- \frac{t_n^2}{2} \int_{\mathbb{R}^3} V_\infty v_n^2 + \int_{\mathbb{R}^3} [F(t_n v_n) - F(v_n)],$$

and using the fact that $\{v_n\}$ is bounded in W_ε by $C > 0$ together with the condition (16.3), we get

$$I_\varepsilon(v_n) - E_{V_\infty}(t_n v_n) \geq o_n(1) - C\xi + \int_{\mathbb{R}^N} [F(t_n v_n) - F(v_n)].$$

Moreover, by the Mean Value Theorem,

$$\int_{\mathbb{R}^N} [F(t_n v_n) - F(v_n)] = o_n(1),$$

therefore (16.34) becomes

$$d + o_n(1) \geq m(V_\infty) - C\xi + o_n(1),$$

and taking the limit in n, by the arbitrariness of ξ, we have $d \geq m(V_\infty)$.

Case 2 $\limsup_{n\to\infty} t_n = t_0 < 1$.

In this case up to a subsequence, still denoted by $\{t_n\}$, we have

$$t_n \to t_0 \text{ and } t_n < 1 \quad \text{for all } n \in \mathbb{N}.$$

Since $u \mapsto \frac{1}{2} f(u)u - F(u)$ is increasing, we have

$$m(V_\infty) \leq \int_{\mathbb{R}^N} \left[\frac{1}{2} f(t_n v_n) t_n v_n - F(t_n v_n) \right] \leq \int_{\mathbb{R}^N} \left[\frac{1}{2} f(v_n) v_n - F(v_n) \right]$$

hence,

$$m(V_\infty) \le I_\varepsilon(v_n) - \frac{1}{2} I'_\varepsilon(v_n)[v_n] = d + o_n(1),$$

and again we easily conclude. □

Now we are ready to give the desired compactness result.

Proposition 16.3.10 *The functional I_ε in W_ε satisfies the $(PS)_c$ condition*

1. *at any level $c < m(V_\infty)$, if $V_\infty < \infty$,*
2. *at any level $c \in \mathbb{R}$, if $V_\infty = \infty$.*

Proof Let $\{u_n\} \subset W_\varepsilon$ be such that $I_\varepsilon(u_n) \to c$ and $I'_\varepsilon(u_n) \to 0$. By standard calculations, we can see that $\{u_n\}$ is bounded in W_ε. Thus there exists $u \in W_\varepsilon$ such that, up to a subsequence, $u_n \rightharpoonup u$ in W_ε and we see that $I'_\varepsilon(u) = 0$.

Defining $v_n := u_n - u$, by [11] we know that

$$\int_{\mathbb{R}^3} F(v_n) = \int_{\mathbb{R}^3} F(u_n) - \int_{\mathbb{R}^3} F(u) + o(1)$$

and arguing as in [3] we have also $I'_\varepsilon(v_n) \to 0$. Then

$$I_\varepsilon(v_n) = I_\varepsilon(u_n) - I_\varepsilon(u) + o_n(1) = c - I_\varepsilon(u) + o_n(1) =: d + o_n(1) \qquad (16.35)$$

and $\{v_n\}$ is a $(PS)_d$ sequence. By (f3),

$$I_\varepsilon(u) = I_\varepsilon(u) - \frac{1}{2} I'_\varepsilon(u)[u] = \int_{\mathbb{R}^3} [\frac{1}{2} f(u)u - F(u)] \ge 0,$$

and then, if $V_\infty < \infty$ and $c < m(V_\infty)$, by (16.35) we obtain

$$d \le c < m(V_\infty).$$

It follows from Lemma 16.3.9 that $v_n \to 0$, that is $u_n \to u$ in W_ε.

In the case $V_\infty = \infty$ by the compact imbedding $W_\varepsilon \hookrightarrow\hookrightarrow L^p(\mathbb{R}^3)$, $2 \le p < 2^*_s$, up to a subsequence, $v_n \to 0$ in $L^p(\mathbb{R}^3)$ and by (f2)

$$\|v_n\|^2_{W_\varepsilon} = \int_{\mathbb{R}^3} f(v_n)v_n = o_n(1).$$

This last equality implies that $u_n \to u$ in W_ε. □

It follows the next

Proposition 16.3.11 *The functional I_ε restricted to $\mathcal{N}_\varepsilon$ satisfies the $(PS)_c$ condition*

1. *at any level $c < m(V_\infty)$, if $V_\infty < \infty$,*
2. *at any level $c \in \mathbb{R}$, if $V_\infty = \infty$.*

Proof Let $\{u_n\} \subset \mathcal{N}_\varepsilon$ be such that $I_\varepsilon(u_n) \to c$ and for some sequence $\{\lambda_n\} \subset \mathbb{R}$,

$$I'_\varepsilon(u_n) = \lambda_n J'_\varepsilon(u_n) + o_n(1), \tag{16.36}$$

where $J_\varepsilon : W_\varepsilon \to \mathbb{R}$ is defined in (16.29). Again we can deduce that $\{u_n\}$ is bounded. Now

a) evaluating (16.36) in u_n we get $\lambda_n J'_\varepsilon(u_n)[u_n] = o_n(1)$,
b) evaluating (16.36) in $v \in \mathrm{T}_{u_n}\mathcal{N}_\varepsilon$ we get $J'_\varepsilon(u_n)[v] = 0$.

Hence $\lambda_n J'_\varepsilon(u_n) = o_n(1)$ and by (16.36) we deduce $I'_\varepsilon(u_n) = o_n(1)$. Then $\{u_n\}$ is a $(PS)_c$ sequence for I_ε and we conclude by Proposition 16.3.10. □

Corollary 16.3.12 *The constrained critical points of the functional I_ε on $\mathcal{N}_\varepsilon$ are critical points of I_ε in W_ε.*

Proof The standard proof follows by using similar arguments explored in the last proposition. □

Now let us pass to the functional related to the autonomous problem (A_μ).

Lemma 16.3.13 (Ground State for the Autonomous Problem) *Let $\{u_n\} \subset \mathcal{M}_\mu$ be a sequence satisfying $E_\mu(u_n) \to m(\mu)$. Then, up to subsequences the following alternative holds:*

a) $\{u_n\}$ strongly converges in $H^s(\mathbb{R}^3)$;
b) there exists a sequence $\{\tilde{y}_n\} \subset \mathbb{R}^3$ such that $u_n(\cdot + \tilde{y}_n)$ strongly converges in $H^s(\mathbb{R}^3)$.

In particular, there exists a minimizer $\mathfrak{w}_\mu \geq 0$ for $m(\mu)$.

This result is known in the literature, but for completeness we give here the proof.

Proof By the Ekeland Variational Principle we may suppose that $\{u_n\}$ is a $(PS)_{m(\mu)}$ sequence for E_μ. Thus going to a subsequence if necessary, we have that $u_n \rightharpoonup u$ weakly in $H^s(\mathbb{R}^3)$ and it is easy to verify that $E'_\mu(u) = 0$.

In case $u \neq 0$, then $\mathfrak{w}_\mu := u$ is a ground state solution of the autonomous problem (A_μ), that is, $E_\mu(\mathfrak{w}_\mu) = m(\mu)$.

In case $u \equiv 0$, applying the same arguments employed in the proof of Lemma 16.3.8, there exists a sequence $\{\tilde{y}_n\} \subset \mathbb{R}^3$ such that

$$v_n \rightharpoonup v \ \text{ in } \ H^s(\mathbb{R}^3)$$

where $v_n := u_n(\cdot + \tilde{y}_n)$. Therefore, $\{v_n\}$ is also a $(PS)_{m(\mu)}$ sequence of E_μ and $v \not\equiv 0$. It follows from the above arguments that setting $\mathfrak{w}_\mu := v$ it is the ground state solution we were looking for.

In both cases, it is easy to see that $\mathfrak{w}_\mu \geq 0$ and the proof of the lemma is finished. □

Now we can complete the proof of Theorem 16.3.1 on the existence of ground state solution for problem (P_ε).

By Lemma 16.3.6, the functional I_ε has the geometry of the Mountain Pass Theorem in W_ε. Then by well-known results there exists $\{u_n\} \subset W_\varepsilon$ satisfying

$$I_\varepsilon(u_n) \to c_\varepsilon \text{ and } I'_\varepsilon(u_n) \to 0.$$

Case I $\mathbf{V}_\infty = \infty$. By Proposition 16.3.10, $\{u_n\}$ strongly converges to some $\mathfrak{u}_\varepsilon$ in $H^s(\mathbb{R}^3)$, which satisfies

$$I_\varepsilon(\mathfrak{u}_\varepsilon) = c_\varepsilon \text{ and } I'_\varepsilon(\mathfrak{u}_\varepsilon) = 0.$$

Case II $\mathbf{V}_\infty < \infty$. In virtue of Proposition 16.3.10 we just need to show that $c_\varepsilon < m(V_\infty)$. Suppose without loss of generality that $0 \in M$, i.e.

$$V(0) = V_0.$$

Let $\mu \in (V_0, V_\infty)$, so that

$$m(V_0) < m(\mu) < m(V_\infty). \tag{16.37}$$

For $r > 0$ let η_r a smooth cut-off function in $\mathbb{R}^3$ which equals 1 on B_r and with support in B_{2r}. Let $w_r := \eta_r \mathfrak{w}_\mu$ and $t_r > 0$ such that $t_r w_r \in \mathcal{M}_\mu$. If it were, for every $r > 0 : E_\mu(t_r w_r) \geq m(V_\infty)$, since $w_r \to \mathfrak{w}_\mu$ in $H^s(\mathbb{R}^3)$ for $r \to +\infty$, we would have $t_r \to 1$ and then

$$m(V_\infty) \leq \liminf_{r \to +\infty} E_\mu(t_r w_r) = E_\mu(\mathfrak{w}_\mu) = m(\mu)$$

which contradicts (16.37). Then there exists $\bar{r} > 0$ such that $\phi := t_{\bar{r}} w_{\bar{r}}$ satisfies $E_\mu(\phi) < m(V_\infty)$. Condition (16.3) implies that for some $\bar{\varepsilon} > 0$

$$V(\varepsilon x) \leq \mu, \quad \text{for all } x \in \operatorname{supp} \phi \text{ and } \varepsilon \leq \bar{\varepsilon},$$

so

$$\int_{\mathbb{R}^N} V(\varepsilon x) \phi^2 \leq \mu \int_{\mathbb{R}^N} \phi^2 \text{ for all } \varepsilon \leq \bar{\varepsilon}$$

and consequently

$$I_\varepsilon(t\phi) \leq E_\mu(t\phi) \leq E_\mu(\phi) \ \text{ for all } \ t > 0.$$

Therefore $\max_{t>0} I_\varepsilon(t\phi) \leq E_\mu(\phi)$, and then

$$c_\varepsilon < m(V_\infty)$$

which conclude the proof.

To get the multiplicity result we need to introduce the machinery of the "barycenter method."

16.3.3 The Barycenter Map

Up to now ε was fixed in our considerations. Now we deal with the case $\varepsilon \to 0^+$. The next result will be fundamental when we implement the "barycenter machinery" below.

Proposition 16.3.14 *Let $\varepsilon_n \to 0$ and $\{u_n\} \subset \mathcal{N}_{\varepsilon_n}$ be such that $I_{\varepsilon_n}(u_n) \to m(V_0)$. Then there exists a sequence $\{\tilde{y}_n\} \subset \mathbb{R}^3$ such that $u_n(\cdot + \tilde{y}_n)$ has a convergent subsequence in $H^s(\mathbb{R}^3)$. Moreover, up to a subsequence, $y_n := \varepsilon_n \tilde{y}_n \to y \in M$.*

Proof Arguing as in the proof of Lemma 16.3.8, we obtain a sequence $\{\tilde{y}_n\} \subset \mathbb{R}^3$ and constants $R, c > 0$ such that

$$\liminf_{n\to\infty} \int_{B_R(\tilde{y}_n)} u_n^2 \geq c > 0.$$

Thus, if $v_n := u_n(\cdot + \tilde{y}_n)$, up to a subsequence, $v_n \rightharpoonup v \not\equiv 0$ in $H^s(\mathbb{R}^3)$. Let $t_n > 0$ be such that $\tilde{v}_n := t_n v_n \in \mathcal{M}_{V_0}$. Then,

$$E_{V_0}(\tilde{v}_n) \to m(V_0).$$

Since $\{t_n\}$ is bounded, so is the sequence $\{\tilde{v}_n\}$, thus for some subsequence, $\tilde{v}_n \rightharpoonup \tilde{v}$ in $H^s(\mathbb{R}^N)$. Moreover, reasoning as in [3], up to some subsequence still denoted with $\{t_n\}$, we can assume that $t_n \to t_0 > 0$, and this limit implies that $\tilde{v} \not\equiv 0$. From Lemma 16.3.13, $\tilde{v}_n \to \tilde{v}$ in $H^s(\mathbb{R}^3)$, and so $v_n \to v$ in $H^s(\mathbb{R}^3)$.

Now, we will show that $\{y_n\} := \{\varepsilon_n \tilde{y}_n\}$ has a subsequence verifying $y_n \to y \in M$. First note that the sequence $\{y_n\}$ is bounded in $\mathbb{R}^3$. Indeed, assume by contradiction that (up to subsequences) $|y_n| \to \infty$.

In case $V_\infty = \infty$, the inequality

$$\int_{\mathbb{R}^3} V(\varepsilon_n x + y_n) v_n^2 \leq \int_{\mathbb{R}^3} |(-\Delta)^{s/2} v_n|^2 + \int_{\mathbb{R}^3} V(\varepsilon_n x + y_n) v_n^2 = \int_{\mathbb{R}^3} f(v_n) v_n,$$

and the Fatou's Lemma imply

$$\infty = \liminf_{n\to\infty} \int_{\mathbb{R}^3} f(v_n)v_n$$

which is an absurd, since the sequence $\{f(v_n)v_n\}$ is bounded in $L^1(\mathbb{R}^3)$.

Now let us consider the case $V_\infty < \infty$. Since $\tilde{v}_n \to \tilde{v}$ in $H^s(\mathbb{R}^3)$ and $V_0 < V_\infty$, we have

$$\begin{aligned} m(V_0) &= \frac{1}{2}\int_{\mathbb{R}^3} |(-\Delta)^{s/2}\tilde{v}|^2 + \frac{V_0}{2}\int_{\mathbb{R}^3} \tilde{v}^2 - \int_{\mathbb{R}^3} F(\tilde{v}) \\ &< \frac{1}{2}\int_{\mathbb{R}^3} |(-\Delta)^{s/2}\tilde{v}|^2 + \frac{V_\infty}{2}\int_{\mathbb{R}^3} \tilde{v}^2 - \int_{\mathbb{R}^3} F(\tilde{v}) \\ &\le \liminf_{n\to\infty}\left[\frac{1}{2}\int_{\mathbb{R}^3} |(-\Delta)^{s/2}\tilde{v}_n|^2 + \frac{1}{2}\int_{\mathbb{R}^3} V(\varepsilon_n x + y_n)\tilde{v}_n^2 - \int_{\mathbb{R}^3} F(\tilde{v}_n)\right], \end{aligned}$$

or equivalently

$$m(V_0) < \liminf_{n\to\infty}\left[\frac{t_n^2}{2}\int_{\mathbb{R}^3} |(-\Delta)^{s/2}u_n|^2 + \frac{t_n^2}{2}\int_{\mathbb{R}^3} V(\varepsilon_n z)u_n^2 - \int_{\mathbb{R}^3} F(t_n u_n)\right].$$

The last inequality implies

$$m(V_0) < \liminf_{n\to\infty} I_{\varepsilon_n}(t_n u_n) \le \liminf_{n\to\infty} I_{\varepsilon_n}(u_n) = m(V_0),$$

which is a contradiction. Hence, $\{y_n\}$ has to be bounded and, up to a subsequence, $y_n \to y \in \mathbb{R}^N$. If $y \notin 3$, then $V(y) > V_0$ and we obtain a contradiction arguing as above. Thus, $y \in M$ and the Proposition is proved. □

Let $\delta > 0$ be fixed and η be a smooth nonincreasing cut-off function defined in $[0, \infty)$ by

$$\eta(s) = \begin{cases} 1 & \text{if } 0 \le s \le \delta/2 \\ 0 & \text{if } s \ge \delta. \end{cases}$$

Let $\mathfrak{w}_{V_0}$ be a ground state solution given in Lemma 16.3.13 of problem (A_μ) with $\mu = V_0$ and for any $y \in M$, let us define

$$\Psi_{\varepsilon,y}(x) := \eta(|\varepsilon x - y|)\mathfrak{w}_{V_0}\left(\frac{\varepsilon x - y}{\varepsilon}\right).$$

Let $t_\varepsilon > 0$ verifying $\max_{t\ge 0} I_\varepsilon(t\Psi_{\varepsilon,y}) = I_\varepsilon(t_\varepsilon\Psi_{\varepsilon,y})$, so that $t_\varepsilon\Psi_{\varepsilon,y} \in \mathcal{N}_\varepsilon$, and let

$$\Phi_\varepsilon : y \in M \mapsto t_\varepsilon\Psi_{\varepsilon,y} \in \mathcal{N}_\varepsilon.$$

By construction, $\Phi_\varepsilon(y)$ has compact support for any $y \in M$ and Φ_ε is a continuous map.

The next result will help us to define a map from M to a suitable sublevel in the Nehari manifold.

Lemma 16.3.15 *The function Φ_ε satisfies*

$$\lim_{\varepsilon\to 0^+} I_\varepsilon(\Phi_\varepsilon(y)) = m(V_0), \ \textit{uniformly in } y \in M.$$

Proof Suppose by contradiction that the lemma is false. Then there exist $\delta_0 > 0$, $\{y_n\} \subset M$ and $\varepsilon_n \to 0^+$ such that

$$|I_{\varepsilon_n}(\Phi_{\varepsilon_n}(y_n)) - m(V_0)| \geq \delta_0. \tag{16.38}$$

Repeating the same arguments explored in [2], it is possible to check that $t_{\varepsilon_n} \to 1$. From Lebesgue's Theorem, we can check that

$$\lim_{n\to\infty} \|\Psi_{\varepsilon_n,y_n}\|^2_{\varepsilon_n} = \|\mathfrak{w}_{V_0}\|^2_{V_0}$$

and

$$\lim_{n\to\infty} \int_{\mathbb{R}^3} F(\Psi_{\varepsilon_n,y_n}) = \int_{\mathbb{R}^3} F(\mathfrak{w}_{V_0}).$$

Now, note that

$$\begin{aligned} I_{\varepsilon_n}(\Phi_{\varepsilon_n}(y_n)) = &\frac{t^2_{\varepsilon_n}}{2}\int_{\mathbb{R}^3}\left|(-\Delta)^{s/2}(\eta(|\varepsilon_n z|)\mathfrak{w}_{V_0}(z))\right|^2 \\ &+\frac{t^2_{\varepsilon_n}}{2}\int_{\mathbb{R}^3} V(\varepsilon_n z + y_n)|\eta(|\varepsilon_n z|)\mathfrak{w}_{V_0}(z)|^2 \\ &-\int_{\mathbb{R}^3} F(t_{\varepsilon_n}\eta(|\varepsilon_n z|)\mathfrak{w}_{V_0}(z)). \end{aligned}$$

Letting $n \to \infty$, we get $\lim_{n\to\infty} I_{\varepsilon_n}(\Phi_{\varepsilon_n}(y_n)) = E_{V_0}(\mathfrak{w}_{V_0}) = m(V_0)$, which contradicts (16.38). Thus the Lemma holds. □

Observe that by Lemma 16.3.15, $h(\varepsilon) := |I_\varepsilon(\Phi_\varepsilon(y)) - m(V_0)| = o(1)$ for $\varepsilon \to 0^+$ uniformly in y, and then $I_\varepsilon(\Phi_\varepsilon(y)) - m(V_0) \leq h(\varepsilon)$. In particular the set

$$\mathcal{N}_\varepsilon^{m(V_0)+h(\varepsilon)} := \left\{u \in \mathcal{N}_\varepsilon : I_\varepsilon(u) \leq m(V_0) + h(\varepsilon)\right\}$$

is not empty, since for sufficiently small ε,

$$\forall\, y \in M : \Phi_\varepsilon(y) \in \mathcal{N}_\varepsilon^{m(V_0)+h(\varepsilon)}. \tag{16.39}$$

We are in a position now to define the barycenter map that will send a convenient sublevel in the Nehari manifold in a suitable neighborhood of M. From now on we fix a $\delta > 0$ in such a way that M and

$$M_{2\delta} := \left\{ x \in \mathbb{R}^3 : d(x, M) \le 2\delta \right\}$$

are homotopically equivalent (d denotes the euclidean distance). Let $\rho = \rho(\delta) > 0$ be such that $M_{2\delta} \subset B_\rho$ and $\chi : \mathbb{R}^3 \to \mathbb{R}^3$ be defined as

$$\chi(x) = \begin{cases} x & \text{if } |x| \le \rho \\ \rho \dfrac{x}{|x|} & \text{if } |x| \ge \rho. \end{cases}$$

Finally, let us consider the so-called *barycenter map* β_ε defined on functions with compact support $u \in W_\varepsilon$ by

$$\beta_\varepsilon(u) := \frac{\displaystyle\int_{\mathbb{R}^3} \chi(\varepsilon x) u^2(x)}{\displaystyle\int_{\mathbb{R}^3} u^2(x)} \in \mathbb{R}^3.$$

Lemma 16.3.16 *The function β_ε satisfies*

$$\lim_{\varepsilon \to 0^+} \beta_\varepsilon(\Phi_\varepsilon(y)) = y, \quad \textit{uniformly in } y \in M.$$

Proof Suppose, by contradiction, that the lemma is false. Then, there exist $\delta_0 > 0$, $\{y_n\} \subset M$ and $\varepsilon_n \to 0^+$ such that

$$|\beta_{\varepsilon_n}(\Phi_{\varepsilon_n}(y_n)) - y_n| \ge \delta_0. \tag{16.40}$$

Using the definition of $\Phi_{\varepsilon_n}(y_n)$, β_{ε_n} and η given above, we have the equality

$$\beta_{\varepsilon_n}(\Phi_{\varepsilon_n}(y_n)) = y_n + \frac{\displaystyle\int_{\mathbb{R}^3} [\chi(\varepsilon_n z + y_n) - y_n] \left| \eta(|\varepsilon_n z|) w(z) \right|^2}{\displaystyle\int_{\mathbb{R}^3} \left| \eta(|\varepsilon_n z|) w(z) \right|^2}.$$

Using the fact that $\{y_n\} \subset M \subset B_\rho$ and the Lebesgue's Theorem, it follows

$$|\beta_{\varepsilon_n}(\Phi_{\varepsilon_n}(y_n)) - y_n| = o_n(1),$$

which contradicts (16.40) and the Lemma is proved. □

Lemma 16.3.17 *We have*

$$\lim_{\varepsilon\to 0^+} \sup_{u\in\mathcal{N}_\varepsilon^{m(V_0)+h(\varepsilon)}} \inf_{y\in M_\delta} \left|\beta_\varepsilon(u)-y\right| = 0.$$

Proof Let $\{\varepsilon_n\}$ be such that $\varepsilon_n \to 0^+$. For each $n \in \mathbb{N}$, there exists $u_n \in \mathcal{N}_{\varepsilon_n}^{m(V_0)+h(\varepsilon_n)}$ such that

$$\inf_{y\in M_\delta} \left|\beta_{\varepsilon_n}(u_n)-y\right| = \sup_{u\in\mathcal{N}_{\varepsilon_n}^{m(V_0)+h(\varepsilon_n)}} \inf_{y\in M_\delta} \left|\beta_{\varepsilon_n}(u)-y\right| + o_n(1).$$

Thus, it suffices to find a sequence $\{y_n\} \subset M_\delta$ such that

$$\lim_{n\to\infty} \left|\beta_{\varepsilon_n}(u_n)-y_n\right| = 0. \tag{16.41}$$

Recalling that $u_n \in \mathcal{N}_{\varepsilon_n}^{m(V_0)+h(\varepsilon_n)} \subset \mathcal{N}_{\varepsilon_n}$ we have

$$m(V_0) \le c_{\varepsilon_n} \le I_{\varepsilon_n}(u_n) \le m(V_0) + h(\varepsilon_n),$$

so $I_{\varepsilon_n}(u_n) \to m(V_0)$. By Proposition 16.3.14, we get a sequence $\{\tilde{y}_n\} \subset \mathbb{R}^3$ such that $v_n := u_n(\cdot + \tilde{y}_n)$ converges in $H^s(\mathbb{R}^3)$ to some v and $\{y_n\} := \{\varepsilon_n \tilde{y}_n\} \subset M_\delta$, for n sufficiently large. Thus

$$\beta_{\varepsilon_n}(u_n) = y_n + \frac{\displaystyle\int_{\mathbb{R}^3} [\chi(\varepsilon_n z + y_n) - y_n] v_n^2(z)}{\displaystyle\int_{\mathbb{R}^N} v_n(z)^2}.$$

Since $v_n \to v$ in $H^s(\mathbb{R}^3)$, it is easy to check that the sequence $\{y_n\}$ verifies (16.41). □

16.3.4 Proof of Theorem 16.3.2

In virtue of Lemma 16.3.17, there exists $\varepsilon^* > 0$ such that

$$\forall\, \varepsilon \in (0, \varepsilon^*] : \sup_{u\in\mathcal{N}_\varepsilon^{m(V_0)+h(\varepsilon)}} d(\beta_\varepsilon(u), M_\delta) < \delta/2.$$

Define now

$$M^+ := \left\{x \in \mathbb{R}^3 : d(x, M) \le 3\delta/2\right\}$$

so that M and M^+ are homotopically equivalent.

Now, reducing $\varepsilon^* > 0$ if necessary, we can assume that Lemmas 16.3.16, 16.3.17 and (16.39) hold. Then by standard arguments the composed map

$$M \xrightarrow{\Phi_\varepsilon} \mathcal{N}_\varepsilon^{m(V_0)+h(\varepsilon)} \xrightarrow{\beta_\varepsilon} M^+ \quad \text{is homotopic to the inclusion map.} \tag{16.42}$$

In case $V_\infty < \infty$, we eventually reduce ε^* in such a way that also the Palais-Smale condition is satisfied in the interval $(m(V_0), m(V_0) + h(\varepsilon))$, see Proposition 16.3.11.

By (16.42) and well-known properties of the category, we get

$$\mathrm{cat}(\mathcal{N}_\varepsilon^{m(V_0)+h(\varepsilon)}) \ge \mathrm{cat}_{M^+}(M),$$

and the Ljusternik-Schnirelman theory (see, e.g., [42]) implies that I_ε has at least $\mathrm{cat}_{M^+}(M) = \mathrm{cat}\, M$ critical points on $\mathcal{N}_\varepsilon$.

To obtain another solution we use the same ideas of [18]. First note that, since M is not contractible, the set $\mathcal{A} := \Phi_\varepsilon(M)$ cannot be contractible in $\mathcal{N}_\varepsilon^{m(V_0)+h(\varepsilon)}$. Moreover $\mathcal{A}$ is compact.

For $u \in W_\varepsilon \setminus \{0\}$ we denote with $t_\varepsilon(u) > 0$ the unique positive number such that $t_\varepsilon(u)u \in \mathcal{N}_\varepsilon$. Let $u^* \in W_\varepsilon$ be such that $u^* \ge 0$, and $I_\varepsilon(t_\varepsilon(u^*)u^*) > m(V_0) + h(\varepsilon)$. Consider the cone

$$\mathfrak{C} := \left\{tu^* + (1-t)u : t \in [0, 1], u \in \mathcal{A}\right\}$$

and note that $0 \notin \mathfrak{C}$, since functions in $\mathfrak{C}$ have to be positive on a set of nonzero measure. Clearly it is compact and contractible. Let

$$t_\varepsilon(\mathfrak{C}) := \left\{t_\varepsilon(w)w : w \in \mathfrak{C}\right\}$$

be its projection on $\mathcal{N}_\varepsilon$, which is compact as well, and

$$c := \max_{t_\varepsilon(\mathfrak{C})} I_\varepsilon > m(V_0) + h(\varepsilon).$$

Since $\mathcal{A} \subset t_\varepsilon(\mathfrak{C}) \subset \mathcal{N}_\varepsilon$ and $t_\varepsilon(\mathfrak{C})$ is contractible in $\mathcal{N}_\varepsilon^c := \{u \in \mathcal{N}_\varepsilon : I_\varepsilon(u) \le c\}$, we infer that also $\mathcal{A}$ is contractible in $\mathcal{N}_\varepsilon^c$.

Summing up, we have a set $\mathcal{A}$ which is contractible in $\mathcal{N}_\varepsilon^c$ but not in $\mathcal{N}_\varepsilon^{m(V_0)+h(\varepsilon)}$, where $c > m(V_0) + h(\varepsilon)$. This is only possible, since I_ε satisfies the Palais-Smale condition, if there is a critical level between $m(V_0) + h(\varepsilon)$ and c.

By Corollary 16.3.12, we conclude the proof of statements about the existence of solutions in Theorem 16.3.2.

16.4 The Case of Unknown Potential: The Fractional Schrödinger-Poisson System

The results shown in this section are taken from [53].

In the last decades a great attention has been given to the following Schrödinger-Poisson type system

$$\begin{cases} -\Delta u + V(x)u + \phi u = |u|^{p-2}u \\ -\Delta \phi = u^2, \end{cases}$$

which arises in nonrelativistic Quantum Mechanics. Such a system is obtained by looking for standing wave solutions in the purely electrostatic case to the Schrödinger-Maxwell system. For a deduction of this system, see, e.g., [19]. Here the unknowns are u, the modulus of the wave function, and ϕ which represents the electrostatic potential. V is a given external potential and $p \geq 2$ a suitable given number.

The system has been studied by many authors, both in bounded and unbounded domains, with different assumptions on the data involved: boundary conditions, potentials, nonlinearities; many different type of solutions have been encountered (minimal energy, sign changing, radial, nonradial, etc.), the behavior of the solutions (e.g., concentration phenomena) has been studied as well as multiplicity results have been obtained. It is really difficult to give a complete list of references: the reader may see [20] and the references therein.

However it seems that results relating the number of positive solutions with topological invariants of the "objects" appearing in the problem are few in the literature. We cite the paper [58] where the system is studied in a (smooth and) bounded domain $\Omega \subset \mathbb{R}^3$ with $u = \phi = 0$ on $\partial\Omega$ and V constant. It is shown, by using variational methods, that whenever p is sufficiently near the critical Sobolev exponent 6, the number of positive solutions is estimated below by the *Ljusternick-Schnirelmann category* of the domain Ω.

On the other hand, it is known that a particular interest has the *semiclassical limit* of the Schrödinger-Poisson system (that is when the Plank constant $\hbar$ appearing in the system, see, e.g., [19], tends to zero) especially due to the fact that this limit describes the transition from Quantum to Classical Mechanics. Such a situation is studied, e.g., in [57], among many other papers. We cite also Fang and Zhang [37] which consider the following doubly perturbed system in the whole space $\mathbb{R}^3$:

$$\begin{cases} -\varepsilon^2 \Delta w + V(x)w + \psi w = |w|^{p-2}w \\ -\varepsilon \Delta \psi = w^2. \end{cases}$$

Here V is a suitable potential, $4 < p < 6$, and ε is a positive parameter proportional to $\hbar$. In this case the authors estimate, whenever ε tends to zero, the number of positive solutions by the Ljusternick-Schnirelamnn category of the set of minima of the potential V, obtaining a result in the same spirit of [58].

Motivated by the previous discussion, we investigate in this paper the existence of positive solutions for the following doubly singularly perturbed fractional Schrödinger-Poisson system in $\mathbb{R}^N$:

$$\begin{cases} \varepsilon^{2s} (-\Delta)^s w + V(x)w + \psi w = f(w) \\ \varepsilon^{\theta} (-\Delta)^{\alpha/2} \psi = \gamma_\alpha w^2, \end{cases} \tag{P_ε}$$

where $\gamma_\alpha := \frac{\pi^{N/2} 2^\alpha \Gamma(\alpha/2)}{\Gamma(N/2-\alpha/2)}$ is a constant (Γ is the Euler function). By a positive solution of (P_ε) we mean a pair (w, ψ) where w is positive. To the best of our knowledge, there are only few recent papers dealing with a system like (P_ε): in [63] the author deals with $\varepsilon = 1$ proving under suitable assumptions on f the existence of infinitely many (but possibly sign changing) solutions by means of the Fountain Theorem. A similar system is studied in [61] and the existence of infinitely many (again, possibly sign changing) solutions is obtained by means of the Symmetric Mountain Pass Theorem.

In this section we assume that

(H) $s \in (0, 1), \alpha \in (3 - 2s, 3), \theta \in (0, \alpha)$,

moreover the potential V and the nonlinearity f satisfy the assumptions listed below:

(V) $V : \mathbb{R}^3 \to \mathbb{R}$ is a continuous function and

$$0 < \min_{\mathbb{R}^3} V := V_0 < V_\infty := \liminf_{|x|\to+\infty} V \in (V_0, +\infty];$$

(f1) $f : \mathbb{R} \to \mathbb{R}$ is a function of class C^1 and $f(t) = 0$ for $t \leq 0$;
(f2) there is $q_0 \in (2, 2^*_s - 1)$ such that $\lim_{t\to\infty} f(t)/t^{q_0} = 0$, where $2^*_s := 6/(3 - 2s)$;
(f3) there is $K > 4$ such that $0 < KF(t) := K \int_0^t f(\tau)d\tau \leq tf(t)$ for all $t > 0$;
(f4) $\dfrac{d}{dt}\dfrac{f(t)}{t^3} > 0$ in $(0, +\infty)$.

The assumptions on the nonlinearity f are quite standard in order to work with variational methods, use the Nehari manifold and the Palais-Smale condition. The assumption (V) will be fundamental in order to estimate the number of positive solutions and also to recover some compactness.

Observe that by (f1) it follows that

$$\lim_{t\to 0} f(t)/t = 0. \tag{16.43}$$

Moreover

$$\forall \varepsilon > 0\, \exists M_\xi > 0 \ : \quad \int_{\mathbb{R}^3} f(u)u \leq \xi \int_{\mathbb{R}^3} u^2 + M_\xi \int_{\mathbb{R}^3} |u|^{q_0+1}, \quad \forall u \in H^s(\mathbb{R}^3). \tag{16.44}$$

which simply follows by (16.43) and (f2). A similar inequality is used in the proof of Lemma 16.3.8 in the previous section.

To state our result let us introduce

$$M := \left\{x \in \mathbb{R}^3 : V(x) = V_0\right\}$$

the set of minima of V. Our results are the following

Theorem 16.4.1 *Under the above assumptions (H), (V), (f1)–(f4), there exists a ground state solution* $\mathrm{u}_\varepsilon \in W_\varepsilon$ *of problem* (P_ε),

1. *for every* $\varepsilon \in (0, \bar{\varepsilon}]$, *for some* $\bar{\varepsilon} > 0$, *if* $V_\infty < \infty$;
2. *for every* $\varepsilon > 0$, *if* $V_\infty = \infty$.

Theorem 16.4.2 *Under the above assumptions (H), (V), (f1)–(f4), there exists an* $\varepsilon^* > 0$ *such that for every* $\varepsilon \in (0, \varepsilon^*]$ *problem* (P_ε) *has at least*

1. cat M *positive solutions.*
2. cat $M + 1$ *positive solutions, if* M *is bounded and* cat $M \geq 2$.

As for the "single" Fractional Schrödinger equation, the proof of Theorem 16.4.2 is carried out by adapting some ideas of Benci, Cerami, and Passaseo [17, 18, 21] and using the Ljusternick-Schnirelmann Theory.

As done in Sect. 16.3 the plan here is the following.

- **The Variational Setting**. Here after a change of variable we introduce an equivalent problem. Then some preliminaries facts are presented and the variational setting for the problem is given.
- **Compactness for $\mathbf{I}_\varepsilon$ and $\mathbf{E}_\mu$: Existence of a Ground State Solution**. Here we prove some compactness properties; as a by-product we prove the existence of a ground state solution for our problem, proving Theorem 16.4.1.
- **The Barycenter Map**. Here barycenter maps are introduced in order to estimate the category of suitable sub levels of the energy functional for $\varepsilon \to 0$.
- **Proof of Theorem 16.4.2**. Here, by using all the previous machinery, the proof of the multiplicity result is completed.

16.4.1 The Variational Setting

It is easily seen that, just performing the change of variables $w(x) := u(x/\varepsilon)$, $\psi(x) := \phi(x/\varepsilon)$, problem ($P_\varepsilon$) can be rewritten as

$$\begin{cases} (-\Delta)^s u + V(\varepsilon x)u + \phi(x)u = f(u) \\ (-\Delta)^{\alpha/2}\phi = \varepsilon^{\alpha-\theta}\gamma_\alpha u^2, \end{cases} \qquad (P_\varepsilon^*)$$

to which we will refer from now on.

A usual "reduction" argument can be used to deal with a single equation involving just u. Indeed for every $u \in H^s(\mathbb{R}^3)$ the second equation in (P_ε^*) is uniquely solved. Actually, for future reference, we will prove a slightly more general fact.

Let us fix two functions $u, w \in H^s(\mathbb{R}^3)$ and consider the problem

$$\begin{cases} (-\Delta)^{\alpha/2}\phi = \varepsilon^{\alpha-\theta}\gamma_\alpha uw, \\ \phi \in \dot{H}^{\alpha/2}(\mathbb{R}^3) \end{cases} \qquad (Q_\varepsilon)$$

whose weak solution is a function $\tilde{\phi} \in \dot{H}^{\alpha/2}(\mathbb{R}^3)$ such that

$$\forall v \in \dot{H}^{\alpha/2}(\mathbb{R}^3): \quad \int_{\mathbb{R}^3}(-\Delta)^{\alpha/4}\tilde{\phi}(-\Delta)^{\alpha/4}v = \varepsilon^{\alpha-\theta}\gamma_\alpha\int_{\mathbb{R}^3} uwv.$$

For every $v \in \dot{H}^{\alpha/2}(\mathbb{R}^3)$, by the Hölder inequality and the continuous embeddings, we have

$$\Big|\int_{\mathbb{R}^3} uwv\Big| \le |u|_{\frac{4\cdot 3}{3+\alpha}}|w|_{\frac{4\cdot 3}{3+\alpha}}|v|_{2^*_{\alpha/2}} \le C\|u\|\|w\|\|v\|_{\dot{H}^{\alpha/2}}$$

deducing that the map

$$T_{u,w}: v \in \dot{H}^{\alpha/2}(\mathbb{R}^3) \longmapsto \int_{\mathbb{R}^3} uwv \in \mathbb{R}$$

is linear and continuous: then there exists a unique solution $\phi_{\varepsilon,u,w} \in \dot{H}^{\alpha/2}(\mathbb{R}^3)$ to (Q_ε). Moreover this solution has the representation by means of the Riesz kernel $\mathcal{K}_\alpha(x) = \gamma_\alpha^{-1}|x|^{\alpha-3}$, hence

$$\phi_{\varepsilon,u,w} = \varepsilon^{\alpha-\theta}\frac{1}{|\cdot|^{3-\alpha}} \star (uw).$$

Furthermore

$$\|\phi_{\varepsilon,u,w}\|_{\dot{H}^{\alpha/2}} = \varepsilon^{\alpha-\theta}\|T_{u,w}\|_{\mathcal{L}(\dot{H}^{\alpha/2};\mathbb{R})} \le \varepsilon^{\alpha-\theta}C\|u\|\|w\| \qquad (16.45)$$

and then, for $\zeta, \eta \in H^s(\mathbb{R}^3)$

$$\int_{\mathbb{R}^3} \phi_{\varepsilon,u,w}\zeta\eta \leq \|\phi_{\varepsilon,u,w}\|_{2^*_{\alpha/2}}|\zeta|_{\frac{4\cdot 3}{3+\alpha}}|\eta|_{\frac{4\cdot 3}{N+\alpha}} \leq \varepsilon^{\alpha-\theta}C_{\mathrm{e}}\|u\|\|w\|\|\zeta\|\|\eta\| \tag{16.46}$$

where C_{e} is a suitable embedding constant. Although its value is not important, we will refer to this constant later on.

A particular case of the previous situation is when $u = w$. In this case we simplify the notation and write

- $T_u(v) := T_{u,u}(v) = \int_{\mathbb{R}^3} u^2 v$, and
- $\phi_{\varepsilon,u}$ for the unique solution of the second equation in (P^*_ε) for fixed $u \in H^s(\mathbb{R}^3)$. Then

  $$\|\phi_{\varepsilon,u}\|_{\dot{H}^{\alpha/2}} \leq \varepsilon^{\alpha-\theta}C\|u\|^2$$

  and the map

  $$u \in H^s(\mathbb{R}^3) \longmapsto \phi_{\varepsilon,u} \in \dot{H}^{\alpha/2}(\mathbb{R}^3)$$

  is bounded.

Observe also that

$$u_n^2 \to u^2 \text{ in } L^{\frac{2\cdot 3}{3+\alpha}}(\mathbb{R}^3) \Longrightarrow T_{u_n} \to T_u \text{ as operators}$$
$$\Longrightarrow \phi_{\varepsilon,u_n} \to \phi_{\varepsilon,u} \text{ in } \dot{H}^{\alpha/2}(\mathbb{R}^3). \tag{16.47}$$

For convenience let us define the map (well defined by (16.46))

$$A : u \in H^s(\mathbb{R}^3) \longmapsto \int_{\mathbb{R}^3} \phi_{\varepsilon,u}u^2 \in \mathbb{R}.$$

Then

$$|A(u)| \leq \varepsilon^{\alpha-\theta}C_{\mathrm{e}}\|u\|^4 \tag{16.48}$$

(where C_{e} is the same constant in (16.46)). Some relevant properties of $\phi_{\varepsilon,u}$ and A are listed below. Although these properties are known to be true, we are not able to find them explicitly in the literature; so we prefer to give a proof here.

Lemma 16.4.3 *The following propositions hold.*

(*i*) *For every* $u \in H^s(\mathbb{R}^3) : \phi_{\varepsilon,u} \geq 0$*;*
(*ii*) *for every* $u \in H^s(\mathbb{R}^3), t \in \mathbb{R} : \phi_{\varepsilon,tu} = t^2\phi_{\varepsilon,u}$*;*
(*iii*) *if* $u_n \rightharpoonup u$ *in* $H^s(\mathbb{R}^3)$*, then* $\phi_{\varepsilon,u_n} \rightharpoonup \phi_{\varepsilon,u}$ *in* $\dot{H}^{\alpha/2}(\mathbb{R}^3)$*;*

(iv) *A is of class* C^2 *and for every* $u, v, w \in H^s(\mathbb{R}^3)$

$$A'(u)[v] = 4\int_{\mathbb{R}^3} \phi_{\varepsilon,u} uv, \quad A''(u)[v,w] = 4\int_{\mathbb{R}^3} \phi_{\varepsilon,u} vw + 8\int_{\mathbb{R}^3} \phi_{\varepsilon,u,w} uv,$$

(v) *if* $u_n \to u$ *in* $L^r(\mathbb{R}^3)$, *with* $2 \le r < 2_s^*$, *then* $A(u_n) \to A(u)$;
(vi) *if* $u_n \rightharpoonup u$ *in* $H^s(\mathbb{R}^3)$ *then* $A(u_n - u) = A(u_n) - A(u) + o_n(1)$.

Proof Items (i) and (ii) follow directly by the definition of $\phi_{\varepsilon,u}$.

To prove (iii), let $v \in C_c^\infty(\mathbb{R}^3)$; we have

$$\begin{aligned}\int_{\mathbb{R}^3} (-\Delta)^{\alpha/4}(\phi_{\varepsilon,u_n} - \phi_{\varepsilon,u})(-\Delta)^{\alpha/4} v &= \int_{\mathbb{R}^3} (u_n^2 - u^2)v \\ &\le |v|_\infty \Big(\int_{\mathrm{supp}\, v} (u_n - u)^2\Big)^{1/2} \\ &\quad \times \Big(\int_{\mathrm{supp}\, v} (u_n + u)^2\Big)^{1/2} \to 0.\end{aligned}$$

The conclusion then follows by density.

The proof of (iv) is straightforward: we refer the reader to [37].

To show (v), recall that $2 < \frac{4\cdot 3}{3+\alpha} < 2_s^*$. Since by assumption $|u_n^2|_{\frac{2\cdot 3}{3+\alpha}} \to |u^2|_{\frac{2\cdot 3}{3+\alpha}}$ and $u_n^2 \to u^2$ a.e. in $\mathbb{R}^3$, using the Brezis-Lieb Lemma, $u_n^2 \to u^2$ in $L^{\frac{2\cdot 3}{3+\alpha}}(\mathbb{R}^3)$. But then using (16.47) we get $\phi_{\varepsilon,u_n} \to \phi_{\varepsilon,u}$ in $L^{2^*_{\alpha/2}}(\mathbb{R}^3)$. Consequently

$$\begin{aligned}|A(u_n) - A(u)| &\le \int_{\mathbb{R}^3} |\phi_{\varepsilon,u_n} u_n^2 - \phi_{\varepsilon,u} u^2| \\ &\le \int_{\mathbb{R}^3} |\left(\phi_{\varepsilon,u_n} - \phi_{\varepsilon,u}\right) u_n^2| + \int_{\mathbb{R}^3} |\phi_{\varepsilon,u}\left(u_n^2 - u^2\right)| \\ &\le |\phi_{\varepsilon,u_n} - \phi_{\varepsilon,u}|_{2^*_{\alpha/2}} |u_n^2|_{\frac{2\cdot 3}{3+\alpha}} + |\phi_{\varepsilon,u}|_{2^*_{\alpha/2}} |u_n^2 - u^2|_{\frac{2\cdot 3}{3+\alpha}}\end{aligned}$$

from which we conclude.

To prove (vi), for the sake of simplicity we drop the factor $\varepsilon^{\alpha-\theta}$ in the expression of $\phi_{\varepsilon,u,v}$. Defining

$$\sigma := \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{u^2(y)u^2(x)}{|x-y|^{3-\alpha}} dydx\,,$$

$$\sigma_n^1 := \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{u_n^2(y)u^2(x)}{|x-y|^{3-\alpha}} dydx\,, \qquad \sigma_n^2 := \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{u_n(y)u(y)u_n(x)u(x)}{|x-y|^{3-\alpha}} dydx$$

$$\sigma_n^3 := \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{u_n^2(y)u_n(x)u(x)}{|x-y|^{3-\alpha}} dydx\,, \qquad \sigma_n^4 := \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{u_n(y)u(y)u^2(x)}{|x-y|^{3-\alpha}} dydx,$$

it is easy to check that

$$A(u_n - u) - A(u_n) + A(u) = 2\sigma + 2\sigma_n^1 + 4\sigma_n^2 - 4\sigma_n^3 - 4\sigma_n^4.$$

Now we claim that, whenever $u_n \rightharpoonup u$ in $H^s(\mathbb{R}^3)$,

$$\lim_{n\to\infty} \sigma_n^i = \sigma, \quad i = 1, 2, 3, 4$$

which readily gives the conclusion.

We prove here only the cases $i = 1, 2$ since the proof of the other cases is very similar. Recall that

$$\phi_{\varepsilon,u}(x) = \int_{\mathbb{R}^3} \frac{u^2(y)}{|x-y|^{3-\alpha}} dy, \qquad \phi_{\varepsilon,u_n}(x) = \int_{\mathbb{R}^3} \frac{u_n^2(y)}{|x-y|^{3-\alpha}} dy.$$

Since $u^2 \in L^{\frac{2\cdot 3}{3+\alpha}}(\mathbb{R}^3) = L^{(2^*_\alpha/2)'}(\mathbb{R}^3)$ and by item (iii) it holds $\phi_{\varepsilon,u_n} \to \phi_{\varepsilon,u}$ in $L^{2^*_\alpha/2}(\mathbb{R}^3)$, we conclude that

$$\sigma_n^1 = \int_{\mathbb{R}^3} \phi_{\varepsilon,u_n} u^2 \to \int_{\mathbb{R}^3} \phi_{\varepsilon,u} u^2 = \sigma$$

and the claim is true for $i = 1$.

For $i = 2$ recall that

$$\phi_{\varepsilon,u_n,u}(x) = \int_{\mathbb{R}^3} \frac{u_n(y)\, u(y)}{|x-y|^{3-\alpha}} dy.$$

First we show that $\phi_{\varepsilon,u_n,u} \to \phi_{\varepsilon,u}$ a.e. in $\mathbb{R}^3$. Given $\xi > 0$ and choosing $R > 1/\xi$, $\frac{N}{2s} < p < \frac{3}{3-\alpha}$ and $\frac{3}{3-\alpha} < q$ (so that $2p', 2q' \in (2, 2^*_s)$), we have, for large n:

$$\begin{aligned} |\phi_{\varepsilon,u_n,u}(x) - \phi_{\varepsilon,u}(x)| \le\ & |u_n - u|_{L^{2p'}(B_R(x))} |u|_{L^{2p'}(B_R(x))} \\ & \times \Big(\int_{|y-x|<R} \frac{dy}{|x-y|^{p(3-\alpha)}} \Big)^{1/p} \\ & + |u_n - u|_{L^{2q'}(B_R^c(x))} |u|_{L^{2q'}(B_R^c(x))} \\ & \times \left(\int_{|y-x|\ge R} \frac{dy}{|x-y|^{q(3-\alpha)}} \right)^{1/q} \\ \le\ & C_1 \xi + C_2 \xi^{3-\alpha}, \end{aligned}$$

concluding the pointwise convergence. Moreover by the Sobolev embedding and using (16.45),

$$|\phi_{\varepsilon,u_n,u} u_n|_2 \le |\phi_{\varepsilon,u_n,u}|_{2^*_{\alpha/2}} |u_n|_{2N/\alpha} \le C_1 \|u_n\|^2 \|u\| \le C_2$$

and therefore, up to subsequence, $\phi_{\varepsilon,u_n,u}u_n \rightharpoonup \phi_{\varepsilon,u}u$ in $L^2(\mathbb{R}^3)$, by Kavian [44, Lemma 4.8]. Since $u \in L^2(\mathbb{R}^3)$

$$\sigma_n^2 = \int_{\mathbb{R}^3} \phi_{\varepsilon,u_n,u}u_n u \to \int_{\mathbb{R}^3} \phi_{\varepsilon,u}u^2 = \sigma$$

and the claim is proved for $i = 2$. □

We introduce now the variational setting for our problem. Let us define the Hilbert space

$$W_\varepsilon := \left\{ u \in H^s(\mathbb{R}^3) : \int_{\mathbb{R}^3} V(\varepsilon x)u^2 < \infty \right\}$$

endowed with scalar product and (squared) norm given by

$$(u, v)_\varepsilon := \int_{\mathbb{R}^3} (-\Delta)^{s/2}u(-\Delta)^{s/2}v + \int_{\mathbb{R}^3} V(\varepsilon x)uv$$

and

$$\|u\|_\varepsilon^2 := \int_{\mathbb{R}^3} |(-\Delta)^{s/2}u|^2 + \int_{\mathbb{R}^3} V(\varepsilon x)u^2.$$

Then it is standard to see that the critical points of the C^2 functional (see Lemma 16.4.3 (iv))

$$I_\varepsilon(u) := \frac{1}{2}\int_{\mathbb{R}^3} |(-\Delta)^{s/2}u|^2 + \frac{1}{2}\int_{\mathbb{R}^3} V(\varepsilon x)u^2 + \frac{1}{4}\int_{\mathbb{R}^3} \phi_{\varepsilon,u}u^2 - \int_{\mathbb{R}^3} F(u),$$

on W_ε are weak solutions of problem (P_ε^*).

By defining

$$\mathcal{N}_\varepsilon := \left\{ u \in W_\varepsilon \setminus \{0\} : J_\varepsilon(u) = 0 \right\},$$

where

$$J_\varepsilon(u) := I'_\varepsilon(u)[u] = \|u\|_\varepsilon^2 + \int_{\mathbb{R}^3} \phi_{\varepsilon,u}u^2 - \int_{\mathbb{R}^3} f(u)u,$$

we have, by standard arguments:

Lemma 16.4.4 *For every $u \in \mathcal{N}_\varepsilon$, $J'_\varepsilon(u)[u] < 0$ and there are positive constants $h_\varepsilon, k_\varepsilon$ such that $\|u\|_\varepsilon \geq h_\varepsilon$, $I_\varepsilon(u) \geq k_\varepsilon$. Furthermore, $\mathcal{N}_\varepsilon$ is diffeomorphic to the set*

$$\mathcal{S}_\varepsilon := \{u \in W_\varepsilon : \|u\|_\varepsilon = 1\} \setminus \{u \in W_\varepsilon : u \leq 0 \text{ a.e. }\}.$$

$\mathcal{N}_\varepsilon$ is the *Nehari manifold* associated with I_ε. By the assumptions on f, the functional I_ε has the Mountain Pass geometry. This is standard but we give the easy proof for completeness.

(MP1) $I_\varepsilon(0) = 0$;
(MP2) since, for every $\xi > 0$ there exists $M_\xi > 0$ such that $F(u) \le \xi u^2 + M_\xi |u|^{q_0+1}$, we have

$$\begin{aligned} I_\varepsilon(u) &\ge \frac{1}{2}\|u\|_\varepsilon^2 - \int_{\mathbb{R}^3} F(u) \\ &\ge \frac{1}{2}\|u\|_\varepsilon^2 - \xi C_1 \|u\|_\varepsilon^2 - M_\xi C_2 \|u\|_\varepsilon^{q_0+1} \end{aligned}$$

and we conclude I_ε has a strict local minimum at $u = 0$;
(MP3) finally, since (f3) implies $F(t) \ge Ct^K$ for $t > 0$, with $K > 4$ (and less than $q_0 + 1$), fixed $v \in C_c^\infty(\mathbb{R}^N)$, $v > 0$ we have

$$\begin{aligned} I_\varepsilon(tv) &= \frac{t^2}{2}\|v\|_\varepsilon^2 + \frac{t^4}{4}\int_{\mathbb{R}^3} \phi_{\varepsilon,v} v^2 - \int_{\mathbb{R}^3} F(tv) \\ &\le \frac{t^2}{2}\|v\|_\varepsilon^2 + \frac{t^4}{4}\int_{\mathbb{R}^N} \phi_{\varepsilon,v} v^2 - Ct^K \int_{\mathbb{R}^3} v^K \end{aligned}$$

concluding that the functional is negative for suitable large t.

Then denoting with

$$\begin{aligned} c_\varepsilon &:= \inf_{\gamma \in \mathcal{H}_\varepsilon} \sup_{t\in[0,1]} I_\varepsilon(\gamma(t)), \\ \mathcal{H}_\varepsilon &= \Big\{ \gamma \in C([0,1], W_\varepsilon) : \gamma(0) = 0,\, I_\varepsilon(\gamma(1)) < 0 \Big\} \end{aligned}$$

the Mountain Pass level, and with

$$m_\varepsilon := \inf_{u\in\mathcal{N}_\varepsilon} I_\varepsilon(u)$$

the *ground state level*, it holds, in a standard way, that

$$c_\varepsilon = m_\varepsilon = \inf_{u\in W_\varepsilon\setminus\{0\}} \sup_{t\ge 0} I_\varepsilon(tu). \tag{16.49}$$

It is known that for "perturbed" problems a major role is played by the problem at infinity that we now introduce.

16.4.1.1 The Problem at "Infinity"

Let us consider the "limit" problem (the autonomous problem) associated with (P_ε^*), that is

$$\begin{cases} (-\Delta)^s u + \mu u = f(u) \\ u \in H^s(\mathbb{R}^3) \end{cases} \tag{A_μ}$$

where $\mu > 0$ is a constant. The solutions are critical points of the functional

$$E_\mu(u) = \frac{1}{2}\int_{\mathbb{R}^3} |(-\Delta)^{s/2}u|^2 + \frac{\mu}{2}\int_{\mathbb{R}^3} u^2 - \int_{\mathbb{R}^3} F(u).$$

in $H^s(\mathbb{R}^3)$. Denoting with $H^s_\mu(\mathbb{R}^3)$ simply the space $H^s(\mathbb{R}^3)$ endowed with the (equivalent squared) norm

$$\|u\|^2_{H^s_\mu} := |(-\Delta)^{s/2}u|_2^2 + \mu|u|_2^2,$$

by the assumptions of the nonlinearity f, it is easy to see that the functional E_μ has the Mountain Pass geometry with Mountain Pass level

$$c_\mu^\infty := \inf_{\gamma\in\mathcal{H}_\mu} \sup_{t\in[0,1]} E_\mu(\gamma(t)),$$

$$\mathcal{H}_\mu := \Big\{\gamma \in C([0,1], H^s_\mu(\mathbb{R}^3)) : \gamma(0) = 0, E_\mu(\gamma(1)) < 0\Big\}.$$

Introducing the set

$$\mathcal{M}_\mu := \Big\{u \in H^s(\mathbb{R}^3)\setminus\{0\} : \|u\|^2_{H^s_\mu} = \int_{\mathbb{R}^3} f(u)u\Big\}$$

it is standard to see that

- $\mathcal{M}_\mu$ has a structure of differentiable manifold (said the *Nehari manifold* associated with E_μ),
- $\mathcal{M}_\mu$ is bounded away from zero and radially homeomorphic to the unit sphere,
- the mountain pass value c_μ^∞ coincides with the *ground state level*

$$m_\mu^\infty := \inf_{u\in\mathcal{M}_\mu} E_\mu(u) > 0.$$

The symbol "∞" in the notations is just to recall we are dealing with the limit problem. In the sequel we will mainly deal with $\mu = V_0$ and $\mu = V_\infty$ (whenever this last one is finite). Of course the inequality

$$m_\varepsilon \geq m^\infty_{V_0}$$

holds.

16.4.2 Compactness for I_ε and E_μ: Existence of a Ground State Solution

We begin by showing the boundedness of the Palais-Smale sequences for E_μ in $H^s_\mu(\mathbb{R}^3)$ and I_ε in W_ε. Let $\{u_n\} \subset H^s_\mu(\mathbb{R}^3)$ be a Palais-Smale sequence for E_μ, that is, $|E_\mu(u_n)| \le C$ and $E'_\mu(u_n) \to 0$. Then, for large n,

$$\begin{aligned} C + \|u_n\|_{H^s_\mu} &> E_\mu(u_n) - \frac{1}{K} E'_\mu(u_n)[u_n] \\ &= \left(\frac{1}{2} - \frac{1}{K}\right) \|u_n\|^2_{H^s_\mu} + \frac{1}{K} \int_{\mathbb{R}^3} (f(u_n)u_n - KF(u_n)) \\ &\ge \left(\frac{1}{2} - \frac{1}{K}\right) \|u_n\|^2_{H^s_\mu}, \end{aligned}$$

and thus $\{u_n\}$ is bounded. Similarly we conclude for I_ε, using that

$$\begin{aligned} I_\varepsilon(u_n) - \frac{1}{K} I'_\varepsilon(u_n)[u_n] &= \left(\frac{1}{2} - \frac{1}{K}\right) ||u_n||^2_\varepsilon + \left(\frac{1}{4} - \frac{1}{K}\right) \int_{\mathbb{R}^3} \phi_{\varepsilon,u_n} u_n^2 \\ &\quad + \frac{1}{K} \int_{\mathbb{R}^3} (f(u_n)u_n - KF(u_n)) \\ &\ge \left(\frac{1}{2} - \frac{1}{K}\right) ||u_n||^2_\varepsilon. \end{aligned}$$

In order to prove compactness, some preliminary work is needed. Let us recall the following Lions type lemma, whose proof can be found in [30, Lemma 2.3].

Lemma 16.4.5 *If $\{u_n\}$ is bounded in $H^s(\mathbb{R}^3)$ and for some $R > 0$ and $2 \le r < 2^*_s$ we have*

$$\sup_{x \in \mathbb{R}^3} \int_{B_R(x)} |u_n|^r \to 0 \quad as \quad n \to \infty,$$

then $u_n \to 0$ in $L^p(\mathbb{R}^3)$ for $2 < p < 2^*_s$.

Then we can prove the following

Lemma 16.4.6 *Let $\{u_n\} \subset W_\varepsilon$ be bounded and such that $I'_\varepsilon(u_n) \to 0$. Then we have either*

a) $u_n \to 0$ in W_ε, or
b) there exist a sequence $\{y_n\} \subset \mathbb{R}^3$ and constants $R, c > 0$ such that

$$\liminf_{n \to +\infty} \int_{B_R(y_n)} u_n^2 \ge c > 0.$$

Proof Suppose that b) does not occur. Using Lemma 16.4.5 it follows

$$u_n \to 0 \ \text{ in } \ L^p(\mathbb{R}^3) \ \text{ for } \ p \in (2, 2_s^*).$$

Using (16.44), the boundedness of $\{u_n\}$ in $L^2(\mathbb{R}^3)$ and the fact that $u_n \to 0$ in $L^{q_0+1}(\mathbb{R}^3)$, we conclude that

$$\int_{\mathbb{R}^3} f(u_n)u_n \to 0.$$

Finally, since

$$||u_n||_\varepsilon^2 - \int_{\mathbb{R}^3} f(u_n)u_n \leq ||u_n||_\varepsilon^2 + \int_{\mathbb{R}^3} \phi_{\varepsilon,u_n} u_n^2 - \int_{\mathbb{R}^3} f(u_n)u_n = I'_\varepsilon(u_n)[u_n] = o_n(1),$$

it follows that $u_n \to 0$ in W_ε. □

In the rest of the section we assume, without loss of generality, that $0 \in M$, that is, $V(0) = V_0$.

Lemma 16.4.7 *Assume that $V_\infty < \infty$ and let $\{v_n\} \subset W_\varepsilon$ be a $(PS)_d$ sequence for I_ε such that $v_n \rightharpoonup 0$ in W_ε. Then*

$$v_n \not\to 0 \ \text{ in } \ W_\varepsilon \implies d \geq m_{V_\infty}^\infty.$$

Proof Observe, preliminarily, that by condition (V) it follows that

$$\forall \xi > 0 \, \exists \widetilde{R} = \widetilde{R}_\xi > 0 : \quad V(\varepsilon x) > V_\infty - \xi, \qquad \forall x \notin B_{\widetilde{R}}. \tag{16.50}$$

Let $\{t_n\} \subset (0, +\infty)$ be such that $\{t_n v_n\} \subset \mathcal{M}_{V_\infty}$. We start by showing the following

Claim The sequence $\{t_n\}$ satisfies $\limsup_{n\to\infty} t_n \leq 1$.

Supposing by contradiction that the claim does not hold, there exists $\delta > 0$ and a subsequence still denoted by $\{t_n\}$, such that

$$t_n \geq 1 + \delta \quad \text{for all} \quad n \in \mathbb{N}. \tag{16.51}$$

Since $\{v_n\}$ is a bounded $(PS)_d$ sequence for I_ε, $I'_\varepsilon(v_n)[v_n] = o_n(1)$, that is,

$$||v_n||_\varepsilon^2 + \int_{\mathbb{R}^3} \phi_{\varepsilon,v_n} v_n^2 = \int_{\mathbb{R}^3} f(v_n)v_n + o_n(1).$$

Moreover, since $\{t_n v_n\} \subset \mathcal{M}_{V_\infty}$, we get

$$||t_n v_n||_{H^s_{V_\infty}}^2 = \int_{\mathbb{R}^3} f(t_n v_n)t_n v_n.$$

These equalities imply that

$$\int_{\mathbb{R}^3}\left(\frac{f(t_nv_n)}{t_n}-f(v_n)\right)v_n=\int_{\mathbb{R}^3}(V_\infty-V(\varepsilon x))v_n^2-\int_{\mathbb{R}^3}\phi_{\varepsilon,v_n}v_n^2+o_n(1),$$

and thus

$$\int_{\mathbb{R}^N}\Big(\frac{f(t_nv_n)}{t_n}-f(v_n)\Big)v_n\leq\int_{\mathbb{R}^N}(V_\infty-V(\varepsilon x))v_n^2+o_n(1). \tag{16.52}$$

Using (16.50), the fact that $v_n\to 0$ in $L^2(B_{\widetilde{R}})$ and that $\{v_n\}$ is bounded in W_ε, let us say by some constant $C>0$, we deduce by (16.52)

$$\forall\xi>0:\int_{\mathbb{R}^3}\left(\frac{f(t_nv_n)}{t_n}-f(v_n)\right)v_n\leq\xi C+o_n(1). \tag{16.53}$$

Since $v_n\not\to 0$ in W_ε, we may invoke Lemma 16.4.6 to obtain $\{y_n\}\subset\mathbb{R}^3$ and $R,c>0$ such that

$$\int_{B_R(y_n)}v_n^2\geq c. \tag{16.54}$$

Defining $\check{v}_n:=v_n(\cdot+y_n)$, we may suppose that, up to a subsequence,

$$\check{v}_n\rightharpoonup\check{v}\ \text{ in }\ H^s(\mathbb{R}^3)$$

and, in view of (16.54), there exists a subset $\Omega\subset\mathbb{R}^N$ with positive measure such that $\check{v}>0$ in Ω. By (f4) and (16.51), (16.53) becomes

$$0<\int_\Omega\left(\frac{f((1+\delta)\check{v}_n)}{(1+\delta)\check{v}_n}-\frac{f(\check{v}_n)}{\check{v}_n}\right)\check{v}_n^2\leq\xi C+o_n(1).$$

Now passing to the limit and applying Fatou's Lemma, it follows that, for every $\xi>0$

$$0<\int_\Omega\left[\frac{f((1+\delta)\check{v})}{(1+\delta)\check{v}}-\frac{f(\check{v})}{\check{v}}\right]\check{v}^2\leq\xi C,$$

which is absurd and proves the claim.

Now we distinguish two cases.

Case 1 $\limsup_{n\to\infty}t_n=1$.

Up to subsequence we can assume that $t_n\to 1$. We have

$$d+o_n(1)=I_\varepsilon(v_n)\geq m_{V_\infty}^\infty+I_\varepsilon(v_n)-E_{V_\infty}(t_nv_n). \tag{16.55}$$

Moreover,

$$I_\varepsilon(v_n) - E_{V_\infty}(t_n v_n) = \frac{(1-t_n^2)}{2}\int_{\mathbb{R}^N} |(-\Delta)^{s/2} v_n|^2 + \frac{1}{2}\int_{\mathbb{R}^N} (V(\varepsilon x) - t_n^2 V_\infty) v_n^2 + \frac{1}{4}\int_{\mathbb{R}^n} \phi_{\varepsilon, v_n} v_n^2 + \int_{\mathbb{R}^N} (F(t_n v_n) - F(v_n)),$$

and due to the boundedness of $\{v_n\}$ we get, for every $\xi > 0$,

$$I_\varepsilon(v_n) - E_{V_\infty}(t_n v_n) \geq \; o_n(1) - C\xi + \int_{\mathbb{R}^N} (F(t_n v_n) - F(v_n)),$$

where we have used again (16.50). By the Mean Value Theorem, it is

$$\int_{\mathbb{R}^3} (F(t_n v_n) - F(v_n)) = o_n(1),$$

therefore (16.55) becomes

$$d + o_n(1) \geq m_{V_\infty}^\infty - C\xi + o_n(1),$$

and taking the limit in n, by the arbitrariness of ξ, we deduce $d \geq m_{V_\infty}^\infty$.

Case 2 $\limsup_{n\to\infty} t_n = t_0 < 1$.

We can assume $t_n \to t_0$ and $t_n < 1$. Since $t \mapsto \frac{1}{4} f(t)t - F(t)$ is increasing in $(0, \infty)$,

$$\begin{aligned} m_{V_\infty}^\infty \leq E_{V_\infty}(t_n v_n) &= \int_{\mathbb{R}^3} \left(\frac{1}{2} f(t_n v_n) t_n v_n - F(t_n v_n)\right) \\ &= \int_{\mathbb{R}^N} \frac{1}{4} f(t_n v_n) t_n v_n + \int_{\mathbb{R}^3} \left(\frac{1}{4} f(t_n v_n) t_n v_n - F(t_n v_n)\right) \\ &= \frac{1}{4} ||t_n v_n||_{H^s_{V_\infty}}^2 + \int_{\mathbb{R}^3} \left(\frac{1}{4} f(t_n v_n) t_n v_n - F(t_n v_n)\right) \\ &\leq \frac{1}{4} ||t_n v_n||_{H^s_{V_\infty}}^2 + \int_{\mathbb{R}^3} \left(\frac{1}{4} f(v_n) v_n - F(v_n)\right). \end{aligned} \tag{16.56}$$

But

$$||t_n v_n||_{V_\infty}^2 \leq \int_{\mathbb{R}^3} \left|(-\Delta)^{s/2} v_n\right|^2 + \int_{\mathbb{R}^3} t_n^2 V_\infty v_n^2. \tag{16.57}$$

Again by (16.50), given $\xi > 0$,

$$t_n^2 V_\infty - \xi < V_\infty - \xi < V(\varepsilon x) \quad \text{for } x \notin B_{\widetilde{R}}$$

and hence

$$\int_{\mathbb{R}^3} t_n^2 V_\infty v_n^2 \le \int_{B_{\widetilde{R}}} V_\infty v_n^2 + \int_{|x|\ge \widetilde{R}} V(\varepsilon x) v_n^2 + \int_{|x|\ge \widetilde{R}} \xi v_n^2$$
$$\le o_n(1) + \int_{\mathbb{R}^3} V(\varepsilon x) v_n^2 + C\xi.$$

From this and (16.57) we have

$$||t_n v_n||^2_{H^s_{V_\infty}} \le ||v_n||^2_\varepsilon + C\xi + o_n(1).$$

Therefore, using (16.56)

$$m^\infty_{V_\infty} \le \frac{1}{4}||v_n||^2_\varepsilon + \int_{\mathbb{R}^3}\left(\frac{1}{4} f(v_n)v_n - F(v_n)\right) + C\xi + o_n(1)$$
$$= I_\varepsilon(v_n) - \frac{1}{4} I'_\varepsilon(v_n)[v_n] + C\xi + o_n(1)$$
$$= d + C\xi + o_n(1).$$

concluding the proof. □

Proposition 16.4.8 *The functional I_ε in W_ε satisfies the $(PS)_c$ condition*

1. *at any level $c < m^\infty_{V_\infty}$, if $V_\infty < \infty$,*
2. *at any level $c \in \mathbb{R}$, if $V_\infty = \infty$.*

Proof Let $\{u_n\} \subset W_\varepsilon$ be such that $I_\varepsilon(u_n) \to c$ and $I'_\varepsilon(u_n) \to 0$. We have already seen that $\{u_n\}$ is bounded in W_ε. Thus there exists $u \in W_\varepsilon$ such that, up to a subsequence, $u_n \rightharpoonup u$ in W_ε. Note that $I'_\varepsilon(u) = 0$, since by Lemma 16.4.3 (iv), we have for every $w \in W_\varepsilon$

$$(u_n, w)_\varepsilon \to (u, w)_\varepsilon, \quad A'(u_n)[w] \to A'(u)[w] \quad \text{and} \int_{\mathbb{R}^N} f(u_n)w \to \int_{\mathbb{R}^N} f(u)w.$$

Defining $v_n := u_n - u$, we have that

$$\int_{\mathbb{R}^3} F(v_n) = \int_{\mathbb{R}^N} F(u_n) - \int_{\mathbb{R}^3} F(u) + o_n(1)$$

(see [11]) and by Lemma 16.4.3 (vi), we have $A(v_n) = A(u_n) - A(u) + o_n(1)$; hence arguing as in [3], we obtain also

$$I'_\varepsilon(v_n) \to 0. \tag{16.58}$$

Moreover

$$I_\varepsilon(v_n) = I_\varepsilon(u_n) - I_\varepsilon(u) + o_n(1) = c - I_\varepsilon(u) + o_n(1) =: d + o_n(1) \tag{16.59}$$

and (16.58) and (16.59) show that $\{v_n\}$ is a $(PS)_d$ sequence. By (f3),

$$\begin{aligned} I_\varepsilon(u) &= I_\varepsilon(u) - \frac{1}{4} I'_\varepsilon(u)[u] \\ &= \frac{1}{4} ||u||^2_\varepsilon + \int_{\mathbb{R}^3} \Big(\frac{1}{4} f(u)u - F(u)\Big) \\ &\geq \frac{1}{4} \int_{\mathbb{R}^3} \Big(f(u)u - 4F(u)\Big) \\ &\geq 0 \end{aligned}$$

and then coming back in (16.59) we have

$$d \leq c. \tag{16.60}$$

Then,

1. if $V_\infty < \infty$, and $c < m^\infty_{V_\infty}$, by (16.60) we obtain
$$d \leq c < m^\infty_{V_\infty}.$$
It follows from Lemma 16.4.7 that $v_n \to 0$, that is $u_n \to u$ in W_ε.
2. If $V_\infty = \infty$, by the compact imbedding $W_\varepsilon \hookrightarrow\hookrightarrow L^r(\mathbb{R}^3)$, $2 \leq r < 2^*_s$, up to a subsequence, $v_n \to 0$ in $L^r(\mathbb{R}^3)$ and since $I'_\varepsilon(v_n) \to 0$, we have
$$I'_\varepsilon(v_n)[v_n] = ||v_n||^2_\varepsilon + \int_{\mathbb{R}^3} \phi_{\varepsilon, v_n} v_n^2 - \int_{\mathbb{R}^3} f(v_n)v_n = o_n(1). \tag{16.61}$$
By Lemma 16.4.3 (*v*), $A(v_n) = \int_{\mathbb{R}^3} \phi_{\varepsilon, v_n} v_n^2 = o_n(1)$, and since by (16.44) it holds again $\int_{\mathbb{R}^3} f(v_n)v_n = o_n(1)$, we have by (16.61) $||v_n||^2_\varepsilon = o_n(1)$, that is $u_n \to u$ in W_ε.

The proof is thereby complete. □

As a consequence it is standard to prove that

Proposition 16.4.9 *The functional I_ε restricted to $\mathcal{N}_\varepsilon$ satisfies the $(PS)_c$ condition*

1. *at any level $c < m^\infty_{V_\infty}$, if $V_\infty < \infty$,*
2. *at any level $c \in \mathbb{R}$, if $V_\infty = \infty$.*

Moreover, the constrained critical points of the functional I_ε on $\mathcal{N}_\varepsilon$ are critical points of I_ε in W_ε, hence solutions of (P_ε^*).

Let us recall the following result (see [39, Lemma 6]) concerning problem (A_μ).

Lemma 16.4.10 (Ground State for the Autonomous Problem) *Let $\{u_n\} \subset \mathcal{M}_\mu$ be a sequence satisfying $E_\mu(u_n) \to m_\mu^\infty$. Then, up to subsequences the following alternative holds:*

a) $\{u_n\}$ strongly converges in $H^s(\mathbb{R}^3)$;
b) there exists a sequence $\{\tilde{y}_n\} \subset \mathbb{R}^3$ such that $u_n(\cdot + \tilde{y}_n)$ strongly converges in $H^s(\mathbb{R}^3)$.

In particular, there exists a minimizer $\mathfrak{w}_\mu \geq 0$ for m_μ^∞.

Now we can prove the existence of a ground state for our problem. Assumption (H) is tacitly assumed.

Proof Since the functional I_ε has the geometry of the Mountain Pass Theorem in W_ε there exists $\{u_n\} \subset W_\varepsilon$ satisfying

$$I_\varepsilon(u_n) \to c_\varepsilon \text{ and } I'_\varepsilon(u_n) \to 0.$$

1. If $V_\infty < \infty$, in virtue of Proposition 16.4.8, we have only to show that $c_\varepsilon < m_{V_\infty}^\infty$ for every positive ε smaller than a certain $\bar{\varepsilon}$.

   Let $\mu \in (V_0, V_\infty)$, so that

$$m_{V_0}^\infty < m_\mu^\infty < m_{V_\infty}^\infty. \tag{16.62}$$

   For $r > 0$ let η_r be a smooth cut-off function in $\mathbb{R}^3$ which equals 1 on B_r and with support in B_{2r}. Let $w_r := \eta_r \mathfrak{w}_\mu$ and $s_r > 0$ such that $s_r w_r \in \mathcal{M}_\mu$. If it were, for every $r > 0 : E_\mu(s_r w_r) \geq m_{V_\infty}^\infty$, since $w_r \to \mathfrak{w}_\mu$ in $H^s(\mathbb{R}^3)$ for $r \to +\infty$, we would have $s_r \to 1$ and then

$$m_{V_\infty}^\infty \leq \liminf_{r \to +\infty} E_\mu(s_r w_r) = E_\mu(\mathfrak{w}_\mu) = m_\mu^\infty$$

   which contradicts (16.62). This means that there exists $\bar{r} > 0$ such that $\omega := s_{\bar{r}} w_{\bar{r}} \in \mathcal{M}_\mu$ satisfies

$$E_\mu(\omega) < m_{V_\infty}^\infty. \tag{16.63}$$

   Given $\varepsilon > 0$, let $t_\varepsilon > 0$ the number such that $t_\varepsilon \omega \in \mathcal{N}_\varepsilon$. Therefore

$$t_\varepsilon^2 ||\omega||_\varepsilon^2 + t_\varepsilon^4 \int_{\mathbb{R}^3} \phi_{\varepsilon,\omega} \omega^2 = t_\varepsilon \int_{\mathbb{R}^3} f(t_\varepsilon \omega)\omega$$

implying that

$$\frac{\|\omega\|_\varepsilon^2}{t_\varepsilon^2} + \int_{\mathbb{R}^3} \phi_{\varepsilon,\omega}\,\omega^2 \geq \int_{B_{\overline{r}}} \frac{f\,(t_\varepsilon\,\omega)}{(t_\varepsilon\,\omega)^3}\omega^4. \tag{16.64}$$

Now we claim that there exists $T > 0$ such that $\limsup_{\varepsilon\to 0^+} t_\varepsilon \leq T$. If by contradiction there exists $\varepsilon_n \to 0^+$ with $t_{\varepsilon_n} \to \infty$, then by (16.64) and (f4) we have

$$\frac{\|\omega\|_{\varepsilon_n}^2}{t_{\varepsilon_n}^2} + \int_{\mathbb{R}^3} \phi_{\varepsilon_n,\omega}\,\omega^2 \geq \frac{f(t_{\varepsilon_n}\,\omega(\overline{x}))}{(t_{\varepsilon_n}\,\omega(\overline{x}))^3}\int_{B_{\overline{r}}} \omega^4, \tag{16.65}$$

where $\omega(\overline{x}) := \min_{B_{\overline{r}}} \omega\,(x)$. The absurd is achieved by passing to the limit in n, since by (f3) the right-hand side of (16.65) tends to ∞, while the left-hand side tends to 0.

Then there exists $\varepsilon_1 > 0$ such that

$$\forall \varepsilon \in (0, \varepsilon_1] : \quad t_\varepsilon \in (0, T]. \tag{16.66}$$

Condition (V) implies also that there exists some $\varepsilon_2 > 0$ such that

$$\forall \varepsilon \in (0, \varepsilon_2] : \quad V(\varepsilon x) \leq \frac{V_0 + \mu}{2}, \quad \text{for all} \quad x \in \operatorname{supp}\omega. \tag{16.67}$$

Finally let

$$\varepsilon_3 := \left(\frac{(\mu - V_0)|\omega|_2^2}{C_{\mathrm{e}}\,T^2\|\omega\|^4}\right)^{1/(\alpha-\theta)},$$

where C_{e} is the same constant appearing in (16.48), hence in particular

$$\forall \varepsilon \in (0, \varepsilon_3] : \int_{\mathbb{R}^3} \phi_{\varepsilon,\omega}\omega^2 \leq \varepsilon^{\alpha-\theta} C_{\mathrm{e}}\|\omega\|^4 \text{ and}$$

$$T^2\varepsilon^{\alpha-\theta} C_{\mathrm{e}}\,\|\omega\|^4 \leq (\mu - V_0)\int_{\mathbb{R}^3} \omega^2. \tag{16.68}$$

Let $\bar{\varepsilon} := \min\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$. By using (16.66)–(16.68) we have, for every $\varepsilon \in (0, \bar{\varepsilon}]$:

$$\int_{\mathbb{R}^3} V(\varepsilon x)\omega^2 + \frac{t_\varepsilon^2}{2}\int_{\mathbb{R}^3} \phi_{\varepsilon,\omega}\,\omega^2 \leq \frac{V_0 + \mu}{2}|\omega|_2^2 + \frac{1}{2}T^2\varepsilon^{\alpha-\theta}C_{\mathrm{e}}\|\omega\|^4 \leq \mu\int_{\mathbb{R}^3} \omega^2,$$

from which we infer $I_\varepsilon(t_\varepsilon\omega) \leq E_\mu(t_\varepsilon\omega)$. Then by (16.49) and (16.63),

$$c_\varepsilon \leq I_\varepsilon(t_\varepsilon\,\omega) \leq E_\mu\,(t_\varepsilon\,\omega) \leq E_\mu(\omega) < m_{V_\infty}^\infty.$$

which concludes the proof in this case.

2. If $V_\infty = \infty$, by Proposition 16.4.8, $\{u_n\}$ strongly converges to some $\mathfrak{u}_\varepsilon$ in $H^s(\mathbb{R}^3)$, which satisfies

$$I_\varepsilon(\mathfrak{u}_\varepsilon) = c_\varepsilon \text{ and } I'_\varepsilon(\mathfrak{u}_\varepsilon) = 0.$$

and $\mathfrak{u}_\varepsilon$ is the ground state we were looking for. □

16.4.3 The Barycenter Map

In this subsection we introduce the barycenter map in order to study the "topological complexity" of suitable sublevels of the functional I_ε in the Nehari manifold. Let us start with the following

Proposition 16.4.11 *Let $\varepsilon_n \to 0^+$ and $u_n \in \mathcal{N}_{\varepsilon_n}$ be such that $I_{\varepsilon_n}(u_n) \to m^\infty_{V_0}$. Then there exists a sequence $\{\tilde{y}_n\} \subset \mathbb{R}^3$ such that $u_n(\cdot + \tilde{y}_n)$ has a convergent subsequence in $H^s(\mathbb{R}^3)$. Moreover, up to a subsequence, $y_n := \varepsilon_n \tilde{y}_n \to y \in M$.*

Recall that M is the set where V achieves the minimum V_0.

Proof We begin by showing that $\{u_n\}$ is bounded in $H^s_{V_0}(\mathbb{R}^3)$. By assumptions, $I'_{\varepsilon_n}(u_n)[u_n] = 0$ and $I_{\varepsilon_n}(u_n) \to m^\infty_{V_0}$ write as

$$||u_n||^2_{\varepsilon_n} + \int_{\mathbb{R}^3} \phi_{\varepsilon_n,u_n} u_n^2 = \int_{\mathbb{R}^3} f(u_n)u_n \tag{16.69}$$

and

$$\frac{1}{2}||u_n||^2_{\varepsilon_n} + \frac{1}{4}\int_{\mathbb{R}^3} \phi_{\varepsilon_n,u_n} u_n^2 - \int_{\mathbb{R}^3} F(u_n) = m^\infty_{V_0} + o_n(1)$$

which combined together give

$$\begin{aligned}\frac{1}{4}\int_{\mathbb{R}^3} f(u_n)u_n - \int_{\mathbb{R}^3} F(u_n) &= \frac{1}{4}\left(||u_n||^2_{\varepsilon_n} + \int_{\mathbb{R}^3} \phi_{\varepsilon_n,u_n} u_n^2\right) \\ &\quad - \int_{\mathbb{R}^3} F(u_n) \le m^\infty_{V_0} + o_n(1).\end{aligned}$$

Using (f3) we get

$$0 \le \left(\frac{1}{4} - \frac{1}{K}\right)\int_{\mathbb{R}^3} f(u_n)u_n \le m^\infty_{V_0} + o_n(1),$$

and therefore, coming back to (16.69), for some positive constant C (independent on n)

$$||u_n||_{H^s_{V_0}} \leq ||u_n||_{\varepsilon_n} \leq C. \tag{16.70}$$

We prove the following

Claim There exists $\{\tilde{y}_n\} \subset \mathbb{R}^N$ and $R, c > 0$ such that $\liminf_{n\to\infty} \int_{B_R(\tilde{y}_n)} u_n^2 \geq c > 0$.

Indeed, if it were not the case, then

$$\lim_{n\to\infty} \sup_{y\in\mathbb{R}^N} \int_{B_R(y)} u_n^2 = 0, \quad \text{for every } R > 0.$$

By Lemma 16.4.6, $u_n \to 0$ in $L^p(\mathbb{R}^N)$, for $2 < p < 2_s^*$ and then

$$\int_{\mathbb{R}^3} f(u_n)u_n \to 0.$$

Therefore $||u_n||_{\varepsilon_n}^2 + \int_{\mathbb{R}^3} \phi_{\varepsilon_n,u_n} u_n^2 = o_n(1)$, and also from

$$0 \leq \int_{\mathbb{R}^3} F(u_n) \leq \frac{1}{K} \int_{\mathbb{R}^3} f(u_n)u_n$$

we have $\int_{\mathbb{R}^3} F(u_n) = o_n(1)$. But then $\lim_{n\to\infty} I_{\varepsilon_n}(u_n) = m^\infty_{V_0} = 0$ which is a contradiction and proves our claim.

Then the sequence $v_n := u_n(\cdot + \tilde{y}_n)$ is also bounded in $H^s(\mathbb{R}^3)$ and

$$v_n \rightharpoonup v \not\equiv 0 \quad \text{in} \quad H^s(\mathbb{R}^3) \tag{16.71}$$

since

$$\int_{B_R} v^2 = \liminf_{n\to\infty} \int_{B_R} v_n^2 = \liminf_{n\to\infty} \int_{B_R(\tilde{y}_n)} u_n^2 \geq c > 0,$$

by the claim.

Let now $t_n > 0$ be such that $\tilde{v}_n := t_n v_n \in \mathcal{M}_{V_0}$; the next step is to prove that

$$E_{V_0}(\tilde{v}_n) \to m^\infty_{V_0}. \tag{16.72}$$

For this, note that

$$
\begin{aligned}
m_{V_0}^{\infty} \le E_{V_0}(\tilde{v}_n) &= \frac{1}{2}\,||\tilde{v}_n||_{V_0}^2 - \int_{\mathbb{R}^3} F(\tilde{v}_n) \\
&= \frac{t_n^2}{2}\int_{\mathbb{R}^3}\Big[\left|(-\Delta)^{s/2}u_n(x+\tilde{y}_n)\right|^2 + V_0u_n^2(x+\tilde{y}_n)\Big]dx \\
&\quad - \int_{\mathbb{R}^3} F(t_nu_n(x+\tilde{y}_n))dx \\
&= \frac{t_n^2}{2}\int_{\mathbb{R}^3}\left|(-\Delta)^{s/2}u_n(z)\right|^2 dz + \frac{t_n^2}{2}\int_{\mathbb{R}^3} V_0u_n^2(z)dz \\
&\quad - \int_{\mathbb{R}^3} F(t_nu_n(z))dz \\
&\le \frac{t_n^2}{2}\int_{\mathbb{R}^3}\left|(-\Delta)^{s/2}u_n\right|^2 + \frac{t_n^2}{2}\int_{\mathbb{R}^3} V(\varepsilon_nz)u_n^2 \\
&\quad + \frac{t_n^4}{4}\int_{\mathbb{R}^3}\phi_{\varepsilon_n,u_n}u_n^2 - \int_{\mathbb{R}^3} F(t_nu_n) \\
&= I_{\varepsilon_n}(t_nu_n)
\end{aligned}
$$

and then

$$
m_{V_0}^{\infty} \le E_{V_0}(\tilde{v}_n) \le I_{\varepsilon_n}(t_nu_n) \le I_{\varepsilon_n}(u_n) = m_{V_0}^{\infty} + o_n(1)
$$

which proves (16.72).

We can prove now that $v_n \to v$ in $H^s(\mathbb{R}^3)$. As in the first part of the proof (where we proved the boundedness of $\{u_n\}$ in $H_{V_0}^s(\mathbb{R}^3)$), it is easy to see that

$$
\{\tilde{v}_n\} \subset \mathcal{M}_{V_0} \quad \text{and} \quad E_{V_0}(\tilde{v}_n) \to m_{V_0}^{\infty} \implies ||\tilde{v}_n||_{H_{V_0}^s} \le C
$$

and an analogous claim as before holds for the sequence $\{\tilde{v}_n\}$. Then $\tilde{v}_n \rightharpoonup \bar{v}$ in $H_{V_0}^s(\mathbb{R}^3)$. Since $\|v_n\|_{H_{V_0}^s} \not\to 0$, there exists $\delta > 0$ such that

$$
0 < \delta \le ||v_n||_{H_{V_0}^s}. \tag{16.73}
$$

This implies

$$
0 < t_n\delta \le ||t_nv_n||_{H_{V_0}^s} = ||\tilde{v}_n||_{H_{V_0}^s} \le C,
$$

showing that, up to subsequence, $t_n \to t_0 \geq 0$. If now $t_0 = 0$ using (16.70) we derive

$$0 \leq ||\tilde{v}_n||_{H^s_{V_0}} = t_n \, ||v_n||_{H^s_{V_0}} \leq t_n C_1 \to 0,$$

so that $\tilde{v}_n \to 0$ in $H^s_{V_0}(\mathbb{R}^3)$. From this and (16.72) it follows $m^\infty_{V_0} = 0$ which is absurd. So $t_0 > 0$. Then $t_n v_n \rightharpoonup t_0 \bar{v} =: \tilde{v}$ in $H^s(\mathbb{R}^3)$ and by (16.73) $\tilde{v} \not\equiv 0$. By Lemma 16.4.10 applied to $\{\tilde{v}_n\}$ we get $\tilde{v}_n \to \tilde{v}$ in $H^s(\mathbb{R}^3)$ and then $v_n \to \bar{v}$. By (16.71) we deduce $v_n \to v$ and the first part of the proposition is proved.

We proceed to prove the second part. We first state that $\{y_n\}$ is bounded in $\mathbb{R}^3$ (here $y_n = \varepsilon_n \tilde{y}_n$ with $\tilde{y}_n$ given in the above claim). Assume the contrary; then

1. if $V_\infty < \infty$, since $\tilde{v}_n \to \tilde{v}$ in $H^s(\mathbb{R}^3)$ and $V_0 < V_\infty$, we have

$$\begin{aligned} m^\infty_{V_0} &= \frac{1}{2} \, ||\tilde{v}||^2_{H^s_{V_0}} - \int_{\mathbb{R}^3} F(\tilde{v}) < \frac{1}{2} \, ||\tilde{v}||^2_{H^s_{V_\infty}} - \int_{\mathbb{R}^3} F(\tilde{v}) \\ &\leq \liminf_{n\to\infty} \frac{1}{2} \int_{\mathbb{R}^3} \left|(-\Delta)^{s/2} \tilde{v}_n\right|^2 \\ &\quad + \lim_{n\to\infty} \left(\frac{1}{2} \int_{\mathbb{R}^3} V(\varepsilon_n x + y_n) \tilde{v}_n^2(x) dx - \int_{\mathbb{R}^3} F(\tilde{v}_n) \right) \\ &= \liminf_{n\to\infty} \left(\frac{t_n^2}{2} \int_{\mathbb{R}^3} \left|(-\Delta)^{s/2} u_n\right|^2 + \frac{t_n^2}{2} \int_{\mathbb{R}^3} V(\varepsilon_n z) u_n^2 - \int_{\mathbb{R}^3} F(t_n u_n) \right) \\ &\leq \liminf_{n\to\infty} \left(\frac{1}{2} \, ||t_n u_n||^2_{\varepsilon_n} - \int_{\mathbb{R}^3} F(t_n u_n) + \frac{t_n^4}{4} \int_{\mathbb{R}^3} \phi_{\varepsilon_n, u_n} u_n^2 \right) \end{aligned}$$

from which

$$m^\infty_{V_0} < \liminf_{n\to\infty} I_{\varepsilon_n}(t_n u_n) \leq \liminf_{n\to\infty} I_{\varepsilon_n}(u_n) = m^\infty_{V_0}$$

which is a contradiction.

2. If $V_\infty = \infty$, we have

$$\begin{aligned} \int_{\mathbb{R}^3} V(\varepsilon_n x + y_n) v_n^2(x) dx \leq \int_{\mathbb{R}^3} & \left|(-\Delta)^{s/2} v_n(x)\right|^2 dx \\ &+ \int_{\mathbb{R}^3} V(\varepsilon_n x + y_n) v_n^2(x) dx \\ &+ \int_{\mathbb{R}^3} \phi_{\varepsilon_n, v_n}(x) v_n^2(x) dx \\ &= \int_{\mathbb{R}^3} f(v_n(x)) v_n(x) dx, \end{aligned}$$

and by the Fatou's Lemma we obtain the absurd

$$\infty = \liminf_{n\to\infty} \int_{\mathbb{R}^3} f(v_n)v_n = \int_{\mathbb{R}^3} f(v)v.$$

Then $\{y_n\}$ has to be bounded and we can assume $y_n \to y \in \mathbb{R}^3$. If $y \notin M$ then $V_0 < V(y)$, and similarly to the computation made in case 1. above (simply replace V_∞ with $V(y)$) we have a contradiction. Hence $y \in M$ and the proof is thereby complete. □

For $\delta > 0$ (later on it will be fixed conveniently) let η be a smooth nonincreasing cut-off function defined in $[0, \infty)$ such that

$$\eta(\xi) = \begin{cases} 1 & \text{if } 0 \le \xi \le \delta/2 \\ 0 & \text{if } \xi \ge \delta. \end{cases}$$

Let $\mathfrak{w}_{V_0}$ be a ground state solution given in Lemma 16.4.10 of problem (A_μ) with $\mu = V_0$ and for any $y \in M$, let us define

$$\Psi_{\varepsilon,y}(x) := \eta(|\varepsilon x - y|)\mathfrak{w}_{V_0}\left(\frac{\varepsilon x - y}{\varepsilon}\right).$$

Let $t_\varepsilon > 0$ verifying $\max_{t\ge 0} I_\varepsilon(t\Psi_{\varepsilon,y}) = I_\varepsilon(t_\varepsilon \Psi_{\varepsilon,y})$, so that $t_\varepsilon \Psi_{\varepsilon,y} \in \mathcal{N}_\varepsilon$, and let

$$\Phi_\varepsilon : y \in M \mapsto t_\varepsilon \Psi_{\varepsilon,y} \in \mathcal{N}_\varepsilon.$$

By construction, $\Phi_\varepsilon(y)$ has compact support for any $y \in M$ and it is easy to see that Φ_ε is a continuous map.

The next result will help us to define a map from M to a suitable sublevel in the Nehari manifold.

Lemma 16.4.12 *The function Φ_ε satisfies*

$$\lim_{\varepsilon\to 0^+} I_\varepsilon(\Phi_\varepsilon(y)) = m^\infty_{V_0}, \text{ uniformly in } y \in M.$$

Proof Suppose by contradiction that the lemma is false. Then there exist $\delta_0 > 0$, $\{y_n\} \subset M$ and $\varepsilon_n \to 0^+$ such that

$$|I_{\varepsilon_n}(\Phi_{\varepsilon_n}(y_n)) - m^\infty_{V_0}| \ge \delta_0. \tag{16.74}$$

Using Lebesgue's Theorem, we have

$$\lim_{n\to\infty} \left\|\Psi_{\varepsilon_n,y_n}\right\|_{\varepsilon_n}^2 = \left\|\mathfrak{w}_{V_0}\right\|_{H^s_{V_0}}^2, \tag{16.75}$$
$$\lim_{n\to\infty} \int_{\mathbb{R}^3} F\left(\Psi_{\varepsilon_n,y_n}\right) = \int_{\mathbb{R}^N} F\left(\mathfrak{w}_{V_0}\right),$$
$$\lim_{n\to\infty} \left\|\Psi_{\varepsilon_n,y_n}\right\|_{H^s_{V_0}}^2 = \left\|\mathfrak{w}_{V_0}\right\|_{H^s_{V_0}}^2.$$

This last convergence implies that $\{\left\|\Psi_{\varepsilon_n,y_n}\right\|\}$ is bounded. From (16.46)

$$\int_{\mathbb{R}^3} \phi_{\varepsilon_n,\Psi_{\varepsilon_n,y_n}} \Psi_{\varepsilon_n,y_n}^2 \le \varepsilon_n^{\alpha-\theta} C_{\mathrm{e}} \left\|\Psi_{\varepsilon_n,y_n}\right\|^4,$$

and then

$$\lim_{n\to\infty} \int_{\mathbb{R}^3} \phi_{\varepsilon_n,\Psi_{\varepsilon_n,y_n}} \Psi_{\varepsilon_n,y_n}^2 = 0. \tag{16.76}$$

Remembering that $t_{\varepsilon_n}\Psi_{\varepsilon_n,y} \in \mathcal{N}_{\varepsilon_n}$ (see few lines before the Lemma), the condition

$$I'_{\varepsilon_n}(t_{\varepsilon_n}\Psi_{\varepsilon_n,y_n})[t_{\varepsilon_n}\Psi_{\varepsilon_n,y_n}] = 0$$

means

$$\left\|\Psi_{\varepsilon_n,y_n}\right\|_{\varepsilon_n}^2 + t_{\varepsilon_n}^2 \int_{\mathbb{R}^3} \phi_{\varepsilon_n,\Psi_{\varepsilon_n,y_n}} \Psi_{\varepsilon_n,y_n}^2 = \int_{\mathbb{R}^3} \frac{f\left(t_{\varepsilon_n}\Psi_{\varepsilon_n,y_n}\right)}{t_{\varepsilon_n}\Psi_{\varepsilon_n,y_n}} \Psi_{\varepsilon_n,y_n}^2. \tag{16.77}$$

We now prove the following

Claim $\lim_{n\to+\infty} t_{\varepsilon_n} = 1$.

We begin by showing the boundedness of $\{t_{\varepsilon_n}\}$. Since $\varepsilon_n \to 0^+$, we can assume $\delta/2 < \delta/(2\varepsilon_n)$ and then from (16.77), using (f4) and making the change of variable $z := (\varepsilon_n x - y_n)/\varepsilon_n$, we get

$$\frac{\left\|\Psi_{\varepsilon_n,y_n}\right\|_{\varepsilon_n}^2}{t_{\varepsilon_n}^2} + \int_{\mathbb{R}^3} \phi_{\varepsilon_n,\Psi_{\varepsilon_n,y_n}} \Psi_{\varepsilon_n,y_n}^2 \ge \frac{f\left(t_{\varepsilon_n}\mathfrak{w}_{V_0}(\overline{z})\right)}{\left(t_{\varepsilon_n}\mathfrak{w}_{V_0}(\overline{z})\right)^3} \int_{B_{\delta/2}} \mathfrak{w}_{V_0}^4(z), \tag{16.78}$$

where $\mathfrak{w}_{V_0}(\overline{z}) := \min_{B_{\delta/2}} \mathfrak{w}_{V_0}(z)$. If $\{t_{\varepsilon_n}\}$ were unbounded, passing to the limit in n in (16.78), the left-hand side would tend to 0 (due to (16.75) and (16.76)), the right-hand side to $+\infty$ (due to (f3)). So we can assume that $t_{\varepsilon_n} \to t_0 \ge 0$.

For given $\xi > 0$, by (16.44), there exists $M_\xi > 0$ such that

$$\int_{\mathbb{R}^3} \frac{f\left(t_{\varepsilon_n}\Psi_{\varepsilon_n,y_n}\right)}{t_{\varepsilon_n}\Psi_{\varepsilon_n,y_n}}\Psi^2_{\varepsilon_n,y_n} \leq \xi \int_{\mathbb{R}^3}\Psi^2_{\varepsilon_n,y_n} + M_\xi t^{q-1}_{\varepsilon_n}\int_{\mathbb{R}^3}\Psi^{q+1}_{\varepsilon_n,y_n}. \tag{16.79}$$

Since $\{\Psi_{\varepsilon_n,y_n}\}$ is bounded in $H^s(\mathbb{R}^3)$, if $t_0 = 0$, from (16.79) we deduce

$$\lim_{n\to\infty}\int_{\mathbb{R}^3} \frac{f\left(t_{\varepsilon_n}\Psi_{\varepsilon_n,y_n}\right)}{t_{\varepsilon_n}\Psi_{\varepsilon_n,y_n}}\Psi^2_{\varepsilon_n,y_n} = 0,$$

which joint with (16.76) and (16.77) led to $\lim_{n\to\infty}\left|\left|\Psi_{\varepsilon_n,y_n}\right|\right|^2_{\varepsilon_n} = 0$ contradicting (16.75). Then $t_{\varepsilon_n} \to t_0 > 0$. Now taking the limit in n in (16.77) we arrive at

$$\left|\left|\mathfrak{w}_{V_0}\right|\right|^2_{H^s_{V_0}} = \int_{\mathbb{R}^3}\frac{f(t_0\mathfrak{w}_{V_0})}{t_0}\mathfrak{w}_{V_0},$$

and since $\mathfrak{w}_{V_0} \in \mathcal{M}_{V_0}$, it has to be $t_0 = 1$, which proves the claim.

Finally, note that

$$\begin{aligned} I_{\varepsilon_n}(\Phi_{\varepsilon_n}(y_n)) = \frac{t^2_{\varepsilon_n}}{2}\int_{\mathbb{R}^3}\left|(-\Delta)^{s/2}\Psi_{\varepsilon_n,y_n}\right|^2 + \frac{t^2_{\varepsilon_n}}{2}\int_{\mathbb{R}^3}V(\varepsilon_n x)\Psi^2_{\varepsilon_n,y_n} \\ + \frac{t^4_{\varepsilon_n}}{4}\int_{\mathbb{R}^3}\phi_{\varepsilon_n,\Psi_{\varepsilon_n,y_n}}\Psi^2_{\varepsilon_n,y_n} - \int_{\mathbb{R}^3}F\left(t_{\varepsilon_n}\Psi_{\varepsilon_n,y_n}\right). \end{aligned}$$

and then (by using the claim) $\lim_{n\to\infty} I_{\varepsilon_n}(\Phi_{\varepsilon_n}(y_n)) = E_{V_0}(\mathfrak{w}_{V_0}) = m^\infty_{V_0}$, which contradicts (16.74). Thus the Lemma holds. □

The remaining part of the paper mainly follows the arguments of [39].

By Lemma 16.4.12, $h(\varepsilon) := |I_\varepsilon(\Phi_\varepsilon(y)) - m^\infty_{V_0}| = o(1)$ for $\varepsilon \to 0^+$ uniformly in y, and then $I_\varepsilon(\Phi_\varepsilon(y)) - m^\infty_{V_0} \leq h(\varepsilon)$. In particular the sublevel set in the Nehari manifold

$$\mathcal{N}_\varepsilon^{m^\infty_{V_0}+h(\varepsilon)} := \left\{u \in \mathcal{N}_\varepsilon : I_\varepsilon(u) \leq m^\infty_{V_0} + h(\varepsilon)\right\}$$

is not empty, since for sufficiently small ε,

$$\forall\, y \in M : \Phi_\varepsilon(y) \in \mathcal{N}_\varepsilon^{m^\infty_{V_0}+h(\varepsilon)}. \tag{16.80}$$

From now on we fix a $\delta > 0$ in such a way that M and

$$M_{2\delta} := \left\{x \in \mathbb{R}^3 : d(x, M) \leq 2\delta\right\}$$

are homotopically equivalent (d denotes the euclidean distance). Take a $\rho = \rho(\delta) > 0$ such that $M_{2\delta} \subset B_\rho$ and $\chi : \mathbb{R}^3 \to \mathbb{R}^3$ be defined as follows:

$$\chi(x) = \begin{cases} x & \text{if } |x| \le \rho \\ \rho \dfrac{x}{|x|} & \text{if } |x| \ge \rho. \end{cases}$$

Define the *barycenter map* β_ε

$$\beta_\varepsilon(u) := \frac{\displaystyle\int_{\mathbb{R}^3} \chi(\varepsilon x) u^2(x)}{\displaystyle\int_{\mathbb{R}^3} u^2(x)} \in \mathbb{R}^3$$

for all $u \in W_\varepsilon$ with compact support.

We will take advantage of the following results from Sect. 16.3.3. They are rewritten for the reader convenience.

Lemma 16.4.13 *The function β_ε satisfies*

$$\lim_{\varepsilon \to 0^+} \beta_\varepsilon(\Phi_\varepsilon(y)) = y, \quad \textit{uniformly in } y \in M.$$

Lemma 16.4.14 *We have*

$$\lim_{\varepsilon \to 0^+} \sup_{u \in \mathcal{N}_\varepsilon^{m_{V_0}^\infty + h(\varepsilon)}} \inf_{y \in M_\delta} \Big| \beta_\varepsilon(u) - y \Big| = 0.$$

16.4.4 Proof of Theorem 16.4.2

In virtue of Lemma 16.4.14, there exists $\varepsilon^* > 0$ such that

$$\forall \varepsilon \in (0, \varepsilon^*] : \quad \sup_{u \in \mathcal{N}_\varepsilon^{m_{V_0}^\infty + h(\varepsilon)}} d(\beta_\varepsilon(u), M_\delta) < \delta/2.$$

Define now

$$M^+ := \Big\{ x \in \mathbb{R}^3 : d(x, M) \le 3\delta/2 \Big\}$$

so that M and M^+ are homotopically equivalent.

Now, reducing $\varepsilon^* > 0$ if necessary, we can assume that Lemmas 16.4.13, 16.4.14 and (16.80) hold. Then by standard arguments the composed map

$$M \xrightarrow{\Phi_\varepsilon} \mathcal{N}_\varepsilon^{m_{V_0}^\infty + h(\varepsilon)} \xrightarrow{\beta_\varepsilon} M^+ \quad \text{is homotopic to the inclusion map.}$$

In case $V_\infty < \infty$, we eventually reduce ε^* in such a way that also the Palais-Smale condition is satisfied in the interval $(m_{V_0}^\infty, m_{V_0}^\infty + h(\varepsilon))$, see Proposition 16.4.9. By well-known properties of the category, it is

$$\operatorname{cat}(\mathcal{N}_\varepsilon^{m_{V_0}^\infty + h(\varepsilon)}) \geq \operatorname{cat}_{M^+}(M)$$

and the Ljusternik-Schnirelmann theory ensures the existence of at least $\operatorname{cat}_{M^+}(M) = \operatorname{cat} M$ constraint critical points of I_ε on $\mathcal{N}_\varepsilon$. The proof of the main Theorem 16.4.2 then follows by Proposition 16.4.9.

If M is bounded and not contractible in itself, then the existence of another critical point of I_ε on $\mathcal{N}_\varepsilon$ follows from some ideas in [21]. We recall here the main steps for completeness.

The goal is to exhibit a subset $\mathcal{A} \subset \mathcal{N}_\varepsilon$ such that

i) $\mathcal{A}$ is not contractible in $\mathcal{N}_\varepsilon^{m_{V_0}^\infty + h(\varepsilon)}$,
ii) $\mathcal{A}$ is contractible in $\mathcal{N}_\varepsilon^{\bar{c}} = \{u \in \mathcal{N}_\varepsilon : I_\varepsilon(u) \leq \bar{c}\}$, for some $\bar{c} > m_{V_0}^\infty + h(\varepsilon)$.

This would imply, since the Palais-Smale holds, that there is a critical level between $m_{V_0}^\infty + h(\varepsilon)$ and $\bar{c}$.

First note that when M is not contractible and bounded the compact set $\mathcal{A} := \Phi_\varepsilon(M)$ cannot be contractible in $\mathcal{N}_\varepsilon^{m_{V_0}^\infty + h(\varepsilon)}$, proving i).

Let us denote, for $u \in W_\varepsilon \setminus \{0\}$, with $t_\varepsilon(u) > 0$ the unique positive number such that $t_\varepsilon(u)u \in \mathcal{N}_\varepsilon$. Choose a function $u^* \in W_\varepsilon$ be such that $u^* \geq 0$, $I_\varepsilon(t_\varepsilon(u^*)u^*) > m_{V_0}^\infty + h(\varepsilon)$ and consider the compact and contractible cone

$$\mathfrak{C} := \Big\{ tu^* + (1-t)u : t \in [0,1], u \in \mathcal{A} \Big\}.$$

Observe that, since the functions in $\mathfrak{C}$ have to be positive on a set of nonzero measure, it is $0 \notin \mathfrak{C}$. Now we project this cone on $\mathcal{N}_\varepsilon$: let

$$t_\varepsilon(\mathfrak{C}) := \Big\{ t_\varepsilon(w)w : w \in \mathfrak{C} \Big\} \subset \mathcal{N}_\varepsilon$$

and set

$$\bar{c} := \max_{t_\varepsilon(\mathfrak{C})} I_\varepsilon > m_{V_0}^\infty + h(\varepsilon)$$

(indeed the maximum is achieved being $t_\varepsilon(\mathfrak{C})$ compact). Of course $\mathcal{A} \subset t_\varepsilon(\mathfrak{C}) \subset \mathcal{N}_\varepsilon$ and $t_\varepsilon(\mathfrak{C})$ is contractible in $\mathcal{N}_\varepsilon^{\bar{c}}$: we deduce ii).

Then there is a critical level for I_ε greater than $m_{V_0}^\infty + h(\varepsilon)$, hence different from the previous ones we have found. The proof of Theorem 16.4.2 is complete.

Acknowledgements G. M. Figueiredo is partially supported by Capes, Fapdf, and CNPq, Brazil. E. G. Murcia is partially supported by Universidad Javeriana, G. Siciliano is partially supported by Capes, Fapesp, and CNPq, Brazil.

References

1. G. Alberti, G. Bouchitté, P. Seppecher, Phase transition with the line-tension effect. Arch. Ration. Mech. Anal. **144**, 1–46 (1998)
2. C.O. Alves, Existence and multiplicity of solution for a class of quasilinear equations. Adv. Nonlinear Stud. **5**, 73–86 (2005)
3. C.O. Alves, G.M. Figueiredo, Existence and multiplicity of positive solutions to a p-Laplacian equation in $\mathbb{R}^N$. Differ. Int. Equ. **19**, 143–162 (2006)
4. C.O. Alves, G.M. Figueiredo, Multiplicity of nontrivial solutions to a biharmonic equation via Lusternik-Schnirelman theory. Math. Methods Appl. Sci. **36**, 683–694 (2013)
5. C.O. Alves, G.M. Figueiredo, Multiple solutions for a semilinear elliptic equation with critical growth and magnetic field. Milan J. Math. **82**, 389–405 (2014)
6. C.O. Alves, G.M. Figueiredo, Multiplicity of solutions for a NLS equations with magnetic fields in $\mathbb{R}^N$. Monatsh. Math. **175**, 1–23 (2014)
7. C.O. Alves, G.M. Figueiredo, Multiple solutions for a quasilinear Schrödinger equation on $\mathbb{R}^N$. Acta Appl. Math. **136**, 91–117 (2015)
8. C.O. Alves, O.H. Miyagaki, Existence and concentration of solution for a class of fractional elliptic equation in $\mathbb{R}^N$ via penalization method Calc. Var. Partial Differ. Equ. **55**, Art. 47, 19 pp. (2016)
9. C.O. Alves, S.H.M. Soares, Multiplicity of positive solutions for a class of nonlinear Schrödinger equations. Adv. Differ. Equ. **15**, 1083–1102 (2010)
10. C.O. Alves, G.M. Figueiredo, G. Siciliano, Ground state solutions for fractional scalar field equations under a general critical nonlinearity. Commun. Pure Appl. Anal. **18**, 2199–2215 (2019)
11. C.O. Alves, P.C. Carrião, E.S. Medeiros, Multiplicity of solutions for a class of quasilinear problem in exterior domains with Newmann conditions. Abstr. Appl. Anal. **03**, 251–268 (2004)
12. C.O. Alves, G.M. Figueiredo, U.B. Severo, Multiplicity of positive solutions for a class of quasilinear problems. Adv. Differ. Equ. **14**, 911–942 (2009)
13. C.O. Alves, G.M. Figueiredo, U.B. Severo, A result of multiplicity of solutions for a class of quasilinear equations. Proc. Edinb. Math. Soc. **55**, 291–309 (2012)
14. A. Ambrosetti, A. Malchiodi, *Nonlinear Analysis and Semilinear Elliptic Problems*. Cambridge Studies in Advanced Mathematics, vol. 104 (Cambridge University Press, Cambridge, 2007)
15. M. Badiale, E. Serra, *Semilinear Elliptic Equations for Beginners. Existence Results via the Variational Approach*, Universitext (Springer, London, 2011)
16. P.W. Bates, On some nonlocal evolution equations arising in materials science, in *Nonlinear Dynamics and Evolution Equations*. Fields Institute Communications, vol. 48, 13–52 (American Mathematical Society, Providence, 2006)
17. V. Benci, G. Cerami, The effect of the domain topology on the number of positive solutions of nonlinear elliptic problems. Arch. Ration. Mech. Anal. **114**, 79–83 (1991)

18. V. Benci, G. Cerami, Multiple positive solutions of some elliptic problems via the Morse theory and the domain topology. Calc. Var. **2**, 29–48 (1994)
19. V. Benci, D. Fortunato, An eigenvalue problem for the Schrödinger-Maxwell equations. Topol. Methods. Nonlinear Anal. **11**, 283–293 (1998)
20. V. Benci, D. Fortunato, Solitons in Schrödinger-Maxwell equations. J. Fixed Point Theory Appl. **15**, 101–132 (2014)
21. V. Benci, G. Cerami, D. Passaseo, *On the number of positive solutions of some nonlinear elliptic problems.* Nonlinear analysis, Sc. Norm. Super. di Pisa Quaderni, Scuola Norm. Sup., Pisa, 93–107 (1991)
22. L. Caffarelli, J.M. Roquejoffre, O. Savin, Nonlocal minimal surfaces. Commun. Pure Appl. Math. **63**, 1111–1144 (2012)
23. L. Caffarelli, L.E. Silvestre, An extension problem related to the fractional Laplacian. Commun. Partial Differ. Equ. **32**, 1245–1260 (2007)
24. L. Caffarelli, E. Valdinoci, Uniform estimates and limiting arguments for nonlocal minimal surfaces. Calc. Var. **32**, 1245–1260 (2007)
25. S.-Y.A. Chang, M. del Mar González, Fractional Laplacian in conformal geometry. Adv. Math. **226**, 1410–1432 (2011)
26. M. Cheng, Bound state for the fractional Schrödingier equation with unbounded potential. J. Math. Phys. **53**, 043507 (2012)
27. S. Cingolani, M. Lazzo, Multiple semiclassical standing waves for a class of nonlinear Schrödinger equations. Topol. Methods Nonlinear Anal. **10**, 1–13 (1997)
28. S. Cingolani, M. Lazzo, Multiple positive solutions to nonlinear Schrödinger equations with competing potential functions. J. Differ. Equ. **160**, 118–138 (2000)
29. R. Cont, P. Tankov, *Financial Modeling with Jump Processes* (Chapman&Hall/CRC Financial Mathematics Series, Boca Raton, 2004)
30. P. d'Avenia, G. Siciliano, M. Squassina, On fractional Choquard equations. Math. Models Methods Appl. Sci. **25**, 1447–1476 (2015)
31. J. Dávila, M. del Pino, J. Wei, Concentrating standing waves for the fractional nonlinear Schrödinger equation. J. Differ. Equ. **256**, 858–892 (2014)
32. J. Dávila, M. del Pino, S. Dipierro, E. Valdinoci, Concentration phenomena for the nonlocal Schrödinger equation with Dirichlet datum. Anal. PDE **8**(5), 1165–1235 (2015)
33. M. del Mar González, R. Monneau, Slow motion of particle system as a limit of a reaction-diffusion equation with half-Laplacian in dimension one. Discret. Contin. Dyn. Syst. **32**, 1255–1286 (2012)
34. E. Di Nezza, G. Palatucci, E. Valdinoci, Hitchhiker's guide to the fractional Sobolev spaces. Bull. Sci. Math. **136**, 512–573 (2012)
35. G. Duvaut, J.L. Lions, *Inequalities in Mechanics and Physics* (Springer, Berlin, 1976)
36. M.M. Fall, F. Mahmoudi, E. Valdinoci, Ground states and concentration phenomena for the fractional Schrödinger equation. Nonlinearity **28**, 1937–1961 (2015)
37. Y. Fang, J. Zhang, Multiplicity of solutions for the nonlinear Schrödinger-Maxwell system. Commun. Pure Appl. Anal. **10**, 1267–1279 (2011)
38. P. Felmer, A. Quaas, J. Tan, Positive solutions of the nonlinear Schrödinger equation with the fractional Laplacian. Proc. Royal Soc. Edimb. **142A**, 1237–1262 (2012)
39. G.M. Figueiredo, G. Siciliano, A multiplicity result via Ljusternick-Schnirelmann category and Morse theory for a fractional Schrödinger equation in $\mathbb{R}^N$. Nonlinear Differ. Equ. Appl. **23**, 12 (2016)
40. G.M. Figueiredo, G. Siciliano, Positive solutions for the fractional Laplacian in the almost critical case in a bounded domain. Nonlinear Anal. Real World Appl. **36**, 89–100 (2017)
41. G.M. Figueiredo, M.T. Pimenta, G. Siciliano, Multiplicity results for the fractional Laplacian in expanded domains. Mediterr. J. Math. **15**, Art. 137, 23pp. (2018)
42. N. Ghoussoub, *Duality and Perturbation Methods in Critical Point Theory* (Cambridge University Press, Cambridge, 1993)
43. I.M. James, On category, in the sense of Lusternik-Schnirelmann. Topology **17**, 331–348 (1978)

44. O. Kavian, *Introduction à la Théorie des Points Critiques* (Springer, Berlin, 1993)
45. N. Laskin, Fractals and quantum mechanics. Chaos **10**, 780–790 (2000)
46. N. Laskin, Fractional quantum mechanics and Lévy path integrals. Phys. Lett. A **268**, 298–305 (2000)
47. N. Laskin, Fractional Schrödinger equation. Phys. Rev. E **66**, 056108 (2002)
48. R. Metzler, J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamic approach. Phys. Rep. **339**, 77 (2000)
49. R. Metzler, J. Klafter, The restaurant at the random walk: recent developments in the description of anomalous transport by fractional dynamics. J. Phys. A **37**, 161–208 (2004)
50. E. Milakis, L. Silvestre, Regularity for the nonlinear Signorini problem. Adv. Math. **217**, 1301–1312 (2008)
51. J. Munkres, *Topology* (Prentice-Hall, Inc., Englewood Cliffs, 2000)
52. E.G. Murcia, Positive solutions for Schrödinger-Poisson type systems, Ph.D. thesis, University of São Paulo, 2016
53. E.G. Murcia, G. Siciliano, Positive semiclassical states for a fractional Schrd̈inger-Poisson system. Differ. Integral Equ. **30**, 231–258 (2017)
54. R.S. Palais, Lusternik-Schnirelman theory on Banach manifolds. Topology **5**, 115–132 (1966)
55. P.H. Rabinowitz, On a class of nonlinear Schrödinger equations. Z. Angew Math. Phys. **43**, 27–42 (1992)
56. H. Royden, *Real Analysis*, 3rd edn. (Macmillan Publishing Company, New York)
57. D. Ruiz, Semiclassical states for coupled Schrödinger-Maxwell equations: concentration around a sphere. Math. Models Methods Appl. Sci. **15**, 141–164 (2005)
58. G. Siciliano, Multiple positive solutions for a Schrödinger-Poisson-Slater system. J. Math. Anal. Appl. **365**, 288–299 (2010)
59. L. Silvestre, Regularity of the obstacle problem for a fractional power of the Laplace operator. Commun. Pure Appl. Math. **13**, 457–468 (1960)
60. Y. Sire, E. Valdinoci, Fractional Laplacian phase transitions and boundary reactions: a geometric inequality and a symmetric result. J. Funct. Anal. **60**, 67–112 (2007)
61. Z. Wei, Existence of infinitely many solutions for the fractional Schrödinger-Maxwell equations. arXiv:1508.03088v1
62. M. Willem, *Minimax Theorems* (Birkhäuser, Basel, 1996)
63. J. Zhang, Existence and multiplicity results for the fractional Schrödinger-Poisson systems. arXiv:1507.01205v1

Chapter 17
Nonlinear Nonhomogeneous Elliptic Problems

Nikolaos S. Papageorgiou, Calogero Vetro, and Francesca Vetro

Abstract We consider nonlinear elliptic equations driven by a nonhomogeneous differential operator plus an indefinite potential. The boundary condition is either Dirichlet or Robin (including as a special case the Neumann problem). First we present the corresponding regularity theory (up to the boundary). Then we develop the nonlinear maximum principle and present some important nonlinear strong comparison principles. Subsequently we see how these results together with variational methods, truncation and perturbation techniques, and Morse theory (critical groups) can be used to analyze different classes of elliptic equations. Special attention is given to $(p, 2)$-equations (these are equations driven by the sum of a p-Laplacian and a Laplacian), where stronger results can be stated.

Keywords Nonlinear regularity · Nonlinear maximum principle · Strong comparison principles · Multiplicity theorems · Nodal solutions · $(p, 2)$-equations

N. S. Papageorgiou (✉)
National Technical University, Department of Mathematics, Athens, Greece
e-mail: npapg@math.ntua.gr

C. Vetro
University of Palermo, Department of Mathematics and Computer Science, Palermo, Italy
e-mail: calogero.vetro@unipa.it

F. Vetro
Nonlinear Analysis Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam
e-mail: francescavetro@tdtu.edu.vn

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_17

17.1 Introduction

In this chapter we study the following nonlinear nonhomogeneous elliptic equation

$$\begin{cases} -\text{div } a(\nabla u(z)) + \xi(z)|u(z)|^{p-2}u(z) = f(z, u(z)) \text{ in } \Omega, \\ u \in \text{ BC}, \quad 1 < p < +\infty. \end{cases}$$

In this problem $\Omega \subseteq \mathbb{R}^N$ is a bounded domain with a C^2-boundary $\partial\Omega$. In the differential operator the map $a\colon \mathbb{R}^N \to \mathbb{R}^N$ is continuous, strictly monotone (hence maximal monotone too) and satisfies certain other regularity and growth conditions listed in hypotheses $H(a)$ (see Sect. 17.2). These conditions are mild and incorporate in our framework many differential operators of interest, such as the p-Laplacian and the (p, q)-Laplacian (that is, the sum of a p-Laplacian and of a q-Laplacian). There is also the potential term $u \to \xi(z)|u|^{p-2}u$ with the potential function $\xi \in L^\infty(\Omega)$ being in general indefinite. The reaction term $f(z, x)$ is in general a Carathéodory function (that is, for all $x \in \mathbb{R}$ $z \to f(z, x)$ is measurable and for a.a. $z \in \Omega$ $x \to f(z, x)$ is continuous). We will consider reactions with different structure. More precisely, we will examine problems with $f(z, \cdot)$ being $(p - 1)$-superlinear but without satisfying the usual in such cases Ambrosetti-Rabinowitz condition (the AR-condition for short). Also, we will consider the case of a reaction which is $(p - 1)$-linear near $\pm\infty$ and of a reaction which exhibits concave nonlinearities near the origin. We will also consider parametric problems and examine the set of positive solutions as the parameter $\lambda > 0$ varies. The particular case of $(p, 2)$-equations (that is, when $a(y) = |y|^{p-2}y + y$ for all $y \in \mathbb{R}^N$ with $2 < p$), will be considered separately because for such equations stronger results can be obtained, combining variational methods and Morse theoretic tools. The notation $u \in$ BC means that we will deal with both the Dirichlet problem, that is, $u\big|_{\partial\Omega} = 0$ or the Robin problem $\dfrac{\partial u}{\partial n_a} + \beta(z)|u|^{p-2}u = 0$ on $\partial\Omega$ with $\beta \geq 0$ on $\partial\Omega$. The case $\beta \equiv 0$ is also included and corresponds to the Neumann problem. By $\dfrac{\partial u}{\partial n_a}$ we denote the conormal derivative defined by extension of the map

$$C^1(\overline{\Omega}) \ni u \to (a(\nabla u), n)_{\mathbb{R}^N},$$

with $n(\cdot)$ being the outward unit normal on $\partial\Omega$.

17.2 Regularity and Auxiliary Results

First let us introduce the hypotheses on the map $a(\cdot)$ involved in the differential operator. These conditions on $a(\cdot)$ are dictated by the nonlinear regularity theory of Lieberman [45] and the nonlinear maximum principle of Pucci and Serrin [72].

So, let $\vartheta \in C^1(0, +\infty)$ and assume that it satisfies the following conditions:

$$0 < \widehat{c} \leq \frac{t\vartheta'(t)}{\vartheta(t)} \leq c_0 \text{ and } c_1 t^{p-1} \leq \vartheta(t) \leq c_2[t^{\tau-1} + t^{p-1}] \text{ for all } t > 0, \tag{17.1}$$

with $0 < c_1, c_2$ and $1 \leq \tau < p$.

$H(a)$ $a(y) = a_0(|y|)y$ for all $y \in \mathbb{R}^N$ with $a_0(t) > 0$ for all $t > 0$ and

(i) $a_0 \in C^1(0, +\infty)$, $t \to a_0(t)t$ is strictly increasing on $(0, +\infty)$, $a_0(t)t \to 0^+$ as $t \to 0^+$ and $\lim\limits_{t\to 0^+} \dfrac{a_0'(t)t}{a_0(t)} > -1$;

(ii) $|\nabla a(y)| \leq c_3 \dfrac{\vartheta(|y|)}{|y|}$ for all $y \in \mathbb{R}^N \setminus \{0\}$, some $c_3 > 0$;

(iii) $(\nabla a(y)\xi, \xi)_{\mathbb{R}^N} \geq \dfrac{\vartheta(|y|)}{|y|}|\xi|^2$ for all $y \in \mathbb{R}^N \setminus \{0\}$, all $\xi \in \mathbb{R}^N$;

(iv) if $G_0(t) = \int_0^t a_0(s)sds$, then $0 \leq pG_0(t) - a_0(t)t^2$ for all $t \geq 0$.

These conditions on $a(\cdot)$ permit the use of the nonlinear regularity theory of Lieberman [45] and of the nonlinear maximum principle of Pucci and Serrin [72].

Some additional conditions will be imposed on $G_0(\cdot)$ according to the needs of our problem. These extra conditions are mild and do not eliminate any of the maps $a(\cdot)$ included in the examples below.

Evidently $G_0(\cdot)$ is strictly increasing and strictly convex. Let $G(y) = G_0(|y|)$ for all $y \in \mathbb{R}^N$. Then

$$\nabla G(y) = G_0'(|y|)\frac{y}{|y|} = a_0(|y|)y = a(y) \text{ for all } y \in \mathbb{R}^N \setminus \{0\},\ \nabla G(0) = 0.$$

Therefore $G(\cdot)$ is the primitive of $a(\cdot)$ and $G(\cdot)$ is convex with $G(0) = 0$. It follows that

$$G(y) \leq (a(y), y)_{\mathbb{R}^N} \text{ for all } y \in \mathbb{R}^N. \tag{17.2}$$

From hypotheses $H(a)$ (i), (ii), (iii) and (17.1), we deduce easily the following properties for the map $a(\cdot)$.

Lemma 17.2.1 *If hypotheses $H(a)$ (i), (ii), (iii) hold, then*

(a) *$y \to a(y)$ is continuous and strictly monotone (hence maximal monotone too);*
(b) *$|a(y)| \leq c_4[1 + |y|^{p-1}]$ for all $y \in \mathbb{R}^N$, some $c_4 > 0$;*
(c) *$(a(y), y)_{\mathbb{R}^N} \geq \dfrac{c_1}{p-1}|y|^p$ for all $y \in \mathbb{R}^N$.*

This lemma and (17.2) lead to the following growth properties for the primitive $G(\cdot)$.

Corollary 17.2.2 *If hypotheses* $H(a)\,(i), (ii), (iii)$ *hold, then* $\frac{c_1}{p(p-1)}|y|^p \leq G(y) \leq c_5[1+|y|^p]$ *for all* $y \in \mathbb{R}^N$*, some* $c_5 > 0$.

Examples The following functions satisfy hypotheses $H(a)$:

(a) $a(y) = |y|^{p-2}y$ with $1 < p < +\infty$.
This map corresponds to the p-Laplace differential operator defined by

$$\Delta_p u = \operatorname{div}\left(|\nabla u|^{p-2}\nabla u\right) \text{ for all } u \in W^{1,p}(\Omega).$$

(b) $a(y) = |y|^{p-2}y + |y|^{q-2}y$ with $1 < q < p < +\infty$.
This map corresponds to the (p, q)-Laplace differential operator defined by

$$\Delta_p u + \Delta_q u \text{ for all } u \in W^{1,p}(\Omega).$$

(c) $a(y) = \left[1 + |y|^2\right]^{\frac{p-2}{2}} y$ with $1 < p < +\infty$.
This map corresponds to the generalized p-mean curvature differential operator defined by

$$\operatorname{div}\left((1 + |\nabla u|^2)^{\frac{p-2}{2}}\nabla u\right) \text{ for all } u \in W^{1,p}(\Omega).$$

(d) $a(y) = |y|^{p-2}y\left[1 + \frac{1}{1+|y|^p}\right]$ with $1 < p < +\infty$.

Let $f_0 \colon \Omega \times \mathbb{R} \to \mathbb{R}$ be a Carathéodory function such that

$$|f_0(z, x)| \leq a_0(z)[1 + |x|^{r-1}] \text{ for a.a. } z \in \Omega, \text{ all } x \in \mathbb{R},$$

with $a_0 \in L^\infty(\Omega)$ and $1 < r \leq p^* = \begin{cases} \frac{Np}{N-p} & \text{if } p < N, \\ +\infty & \text{if } N \leq p \end{cases}$ (the critical Sobolev exponent). We consider the following nonlinear elliptic boundary value problem

$$\begin{cases} -\operatorname{div}\, a(\nabla u(z)) = f_0(z, u(z)) \text{ in } \Omega, \\ u \in \mathrm{BC}. \end{cases} \tag{17.3}$$

In the case of Robin boundary condition, we assume the following for the boundary coefficient $\beta(\cdot)$:

$H(\beta)$ $\beta \in C^{0,\alpha}(\partial\Omega)$ with $\alpha \in (0, 1)$, $\beta(z) \geq 0$ for all $z \in \partial\Omega$.

From Guedda and Véron [38] (Dirichlet case) and Papageorgiou and Rădulescu [59] (Robin case), we have:

Proposition 17.2.3 *If* $u \in W^{1,p}(\Omega)$ *is a weak solution of* (17.3)*, then* $u \in L^\infty(\Omega)$ *and there exists* $M = M(p, N, \Omega, \|u\|_{p^*}) > 0$ *such that* $\|u\|_\infty \le M$.

Using this proposition, we can now apply the nonlinear regularity theory of Lieberman [45] and have:

Proposition 17.2.4 *If* $u \in W^{1,p}(\Omega)$ *is a weak solution of* (17.3)*, then there exists* $\eta = \eta(p, N) > 0$ *and* $\widehat{M} = \widehat{M}(p, N, \Omega, \|u\|_\infty) > 0$ *such that* $u \in C^{1,\eta}(\overline{\Omega})$ *and* $\|u\|_{C^{1,\eta}(\overline{\Omega})} \le \widehat{M}$.

This regularity up to the boundary result is a very powerful tool in our disposal. In particular it leads to the following result relating local minimizers for the energy functional of (17.3). So, let $F_0(z, x) = \int_0^x f_0(z, s)ds$ and consider the following C^1-functionals:

$$\varphi_0(u) = \int_\Omega G(\nabla u)dz - \int_\Omega F_0(z, u)dz \text{ for all } u \in W_0^{1,p}(\Omega) \text{ (Dirichlet problem)},$$

$$\varphi_0(u) = \int_\Omega G(\nabla u)dz + \frac{1}{p}\int_{\partial\Omega} \beta(z)|u|^p d\sigma - \int_\Omega F_0(z, u)dz \text{ for all } u \in W^{1,p}(\Omega)$$

$$\text{(Robin problem)}.$$

Here $\sigma(\cdot)$ denotes the Hausdorff (surface) $(N-1)$-dimensional measure on $\partial\Omega$. Using Proposition 17.2.4 and the Lagrange multiplier rule, we can prove the following result relating local $C_0^1(\overline{\Omega})$ or $C^1(\overline{\Omega})$ and $W_0^{1,p}(\Omega)$ or $W^{1,p}(\Omega)$ minimizers of φ_0. In what follows for notational economy, we set $V = C_0^1(\overline{\Omega})$ or $C^1(\overline{\Omega})$ and $X = W_0^{1,p}(\Omega)$ or $W^{1,p}(\Omega)$.

Proposition 17.2.5 *If* $u_0 \in X$ *is a local* V*-minimizer of* φ_0*, that is, there exists* $\rho_1 > 0$ *such that* $\varphi_0(u_0) \le \varphi_0(u_0 + h)$ *for all* $h \in V$ *with* $\|h\|_V \le \rho_1$*, then* $u_0 \in C^{1,\eta}(\overline{\Omega})$ *for some* $\eta \in (0, 1)$ *and* u_0 *is a local* X*-minimizer of* φ_0*, that is, there exists* $\rho_2 > 0$ *such that* $\varphi_0(u_0) \le \varphi_0(u_0 + h)$ *for all* $h \in X$ *with* $\|h\|_X \le \rho_2$.

Remark 17.2.6 The first such result was proved by Brezis and Nirenberg [15] for the space $H_0^1(\Omega)$ with $G(y) = \frac{1}{2}|y|^2$. It was extended to the space $W_0^{1,p}(\Omega)$ with $G(y) = \frac{1}{p}|y|^p$ by Garcia Azorero et al. [30]. The version for the space $W^{1,p}(\Omega)$ and general $G(\cdot)$ can be found in Papageorgiou and Rădulescu [59].

Consider the nonlinear map $A : X \to X^*$ defined by

$$\langle A(u), h\rangle = \int_\Omega (a(\nabla u), \nabla h)_{\mathbb{R}^N} dz \text{ for all } u, h \in X. \tag{17.4}$$

Recall $X = W_0^{1,p}(\Omega)$ or $W^{1,p}(\Omega)$ and that $W_0^{1,p}(\Omega)^* = W^{-1,p'}(\Omega)$ with $\frac{1}{p} + \frac{1}{p'} = 1$. The following proposition determines the main properties of $A(\cdot)$.

Proposition 17.2.7 *If hypotheses $H(a)$ hold and $A\colon X \to X^*$ is defined by (17.4), then A is bounded (that is maps bounded sets to bounded sets), continuous, monotone (hence maximal monotone too) and of type $(S)_+$, that is, if $u_n \xrightarrow{w} u$ in X and $\limsup_{n\to+\infty}\langle A(u_n), u_n - u\rangle \le 0$, then $u_n \to u$ in X.*

Finally by $\delta_{k,i}$ with $k, i \in \mathbb{N}_0$, we denote the Kronecker symbol defined by

$$\delta_{k,i} = \begin{cases} 1 & \text{if } k = i, \\ 0 & \text{if } k \neq i. \end{cases}$$

17.3 Maximum Principle: Comparison Results

Let $\gamma(t) = a_0(t)t$ for all $t > 0$. Hypothesis $H(a)$ (iii) and (17.1) ensure that

$$\gamma'(t)t = a_0'(t)t^2 + a_0(t)t \ge c_1 t^{p-1}.$$

Integrating by parts, we obtain

$$\int_0^t \gamma'(s)s\,ds = \gamma(t)t - \int_0^t \gamma(s)ds = a_0(t)t^2 - G_0(t) \ge \frac{c_1}{p}t^p \text{ for all } t > 0.$$

We set $H(t) = a_0(t)t^2 - G_0(t)$ and $H_0(t) = \frac{c_1}{p}t^p$ for all $t > 0$. Let $\delta \in (0, 1)$ and $s > 0$. We consider the following two sets:

$$C = \{t \in (0,1)\colon H(t) \ge s\} \text{ and } C_0 = \{t \in (0,1)\colon H_0(t) \ge s\}.$$

Evidently $C \supseteq C_0$ and so $\inf C \le \inf C_0$. Then Proposition 1.55, p. 12, of Gasiński and Papageorgiou [34] implies that $H^{-1}(s) \le H_0^{-1}(s)$. Let $k\colon \mathbb{R}_+ \to \mathbb{R}_+$ a continuous function such that $k(0) = 0$, $k(\cdot)$ is nondecreasing on $[0, \delta)$ and $\int_0^\delta \frac{1}{H_0(k(s))}ds = +\infty$. Then, the nonlinear strong maximum principle of Pucci and Serrin [72] (pp. 111, 120) applies and we can state the following result.

Theorem 17.3.1 *If hypotheses $H(a)$ hold and $u_0 \in C^1(\overline{\Omega})$, $u_0 \ge 0$, $u_0 \ne 0$ satisfies* div $a(\nabla u_0(z)) \le k(u_0(z))$ *in Ω, then $u_0(z) > 0$ for all $z \in \Omega$ and if $u_0(\widehat{z}) = 0$ for some $\widehat{z} \in \partial\Omega$, then $\frac{\partial u_0}{\partial n}(\widehat{z}) < 0$.*

The Banach spaces

$$C_0^1(\overline{\Omega}) = \left\{u \in C^1(\overline{\Omega}) : u\big|_{\partial\Omega} = 0\right\} \text{ and } C^1(\overline{\Omega})$$

are ordered Banach spaces with positive (order) cones given by

$$C_+^0 = \{u \in C_0^1(\overline{\Omega}) : u(z) \geq 0 \text{ for all } z \in \overline{\Omega}\},$$
$$C_+ = \{u \in C^1(\overline{\Omega}) : u(z) \geq 0 \text{ for all } z \in \overline{\Omega}\}.$$

Both cones have nonempty interiors given by

$$\operatorname{int} C_+^0 = \left\{u \in C_+^0 : u(z) > 0 \text{ for all } z \in \Omega,\ \frac{\partial u}{\partial n}\Big|_{\partial\Omega} < 0\right\},$$
$$\operatorname{int} C_+ = \left\{u \in C_+ : u(z) > 0 \text{ for all } z \in \overline{\Omega}\right\}.$$

Also let

$$\operatorname{int} \widehat{C}_+ = \{u \in C_+ : u(z) > 0 \text{ for all } z \in \Omega,\ \frac{\partial u}{\partial n}\Big|_{\partial\Omega \cap u^{-1}(0)} < 0\}.$$

In general for nonlinear differential operators it is very difficult to have strong comparison principles. This makes it significantly more difficult to prove multiplicity theorems for positive solutions of nonlinear boundary value problems. Next we present some such strong comparison results, which are helpful in this respect.

If $h_1, h_2 \in L^\infty(\Omega)$, then we write that $h_1 \prec h_2$ if and only if for every $K \subseteq \Omega$ compact we have

$$0 < c_K \leq h_2(z) - h_1(z) \text{ for a.a. } z \in K.$$

Clearly if $h_1, h_2 \in C(\Omega)$ and $h_1(z) < h_2(z)$ for all $z \in \Omega$, then $h_1 \prec h_2$.

The next proposition was first proved by Arcoya and Ruiz [8] for the p-Laplacian with $\xi \equiv \vartheta \geq 0$. The more general version stated here is due to Papageorgiou and Winkert [66].

Proposition 17.3.2 *If hypotheses* $H(a)$ *hold,* $\xi \in L^\infty(\Omega)$*,* $\xi(z) \geq 0$ *for a.a.* $z \in \Omega$*,* $h_1, h_2 \in L^\infty(\Omega)$ *with* $h_1 \prec h_2$ *and* $u \in C_0^1(\overline{\Omega})$*,* $v \in \operatorname{int} C_+^0$ *satisfy*

$$-\operatorname{div}\ a(\nabla u(z)) + \xi(z)|u(z)|^{p-2}u(z) = h_1(z) \text{ for a.a. } z \in \Omega,$$
$$-\operatorname{div}\ a(\nabla v(z)) + \xi(z)v(z)^{p-1} = h_2(z) \text{ for a.a. } z \in \Omega,$$

then $v - u \in \operatorname{int} C_+^0$.

Remark 17.3.3 If the hypothesis $h_1 \prec h_2$ is replaced by the condition $h_1(z) \leq h_2(z)$ for a.a. $z \in \Omega$ with strict inequality on a set of positive measure, then the result fails (see Cuesta and Takáč [19]). However, see Proposition 17.3.6 below.

There is the "Robin" counterpart of Proposition 17.3.2 due to Fragnelli et al. [28].

Proposition 17.3.4 *If hypotheses $H(a)$ hold, $\xi \in L^\infty(\Omega)$, $\xi(z) \geq 0$ for a.a. $z \in \Omega$, $h_1, h_2 \in L^\infty(\Omega)$ with $h_1 \prec h_2$, $u \in C^1(\overline{\Omega})$, $v \in \text{int } C_+$, $u \leq v$ satisfy*

$$- \text{div } a(\nabla u(z)) + \xi(z)|u(z)|^{p-2}u(z) = h_1(z) \text{ for a.a. } z \in \Omega,$$

$$- \text{div } a(\nabla v(z)) + \xi(z)v(z)^{p-1} = h_2(z) \text{ for a.a. } z \in \Omega,$$

and $\left.\frac{\partial u}{\partial n_a}\right|_{\partial\Omega} < 0$ *or* $\left.\frac{\partial v}{\partial n_a}\right|_{\partial\Omega} < 0$, *then* $v - u \in \text{int } \widehat{C}_+$.

We can drop the hypothesis on the conormal derivatives used in the above result, by strengthening the hypothesis on $h_2 - h_1$.

Proposition 17.3.5 *If hypotheses $H(a)$ hold, $\xi \in L^\infty(\Omega)$, $\xi(z) \geq 0$ for a.a. $z \in \Omega$, $h_1, h_2 \in L^\infty(\Omega)$,*

$$0 < \widetilde{c} \leq h_2(z) - h_1(z) \text{ for a.a. } z \in \Omega \tag{17.5}$$

and $u, v \in C^{1,\alpha}(\overline{\Omega})$ with $\alpha \in (0, 1]$, $u \leq v$ and

$$- \text{div } a(\nabla u(z)) + \xi(z)|u(z)|^{p-2}u(z) = h_1(z) \text{ for a.a. } z \in \Omega,$$

$$- \text{div } a(\nabla v(z)) + \xi(z)|v(z)|^{p-2}v(z) = h_2(z) \text{ for a.a. } z \in \Omega,$$

then $v - u \in \text{int } \widehat{C}_+$.

Proof By hypothesis we have

$$\begin{aligned} -\text{div } (a(\nabla v(z)) - a(\nabla u(z))) = h_2(z) - h_1(z) \\ - \xi(z)\left[|v(z)|^{p-2}v(z) - |u(z)|^{p-2}u(z)\right] \end{aligned} \tag{17.6}$$

for a.a. $z \in \Omega$. Let $a = (a_k)_{k=1}^N$ denote the components of $a(\cdot)$. We consider $y = (y_i)_{i=1}^N \in \mathbb{R}^N$ and $y' = (y'_i)_{i=1}^N \in \mathbb{R}^N$. We have

$$\begin{aligned} a(y) - a(y') &= \int_0^1 \frac{d}{dt} a\left(y' + t(y - y')\right) dt \\ &= \int_0^1 \nabla a\left(y' + t(y - y')\right)(y - y')dt \text{ (by the chain rule),} \end{aligned}$$

$$\Rightarrow \quad a_k(y) - a_k(y') = \sum_{i=1}^{N} \int_0^1 \frac{\partial a_k}{\partial y_i}\left(y' + t(y - y')\right)(y_i - y_i')dt$$
$$\text{for all } k \in \{1, \ldots, N\}. \tag{17.7}$$

Let $\nabla_i = \dfrac{\partial}{\partial z_i}$ and consider the coefficients

$$e_{k,i}(z) = \int_0^1 \frac{\partial a_k}{\partial y_i}(\nabla u(z) + t(\nabla v(z) - \nabla u(z)))dt. \tag{17.8}$$

Using these coefficients, we introduce the following second order differential operator in divergence form

$$L(w) = -\text{div}\left(\sum_{i=1}^{N} e_{k,i}(z)\nabla_i w\right) = -\sum_{k,i=1}^{N} \nabla_k\left(e_{k,i}(z)\nabla_i w\right).$$

We set $\beta = v - u \in C^{1,\alpha}(\overline{\Omega})$. By hypothesis

$$\beta(z) \geq 0 \text{ for all } z \in \overline{\Omega} \text{ and } \beta \not\equiv 0.$$

Let $z_0 \in \Omega$ be such that $\beta(z_0) = 0$, hence $u(z_0) = v(z_0)$.

The function $x \to |x|^{p-2}x$ is a uniformly continuous homeomorphism on $\mathbb{R}$ (in fact the map is Hölder continuous if $1 < p < 2$, locally Lipschitz if $p \geq 2$). Since $\xi \in L^\infty(\Omega)$, we can find $\rho > 0$ small such that

$$h_2(z) - h_1(z) - \xi(z)\left[|v(z)|^{p-2}v(z) - |u(z)|^{p-2}u(z)\right] \geq c_6 > 0 \tag{17.9}$$
$$\text{for a.a. } z \in \overline{B}_\rho(z_0) \text{ (see (17.5))},$$

where $B_\rho(z_0) = \{z \in \Omega\colon |z - z_0| < \rho\}$. On account of Theorem 1.1 of Lucia and Prashanth [47], we have that $e_{k,i} \in W^{1,\infty}(B_\rho(z_0))$ (see (17.8) and choose $\rho > 0$ even smaller if necessary). From (17.6), (17.7), and (17.9), we have

$$L(\beta)(z) \geq c_7 > 0 \text{ for a.a. } z \in \overline{B}_\rho(z_0), \text{ some } c_7 > 0.$$

Then Theorem 4 of Vázquez [76] (alternatively using the Harnack inequality, see Pucci and Serrin [72], p. 163), we have

$$\beta(z) > 0 \text{ for all } z \in B_\rho(z_0),$$

a contradiction since $\beta(z_0) = 0$. Therefore we infer that

$$\beta(z) > 0 \text{ for all } z \in \Omega.$$

Now, let $\Sigma_0 = \{z \in \partial\Omega \colon \beta(z) = 0\}$. If $\Sigma_0 = \emptyset$, then $v - u \in \operatorname{int} \widehat{C}_+$. So, we assume that $\Sigma_0 \neq \emptyset$ and let $z_0 \in \Sigma_0$. Since by hypothesis $\partial\Omega$ is a C^2-manifold, we can find $r > 0$ small and an open r-ball B_r such that

$$B_r \subseteq \Omega, \quad z_0 \in \partial\Omega \cap \partial B_r.$$

On account of (17.5), (17.6), we can choose $r > 0$ small enough so that $L\big|_{B_r}$ is strictly elliptic and $e_{k,i} \in W^{1,\infty}(B_r)$. So, the Hopf boundary point lemma implies

$$\frac{\partial\beta}{\partial n}(z_0) < 0,$$
$$\Rightarrow \quad \beta = v - u \in \operatorname{int} C_+.$$

□

When $\xi \equiv 0$, we can extend Propositions 17.3.2, 17.3.4, and 17.3.5 as follows. The result can be found in Gasiński and Papageorgiou [36].

Proposition 17.3.6 *If hypotheses $H(a)$ hold, $h_1, h_2 \in L^\infty(\Omega)$, $h_1(z) \le h_2(z)$ for a.a. $z \in \Omega$ and the inequality is strict on a set of positive measure, $u, v \in C^1(\overline{\Omega})$ satisfy $u \le v$ and*

$$-\operatorname{div}\, a(\nabla u(z)) = h_1(z) \textit{ for a.a. } z \in \Omega,$$
$$-\operatorname{div}\, a(\nabla v(z)) = h_2(z) \textit{ for a.a. } z \in \Omega,$$
$$\frac{\partial u}{\partial n}\Big|_{\partial\Omega} < 0, \quad \frac{\partial v}{\partial n}\Big|_{\partial\Omega} < 0,$$

then $v - u \in \operatorname{int} \widehat{C}_+$.

Continuing with the case $\xi \equiv 0$, we have the following result which does not require any conditions on u, v on $\partial\Omega$.

Proposition 17.3.7 *If hypotheses $H(a)$ hold, $h_1, h_2 \in L^\infty(\Omega)$, $h_1(z) \le h_2(z)$ for a.a. $z \in \Omega$, $u, v \in C^1(\overline{\Omega})$ satisfy $u \le v$ and*

$$-\operatorname{div}\, a(\nabla u(z)) = h_1(z) \textit{ for a.a. } z \in \Omega,$$
$$-\operatorname{div}\, a(\nabla v(z)) = h_2(z) \textit{ for a.a. } z \in \Omega,$$

and the set $E = \{z \in \Omega \colon u(z) = v(z)\}$ is either compact or discrete, then $0 < (v - u)(z)$ for all $z \in \Omega$.

Proof Suppose that $E \neq \emptyset$. If E is compact, then we can find U open set such that $E \subseteq U \subseteq \overline{U} \subseteq \Omega$. If E is discrete, then for each $z \in E$ we can find $U \subseteq \Omega$ open such that $z \in U \subseteq \overline{U} \subseteq \Omega$ and $E \cap \overline{U} = \{z\}$. Therefore in both cases we have $E \cap \partial U = \emptyset$. Then we have $(v-u)\big|_{\partial U} > 0$ and we can find $\varepsilon > 0$ such that $(v-u)\big|_{\partial U} \geq \varepsilon > 0$ (recall that $\partial U \subseteq \mathbb{R}^N$ is compact). We set $v^\varepsilon = v - \varepsilon \in C^1(\Omega)$ and have

$$- \operatorname{div} a(\nabla u) = h_1 \leq h_2 = -\operatorname{div} a(\nabla v) = -\operatorname{div} a(\nabla v^\varepsilon). \tag{17.10}$$

On (17.10) we act with $(u - v^\varepsilon)^+ \in W_0^{1,p}(U)$. Using the nonlinear Green's identity (see Gasiński and Papageorgiou [31], p. 211), we have

$$\begin{aligned} &\int_{\{u > v^\varepsilon\} \cap U} (a(\nabla u) - a(\nabla v), \nabla u - \nabla v)_{\mathbb{R}^N} dz \leq 0, \\ \Rightarrow \quad & u \leq v^\varepsilon \text{ on } U, \end{aligned}$$

a contradiction since $E \subseteq U$. Therefore $E = \emptyset$ and we have $u(z) < v(z)$ for all $z \in \Omega$. □

17.4 Eigenvalue Problems

We start with a quick review of the spectral properties of the p-Laplacian plus an indefinite potential under Dirichlet and Robin boundary conditions. So, we consider the following nonlinear eigenvalue problems

$$\begin{cases} -\Delta_p u(z) + \xi(z)|u(z)|^{p-2}u(z) = \widehat{\lambda}|u(z)|^{p-2}u(z) \text{ in } \Omega, \\ u\big|_{\partial\Omega} = 0, \quad 1 < p < +\infty, \end{cases} \tag{17.11}$$

and

$$\begin{cases} -\Delta_p u(z) + \xi(z)|u(z)|^{p-2}u(z) = \widehat{\lambda}|u(z)|^{p-2}u(z) \text{ in } \Omega, \\ \dfrac{\partial u}{\partial n_p} + \beta(z)|u|^{p-2}u = 0 \text{ on } \partial\Omega, \quad 1 < p < +\infty. \end{cases} \tag{17.12}$$

Here $\dfrac{\partial u}{\partial n_p} = |\nabla u|^{p-2}(\nabla u, n)_{\mathbb{R}^N}$ with $n(\cdot)$ being the outward unit normal on $\partial\Omega$. If $\beta \equiv 0$, then the boundary condition in (17.12) becomes $\dfrac{\partial u}{\partial n} = 0$ on $\partial\Omega$ (Neumann problem). The boundary coefficient $\beta(\cdot)$ satisfies hypothesis $H(\beta)$, while for the potential function $\xi(\cdot)$ we assume the following:

$H(\xi)$ $\xi \in L^\infty(\Omega)$.

We say that $\widehat{\lambda} \in \mathbb{R}$ is an "eigenvalue" of (17.11) or (17.12), if the problem admits a nontrivial solution $\widehat{u} \in X$ (recall $X = W_0^{1,p}(\Omega)$ for (17.11) and $X = W^{1,p}(\Omega)$ for (17.12)) known as an "eigenfunction" corresponding to $\widehat{\lambda}$. On account of Propositions 17.2.3 and 17.2.4, we have $\widehat{u} \in C_0^1(\overline{\Omega})$ or $\widehat{u} \in C^1(\overline{\Omega})$. We know that both eigenvalue problems have a smallest eigenvalue $\widehat{\lambda}_1 = \widehat{\lambda}_1(p, \xi)$ for (17.11) and $\widehat{\lambda}_1 = \widehat{\lambda}_1(p, \xi, \beta)$ for (17.12). In what follows let $\mu_0 \colon W_0^{1,p}(\Omega) \to \mathbb{R}$ and $\mu \colon W^{1,p}(\Omega) \to \mathbb{R}$ be the C^1-functionals defined by

$$\mu_0(u) = \|\nabla u\|_p^p + \int_\Omega \xi(z)|u|^p dz \text{ for all } u \in W_0^{1,p}(\Omega),$$

$$\mu(u) = \|\nabla u\|_p^p + \int_\Omega \xi(z)|u|^p dz + \int_{\partial\Omega} \beta(z)|u|^p d\sigma \text{ for all } u \in W^{1,p}(\Omega).$$

The first eigenvalue exhibits some important properties, which we list below:

- $\widehat{\lambda}_1$ is isolated (that is, if $\widehat{\sigma}(p)$ denotes the spectrum of (17.11) or (17.12), then we can find $\varepsilon > 0$ such that $(\widehat{\lambda}_1, \widehat{\lambda}_1 + \varepsilon) \cap \widehat{\sigma}(p) = \emptyset$).
- $\widehat{\lambda}_1$ is simple (that is, if $\widehat{u}, \widehat{v}$ are eigenfunctions corresponding to $\widehat{\lambda}_1$, then $\widehat{u} = \eta\widehat{v}$ for some $\eta \in \mathbb{R} \setminus \{0\}$).
- $\widehat{\lambda}_1$ admits the following variational characterization

$$\widehat{\lambda}_1 = \inf\left[\frac{\mu_0(u)}{\|u\|_p^p} \colon u \in W_0^{1,p}(\Omega),\ u \neq 0\right] \text{ (for (17.11))}, \tag{17.13}$$

$$\widehat{\lambda}_1 = \inf\left[\frac{\mu(u)}{\|u\|_p^p} \colon u \in W^{1,p}(\Omega),\ u \neq 0\right] \text{ (for (17.12))}. \tag{17.14}$$

In both (17.13) and (17.14) the infimum is realized on the corresponding one-dimensional eigenspace. From the above properties, we infer easily that the elements of this eigenspace do not change sign. By $\widehat{u}_1 = \widehat{u}_1(p, \xi)$ (for (17.11)) and $\widehat{u}_1 = \widehat{u}_1(p, \xi, \beta)$ (for (17.12)), we denote the positive, L^p-normalized (that is, $\|\widehat{u}_1\|_p = 1$) eigenfunction corresponding to $\widehat{\lambda}_1$. Then Theorem 17.3.1 implies that

$$\widehat{u}_1 \in \text{int } C_+^0 \text{ (for (17.11)) and } \widehat{u}_1 \in \text{int } C_+ \text{ (for (17.12))}.$$

The second eigenvalue of (17.11) or (17.12) is then defined by

$$\widehat{\lambda}_2 = \min\{\widehat{\lambda} \in \widehat{\sigma}(p) \colon \widehat{\lambda} > \widehat{\lambda}_1\}.$$

In general eigenvalues of (17.11) (resp. (17.12)) are the critical values of the C^1-functional $\mu_0(\cdot)$ (resp. $\mu(\cdot)$) on $M = X \cap \partial B_1^{L^p}$ where $\partial B_1^{L^p} = \{u \in L^p(\Omega) \colon \|u\|_p = 1\}$. Using the Ljusternik-Schnirelmann minimax scheme (see Gasiński and Papageorgiou [31], Section 5.5), we can have a whole increasing sequence of distinct eigenvalues $\{\widehat{\lambda}_k\}_{k \geq 1}$ such that $\widehat{\lambda}_k \to +\infty$. These eigenvalues are known as "variational eigenvalues," but in general we do not know if they exhaust $\widehat{\sigma}(p)$. Depending on the index used in the Ljusternik-Schnirelmann scheme, we can generate three such sequences of variational eigenvalues. Let $\mathcal{S}$ be the closed, symmetric subsets of M. For all $A \in \mathcal{S}$ we define

$$\gamma_0^*(A) = \sup\{k \in \mathbb{N}\colon \text{ there exists a continuous, odd map } h\colon A \to \mathbb{R}^k \setminus \{0\}\},$$

$$\gamma_1^*(A) = \sup\{k \in \mathbb{N}\colon \text{ there exists a continuous, odd map } h\colon \mathbb{R}^k \setminus \{0\} \to A\}.$$

In the above definitions we can replace $\mathbb{R}^k \setminus \{0\}$ by S^{k-1} = the unit sphere of $\mathbb{R}^k$. In the literature $\gamma_0^*(\cdot)$ is known as the "Krasnosel'skii genus." Also we set

$$\gamma^*(A) = \operatorname{ind}(A)$$

where ind $(\cdot)$ denotes the Fadell and Rabinowitz [25] cohomological index. We have

$$\gamma_1^*(A) \le \gamma^*(A) \le \gamma_0^*(A) \text{ for all } A \in \mathcal{S}. \tag{17.15}$$

We define

$$\widehat{\lambda}_k^i = \inf_{\substack{A \in \mathcal{S} \\ \gamma_i^*(A) \ge k}} \sup_{u \in A} \mu_0(u) \text{ (or } \mu(u)\text{) for } i = 0, 1, \tag{17.16}$$

$$\widehat{\lambda}_k = \inf_{\substack{A \in \mathcal{S} \\ \gamma^*(A) \ge k}} \sup_{u \in A} \mu_0(u) \text{ (or } \mu(u)\text{) for all } k \in \mathbb{N}. \tag{17.17}$$

From (17.15) it is clear that

$$\widehat{\lambda}_k^0 \le \widehat{\lambda}_k \le \widehat{\lambda}_k^1 \text{ for all } k \in \mathbb{N}.$$

For $k = 1, 2$, we have $\widehat{\lambda}_k^0 = \widehat{\lambda}_k = \widehat{\lambda}_k^1$ and the common values are defined as above. For $k \ge 3$ we do not know if the three sequences of variational eigenvalues coincide. Depending on the problem, we use one of the above sequences.

Evidently (17.16) (= (17.17)) provide a variational (minimax) characterization of $\widehat{\lambda}_2$. However, for the study of nonlinear elliptic problems, another minimax characterization is more convenient. This characterization was proved by Cuesta-de Figueiredo and Gossez [20] (Dirichlet case), Mugnai and Papageorgiou [53] (Neumann case), and Papageorgiou and Rădulescu [57] (Robin case).

Proposition 17.4.1 $\widehat{\lambda}_2 = \inf_{\widehat{\gamma} \in \widehat{\Gamma}} \max_{-1 \le t \le 1} \mu_0(\widehat{\gamma}(t))$ *(or* $\mu(\widehat{\gamma}(t))$*) where*

$$\widehat{\Gamma} = \{\widehat{\gamma} \in C([-1, 1], M)\colon \widehat{\gamma}(-1) = -\widehat{u}_1,\ \widehat{\gamma}(1) = \widehat{u}_1\}.$$

Remark 17.4.2 In fact in Proposition 17.4.1 we can replace $\widehat{\Gamma}$ by

$$\widehat{\Gamma}_0 = \{\widehat{\gamma} \in C([-1,1], M) \colon \widehat{\gamma}(-1) \leq 0 \leq \widehat{\gamma}(1)\}\,.$$

Let $V = \{u \in X\colon \int_\Omega \widehat{u}_1 u dz = 0\}$ and set

$$\widehat{\lambda}_V^0 = \inf\left[\mu_0(u)\colon u \in M \cap V\right],$$
$$\widehat{\lambda}_V = \inf\left[\mu(u)\colon u \in M \cap V\right].$$

Proposition 17.4.3 $\widehat{\lambda}_1 < \widehat{\lambda}_V^0 \leq \widehat{\lambda}_2$ *and* $\widehat{\lambda}_1 < \widehat{\lambda}_V \leq \widehat{\lambda}_2$.

In fact, let $\mathcal{D}$ be the set of all topological complements of $\mathbb{R}$ in X and set

$$\widehat{\lambda}_{\mathcal{D}}^0 = \inf_{v \in \mathcal{D}} \sup_{v \in V} \mu_0(v),$$
$$\widehat{\lambda}_{\mathcal{D}} = \inf_{v \in \mathcal{D}} \sup_{v \in V} \mu(v).$$

It is an open question whether $\widehat{\lambda}_{\mathcal{D}} = \widehat{\lambda}_2$. This is the case for the linear eigenvalue problem (that is, $p = 2$). In this case the spectral theorem for compact, self-adjoint operators on a Hilbert space gives us a full description of the spectrum $\widehat{\sigma}(2)$ which consists of a sequence $\{\widehat{\lambda}_k\}_{k\geq 1}$ of distinct eigenvalues such that $\widehat{\lambda}_k \to +\infty$. Again we have that $\widehat{\lambda}_1$ is simple and is described by (17.13) and (17.14) with $p = 2$. In the nonlinear case ($p \neq 2$), the eigenspaces are not linear subspaces (they are only cones) and we do not have a direct sum decomposition of X in terms of them. This is the main source of difficulties when we deal with resonant nonlinear problems. For the linear eigenvalue problem, we do not have this difficulty. The eigenspaces are linear subspaces of X (now $X = H_0^1(\Omega)$ or $X = H^1(\Omega)$). Let $E(\widehat{\lambda}_k)$ be the eigenspace corresponding to the eigenvalue $\widehat{\lambda}_k$. Then $E(\widehat{\lambda}_k)$ is a finite dimensional subspace of X and $E(\widehat{\lambda}_k) \subseteq C_0^1(\overline{\Omega})$ or $C^1(\overline{\Omega})$ (standard regularity theory). Also, all eigenvalues admit variational characterizations. Now $\mu_0\colon H_0^1(\Omega) \to \mathbb{R}$ and $\mu\colon H^1(\Omega) \to \mathbb{R}$ are the C^1-functionals defined by

$$\mu_0(u) = \|\nabla u\|_2^2 + \int_\Omega \xi(z) u^2 dz \text{ for all } u \in H_0^1(\Omega),$$
$$\mu(u) = \|\nabla u\|_2^2 + \int_\Omega \xi(z) u^2 dz + \int_{\partial\Omega} \beta(z) u^2 d\sigma \text{ for all } u \in H^1(\Omega).$$

We have

$$\widehat{\lambda}_1 = \inf\left[\frac{\mu_0(u)}{\|u\|_2^2} : u \in H_0^1(\Omega),\ u \neq 0\right], \tag{17.18}$$

$$\widehat{\lambda}_1 = \inf\left[\frac{\mu(u)}{\|u\|_2^2} : u \in H^1(\Omega),\ u \neq 0\right], \tag{17.19}$$

$$\begin{aligned}\widehat{\lambda}_m &= \inf\left[\frac{\mu_0(u)}{\|u\|_2^2} : u \in \widehat{H}_m = \overline{\oplus_{k\geq m} E(\widehat{\lambda}_k)},\ u \neq 0\right] \\ &= \sup\left[\frac{\mu_0(u)}{\|u\|_2^2} : u \in \overline{H}_m = \oplus_{k=1}^m E(\widehat{\lambda}_k),\ u \neq 0\right],\end{aligned} \tag{17.20}$$

$$\begin{aligned}\widehat{\lambda}_m &= \inf\left[\frac{\mu(u)}{\|u\|_2^2} : u \in \widehat{H}_m = \overline{\oplus_{k\geq m} E(\widehat{\lambda}_k)},\ u \neq 0\right] \\ &= \sup\left[\frac{\mu(u)}{\|u\|_2^2} : u \in \overline{H}_m = \oplus_{k=1}^m E(\widehat{\lambda}_k),\ u \neq 0\right] \text{ for } m \geq 2.\end{aligned} \tag{17.21}$$

As before in (17.18) and (17.19) the infimum is realized on $E(\widehat{\lambda}_1)$, the elements of which do not change sign. Also, in (17.20) and in (17.21) both the infimum and the supremum are realized on $E(\widehat{\lambda}_m) \subseteq C^1(\overline{\Omega})$. The eigenspaces $E(\widehat{\lambda}_k)$, $k \in \mathbb{N}$, have the so-called Unique Continuation Property (UCP for short), that is, if $u \in E(\widehat{\lambda}_k)$ and u vanishes on a set of positive measure, then $u \equiv 0$.

For both the linear and nonlinear eigenvalue problems, every eigenfunction corresponding to an eigenvalue $\widehat{\lambda} \neq \widehat{\lambda}_1$ is nodal (that is, sign changing). Also we can have weighted versions of the eigenvalue problems (17.11) and (17.12). So, let $m \in L^\infty(\Omega)$, $m(z) \geq 0$ for a.a. $z \in \Omega$, $m \not\equiv 0$. We consider the following nonlinear eigenvalue problems:

$$\begin{cases} -\Delta_p u(z) + \xi(z)|u(z)|^{p-2}u(z) = \widetilde{\lambda}|u(z)|^{p-2}u(z) \text{ in } \Omega, \\ u\big|_{\partial\Omega} = 0, \quad 1 < p < +\infty, \end{cases} \tag{17.22}$$

$$\begin{cases} -\Delta_p u(z) + \xi(z)|u(z)|^{p-2}u(z) = \widetilde{\lambda}|u(z)|^{p-2}u(z) \text{ in } \Omega, \\ \dfrac{\partial u}{\partial n_p} + \beta(z)|u|^{p-2}u = 0 \text{ on } \partial\Omega, \quad 1 < p < +\infty. \end{cases} \tag{17.23}$$

The previous analysis remains valid also for these weighted eigenvalue problems. So, we have a smallest eigenvalue $\widetilde{\lambda}_1(m) = \widetilde{\lambda}_1(p, \xi, m)$ (for (17.22)) and $\widetilde{\lambda}_1(m) = \widetilde{\lambda}_1(p, \xi, \beta, m)$ (for (17.23)) which is isolated, simple and admits the following variational characterization:

$$\widetilde{\lambda}_1(m) = \inf\left[\frac{\mu_0(u)}{\int_\Omega m(z)|u|^p dz} : u \in W_0^{1,p}(\Omega),\ u \neq 0\right] \text{(for (17.22))}, \tag{17.24}$$

$$\widetilde{\lambda}_1(m) = \inf\left[\frac{\mu(u)}{\int_\Omega m(z)|u|^p dz} : u \in W^{1,p}(\Omega),\ u \neq 0\right] \text{(for (17.23))}. \tag{17.25}$$

In both (17.24) and (17.25) the infimum is realized on the corresponding one-dimensional eigenspace, the elements of which have fixed sign. Again all eigenfunctions corresponding to an eigenvalue $\widetilde{\lambda} \neq \widetilde{\lambda}_1(m)$ are nodal. By $\widetilde{u}_1$ we denote the positive, L^p-normalized eigenfunction for (17.22) or (17.23). We have

$$\widetilde{u}_1 \in \text{int } C_+^0 \text{ (for (17.22)) and } \widetilde{u}_1 \in \text{int}\,\widehat{C}_+ \text{ (for (17.23))}.$$

From (17.24) and (17.25) we easily deduce the following monotonicity property for the map $m \to \widetilde{\lambda}_1(m)$.

Proposition 17.4.4 *If $m, m^* \in L^\infty(\Omega)$, $0 \leq m(z) \leq m^*(z)$ for a.a. $z \in \Omega$, $m \not\equiv 0$, $m \not\equiv m^*$, then $\widetilde{\lambda}_1(m^*) < \widetilde{\lambda}_1(m)$.*

For the second eigenvalue $\widetilde{\lambda}_2(m)$ we have (see [7], [52]).

Proposition 17.4.5 *If $m, m^* \in L^\infty(\Omega)$, $0 \leq m(z) < m^*(z)$ for a.a. $z \in \Omega$, $m \not\equiv 0$, then $\widetilde{\lambda}_2(m^*) < \widetilde{\lambda}_2(m)$.*

Also from the properties of $\widehat{\lambda}_1$ and $\widehat{u}_1$, we have the following result. In what follows $\|\cdot\|$ denotes the norm of the space X. So

$$\|u\| = \|\nabla u\|_p \text{ for all } u \in X = W_0^{1,p}(\Omega) \text{ (by Poincaré's inequality)},$$

$$\|u\| = \left[\|u\|_p^p + \|\nabla u\|_p^p\right] \text{ for all } u \in X = W^{1,p}(\Omega).$$

Proposition 17.4.6 *If $\eta \in L^\infty(\Omega)$, $\eta(z) \leq \widehat{\lambda}_1$ for a.a. $z \in \Omega$, $\eta \not\equiv \widehat{\lambda}_1$, then*

(a) $c_8\|u\|^p \leq \mu_0(u) - \int_\Omega \eta(z)|u|^p dz$ for some $c_8 > 0$, all $u \in W_0^{1,p}(\Omega)$;
(b) $c_9\|u\|^p \leq \mu(u) - \int_\Omega \eta(z)|u|^p dz$ for some $c_9 > 0$, all $u \in W^{1,p}(\Omega)$.

In the "linear" case ($p = 2$), using (17.20), (17.21), and the UCP of the eigenspaces, we have the following useful inequalities.

Proposition 17.4.7 *We have:*

(a) If $\eta \in L^\infty(\Omega)$, $\eta(z) \geq \widehat{\lambda}_m$ for a.a. $z \in \Omega$, $\eta \not\equiv \widehat{\lambda}_m$, $m \in \mathbb{N}$, then

$$\mu_0(u) - \int_\Omega \eta(z)u^2 dz \leq -c_{10}\|u\|^2 \text{ for some } c_{10} > 0, \text{ all } u \in \overline{H}_m;$$

$$\mu(u) - \int_\Omega \eta(z)u^2 dz \leq -c_{11}\|u\|^2 \text{ for some } c_{11} > 0, \text{ all } u \in \overline{H}_m.$$

(*b*) *If* $\eta \in L^{\infty}(\Omega)$, $\eta(z) \le \widehat{\lambda}_m$ *for a.a.* $z \in \Omega$, $\eta \not\equiv \widehat{\lambda}_m$, $m \in \mathbb{N}$, *then*

$$c_{12}\|u\|^2 \le \mu_0(u) - \int_{\Omega} \eta(z)u^2 dz \text{ for some } c_{12} > 0, \text{ all } u \in \widehat{H}_m;$$

$$c_{13}\|u\|^2 \le \mu(u) - \int_{\Omega} \eta(z)u^2 dz \text{ for some } c_{13} > 0, \text{ all } u \in \widehat{H}_m.$$

Another consequence of the UCP is the following monotonicity property of the map $m \to \widetilde{\lambda}_k(m)$ for all positive eigenvalues.

Proposition 17.4.8 *If* $m, m^* \in L^{\infty}(\Omega)$, $0 \le m(z) \le m^*(z)$ *for a.a.* $z \in \Omega$, $m \not\equiv 0$, $m \not\equiv m^*$, *then* $\widetilde{\lambda}_k(m^*) < \widetilde{\lambda}_k(m)$ *for all* $k \ge k_0$.

Next we consider the following nonlinear, nonhomogeneous parametric Neumann problem. The choice of the Neumann boundary condition was made in order to simplify the exposition. A similar analysis can be done for the general Robin problem.

So, the problem under consideration is the following:

$$\begin{cases} -\text{div } a(\nabla u(z)) + \xi(z)u(z)^{p-1} = \lambda u(z)^{p-1} + f(z, u(z)) \text{ in } \Omega, \\ \dfrac{\partial u}{\partial n} = 0 \text{ on } \partial\Omega, \ u > 0, \ 1 < p < +\infty, \ \lambda \in \mathbb{R}. \end{cases} \tag{17.26}$$

The hypotheses on the map $a(\cdot)$ are the following:

$H(a)'$ $a(y) = a_0(|y|)y$ for all $y \in \mathbb{R}^N$ with $a_0(t) > 0$ for all $t > 0$, hypotheses $H(a)'$ (i), (ii), (iii), (iv) are the same as the corresponding hypotheses $H(a)$ (i), (ii), (iii), (iv) and

(v) there exists $q \in (1, p]$ such that

$$t \to G_0(t^{\frac{1}{q}}) \text{ is convex,}$$

$$\lim_{t \to 0^+} \frac{qG_0(t)}{t^q} = \widetilde{c} > 0.$$

The hypotheses on the perturbation $f(z, x)$ are the following:

$H(f)_1$ $f: \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function such that for a.a. $z \in \Omega$ $f(z, 0) = 0$ and

- (i) for every $\rho > 0$, there exists $a_\rho \in L^{\infty}(\Omega)_+$ such that $|f(z, x)| \le a_\rho(z)$ for a.a. $z \in \Omega$, all $0 \le x \le \rho$;
- (ii) $\displaystyle\lim_{x \to +\infty} \frac{f(z, x)}{x^{p-1}} = 0$ uniformly for a.a. $z \in \Omega$;
- (iii) $\displaystyle\lim_{x \to 0^+} \frac{f(z, x)}{x^{q-1}} = +\infty$ uniformly for a.a. $z \in \Omega$;
- (iv) for every $\rho > 0$, there exists $\widehat{\xi}_\rho > 0$ such that for a.a. $z \in \Omega$ the function $x \to f(z, x) + \widehat{\xi}_\rho x^{p-1}$ is nondecreasing on $[0, \rho]$.

Remark 17.4.9 Since we are looking for positive solutions and all the above hypotheses concern the positive semiaxis $\mathbb{R}_+ = [0, +\infty)$, without any loss of generality we may assume that $f(z, x) = 0$ for a.a. $z \in \Omega$, all $x \leq 0$. Hypothesis $H(f)_1(ii)$ implies that $f(z, \cdot)$ is $(p-1)$-sublinear near $+\infty$, while $H(f)_1(iii)$ implies that $f(z, \cdot)$ is $(q-1)$-sublinear near 0^+.

Let

$$\mathcal{L} = \{\lambda \in \mathbb{R} \text{ such that problem (17.26) has a positive solution}\},$$
$$S(\lambda) = \{\text{set of positive solutions for (17.26)}\}.$$

Proposition 17.4.10 *If hypotheses $H(a)'$, $H(\xi)$, $H(f)_1$ hold, then $\mathcal{L} \neq \emptyset$.*

Proof Let $\eta > \|\xi\|_\infty$ and consider the following Carathéodory function:

$$k_\lambda(z, x) = \begin{cases} 0 & \text{if } x \leq 0, \\ [\lambda + \eta]x^{p-1} + f(z, x) & \text{if } 0 < x. \end{cases} \tag{17.27}$$

We set $K_\lambda(z, x) = \int_0^x k_\lambda(z, s)ds$ and consider the C^1-functional $\varphi_\lambda : W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\varphi_\lambda(u) = \frac{1}{p}\widehat{\mu}(u) + \frac{\eta}{p}\|u\|_p^p - \int_\Omega K_\lambda(z, u)dz \text{ for all } u \in W^{1,p}(\Omega),$$

with $\widehat{\mu} : W^{1,p}(\Omega) \to \mathbb{R}$ being the C^1-functional defined by

$$\widehat{\mu}(u) = \int_\Omega pG(\nabla u)dz + \int_\Omega \xi(z)|u|^p dz \text{ for all } u \in W^{1,p}(\Omega).$$

Hypotheses $H(f)_1(i)$, (ii) imply that given $\varepsilon > 0$, we can find $c_{14} = c_{14}(\varepsilon) > 0$ such that

$$F(z, x) \leq \frac{\varepsilon}{p}x^p + c_{14} \text{ for a.a. } z \in \Omega, \text{ all } x \geq 0. \tag{17.28}$$

Using (17.27), (17.28), and Corollary 17.2.2, we have

$$\begin{aligned}
\varphi_\lambda(u) &\geq \frac{c_1}{p(p-1)}\|\nabla u\|_p^p + \frac{1}{p}\int_\Omega [\xi(z) + \eta - (\lambda + \varepsilon)]\,|u|^p dz - c_{15} \text{ for some } c_{15} > 0 \\
&\geq \frac{c_1}{p(p-1)}\|\nabla u\|_p^p + c_{16}\|u\|_p^p - c_{15} \text{ for some } c_{16} > 0 \\
&\qquad (\text{choosing } \lambda \in \mathbb{R}, \varepsilon > 0 \text{ such that } \lambda + \varepsilon < \eta - \|\xi\|_\infty),
\end{aligned}$$

$$\Rightarrow \quad \varphi_\lambda(\cdot) \text{ is coercive.}$$

Also, using the Sobolev embedding theorem, we see that φ_λ is sequentially weakly lower semicontinuous. Then by the Weierstrass-Tonelli theorem, we can find $u_\lambda \in W^{1,p}(\Omega)$ such that

$$\varphi_\lambda(u_\lambda) = \inf\Big[\varphi_\lambda(u)\colon u \in W^{1,p}(\Omega)\Big]. \tag{17.29}$$

Hypothesis $H(a)'(v)$ implies that given $\widetilde{c}_0 > \widetilde{c}$ we can find $\delta > 0$ such that

$$G(y) \le \frac{\widetilde{c}_0}{q}|y|^q \text{ for all } |y| \le \delta. \tag{17.30}$$

Moreover, on account of hypothesis $H(f)_1(iii)$ given any $\vartheta > 0$, by choosing $\delta > 0$ even smaller if necessary, we can also have

$$F(z,x) \ge \frac{\vartheta}{q}x^q \text{ for a.a. } z \in \Omega,\ \text{all } 0 \le x \le \delta. \tag{17.31}$$

Let $\widehat{\lambda}_1 = \widehat{\lambda}_1(q,\xi_0)$, $\widehat{u}_1 = \widehat{u}_1(q,\xi_0)$ with $\xi_0 = \frac{1}{\widetilde{c}_0}\xi$. We have $\widehat{u}_1 \in \operatorname{int}\widehat{C}_+$. We choose $t \in (0,1)$ small such that

$$0 < t\widehat{u}_1(z) \le \delta, \quad |\nabla(t\widehat{u}_1)(z)| \le \delta \text{ for all } z \in \overline{\Omega}. \tag{17.32}$$

Then using (17.27), (17.30), (17.31), (17.32), we have

$$\begin{aligned}\varphi_\lambda(t\widehat{u}_1) &\le \frac{\widetilde{c}_0 t^q}{q}\left[\|\nabla\widehat{u}_1\|_q^q + \int_\Omega \xi_0(z)\widehat{u}_1^q dz\right] \\ &\quad - \frac{\lambda t^p}{p}\|\widehat{u}_1\|_p^p - \frac{\vartheta t^q}{q} \text{ (recall } \|\widehat{u}_1\|_p = 1\text{)} \\ &\le \frac{t^q}{q}\left[\widetilde{c}_0\widehat{\lambda}_1 - \vartheta\right] - \frac{\lambda t^p}{p}.\end{aligned}$$

Since $\vartheta > 0$ is arbitrary, choosing $\vartheta > 0$ appropriately big, we have that

$$\begin{aligned}&\varphi_\lambda(t\widehat{u}_1) < 0 \text{ (recall } t \in (0,1) \text{ is small)}, \\ \Rightarrow\quad &\varphi_\lambda(u_\lambda) < 0 = \varphi_\lambda(0) \text{ (see (17.29))}, \\ \Rightarrow\quad &u_\lambda \ne 0.\end{aligned}$$

From (17.29) we have that

$$\begin{aligned}&\varphi_\lambda'(u_\lambda)=0,\\ \Rightarrow\quad &\langle A(u_\lambda),h\rangle+\int_\Omega[\xi(z)+\eta]|u_\lambda|^{p-2}u_\lambda h dz=\int_\Omega k_\lambda(z,u_\lambda)h dz \\ &\text{for all } h\in W^{1,p}(\Omega).\end{aligned} \tag{17.33}$$

In (17.33) we choose $h=-u_\lambda^-\in W^{1,p}(\Omega)$. Using Lemma 17.2.1 we obtain

$$\begin{aligned}&\frac{c_1}{p-1}\|\nabla u_\lambda^-\|_p^p+\int_\Omega[\xi(z)+\eta](u_\lambda^-)^p dz=0 \text{ (see (17.27))},\\ \Rightarrow\quad & u_\lambda\geq 0,\ u_\lambda\neq 0.\end{aligned}$$

From (17.27) and (17.33) we infer that

$$\begin{cases}-\mathrm{div}\ a(\nabla u_\lambda(z))+\xi(z)u_\lambda(z)^{p-1}=\lambda u_\lambda(z)^{p-1}+f(z,u_\lambda(z)) \text{ for a.a. } z\in\Omega,\\ \dfrac{\partial u_\lambda}{\partial n}=0 \text{ on } \partial\Omega.\end{cases} \tag{17.34}$$

Then Propositions 17.2.3 and 17.2.4 imply that $u_\lambda\in C_+\setminus\{0\}$.

From hypothesis $H(f)_1(iv)$ we know that, if $\rho=\|u_\lambda\|_\infty$, then we can find $\widehat{\xi}_\rho>0$ such that $f(z,x)+\widehat{\xi}_\rho x^{p-1}\geq 0$ for a.a. $z\in\Omega$, all $0\leq x\leq\rho$. Then from (17.34) we have

$$\begin{aligned}&\mathrm{div}\ a(\nabla u_\lambda(z))\leq\left[\|\xi\|_\infty+\widehat{\xi}_\rho\right]u_\lambda(z)^{p-1} \text{ for a.a. } z\in\Omega,\\ \Rightarrow\quad & u_\lambda\in \mathrm{int}\,\widehat{C}_+ \text{ (see Theorem 17.3.1) and solves (17.26).}\end{aligned}$$

Therefore $\lambda\in\mathcal{L}$ and so $\mathcal{L}\neq\emptyset$. □

Remark 17.4.11 A by-product of the above proof is that $S(\lambda)\subseteq\mathrm{int}\,\widehat{C}_+$. Also in the case of the p-Laplacian (that is, $a(y)=|y|^{p-2}y$ for all $y\in\mathbb{R}^N$), from the above proof we see that for all $\lambda<\widehat{\lambda}_1=\widehat{\lambda}_1(p,\xi)$, $S(\lambda)\neq\emptyset$.

Next we show that $\mathcal{L}$ is a half line.

Proposition 17.4.12 *If hypotheses $H(a)'$, $H(\xi)$, $H(f)_1$ hold, $\lambda\in\mathcal{L}$ and $\vartheta<\lambda$, then $\vartheta\in\mathcal{L}$.*

Proof Since $\lambda \in \mathcal{L}$ we can find $u_\lambda \in S(\lambda) \subseteq \operatorname{int} \widehat{C}_+$. We consider the following truncation-perturbation of the reaction in problem (17.26)$_\vartheta$ (again $\eta > \|\xi\|_\infty$):

$$e_\vartheta(z,x) = \begin{cases} 0 & \text{if } x < 0, \\ \vartheta x^{p-1} + f(z,x) + \eta x^{p-1} & \text{if } 0 \le x \le u_\lambda(z), \\ \vartheta u_\lambda(z)^{p-1} + f(z,u_\lambda(z)) + \eta u_\lambda(z)^{p-1} & \text{if } u_\lambda(z) < x. \end{cases} \tag{17.35}$$

This is a Carathéodory function. We set $E_\vartheta(z,x) = \int_0^x e_\vartheta(z,s)ds$ and consider the C^1-functional $\psi_\vartheta : W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\psi_\vartheta(u) = \frac{1}{p}\widehat{\mu}(u) + \frac{\eta}{p}\|u\|_p^p - \int_\Omega E_\vartheta(z,u)dz \text{ for all } u \in W^{1,p}(\Omega).$$

Evidently $\psi_\vartheta(\cdot)$ is coercive and sequentially weakly lower semicontinuous. So, we can find $u_\vartheta \in W^{1,p}(\Omega)$ such that

$$\psi_\vartheta(u_\vartheta) = \inf\left[\psi_\vartheta(u) : u \in W^{1,p}(\Omega)\right]. \tag{17.36}$$

As in the proof of Proposition 17.4.10, using hypotheses $H(a)'(v)$ and $H(f)_1(iii)$ we show that $\psi_\vartheta(u_\vartheta) < 0 = \psi_\vartheta(0)$, hence $u_\vartheta \ne 0$. From (17.36) we have

$$\begin{aligned} & \psi'_\vartheta(u_\vartheta) = 0, \\ \Rightarrow\quad & \langle A(u_\vartheta), h\rangle + \int_\Omega [\xi(z)+\eta]|u_\vartheta|^{p-2}u_\vartheta h dz = \int_\Omega e_\vartheta(z,u_\vartheta)h dz \\ & \qquad\qquad \text{for all } h \in W^{1,p}(\Omega). \end{aligned} \tag{17.37}$$

In (17.37) we choose $h = -u_\vartheta^- \in W^{1,p}(\Omega)$. Then using Lemma 17.2.1 we have

$$\begin{aligned} & \frac{c_1}{p-1}\|\nabla u_\vartheta^-\|_p^p + \int_\Omega [\xi(z)+\eta](u_\vartheta^-)^p dz = 0 \text{ (see (17.35))}, \\ \Rightarrow\quad & c_{16}\|u_\vartheta^-\|^p \le 0 \text{ for some } c_{16} > 0 \text{ (recall } \eta > \|\xi\|_\infty), \\ \Rightarrow\quad & u_\vartheta \ge 0,\ u_\vartheta \ne 0. \end{aligned}$$

Also in (17.37) we choose $h = (u_\vartheta - u_\lambda)^+ \in W^{1,p}(\Omega)$. Then

$$\begin{aligned} & \langle A(u_\vartheta), (u_\vartheta - u_\lambda)^+\rangle + \int_\Omega [\xi(z)+\eta]u_\vartheta^{p-1}(u_\vartheta - u_\lambda)^+ dz \\ & = \int_\Omega \left[\vartheta u_\lambda^{p-1} + f(z,u_\lambda) + \eta u_\lambda^{p-1}\right](u_\vartheta - u_\lambda)^+ dz \text{ (see (17.35))} \end{aligned}$$

$$\leq \int_\Omega \left[\lambda u_\lambda^p + f(z,u_\lambda) + \eta u_\lambda^{p-1}\right](u_\vartheta - u_\lambda)^+ dz \text{ (recall } \vartheta < \lambda)$$

$$= \langle A(u_\lambda), (u_\vartheta - u_\lambda)^+\rangle + \int_\Omega [\xi(z)+\eta] u_\lambda^{p-1}(u_\vartheta - u_\lambda)^+ dz \text{ (since } u_\lambda \in S(\lambda))$$

$$\Rightarrow \quad u_\vartheta \leq u_\lambda.$$

So, we have proved that

$$u_\vartheta \in [0, u_\lambda] = \{u \in W^{1,p}(\Omega) : 0 \leq u(z) \leq u_\lambda(z) \text{ for a.a. } z \in \Omega\},\ u_\vartheta \neq 0. \tag{17.38}$$

Then from (17.35), (17.37), (17.38) we conclude that $u_\vartheta \in S(\vartheta) \subseteq \operatorname{int} \widehat{C}_+$, hence $\vartheta \in \mathcal{L}$. □

Let $\lambda^* = \sup \mathcal{L}$.

Proposition 17.4.13 *If hypotheses $H(a)'$, $H(\xi)$, $H(f)_1$ hold, then $\lambda^* < +\infty$.*

Proof Hypotheses $H(\xi)$, $H(f)_1$ imply that we can find $\widetilde{\lambda} > 0$ big such that

$$\widetilde{\lambda} x^{p-1} + f(z,x) - \xi(z)x^{p-1} \geq x^{p-1} \text{ for a.a. } z \in \Omega,\ \text{ all } x \geq 0. \tag{17.39}$$

Let $\lambda > \widetilde{\lambda}$ and assume that $\lambda \in \mathcal{L}$. Then we can find $u \in S(\lambda) \subseteq \operatorname{int} \widehat{C}_+$. Let

$$m = \min_{\overline{\Omega}} u > 0.$$

For $\delta > 0$ we set $m_\delta = m + \delta > 0$ and for $\rho = \|u\|_\infty$ let $\widehat{\xi}_\rho > 0$ be as in $H(f)_1(iv)$. We can always take $\widehat{\xi}_\rho > \max\{\lambda, \|\xi\|_\infty\}$ so that $x \to \lambda x^{p-1} + f(z,x) + \widehat{\xi}_\rho x^{p-1}$ is nondecreasing on $[0,\rho]$. Then we have

$$\begin{aligned}
& -\operatorname{div}\, a(\nabla m_\delta) + [\xi(z) + \widehat{\xi}_\rho] m_\delta^{p-1} \\
&\leq [\xi(z) + \widehat{\xi}_\rho] m^{p-1} + \tau(\delta) \text{ with } \tau(\delta) \to 0^+ \text{ as } \delta \to 0^+ \\
&\leq [\xi(z) + \widehat{\xi}_\rho] m^{p-1} + m^{p-1} + \tau(\delta) \\
&\leq \widetilde{\lambda} m^{p-1} + f(z,m) + \widehat{\xi}_\rho m^{p-1} + \tau(\delta) \text{ (see (17.39))} \\
&= \lambda m^{p-1} + f(z,m) + \widehat{\xi}_\rho m^{p-1} - (\lambda - \widetilde{\lambda}) m^{p-1} + \tau(\delta) \\
&\leq \lambda m^{p-1} + f(z,m) + \widehat{\xi}_\rho m^{p-1} \text{ for } \delta > 0 \text{ small} \\
&\qquad \text{(recall } \lambda > \widetilde{\lambda} \text{ and } \tau(\delta) \to 0^+ \text{ as } \delta \to 0^+) \\
&\leq \lambda u^{p-1} + f(z,u) + \widehat{\xi}_\rho u^{p-1} \\
&= -\operatorname{div}\, a(\nabla u) + [\xi(z) + \widehat{\xi}_\rho] u^{p-1} \text{ for a.a. } z \in \Omega.
\end{aligned} \tag{17.40}$$

Acting on (17.40) with $(m_\delta - u)^+ \in W^{1,p}(\Omega)$, we obtain

$$m_\delta \leq u \text{ for } \delta > 0 \text{ small,}$$

a contradiction to the definition of m (alternatively, we can use Proposition 17.3.5). Therefore $\lambda \notin \mathcal{L}$ and so $\lambda^* \leq \widetilde{\lambda} < +\infty$. □

Next we show that for $\lambda < \lambda^*$, the solution set $S(\lambda) \subseteq \operatorname{int} \widehat{C}_+$ admits a smallest element. Fix $\lambda < \lambda^*$ and $r \in (p, p^*)$. Hypotheses $H(f)_1$ imply that we can find $c_{17} = c_{17}(\lambda) > 0$ and $c_{18} = c_{18}(\lambda) > 0$ such that

$$\lambda x^{p-1} + f(z,x) \geq c_{17} x^{q-1} - c_{18} x^{r-1} \text{ for a.a. } z \in \Omega,\ x \geq 0.$$

This unilateral growth condition on the reaction of problem (17.26) leads to the following auxiliary Neumann equation

$$\begin{cases} -\operatorname{div}\, a(\nabla u(z)) + |\xi(z)| u(z)^{p-1} = c_{17} u(z)^{q-1} - c_{18} u(z)^{r-1} \text{ in } \Omega, \\ \dfrac{\partial u}{\partial n} = 0 \text{ on } \partial\Omega,\ u > 0. \end{cases} \tag{17.41}$$

Proposition 17.4.14 *If hypotheses $H(a)'$, $H(\xi)$ hold, then problem (17.41) admits a unique solution $u_*^\lambda \in \operatorname{int} \widehat{C}_+$.*

Proof Consider the energy functional for problem (17.41) $\psi\colon W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\psi(u) = \int_\Omega G(\nabla u)dz + \frac{1}{p}\int_\Omega |\xi(z)||u|^p dz + \frac{1}{p}\|u^-\|_p^p + \frac{c_{18}}{r}\|u^+\|_r^r - \frac{c_{17}}{q}\|u^+\|_q^q$$
$$\text{for all } u \in W^{1,p}(\Omega).$$

Evidently $\psi(\cdot)$ is coercive and sequentially weakly lower semicontinuous. So, we can find $u_*^\lambda \in W^{1,p}(\Omega)$ such that

$$\psi(u_*^\lambda) = \inf\left[\psi(u)\colon u \in W^{1,p}(\Omega)\right]. \tag{17.42}$$

Since $q < p < r$, we have $\psi(u_*^\lambda) < 0 = \psi(0)$, hence $u_*^\lambda \neq 0$. From (17.42) we have

$$\begin{aligned} &\psi'(u_*^\lambda) = 0, \\ &\quad \Rightarrow \langle A(u_*^\lambda), h\rangle + \int_\Omega |\xi(z)||u_*^\lambda|^{p-2} u_*^\lambda h dz - \int_\Omega ((u_*^\lambda)^-)^{p-1} h dz \\ &\quad = c_{17}\int_\Omega ((u_*^\lambda)^+)^{q-1} h dz - c_{18}\int_\Omega ((u_*^\lambda)^+)^{r-1} h dz \text{ for all } h \in W^{1,p}(\Omega). \end{aligned} \tag{17.43}$$

In (17.43) we choose $h = -(u_*^\lambda)^- \in W^{1,p}(\Omega)$. Then

$$\frac{c_1}{p-1}\|\nabla(u_*^\lambda)^-\|_p^p + \int_\Omega [|\xi(z)| + 1]((u_*^\lambda)^-)^p dz \le 0 \text{ (see Lemma 17.2.1)},$$
$$\Rightarrow \quad u_*^\lambda \ge 0,\ u_*^\lambda \ne 0.$$

From (17.43) it follows that u_*^λ is a positive solution of (17.41) and $u_*^\lambda \in \operatorname{int} \widehat{C}_+$.

Next we show that this positive solution is unique. To this end, we introduce the integral function $j\colon L^1(\Omega) \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ defined by

$$j(u) = \begin{cases} \int_\Omega G(\nabla u^{\frac{1}{q}})dz + \frac{1}{p}\int_\Omega |\xi(z)|u^{\frac{p}{q}}dz & \text{if } u \ge 0,\ u^{\frac{1}{q}} \in W^{1,p}(\Omega), \\ +\infty & \text{otherwise.} \end{cases}$$

Here $q \le p$ is as in hypothesis $H(a)'(v)$. Let $\operatorname{dom}\ j = \{u \in L^1(\Omega)\colon j(u) < +\infty\}$ (the effective domain of $j(\cdot)$). Let $u_1, u_2 \in \operatorname{dom}\ j$ and set $u = [(1-t)u_1 + tu_2]^{\frac{1}{q}}$ with $t \in [0,1]$. Using Lemma 1 of Díaz and Saá [23]

$$|\nabla u(z)| \le \left[(1-t)|\nabla u_1(z)^{\frac{1}{q}}|^q + t|\nabla u_2(z)^{\frac{1}{q}}|^q\right]^{\frac{1}{q}} \text{ for a.a. } z \in \Omega,$$
$$\Rightarrow \quad G_0(|\nabla u(z)|) \le G_0\left(\left[(1-t)|\nabla u_1(z)^{\frac{1}{q}}|^q + t|\nabla u_2(z)^{\frac{1}{q}}|^q\right]^{\frac{1}{q}}\right)$$
$$\text{(since } G_0(\cdot) \text{ is increasing)}$$
$$\le (1-t)G_0\left(|\nabla u_1(z)^{\frac{1}{q}}|\right) + tG_0\left(|\nabla u_2(z)^{\frac{1}{q}}|\right) \text{ for a.a. } z \in \Omega$$
$$\text{(see hypothesis } H(a)'(v)),$$
$$\Rightarrow \quad G(\nabla u(z)) \le (1-t)G\left(\nabla u_1(z)^{\frac{1}{q}}\right) + tG\left(\nabla u_2(z)^{\frac{1}{q}}\right) \text{ for a.a. } z \in \Omega,$$
$$\Rightarrow \quad \operatorname{dom}\ j \ni u \to \int_\Omega G\left(\nabla u^{\frac{1}{q}}\right)dz \text{ is convex.}$$

Since $q \le p$, it follows that $\operatorname{dom} j \ni u \to \int_\Omega |\xi(z)|u^{\frac{p}{q}}dz$ is convex too. Therefore $u \to j(u)$ is convex and by Fatou's lemma it is also lower semicontinuous.

Suppose that v_*^λ is another positive solution of (17.41). Again we show that $v_*^\lambda \in \operatorname{int} \widehat{C}_+$. Given $h \in C^1(\overline{\Omega})$, for $|t| \le 1$ small we have

$$(u_*^\lambda)^q + th \in \operatorname{dom} j \text{ and } (v_*^\lambda)^q + th \in \operatorname{dom}\ j.$$

We can easily see that $j(\cdot)$ is Gateaux differentiable at $(u_*^\lambda)^q$ and at $(v_*^\lambda)^q$ in the direction h. Moreover, using the chain rule and Green's identity (see Gasiński and Papageorgiou [31], p. 210), we have

$$j'((u_*^\lambda)^q)(h) = \frac{1}{q}\int_\Omega \frac{-\operatorname{div} a(\nabla u_*^\lambda) + |\xi(z)|(u_*^\lambda)^{p-1}}{(u_*^\lambda)^{q-1}} h dz,$$

$$j'((v_*^\lambda)^q)(h) = \frac{1}{q}\int_\Omega \frac{-\operatorname{div} a(\nabla v_*^\lambda) + |\xi(z)|(v_*^\lambda)^{p-1}}{(v_*^\lambda)^{q-1}} h dz.$$

The convexity of $j(\cdot)$ implies the monotonicity of $j'(\cdot)$. So

$$0 \le \int_\Omega \left(\frac{-\operatorname{div} a(\nabla u_*^\lambda) + |\xi(z)|(u_*^\lambda)^{p-1}}{(u_*^\lambda)^{q-1}} - \frac{-\operatorname{div} a(\nabla v_*^\lambda) + |\xi(z)|(v_*^\lambda)^{p-1}}{(v_*^\lambda)^{q-1}} \right) \times \left(\left(u_*^\lambda\right)^q - \left(v_*^\lambda\right)^q \right) dz$$

$$= \int_\Omega c_{18} \left[\left(v_*^\lambda\right)^{r-q} - \left(u_*^\lambda\right)^{r-q} \right] \left(\left(u_*^\lambda\right)^q - \left(v_*^\lambda\right)^q \right) dz$$

$$\Rightarrow \quad u_*^\lambda = v_*^\lambda \text{ (recall } q \le p < r).$$

This proves the uniqueness of the positive solution of problem (17.41). □

Proposition 17.4.15 *If hypotheses $H(a)'$, $H(\xi)$, $H(f)_1$ hold and $\lambda < \lambda^*$, then $u_*^\lambda \le u$ for all $u \in S(\lambda)$.*

Proof Let $u \in S(\lambda) \subseteq \operatorname{int} \widehat{C}_+$. With $\eta > \|\xi\|_\infty$ as before, we introduce the Carathéodory function $e\colon \Omega \times \mathbb{R} \to \mathbb{R}$ defined by:

$$e(z,x) = \begin{cases} 0 & \text{if } x < 0, \\ c_{17}x^{q-1} - c_{18}x^{r-1} + \eta x^{p-1} & \text{if } 0 \le x \le u(z), \\ c_{17}u(z)^{q-1} - c_{18}u(z)^{r-1} + \eta u(z)^{p-1} & \text{if } u(z) < x. \end{cases} \tag{17.44}$$

We set $E(z,x) = \int_0^x e(z,s)ds$ and consider the C^1-functional $\tau\colon W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\tau(u) = \frac{1}{p}\widehat{\mu}(u) + \frac{\eta}{p}\|u\|_p^p - \int_\Omega E(z,u)dz \text{ for all } u \in W^{1,p}(\Omega).$$

Since $\eta > \|\xi\|_\infty$, $\tau(\cdot)$ is coercive (see (17.44)). Also, it is sequentially weakly lower semicontinuous. So, we can find $\widetilde{u}_*^\lambda \in W^{1,p}(\Omega)$ such that

$$\tau(\widetilde{u}_*^\lambda) = \inf\Big[\tau(u)\colon u \in W^{1,p}(\Omega)\Big]. \tag{17.45}$$

As before, using the fact that $q \leq p < r$, we see that

$$\tau(\widetilde{u}_*^\lambda) < 0 = \tau(0),$$
$$\Rightarrow \quad \widetilde{u}_*^\lambda \neq 0.$$

From (17.45) we have

$$\tau'(\widetilde{u}_*^\lambda) = 0,$$
$$\Rightarrow \quad \langle A(\widetilde{u}_*^\lambda), h\rangle + \int_\Omega [\xi(z) + \eta]|\widetilde{u}_*^\lambda|^{p-2}\widetilde{u}_*^\lambda h dz = \int_\Omega e(z, \widetilde{u}_*^\lambda) h dz \quad \text{for all } h \in W^{1,p}(\Omega). \tag{17.46}$$

In (17.46) first we choose $h = -(\widetilde{u}_*^\lambda)^- \in W^{1,p}(\Omega)$. Then

$$\frac{c_1}{p-1}\|\nabla(\widetilde{u}_*^\lambda)^-\|_p^p + \int_\Omega [\xi(z) + \eta]((\widetilde{u}_*^\lambda)^-)^p dz \leq 0 \text{ (see (17.44))},$$
$$\Rightarrow \quad \widetilde{u}_*^\lambda \geq 0, \ \widetilde{u}_*^\lambda \neq 0.$$

Next in (17.46) we choose $h = (\widetilde{u}_*^\lambda - u)^+ \in W^{1,p}(\Omega)$. Then

$$\langle A(\widetilde{u}_*^\lambda), (\widetilde{u}_*^\lambda - u)^+\rangle + \int_\Omega [\xi(z) + \eta](\widetilde{u}_*^\lambda)^{p-1}(\widetilde{u}_*^\lambda - u)^+ dz$$
$$= \int_\Omega [c_{17}u^{q-1} - c_{18}u^{r-1} + \eta u^{p-1}](\widetilde{u}_*^\lambda - u)^+ dz \text{ (see (17.44))}$$
$$= \langle A(u), (\widetilde{u}_*^\lambda - u)^+\rangle + \int_\Omega [\xi(z) + \eta]u^{p-1}(\widetilde{u}_*^\lambda - u)^+ dz$$
$$\Rightarrow \quad \widetilde{u}_*^\lambda \leq u.$$

Therefore

$$\widetilde{u}_*^\lambda \in [0, u], \quad \widetilde{u}_*^\lambda \neq 0. \tag{17.47}$$

From (17.44), (17.46), (17.47) we infer that

$$\widetilde{u}_*^\lambda \text{ is a positive solution of (17.41)},$$
$$\Rightarrow \quad \widetilde{u}_*^\lambda = u_*^\lambda \in \operatorname{int}\widehat{C}_+ \text{ (see Proposition 17.4.14)},$$
$$\Rightarrow \quad \widetilde{u}_*^\lambda \leq u \text{ for all } u \in S(\lambda).$$

□

The set $S(\lambda)$ is downward directed, that is, if $u_1, u_2 \in S(\lambda)$ we can find $u \in S(\lambda)$ such that $u \leq u_1$, $u \leq u_2$ (see Papageorgiou et al. [69], proof of Proposition 7)

Proposition 17.4.16 *If hypotheses $H(a)'$, $H(\xi)$, $H(f)_1$ hold and $\lambda < \lambda^*$, then $S(\lambda)$ admits a smallest element $\overline{u}_\lambda \in S(\lambda)$, that is, $\overline{u}_\lambda \leq u$ for all $u \in S(\lambda)$.*

Proof Using Lemma 3.10, p. 178, of Hu and Papageorgiou [43], we can find $\{u_n\}_{n\geq 1} \subseteq S(\lambda)$ decreasing such that $\inf S(\lambda) = \inf_{n\geq 1} u_n$. We have

$$\langle A(u_n), h\rangle + \int_\Omega \xi(z) u_n^{p-1} h dz = \lambda \int_\Omega u_n^{p-1} h dz + \int_\Omega f(z, u_n) h dz \tag{17.48}$$
$$\text{for all } h \in W^{1,p}(\Omega), \text{ all } n \in \mathbb{N}.$$

Since $0 \leq u_n \leq u_1 \in \operatorname{int} \widehat{C}_+$ for all $n \in \mathbb{N}$, choosing $h = u_n \in W^{1,p}(\Omega)$, we show that $\{u_n\}_{n\geq 1} \subseteq W^{1,p}(\Omega)$ is bounded. So, we may assume that

$$u_n \xrightarrow{w} \overline{u}_\lambda \text{ in } W^{1,p}(\Omega) \text{ and } u_n \to \overline{u}_\lambda \text{ in } L^p(\Omega) \text{ as } n \to +\infty. \tag{17.49}$$

In (17.48) we choose $h = u_n - \overline{u}_\lambda \in W^{1,p}(\Omega)$, pass to the limit as $n \to +\infty$ and use (17.49). Then

$$\lim_{n\to+\infty} \langle A(u_n), u_n - \overline{u}_\lambda\rangle = 0,$$
$$\Rightarrow \quad u_n \to \overline{u}_\lambda \text{ in } W^{1,p}(\Omega) \text{ (see Proposition 17.2.7)}. \tag{17.50}$$

If in (17.48) we pass to the limit as $n \to +\infty$ and use (17.50), then

$$\langle A(\overline{u}_\lambda), h\rangle + \int_\Omega \xi(z)\overline{u}_\lambda^{p-1} h dz = \int_\Omega [\lambda \overline{u}_\lambda^{p-1} + f(z, \overline{u}_\lambda)] h dz \text{ for all } h \in W^{1,p}(\Omega). \tag{17.51}$$

From Proposition 17.4.15, we know that

$$u_*^\lambda \leq u_n \text{ for all } n \in \mathbb{N},$$
$$\Rightarrow \quad u_*^\lambda \leq \overline{u}_\lambda. \tag{17.52}$$

From (17.51) and (17.52) it follows that $\overline{u}_\lambda \in S(\lambda)$, $\overline{u}_\lambda = \inf S(\lambda)$. □

We examine the map $\lambda \to \overline{u}_\lambda$ from $\mathcal{L}$ into $C^1(\overline{\Omega})$.

Proposition 17.4.17 *If hypotheses $H(a)'$, $H(\xi)$, $H(f)_1$ hold, then the map $\lambda \to \overline{u}_\lambda$ from $\mathcal{L}$ into $C^1(\overline{\Omega})$ is*

(*a*) *strictly increasing, that is, if $\vartheta < \lambda \in \mathcal{L}$, then $\overline{u}_\lambda - \overline{u}_\vartheta \in \operatorname{int} C_+$;*
(*b*) *left continuous.*

Proof

(*a*) Let $\vartheta < \lambda \in \mathcal{L}$ and let $\overline{u}_\lambda \in S(\lambda) \subseteq \operatorname{int} \widehat{C}_+$ be the minimal solution of problem (17.26) produced in Proposition 17.4.16. Again we choose $\eta > \|\xi\|_\infty$ and introduce the Carathéodory function $\gamma_\vartheta : \Omega \times \mathbb{R} \to \mathbb{R}$ defined by

$$\gamma_\vartheta(z,x) = \begin{cases} 0 & \text{if } x < 0, \\ [\vartheta + \eta]x^{p-1} + f(z,x) & \text{if } 0 \leq x \leq \overline{u}_\lambda(z), \\ [\vartheta + \eta]\overline{u}_\lambda(z)^{p-1} + f(z,\overline{u}_\lambda(z)) & \text{if } \overline{u}_\lambda(z) < x. \end{cases} \tag{17.53}$$

We set $\Gamma_\vartheta(z,x) = \int_0^x \gamma_\vartheta(z,s)ds$ and consider the C^1-functional $\widetilde{\varphi}_\vartheta : W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\widetilde{\varphi}_\vartheta(u) = \frac{1}{p}\widehat{\mu}(u) + \frac{\eta}{p}\|u\|_p^p - \int_\Omega \Gamma_\vartheta(z,u)dz \text{ for all } u \in W^{1,p}(\Omega).$$

This functional too is coercive (see (17.53)) and sequentially weakly lower semicontinuous. So, we can find $u_\vartheta \in W^{1,p}(\Omega)$ such that

$$\widetilde{\varphi}_\vartheta(u_\vartheta) = \inf[\widetilde{\varphi}_\vartheta(u) : u \in W^{1,p}(\Omega)]. \tag{17.54}$$

Using hypotheses $H(a)'(v)$ and $H(f)_1(iii)$, as in the proof of Proposition 17.4.10, we show that

$$\widetilde{\varphi}_\vartheta(u_\vartheta) < 0 = \widetilde{\varphi}_\vartheta(0), \text{ hence } u_\vartheta \neq 0.$$

From (17.54) we have

$$\langle A(u_\vartheta), h\rangle + \int_\Omega [\xi(z) + \eta]|u_\vartheta|^{p-2}u_\vartheta h dz = \int_\Omega \gamma_\vartheta(z,u_\vartheta)h dz \text{ for all } h \in W^{1,p}(\Omega). \tag{17.55}$$

If in (17.55) we choose first $h = -u_\vartheta^- \in W^{1,p}(\Omega)$ and then $h = (u_\vartheta - \overline{u}_\lambda)^+ \in W^{1,p}(\Omega)$, we obtain

$$u_\vartheta \in [0, \overline{u}_\lambda],\ u_\vartheta \neq 0,\ u_\vartheta \neq \overline{u}_\lambda \text{ (since } \theta < \lambda). \tag{17.56}$$

From (17.53), (17.55), (17.56) it follows that $u_\vartheta \in S(\vartheta) \subseteq \operatorname{int} \widehat{C}_+$. Then

$$\overline{u}_\vartheta \leq u_\vartheta \leq \overline{u}_\lambda.$$

Now let $\rho = \|\overline{u}_\lambda\|_\infty$ and let $\widehat{\xi}_\rho > 0$ be as postulated by hypothesis $H(f)_1(iv)$. In fact we can always take $\widehat{\xi}_\rho > \max\{\lambda, \|\xi\|_\infty\}$ so that for a.a. $z \in \Omega$ the function $x \to [\lambda + \widehat{\xi}_\rho]x^{p-1} + f(z,x)$ is nondecreasing on $[0,\rho]$. Also let $\overline{m}_\vartheta = \min_{\overline{\Omega}} \overline{u}_\vartheta > 0$ (recall $\overline{u}_\vartheta \in \operatorname{int} \widehat{C}_+$). We have

$$\begin{aligned}
&-\operatorname{div}\, a(\nabla \overline{u}_\vartheta^\delta) + \left[\xi(z) + \widehat{\xi}_\rho\right](\overline{u}_\vartheta^\delta)^{p-1} \\
&\quad\le -\operatorname{div}\, a(\nabla \overline{u}_\vartheta) + \left[\xi(z) + \widehat{\xi}_\rho\right]\overline{u}_\vartheta^{p-1} + \varepsilon(\delta) \text{ with } \varepsilon(\delta) \to 0^+ \text{ as } \delta \to 0^+ \\
&\quad= \vartheta \overline{u}_\vartheta^{p-1} + f(z, \overline{u}_\vartheta) + \widehat{\xi}_\rho \overline{u}_\vartheta^{p-1} + \varepsilon(\delta) \\
&\quad= \lambda \overline{u}_\vartheta^{p-1} + f(z, \overline{u}_\vartheta) + \widehat{\xi}_\rho \overline{u}_\vartheta^{p-1} - (\lambda - \vartheta)\overline{u}_\vartheta^{p-1} + \varepsilon(\delta) \\
&\quad\le \lambda \overline{u}_\lambda^{p-1} + f(z, \overline{u}_\lambda) + \widehat{\xi}_\rho \overline{u}_\lambda^{p-1} - (\lambda - \theta)\overline{m}_\vartheta + \varepsilon(\delta) \text{ (recall the choice of } \widehat{\xi}_\rho) \\
&\quad\le \lambda \overline{u}_\lambda^{p-1} + f(z, \overline{u}_\lambda) + \widehat{\xi}_\rho \overline{u}_\lambda^{p-1} \text{ for } \delta > 0 \text{ small} \\
&\quad= -\operatorname{div}\, a(\nabla \overline{u}_\lambda) + \left[\xi(z) + \widehat{\xi}_\rho\right]\overline{u}_\lambda^{p-1}.
\end{aligned} \tag{17.57}$$

Acting on (17.57) with $(\overline{u}_\vartheta^\delta - \overline{u}_\lambda)^+ \in W^{1,p}(\Omega)$, using the nonlinear Green's identity and the Neumann boundary condition (note $\dfrac{\partial \overline{u}_\vartheta^\delta}{\partial n} = \dfrac{\partial \overline{u}}{\partial n} = 0$), we obtain

$$\begin{aligned}
&\overline{u}_\lambda - \overline{u}_\vartheta^\delta \in C_+ \text{ for } \delta > 0 \text{ small,} \\
\Rightarrow\ &\overline{u}_\lambda - \overline{u}_\vartheta \in \operatorname{int} \widehat{C}_+, \\
\Rightarrow\ &\lambda \to \overline{u}_\lambda \text{ is strictly increasing.}
\end{aligned}$$

(*b*) Let $\lambda_n \to \lambda^-$ with $\lambda \in \mathcal{L}$. We set $\overline{u}_n = \overline{u}_{\lambda_n}$ for all $n \in \mathbb{N}$. Evidently $\{\overline{u}_n\}_{n \ge 1} \subseteq W^{1,p}(\Omega)$ is bounded (see hypothesis $H(f)_1(i)$ and note that from part (*a*) we have $0 \le \overline{u}_n \le \overline{u}_\lambda \in \operatorname{int} \widehat{C}_+$ for all $n \in \mathbb{N}$). So, we may assume that

$$\overline{u}_n \xrightarrow{w} \widetilde{u}_\lambda \text{ in } W^{1,p}(\Omega) \text{ and } u_n \to \widetilde{u}_\lambda \text{ in } L^p(\Omega) \text{ as } n \to +\infty. \tag{17.58}$$

We know that

$$\langle A(\overline{u}_n), h\rangle + \int_\Omega \xi(z)\overline{u}_n^{p-1} h dz = \int_\Omega \left[\lambda_n \overline{u}_n^{p-1} + f(z, \overline{u}_n)\right] h dz \tag{17.59}$$
$$\text{for all } h \in W^{1,p}(\Omega), \text{ all } n \in \mathbb{N}.$$

Choosing $h = \overline{u}_n - \widetilde{u}_\lambda \in W^{1,p}(\Omega)$ in (17.59), passing to the limit as $n \to +\infty$ and using Proposition 17.2.7, we have

$$\overline{u}_n \to \widetilde{u}_\lambda \text{ in } W^{1,p}(\Omega). \tag{17.60}$$

So, if we pass to the limit as $n \to +\infty$ in (17.59) and use (17.60), then we infer that $\widetilde{u}_\lambda \in S(\lambda)$. We claim that $\widetilde{u}_\lambda = \overline{u}_\lambda$. On account of (17.60) and Proposition 17.2.3, we can find $c_{19} > 0$ such that

$$\|\overline{u}_n\|_\infty \le c_{19} \text{ for all } n \in \mathbb{N}. \tag{17.61}$$

Then (17.61) and Proposition 17.2.4 imply that we can find $s \in (0, 1)$ and $c_{20} > 0$ such that

$$\overline{u}_n \in C^{1,s}(\overline{\Omega}) \text{ and } \|\overline{u}_n\|_{C^{1,s}(\overline{\Omega})} \le c_{20} \text{ for all } n \in \mathbb{N}. \tag{17.62}$$

Recalling that $C^{1,s}(\overline{\Omega})$ is embedded compactly in $C^1(\overline{\Omega})$ from (17.62) and (17.60) we infer that

$$\overline{u}_n \to \widetilde{u}_\lambda \text{ in } C^1(\overline{\Omega}). \tag{17.63}$$

If $\widetilde{u}_\lambda \neq \overline{u}_\lambda$, then we can find $z_0 \in \overline{\Omega}$ such that

$$\begin{aligned} & \overline{u}_\lambda(z_0) < \widetilde{u}_\lambda(z_0), \\ \Rightarrow \quad & \overline{u}_\lambda(z_0) < \overline{u}_n(z_0) \text{ for all } n \ge n_0 \text{ (see (17.63))}. \end{aligned}$$

But this contradicts (a). Therefore $\widetilde{u}_\lambda = \overline{u}_\lambda$ and so we conclude that $\lambda \to \overline{u}_\lambda$ is left continuous. □

Now we show that λ^* is not admissible.

Proposition 17.4.18 *If hypotheses $H(a)'$, $H(\xi)$, $H(f)_1$ hold, then $\lambda^* \notin \mathcal{L}$.*

Proof Arguing by contradiction suppose that $\lambda^* \in \mathcal{L}$. By Proposition 17.4.16 we have $\overline{u}_{\lambda^*} = \overline{u}_* \in \operatorname{int} \widehat{C}_+$. Let $\lambda > \lambda^*$ and consider the following Carathéodory function

$$\theta_\lambda(z, x) = \begin{cases} [\lambda + \eta]\overline{u}_*^{p-1}(z) + f(z, \overline{u}_*(z)) & \text{if } x \le \overline{u}_*(z), \\ [\lambda + \eta]x^{p-1} + f(z, x) & \text{if } \overline{u}_*(z) < x, \end{cases} \tag{17.64}$$

(as before $\eta > \|\xi\|_\infty$). We set $\Theta_\lambda(z, x) = \int_0^x \theta_\lambda(z, s)ds$ and consider the C^1-functional $\varphi_\lambda^* : W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\varphi_\lambda^*(u) = \frac{1}{p}\widehat{\mu}(u) + \frac{\eta}{p}\|u\|_p^p - \int_\Omega \Theta_\lambda(z, u)dz \text{ for all } u \in W^{1,p}(\Omega).$$

An application of the mountain pass theorem gives

$$\begin{aligned}
& u_\lambda \in K_{\varphi_\lambda^*}, \\
\Rightarrow \quad & (\varphi_\lambda^*)'(u_\lambda) = 0 \\
\Rightarrow \quad & \langle A(u_\lambda), h\rangle + \int_\Omega [\xi(z) + \eta]|u_\lambda|^{p-2}u_\lambda h dz = \int_\Omega \theta_\lambda(z, u_\lambda)h dz \qquad (17.65) \\
& \qquad \text{for all h} \in W^{1,p}(\Omega).
\end{aligned}$$

In (17.65) we choose $h = (\overline{u}_* - u_\lambda)^+ \in W^{1,p}(\Omega)$. Then

$$\begin{aligned}
& \langle A(u_\lambda), (\overline{u}_* - u_\lambda)^+\rangle + \int_\Omega [\xi(z) + \eta]|u_\lambda|^{p-2}u_\lambda(\overline{u}_* - u_\lambda)^+ dz \\
& = \int_\Omega ([\lambda + \eta]\, \overline{u}_*^{p-1} + f(z, \overline{u}_*))(\overline{u}_* - u_\lambda)^+ dz \text{ (see (17.64))} \\
& \geq \int_\Omega ([\lambda^* + \eta]\overline{u}_*^{p-1} + f(z, \overline{u}_*))(\overline{u}_* - u_\lambda)^+ dz \\
& = \langle A(\overline{u}_*), (\overline{u}_* - u_\lambda)^+\rangle + \int_\Omega [\xi(z) + \eta]\overline{u}_*^{p-1}(\overline{u}_* - u_\lambda)^+ dz, \\
\Rightarrow \quad & \overline{u}_* \leq u_\lambda. \qquad (17.66)
\end{aligned}$$

From (17.64), (17.65), (17.66) we conclude $u_\lambda \in S(\lambda) \subseteq \operatorname{int} \widehat{C}_+$, a contradiction since $\lambda > \lambda^*$. Therefore $\lambda^* \notin \mathcal{L}$. ☐

So, finally we have $\mathcal{L} = (-\infty, \lambda^*)$ and summarizing the situation for problem (17.26), we can state the following theorem.

Theorem 17.4.19 *If hypotheses $H(a)'$, $H(\xi)$, $H(f)_1$ hold, then there exists $\lambda^* < +\infty$ such that*

(*a*) *for all $\lambda \geq \lambda^*$ problem* (17.26) *has no positive solution;*
(*b*) *for all $\lambda < \lambda^*$ problem* (17.26) *has at least one positive solution $u_\lambda \in \operatorname{int} \widehat{C}_+$;*
(*c*) *for every $\lambda < \lambda^*$ problem* (17.26) *has a smallest positive solution $\overline{u}_\lambda \in \operatorname{int} \widehat{C}_+$ and the map $\lambda \to \overline{u}_\lambda$ from $\mathcal{L} = (-\infty, \lambda^*)$ into $C^1(\overline{\Omega})$ is strictly increasing (that is, $\vartheta < \lambda < \lambda^* \Rightarrow \overline{u}_\lambda - \overline{u}_\vartheta \in \operatorname{int} \widehat{C}_+$) and left continuous.*

The particular case of the p-Laplacian (that is, $a(y) = |y|^{p-2}y$ for all $y \in \mathbb{R}^N$) presents special interest because we can identify precisely λ^*.

So, we consider the following problem:

$$\begin{cases} -\Delta_p u(z) + \xi(z)u(z)^{p-1} = \lambda u(z)^{p-1} + f(z, u(z)) \text{ in } \Omega, \\ \dfrac{\partial u}{\partial n} = 0 \text{ on } \partial\Omega,\ u > 0. \end{cases} \qquad (17.67)$$

Proposition 17.4.20 *If hypothesis $H(\xi)$ holds and $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function such that for a.a. $z \in \Omega$, $f(z,0) = 0$, $f(z,x) > 0$ for all $x > 0$, $f(z,x) \leq a(z)[1 + x^{p^*-1}]$ for a.a. $z \in \Omega$, all $x \geq 0$, with $a \in L^\infty(\Omega)_+$ and $\lambda \geq \widehat{\lambda}_1(p,\xi)$, then $S(\lambda) = \emptyset$.*

Proof Suppose that $S(\lambda) \neq \emptyset$. Let $u \in S(\lambda) \subseteq \operatorname{int} \widehat{C}_+$. Consider $\widehat{u}_1 = \widehat{u}_1(p,\xi) \in \operatorname{int} \widehat{C}_+$ and introduce the function

$$R(\widehat{u}_1, u)(z) = |\nabla \widehat{u}_1(z)|^p - |\nabla u(z)|^{p-2} \left(\nabla u(z), \nabla \left(\frac{\widehat{u}_1^p}{u^{p-1}} \right)(z) \right)_{\mathbb{R}^N}.$$

From the nonlinear Picone's identity of Allegretto and Huang [5] (see also Motreanu et al. [52], p. 255), we have

$$\begin{aligned}
& 0 \leq R(\widehat{u}_1, u)(z) \text{ for a.a. } z \in \Omega, \\
\Rightarrow \quad & 0 \leq \int_\Omega R(\widehat{u}_1, u)dz \\
& = \|\nabla \widehat{u}_1\|_p^p - \int_\Omega |\nabla u|^{p-2} \left(\nabla u, \nabla \left(\frac{\widehat{u}_1^p}{u^{p-1}} \right) \right)_{\mathbb{R}^N} dz \\
& = \|\nabla \widehat{u}_1\|_p^p - \int_\Omega (-\Delta_p u) \frac{\widehat{u}_1^p}{u^{p-1}} dz \\
& \qquad \text{(by the nonlinear Green's identity, see [31], p. 211)} \\
& = \|\nabla \widehat{u}_1\|_p^p + \int_\Omega \xi(z) \widehat{u}_1^p dz - \int_\Omega \left[\lambda \widehat{u}_1^p + f(z,u) \frac{\widehat{u}_1^p}{u^{p-1}} \right] dz \\
& < \mu(\widehat{u}_1) - \lambda \text{ (recall } \|\widehat{u}_1\|_p = 1) \\
& = \widehat{\lambda}_1 - \lambda \leq 0,
\end{aligned}$$

a contradiction. Therefore $S(\lambda) = \emptyset$ for all $\lambda \geq \widehat{\lambda}_1$. □

On the other hand, we already observed from the proof of Proposition 17.4.10 that for $\lambda < \widehat{\lambda}_1 = \widehat{\lambda}_1(p,\xi)$, we have $S(\lambda) \neq \emptyset$. So, for problem (17.67), we can say that $\mathcal{L} = (-\infty, \widehat{\lambda}_1)$. Note that the argument with the nonlinear Picone's identity can be used also to show that $\lambda^* \notin \mathcal{L}$.

17.5 Superlinear Problems

We continue with $\Omega \subseteq \mathbb{R}^N$ a bounded domain with a C^2-boundary $\partial\Omega$ and $\xi \in L^\infty(\Omega)$ an indefinite potential function. In this section we deal with the following nonlinear Robin problem

$$\begin{cases} -\text{div }\, a(\nabla u(z)) + \xi(z)|u(z)|^{p-2}u(z) = f(z, u(z)) \text{ in } \Omega, \\ \dfrac{\partial u}{\partial n} + \beta(z)|u|^{p-2}u = 0 \text{ on } \partial\Omega, \quad 1 < p < +\infty. \end{cases} \tag{17.68}$$

Again the differential operator $u \to \text{div }\, a(\nabla u)$ need not be homogeneous and the hypotheses on $a(\cdot)$ are the same and so we incorporate as special cases the p-Laplacian and the (p, q)-Laplacian. The result also holds for the Dirichlet problem. In problem (17.68) the reaction term $f(z, x)$ is a Carathéodory function (that is, for all $x \in \mathbb{R}$ $z \to f(z, x)$ is measurable and for a.a. $z \in \Omega$ $x \to f(z, x)$ is continuous) and we assume that $f(z, \cdot)$ exhibits $(p-1)$-superlinear growth near $\pm\infty$ but without satisfying the usual in such cases Ambrosetti-Rabinowitz condition (the AR-condition for short). Here instead we employ a less restrictive condition which involves the function $e(z, x) = f(z, x)x - pF(z, x)$ with $F(z, x) = \int_0^x f(z, s)ds$.

We recall that the AR-condition says that there exist $\tau > p$ and $M > 0$ such that

$$0 < \tau F(z, x) \leq f(z, x)x \text{ for a.a. } z \in \Omega, \text{ all } |x| \geq M \text{ and } \text{essinf}_\Omega F(\cdot, \pm M) > 0. \tag{17.69}$$

A direct integration of (17.69) leads to the following weaker condition

$$c_{22}|x|^\tau \leq F(z, x) \text{ for a.a. } z \in \Omega, \text{ all } |x| \geq M, \text{ some } c_{22} > 0. \tag{17.70}$$

It is clear from (17.69) and (17.70) that the AR-condition dictates a $(p-1)$-superlinear growth for $f(z, \cdot)$ near $\pm\infty$. The AR-condition, although very useful in verifying the compactness condition (the Palais-Smale (PS-) condition) for the energy functional of the problem, is somewhat restrictive and excludes from consideration of superlinear nonlinearities with "slower" growth near $\pm\infty$.

Here combining variational methods with suitable truncation and perturbation techniques and Morse theory (critical groups), we prove a multiplicity theorem producing three nontrivial smooth solutions.

We start by recalling some notions which we will use in the sequel. So, let X be a Banach space and let X^* be its topological dual. Given $\varphi \in C^1(X, \mathbb{R})$, we say that φ satisfies the "Cerami condition" (the "C-condition" for short), if the following property holds:

Every sequence $\{u_n\}_{n\geq 1} \subseteq X$ such that

$$|\varphi(u_n)| \leq M \text{ for some } M > 0, \text{ all } n \in \mathbb{N},$$
$$(1 + \|u_n\|)\varphi'(u_n) \to 0 \text{ in } X^* \text{ as } n \to +\infty,$$

admits a strongly convergent subsequence.

This compactness-type condition on the functional $\varphi(\cdot)$ is more general than the usual PS-condition. Nevertheless practically all the critical point theory remains valid if the PS-condition is replaced by the C-condition (see Gasiński and Papageorgiou [31]).

Given $\varphi \in C^1(X, \mathbb{R})$ and $c \in \mathbb{R}$, we introduce the following sets

$$K_\varphi = \{u \in X : \varphi'(u) = 0\},\ K_\varphi^c = \{u \in K_\varphi : \varphi(u) = c\},$$

$$\varphi^c = \{u \in X : \varphi(u) \le c\},\ \dot{\varphi}^c = \{u \in X : \varphi(u) < c\}.$$

If (Y_1, Y_2) is a topological pair such that $Y_2 \subseteq Y_1 \subseteq X$ and $k \in \mathbb{N}_0$, then by $H_k(Y_1, Y_2)$ we denote the k^{th}-relative singular homology group with integer coefficients for the pair (Y_1, Y_2). The critical groups of φ at an isolated $u_0 \in K_\varphi$ with $\varphi(u_0) = c$ (that is, $u_0 \in K_\varphi^c$) are defined by

$$C_k(\varphi, u_0) = H_k(\varphi^c \cap U, \varphi^c \cap U \setminus \{u_0\}) \text{ for all } k \in \mathbb{N}_0,$$

with U being a neighborhood of u_0 such that $K_\varphi \cap \varphi^c \cap U = \{u_0\}$. The excision property of singular homology implies that this definition is independent of the particular choice of the neighborhood U.

Suppose that $\varphi \in C^1(X, \mathbb{R})$ satisfies the C-condition and $\inf \varphi(K_\varphi) > -\infty$. Let $c < \inf \varphi(K_\varphi)$. The critical groups of φ at infinity are defined by

$$C_k(\varphi, \infty) = H_k(X, \varphi^c) \text{ for all } k \in \mathbb{N}_0.$$

The second deformation theorem (see, for example, Gasiński and Papageorgiou [31], p. 628) implies that this definition is independent of the choice of $c < \inf \varphi(K_\varphi)$.

Moreover, for every $c < \inf \varphi(K_\varphi)$, we have

$$C_k(\varphi, \infty) = H_k(X, \dot{\varphi}^c) \text{ for all } k \in \mathbb{N}_0. \tag{17.71}$$

Indeed if $\vartheta < c < \inf \varphi(K_\varphi)$, then from Granas and Dugundji [37, p. 407], we have that φ^ϑ is a strong deformation retract of $\dot{\varphi}^c$, hence

$$H_k(X, \varphi^\vartheta) = H_k(X, \dot{\varphi}^c) \text{ for all } k \in \mathbb{N}_0$$

(see Motreanu et al. [52, p. 145]), hence (17.71) follows.

With K_φ we define

$$M(t, u) = \sum_{k \ge 0} \operatorname{rank} C_k(\varphi, u) t^k \text{ for all } t \in \mathbb{R}, u \in K_\varphi,$$

$$P(t, \infty) = \sum_{k \ge 0} \operatorname{rank} C_k(\varphi, \infty) t^k \quad t \in \mathbb{R}.$$

Then the "Morse relation" says

$$\sum_{u\in K_\varphi} M(t,u) = P(t,\infty) + (1+t)Q(t) \text{ for all } t \in \mathbb{R}, \tag{17.72}$$

with $Q(t) = \sum_{k\geq 0} \beta_k t^k$ being a formal series with nonnegative integer coefficients.

The hypotheses on the reaction $f(z,x)$ are the following:

$H(f)_2$ $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function such that $f(z,0) = 0$ for a.a. $z \in \Omega$ and

- (i) $|f(z,x)| \leq a(z)[1 + |x|^{r-1}]$ for a.a. $z \in \Omega$, all $x \in \mathbb{R}$, with $a \in L^\infty(\Omega)$, $p < r < p^*$;
- (ii) if $F(z,x) = \int_0^x f(z,s)ds$, then $\lim_{x\to\pm\infty} \frac{F(z,x)}{|x|^p} = +\infty$ uniformly for a.a. $z \in \Omega$;
- (iii) if $e(z,x) = f(z,x)x - pF(z,x)$, then $e(z,x) \leq e(z,y) + \ell(z)$ for a.a. $z \in \Omega$, all $0 \leq x \leq y$ or $y \leq x \leq 0$, with $\ell \in L^1(\Omega)$;
- (iv) there exist a function $\vartheta \in L^\infty(\Omega)$ and $\gamma > 0$ such that

$$\vartheta(z) \leq \frac{c_1}{p-1}\widehat{\lambda}_1(p,\xi_0,\beta_0) \text{ for a.a. } z \in \Omega,\ \vartheta \not\equiv \widehat{\lambda}_1(p,\xi_0,\beta_0),$$

where

$$\xi_0 = \frac{p-1}{c_1}\xi,\ \beta_0 = \frac{p-1}{c_1}\beta,$$

$$\limsup_{x\to 0} \frac{pF(z,x)}{|x|^p} \leq \vartheta(z) \text{ uniformly for a.a. } z \in \Omega,$$

$$\liminf_{x\to 0} \frac{f(z,x)}{|x|^{p-2}x} > -\gamma \text{ uniformly for a.a. } z \in \Omega.$$

Remark 17.5.1 Evidently hypotheses $H(f)_2$ (ii), (iii) imply that

$$\lim_{x\to\pm\infty} \frac{f(z,x)}{|x|^{p-2}x} = +\infty \text{ uniformly for a.a. } z \in \Omega,$$

that is, the reaction $f(z,\cdot)$ is $(p-1)$-superlinear. Hypothesis $H(f)_2$ (iii) is a quasi-monotonicity condition on e and it is satisfied if there exists $M > 0$ such that for a.a. $z \in \Omega$

$$x \to \frac{f(z,x)}{x^{p-1}} \text{ is increasing on } x \geq M \text{ and } x \to \frac{f(z,x)}{|x|^{p-2}x} \text{ is decreasing on } x \leq -M.$$

In what follows $\widetilde{\mu} : W^{1,p}(\Omega) \to \mathbb{R}$ is the C^1-functional defined by

$$\widetilde{\mu}(u) = \int_\Omega pG(\nabla(u))dz + \int_\Omega \xi(z)|u|^p dz + \int_{\partial\Omega} \beta(z)|u|^p d\sigma \text{ for all } u \in W^{1,p}(\Omega).$$

Also let $f_\pm(z,x)$ be the positive (resp. negative) truncations of $f(z,\cdot)$, that is, $f_\pm(z,x) = f(z,x^\pm)$ (recall $x^+ = \max\{x,0\}$, $x^- = \max\{-x,0\}$). These are Carathéodory functions. We set $F_\pm(z,x) = \int_0^x f_\pm(z,s)ds$ and consider the C^1-functionals $\varphi_\pm : W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\varphi_\pm(u) = \frac{1}{p}\widetilde{\mu}(u) + \frac{\eta}{p}\|u^\mp\|_p^p - \int_\Omega F_\pm(z,u)dz \text{ for all } u \in W^{1,p}(\Omega) \quad (\eta > \|\xi\|_\infty).$$

In addition we consider the energy functional $\varphi : W^{1,p}(\Omega) \to \mathbb{R}$ for problem (17.68) defined by

$$\varphi(u) = \frac{1}{p}\widetilde{\mu}(u) - \int_\Omega F(z,u)dz \text{ for all } u \in W^{1,p}(\Omega).$$

Then $\varphi \in C^1(W^{1,p}(\Omega))$.

Proposition 17.5.2 *If hypotheses $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_2$ hold, then the functionals $\varphi_\pm$ satisfy the C-condition.*

Proof We do the proof for φ_+, the proof for φ_- being similar. So, we consider a sequence $\{u_n\}_{n\geq 1} \subseteq W^{1,p}(\Omega)$ such that

$$|\varphi_+(u_n)| \leq M_1 \text{ for some } M_1 > 0, \text{ all } n \in \mathbb{N}, \tag{17.73}$$

$$(1 + \|u_n\|)\varphi_+'(u_n) \to 0 \text{ in } W^{1,p}(\Omega)^* \text{ as } n \to +\infty. \tag{17.74}$$

From (17.74) we have

$$\Big|\langle A(u_n), h\rangle + \int_\Omega \xi(z)(u_n^+)^{p-1}hdz + \int_\Omega [\xi(z)+\eta](u_n^-)^{p-1}hdz + \int_{\partial\Omega} \beta(z)|u_n|^{p-2}u_n hd\sigma - \int_\Omega f_+(z,u_n)hdz\Big| \leq \frac{\varepsilon_n\|h\|}{1+\|u_n\|} \tag{17.75}$$
$$\text{for all } h \in W^{1,p}(\Omega) \text{ with } \varepsilon_n \to 0^+.$$

In (17.75) we choose $h = -u_n^- \in W^{1,p}(\Omega)$. Then

$$\frac{c_1}{p-1}\|\nabla u_n^-\|_p^p + \int_\Omega [\xi(z)+\eta](u_n^-)^p dz \leq \varepsilon_n \text{ for all } n \in \mathbb{N},$$
$$\Rightarrow \quad u_n^- \to 0 \text{ in } W^{1,p}(\Omega) \text{ as } n \to +\infty. \tag{17.76}$$

Now in (17.75) we choose $h = u_n^+ \in W^{1,p}(\Omega)$. Then

$$-\int_\Omega \left(a\left(\nabla u_n^+\right), \nabla u_n^+\right)_{\mathbb{R}^N} dz - \int_\Omega \xi(z)\left(u_n^+\right)^p dz - \int_{\partial\Omega} \beta(z)\left(u_n^+\right)^p d\sigma$$
$$+ \int_\Omega f\left(z, u_n^+\right) u_n^+ dz \le \varepsilon_n \text{ for all } n \in \mathbb{N}. \tag{17.77}$$

Also from (17.73) and (17.76), we have

$$\int_\Omega pG\left(\nabla u_n^+\right) dz + \int_\Omega \xi(z)\left(u_n^+\right)^p dz + \int_{\partial\Omega} \beta(z)\left(u_n^+\right)^p d\sigma - \int_\Omega pF\left(z, u_n^+\right) dz$$
$$\le M_2 \text{ for some } M_2 > 0, \text{ all } n \in \mathbb{N}. \tag{17.78}$$

Adding (17.77), (17.78) and using hypothesis $H(a)'(iv)$, we obtain

$$\int_\Omega e(z, u_n^+) dz \le M_3 \text{ for some } M_3 > 0, \text{ all } n \in \mathbb{N}. \tag{17.79}$$

Claim $\{u_n^+\}_{n\ge1} \subseteq W^{1,p}(\Omega)$ is bounded.

Arguing by contradiction, suppose that the claim is not true. Then we may assume that $\|u_n^+\| \to +\infty$. We set $y_n = \dfrac{u_n^+}{\|u_n^+\|}$, $n \in \mathbb{N}$. Then $\|y_n\| = 1$ for all $n \in \mathbb{N}$ and so we may assume that

$$y_n \xrightarrow{w} y \text{ in } W^{1,p}(\Omega) \text{ and } y_n \to y \text{ in } L^r(\Omega) \text{ and in } L^p(\partial\Omega). \tag{17.80}$$

First we assume that $y \neq 0$. So, if $\Omega_0 = \{z \in \Omega : y(z) = 0\}$, then $|\Omega \setminus \Omega_0|_N > 0$ (by $|\cdot|_N$ we denote the Lebesgue measure on $\mathbb{R}^N$). We have $u_n^+(z) \to +\infty$ for a.a. $z \in \Omega \setminus \Omega_0$. Hence by hypothesis $H(f)_2(ii)$ implies that

$$\frac{F\left(z, u_n^+(z)\right)}{\|u_n^+\|^p} = \frac{F\left(z, u_n^+(z)\right)}{(u_n^+(z))^p} y_n(z)^p \to +\infty \text{ for a.a. } z \in \Omega \setminus \Omega_0,$$
$$\Rightarrow \int_\Omega \frac{F\left(z, u_n^+\right)}{\|u_n^+\|^p} dz \to +\infty \text{ as } n \to +\infty \text{ (by Fatou's lemma).} \tag{17.81}$$

From Corollary 17.2.2 and hypothesis $H(a)'(v)$, we have

$$G(y) \le c_{23}\left[|y|^q + |y|^p\right] \text{ for some } c_{23} > 0, \text{ all } y \in \mathbb{R}^N.$$

From (17.73) and (17.76) it follows that

$$-\int_\Omega G(\nabla u_n^+)dz - \frac{1}{p}\int_\Omega \xi(z)(u_n^+)^p dz - \frac{1}{p}\int_{\partial\Omega} \beta(z)(u_n^+)^p d\sigma$$
$$+\int_\Omega F(z, u_n^+)dz \le M_4 \text{ for some } M_4 > 0, \text{ all } n \in \mathbb{N},$$
$$\Rightarrow \quad \int_\Omega \frac{F(z, u_n^+)}{\|u_n^+\|^p} dz \le M_5 \text{ for some } M_5 > 0, \text{ all } n \in \mathbb{N} \text{ (see (17.80))}. \tag{17.82}$$

Comparing (17.81) and (17.82) we reach a contradiction.

Next assume that $y = 0$. Consider the C^1-functional $\widehat{\varphi}_+ : W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\widehat{\varphi}_+(u) = \frac{c_1}{p(p-1)}\|\nabla u\|_p^p + \frac{1}{p}\int_\Omega [\xi(z) + \eta]|u|^p dz + \frac{1}{p}\int_{\partial\Omega} \beta(z)|u|^p d\sigma$$
$$-\int_\Omega F(z, u^+)dz - \frac{\eta}{p}\|u^+\|_p^p \text{ for all } u \in W^{1,p}(\Omega).$$

Let $k_n(t) = \widehat{\varphi}_+(tu_n^+)$ for all $t \in [0, 1]$, all $n \in \mathbb{N}$. We can find $t_n \in [0, 1]$ such that

$$k_n(t_n) = \max\{k_n(t) : 0 \le t \le 1\}. \tag{17.83}$$

For $\lambda > 0$, let $v_n = (2\lambda)^{\frac{1}{p}} y_n \in W^{1,p}(\Omega)$. Then $v_n \to 0$ in $L^r(\Omega)$ (see (17.80)). Therefore

$$\int_\Omega F(z, v_n)dz \to 0 \text{ as } n \to +\infty. \tag{17.84}$$

Recall that $\|u_n^+\| \to +\infty$. So, we can find $n_0 \in \mathbb{N}$ such that

$$(2\lambda)^{\frac{1}{p}} \frac{1}{\|u_n^+\|} \in (0, 1) \text{ for all } n \ge n_0. \tag{17.85}$$

Then from (17.83) and (17.85) we have

$$k_n(t_n) \ge k_n\left(\frac{(2\lambda)^{\frac{1}{p}}}{\|u_n^+\|}\right) \text{ for all } n \ge n_0,$$
$$\Rightarrow \quad \widehat{\varphi}_+(t_n u_n^+) \ge \widehat{\varphi}_+((2\lambda)^{\frac{1}{p}} y_n) = \widehat{\varphi}_+(v_n)$$

$$\geq \frac{2\lambda c_1}{p(p-1)}\left[\|\nabla y_n\|_p^p + \int_\Omega [\widetilde{\xi}(z)+\eta] y_n^p dz\right] - \int_\Omega F(z, v_n)dz - \frac{\eta}{p}\|v_n\|_p^p$$

$$\geq \frac{2\lambda c_{24}}{p(p-1)} - \left[\int_\Omega F(z, v_n)dz + \frac{\eta}{p}\|v_n\|_p^p\right] \text{ for some } c_{24} > 0$$

$$\text{(recall } \eta > \|\xi\|_\infty,\ \|y_n\| = 1)$$

$$\geq \frac{\lambda c_{24}}{p(p-1)} \text{ for all } n \geq n_1 \geq n_0$$

$$\text{(see (17.84) and recall } v_n \to 0 \text{ in } L^r(\Omega), r > p).$$

But recall that $\lambda > 0$ is arbitrary. Therefore from the last inequality, we infer that

$$\widehat{\varphi}_+(t_n u_n^+) \to +\infty \text{ as } n \to +\infty. \tag{17.86}$$

We have $0 \leq t_n u_n^+ \leq u_n^+$ for all $n \in \mathbb{N}$. Hence hypothesis $H(f)_2(iii)$ implies that

$$\int_\Omega e(z, t_n u_n^+)dz \leq \int_\Omega e(z, u_n^+)dz + \|\ell\|_1 \leq M_6 \tag{17.87}$$

$$\text{for some } M_6 > 0,\ \text{all } n \in \mathbb{N} \text{ (see (17.79)).}$$

We know that

$$\widehat{\varphi}_+(0) = 0 \text{ and } \widehat{\varphi}_+(u_n^+) \leq M_7 \text{ for some } M_7 > 0,\ \text{all } n \in \mathbb{N} \text{ (see (17.73), (17.76)).}$$

Therefore $t_n \in (0, 1)$ for all $n \geq n_2$ (see (17.86)). Hence

$$0 = t_n \frac{d}{dt}\widehat{\varphi}_+(t u_n)\Big|_{t=t_n} = \langle \widehat{\varphi}'_+(t_n u_n), t_n u_n \rangle. \tag{17.88}$$

From (17.87) and (17.88) it follows that

$$p\widehat{\varphi}_+(t_n u_n) \leq M_6 \text{ for all } n \in \mathbb{N}. \tag{17.89}$$

Comparing (17.86) and (17.89), we have a contradiction. This proves the claim.

From (17.76) and the claim it follows that $\{u_n\}_{n\geq 1} \subseteq W^{1,p}(\Omega)$ is bounded. So, we may assume that

$$u_n \xrightarrow{w} u \text{ in } W^{1,p}(\Omega) \text{ and } u_n \to u \text{ in } L^r(\Omega) \text{ and in } L^p(\partial\Omega). \tag{17.90}$$

If in (17.75) we choose $h = u_n - u \in W^{1,p}(\Omega)$, pass to the limit as $n \to +\infty$ and use (17.90), then

$$\lim_{n\to+\infty} \langle A(u_n), u_n - u\rangle = 0,$$
$$\Rightarrow \quad u_n \to u \text{ in } W^{1,p}(\Omega) \text{ (see Proposition 17.2.7).}$$

Therefore φ_+ satisfies the C-condition. Similarly for φ_-. □

Minor changes in the above proof lead to the following result.

Proposition 17.5.3 *If hypotheses $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_2$ hold, then the functional φ satisfies the C-condition.*

Proposition 17.5.4 *If hypotheses $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_2$ hold, then $u = 0$ is a local minimizer of $\varphi_\pm$ and φ.*

Proof We do the proof for the functional φ_+, the proofs for φ_- and φ being similar. Hypothesis $H(f)_2(iv)$ implies that given $\varepsilon > 0$, we can find $\delta > 0$ such that

$$F(z,x) \le \frac{1}{p}[\vartheta(z) + \varepsilon]|x|^p \text{ for a.a } z \in \Omega, \text{ all } |x| \le \delta. \tag{17.91}$$

Consider $u \in C^1(\overline{\Omega})$ with $\|u\|_{C^1(\overline{\Omega})} \le \delta$. We have

$$\begin{aligned}
\varphi_+(u) \ge\ & \frac{c_1}{p(p-1)}\left[\|\nabla u\|_p^p + \int_\Omega \xi_0(z)|u|^p dz + \int_\Omega \beta_0(z)|u|^p d\sigma\right] \\
& + \frac{\eta}{p}\|u^-\|_p^p - \frac{1}{p}\int_\Omega \vartheta(z)\left(u^+\right)^p dz - \frac{\varepsilon}{p}\|u\|^p \text{ (see Corollary 17.2.2 and (17.91))} \\
\ge\ & \frac{c_1}{p(p-1)}\Bigg[\|\nabla u^+\|_p^p + \int_\Omega \xi_0(z)\left(u^+\right)^p dz + \int_{\partial\Omega} \beta_0(z)\left(u^+\right)^p d\sigma \\
& - \int_\Omega \vartheta_0(z)\left(u^+\right)^p dz\Bigg] \\
& + \frac{c_1}{p(p-1)}\|\nabla u^-\|_p^p + \frac{1}{p}\int_\Omega [\xi(z) + \eta](u^-)^p dz - \frac{\varepsilon}{p}\|u\|^p \\
\ge\ & \frac{1}{p}[c_{25} - \varepsilon]\|u\|^p \text{ for some } c_{25} > 0 \text{ (see Proposition 17.4.6).}
\end{aligned}$$

Choosing $\varepsilon \in (0, c_{25})$ we infer that

$$\begin{aligned}
& \varphi_+(u) \ge 0 = \varphi_+(0) \text{ for all } u \in C^1(\overline{\Omega}) \text{ with } \|u\|_{C^1(\overline{\Omega})} \le \delta, \\
\Rightarrow\quad & u = 0 \text{ is a local } C^1(\overline{\Omega})\text{-minimizer of } \varphi_+, \\
\Rightarrow\quad & u = 0 \text{ is a local } W^{1,p}(\Omega)\text{-minimizer of } \varphi_+ \text{ (see Proposition 17.2.5).}
\end{aligned}$$

Similarly for φ_- and φ. $\square$

Now we are ready to produce solutions of constant sign.

Proposition 17.5.5 *If hypotheses* $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_2$ *hold, then problem* (17.68) *admits two nontrivial constant sign solutions* $u_0 \in \operatorname{int}\widehat{C}_+$ *and* $v_0 \in -\operatorname{int}\widehat{C}_+$.

Proof We easily show that $K_{\varphi_+} \subseteq C_+$. So, we can assume that K_{φ_+} is finite. Then Proposition 17.5.4 implies that we can find $\rho \in (0, 1)$ small such that

$$\varphi_+(0) = 0 < \inf[\varphi_+(u) : \|u\| = \rho] = m_\rho^+. \tag{17.92}$$

On account of hypothesis $H(f)_2(ii)$, given $\widetilde{u} \in \operatorname{int}\widehat{C}_+$, we have

$$\varphi_+(t\widetilde{u}) \to -\infty \text{ as } t \to +\infty. \tag{17.93}$$

From Proposition 17.5.2 we know that

$$\varphi_+ \text{ satisfies the C-condition.} \tag{17.94}$$

Then (17.92), (17.93), (17.94) permit the use of the mountain pass theorem. So, we can find $u_0 \in W^{1,p}(\Omega)$ such that

$$u_0 \in K_{\varphi_+} \subseteq C_+,\ m_\rho^+ \leq \varphi_+(u_0),\ \text{ hence } u_0 \neq 0.$$

Therefore $u_0 \in C_+ \setminus \{0\}$. Evidently hypotheses $H(f)_2$ imply that given $\rho_0 > 0$, we can find $\widehat{\xi}_0 > 0$ such that $f(z, x) + \widehat{\xi}_0 x^{p-1} \geq 0$ for a.a. $z \in \Omega$. Hence

$$\begin{aligned} & \operatorname{div} a(\nabla u_0(z)) \leq [\|\xi\|_\infty + \widehat{\xi}_0] u_0(z)^{p-1} \text{ for a.a. } z \in \Omega, \\ \Rightarrow \quad & u_0 \in \operatorname{int}\widehat{C}_+ \text{ (see Theorem 17.3.1).} \end{aligned}$$

Similarly working this time with φ_-, we produce a negative solution $v_0 \in -\operatorname{int}\widehat{C}_+$ for problem (17.68). $\square$

To produce a third solution for problem (17.68), we will use critical groups.

Proposition 17.5.6 *If hypotheses* $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_2$ *hold, then* $C_k(\varphi, \infty) = 0$ *for all* $k \in \mathbb{N}_0$.

Proof Given $u \in W^{1,p}(\Omega)$, $u \neq 0$, hypothesis $H(f)_2(ii)$ implies that

$$\varphi(tu) \to -\infty \text{ as } t \to +\infty. \tag{17.95}$$

Moreover, hypothesis $H(f)_2(iii)$ implies that

$$0 = e(z, 0) \leq e(z, u^+(z)) + \ell(z) \text{ and } 0 = e(z, 0) \leq e(z, -u^-(z)) + \ell(z) \text{ for a.a. } z \in \Omega.$$

It follows that

$$\begin{aligned} & 0 = e(z,0) \le e(z,u(z)) + \ell(z) \text{ for a.a. } z \in \Omega, \\ \Rightarrow \quad & pF(z,u(z)) - f(z,u(z))u(z) \le \ell(z) \text{ for a.a. } z \in \Omega. \end{aligned} \tag{17.96}$$

For $t > 0$, we have

$$\begin{aligned} \frac{d}{dt}\varphi(tu) &= \langle \varphi'(tu), u\rangle = \frac{1}{t}\langle \varphi'(tu), tu\rangle \\ &= \frac{1}{t}\Big[\int_\Omega (a(\nabla(tu)), \nabla(tu))_{\mathbb{R}^N} dz + \int_\Omega \xi(z)|tu|^p dz + \int_{\partial\Omega} \beta(z)|tu|^p d\sigma \\ &\quad - \int_\Omega f(z,tu)(tu)dz\Big] \\ &\le \frac{1}{t}\Big[\int_\Omega pG(\nabla(tu))dz + \int_\Omega \xi(z)|tu|^p dz + \int_{\partial\Omega} \beta(z)|tu|^p d\sigma \\ &\quad - \int_\Omega pF(z,tu) + \|\ell\|_1\Big] \\ &\qquad\qquad \text{(see hypothesis } H(a)'(iv) \text{ and (17.96))} \\ &= \frac{1}{t}\left[p\varphi(tu) + \|\ell\|_1\right]. \end{aligned}$$

On account of (17.95) we see that for $t > 0$ big, we have

$$\begin{aligned} & \varphi(tu) \le \tau_0 < -\frac{\|\ell\|_1}{p}, \\ \Rightarrow \quad & \frac{d}{dt}\varphi(tu) < 0 \text{ for } t > 0 \text{ big.} \end{aligned} \tag{17.97}$$

Let $\partial B_1 = \{u \in W^{1,p}(\Omega) \colon \|u\| = 1\}$. Then for $u \in \partial B_1$ we can find a unique $t_0(u) > 0$ such that $\varphi(t_0(u)u) = \tau_0$. On account of the implicit function theorem (it can be used thanks to (17.97)), we have $t_0 \in C(\partial B_1)$. Extend t_0 on $W^{1,p}(\Omega) \setminus \{0\}$ by setting

$$\widehat{t_0}(u) = \frac{1}{\|u\|} t_0\left(\frac{u}{\|u\|}\right) \text{ for all } u \in W^{1,p}(\Omega) \setminus \{0\}.$$

Then $\widehat{t_0} \in C(W^{1,p}(\Omega) \setminus \{0\})$ and $\varphi(\widehat{t_0}(u)u) = \tau_0$. Moreover, if $\varphi(u) = \tau_0$, then $\widehat{t_0}(u) = 1$. So, we set

$$\widehat{s}(u) = \begin{cases} 1 & \text{if } \varphi(u) < \tau_0, \\ t_0(u) & \text{if } \tau_0 \le \varphi(u). \end{cases} \tag{17.98}$$

Evidently $\widehat{s} \in C(W^{1,p}(\Omega) \setminus \{0\})$. Consider the deformation $h\colon [0,1] \times (W^{1,p}(\Omega) \setminus \{0\}) \to W^{1,p}(\Omega) \setminus \{0\}$ defined by

$$h(t,u) = (1-t)u + t\widehat{s}(u)u.$$

We have

$$h(0,u) = u,\ h(1,u) = \widehat{s}(u)u \in \varphi^{\tau_0} \text{ for all } u \in W^{1,p}(\Omega) \setminus \{0\}$$

and $h(t,\cdot)\big|_{\varphi^{\tau_0}} = \mathrm{id}\,\big|_{\varphi^{\tau_0}}$ (see (17.98)).

From these facts we infer that

$$\varphi^{\tau_0} \text{ is a strong deformation retract of } W^{1,p}(\Omega) \setminus \{0\}. \tag{17.99}$$

Via the radial retraction we see that $W^{1,p}(\Omega) \setminus \{0\}$ is deformable into ∂B_1. Hence by Theorem 6.5, p. 325, of Dugundji [24] we have

$$\partial B_1 \text{ is a deformation retract of } W^{1,p}(\Omega) \setminus \{0\}. \tag{17.100}$$

Then (17.99) and (17.100) imply that

$$\varphi^{\tau_0} \text{ and } \partial B_1 \text{ are homotopy equivalent,}$$
$$\Rightarrow \quad H_k(W^{1,p}(\Omega), \varphi^{\tau_0}) = H_k(W^{1,p}(\Omega), \partial B_1) \text{ for all } k \in \mathbb{N}_0. \tag{17.101}$$

Since $W^{1,p}(\Omega)$ is infinite dimensional, from Gasiński and Papageorgiou [34] (Problems 4.154, 4.159, pp. 677–678), we have

$$H_k(W^{1,p}(\Omega), \partial B_1) = 0 \text{ for all } k \in \mathbb{N}_0,$$
$$\Rightarrow \quad H_k(W^{1,p}(\Omega), \varphi^{\tau_0}) = 0 \text{ for all } k \in \mathbb{N}_0 \text{ (see (17.101))},$$
$$\Rightarrow \quad C_k(\varphi, \infty) = 0 \text{ for all } k \in \mathbb{N}_0.$$

□

We can also compute the critical groups at infinity for the functional $\varphi_\pm$.

Proposition 17.5.7 *If hypotheses* $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_2$ *hold, then* $C_k(\varphi_\pm, \infty) = 0$ *for all* $k \in \mathbb{N}_0$.

Proof Let $\widehat{\varphi}_+ = \varphi_+\big|_{C^1(\overline{\Omega})}$. We have $K_{\varphi_+} \subseteq C_+$ (Proposition 17.2.4). Hence $K_{\widehat{\varphi}_+} = K_{\varphi_+} = K_\varphi \subseteq C_+$. Recall that $C^1(\overline{\Omega})$ is dense in $W^{1,p}(\Omega)$ and so

$$H_k(W^{1,p}(\Omega), \dot{\varphi}_+^c) = H_k(C^1(\overline{\Omega}), \dot{\widehat{\varphi}}_+^c)$$

for $c < \inf_k \widehat{\varphi}_+ = \inf_k \varphi_+$. Therefore

$$C_k(\varphi_+, \infty) = C_k(\widehat{\varphi}_+, \infty) \text{ for all } k \in \mathbb{N}_0 \text{ (see (17.71))}. \tag{17.102}$$

Based on (17.102) we see that it suffices to show that

$$C_k(\varphi_+, \infty) = H_k\left(C^1(\overline{\Omega}), \widehat{\varphi}_+^c\right) = 0 \text{ for all } k \in \mathbb{N}_0.$$

To this end we introduce the following sets

$$\partial B_1^C = \left\{u \in C^1(\overline{\Omega}) : \|u\|_{C^1(\overline{\Omega})} = 1\right\} \text{ and } \partial B_{1,+}^C = \{u \in \partial B_1^C : u^+ \neq 0\}.$$

We consider the deformation $h_+ : [0, 1] \times \partial B_{1,+}^C \to \partial B_{1,+}^C$ defined by

$$h_+(t, u) = \frac{(1-t)u + t\widehat{u}_1}{\|(1-t)u + t\widehat{u}_1\|_{C^1(\overline{\Omega})}} \text{ for all } (t, u) \in [0, 1] \times \partial B_{1,+}^C,$$

with $\widehat{u}_1 = \widehat{u}_1(p, \xi, \beta) \in \operatorname{int} \widehat{C}_+$. Note that

$$h_+(1, u) = \frac{\widehat{u}_1}{\|\widehat{u}_1\|_{C^1(\overline{\Omega})}}$$

and so we see that $\partial B_{1,+}^C$ is contractible in itself.

As before, hypothesis $H(f)_2(ii)$ implies that for all $u \in \partial B_{1,+}^C$ we have

$$\widehat{\varphi}_+(tu) \to -\infty \text{ as } t \to +\infty. \tag{17.103}$$

Hypothesis $H(f)_2(iii)$ implies that

$$-f_+(z, u(z))u(z) \leq \ell(z) - pF_+(z, u(z)) \text{ for a.a. } z \in \Omega. \tag{17.104}$$

In what follows by $\langle \cdot, \cdot \rangle_C$ we denote the duality brackets for the pair $(C^1(\overline{\Omega})^*, C^1(\overline{\Omega}))$. Then for $u \in \partial B_{1,+}^C$ and $t > 0$, we have

$$\begin{aligned}
\frac{d}{dt}\widehat{\varphi}_+(tu) &= \langle \widehat{\varphi}_+'(tu), u\rangle_C \quad \text{(by the chain rule)} \\
&= \frac{1}{t}\langle \widehat{\varphi}_+'(tu), tu\rangle \\
&= \frac{1}{t}\Big[\langle A(tu), tu\rangle + \int_\Omega \xi(z)|tu|^p dz + \int_{\partial\Omega} \beta(z)|tu|^p d\sigma
\end{aligned}$$

$$- \eta \|tu^-\|_p^p - \int_\Omega f_+(z, tu)(tu)dz\Big]$$

$$\leq \frac{1}{t}\Big[\int_\Omega pG(tu)dz + \int_\Omega \xi(z)|tu|^p dz + \int_{\partial\Omega} \beta(z)|tu|^p d\sigma$$

$$- \eta \|tu^-\|_p^p - \int_\Omega pF_+(z, tu)dz + \|\ell\|_1\Big]$$

(see hypothesis $H(a)'(iv)$ and (17.104))

$$= \frac{1}{t}\left[p\varphi_+(tu) + \|\ell\|_1\right]. \tag{17.105}$$

From (17.103) and (17.105) it follows that

$$\frac{d}{dt}\widehat{\varphi}_+(tu) < 0 \text{ for } t > 0 \text{ big.} \tag{17.106}$$

Let $\overline{B_1^C} = \{u \in C^1(\overline{\Omega}) : \|u\|_{C^1(\overline{\Omega})} \leq 1\}$ and choose $\lambda \in \mathbb{R}$ such that

$$\lambda < \min\left\{-\frac{1}{p}\|\ell\|_1, \min_{\overline{B_1^C}} \widehat{\varphi}_+\right\}. \tag{17.107}$$

From (17.106) we infer that there exists unique $\tau(u) \geq 1$ such that

$$\widehat{\varphi}_+(tu) = \begin{cases} > \lambda & \text{if } t \in [0, \tau(u)), \\ = \lambda & \text{if } t = \tau(u), \\ < \lambda & \text{if } \tau(u) < t. \end{cases} \tag{17.108}$$

By the implicit function theorem $\tau \in C(\partial B_{1,+}^C)$. Also we have

$$\widehat{\varphi}_+^\lambda = \left\{tu : u \in \partial B_{1,+}^C,\ t \geq \tau(u)\right\} \quad \text{(see (17.107), (17.108))}. \tag{17.109}$$

We set

$$E_+ = \left\{tu : u \in \partial B_{1,+}^C,\ t \geq 1\right\}.$$

It is clear from (17.109) that $\widehat{\varphi}_+^\lambda \subseteq E_+$. Consider the deformation $\widehat{h}_+ : [0, 1] \times E_+ \to E_+$ defined by

$$\widehat{h}_+(s, tu) = \begin{cases} (1-s)tu + s\tau(u)u & \text{if } t \in [1, \tau(u)), \\ tu & \text{if } \tau(u) < t, \end{cases}$$

for all $s \in [0, 1]$, all $t \geq 1$ and all $u \in \partial B^C_{1,+}$. We have

$$\widehat{h}_+(0, tu) = tu,\ \widehat{h}_+(1, tu) = \tau(u)u \in \widehat{\varphi}^{\lambda}_+ \text{ (see (17.108))}$$

and

$$\widehat{h}_+(s, \cdot)\Big|_{\widehat{\varphi}^{\lambda}_+} = \mathrm{id}\Big|_{\widehat{\varphi}^{\lambda}_+} \text{ for all } s \in [0, 1].$$

These properties imply that $\widehat{\varphi}^{\lambda}_+$ is a strong deformation retract of E_+. This implies that

$$H_k\left(C^1(\overline{\Omega}), E_+\right) = H_k\left(C^1(\overline{\Omega}), \widehat{\varphi}^{\lambda}_+\right) \text{ for all } k \in \mathbb{N}_0 \tag{17.110}$$

(see Motreanu et al. [52, p. 145]). Next consider the deformation $\widetilde{h}_+ : [0, 1]\times E_+ \to E_+$ defined by

$$\widetilde{h}_+(s, tu) = (1-s)tu + s\frac{tu}{\|tu\|_{C^1(\overline{\Omega})}} \text{ for all } s \in [0, 1],\ \text{ all } t \geq 1 \text{ and all } u \in \partial B^C_{1,+}.$$

As before, using Theorem 6.5, p. 325 of Dugundji [24], we have that $\partial B^C_{1,+}$ is a deformation retract of E_+. Therefore

$$H_k(C^1(\overline{\Omega}), E_+) = H_k(C^1(\overline{\Omega}), \partial B^C_{1,+}) \text{ for all } k \in \mathbb{N}_0. \tag{17.111}$$

From (17.110) and (17.111) it follows that

$$H_k(C^1(\overline{\Omega}), \widehat{\varphi}^{\lambda}_+) = H_k(C^1(\overline{\Omega}), \partial B^C_{1,+}) \text{ for all } k \in \mathbb{N}_0. \tag{17.112}$$

But we have seen that $\partial B^C_{1,+}$ is contractible in itself. Hence

$$H_k(C^1(\overline{\Omega}), \partial B^C_{1,+}) = 0 \text{ for all } k \in \mathbb{N}_0$$

(see Motreanu et al. [52, p.147]). Then

$$\begin{aligned} & H_k(C^1(\overline{\Omega}), \widehat{\varphi}^{\lambda}_+) = 0 \text{ for all } k \in \mathbb{N}_0 \text{ (see (17.112))}, \\ \Rightarrow\quad & C_k(\widehat{\varphi}_+, \infty) = 0 \text{ for all } k \in \mathbb{N}_0, \\ \Rightarrow\quad & C_k(\varphi_+, \infty) = 0 \text{ for all } k \in \mathbb{N}_0 \text{ (see (17.102))}. \end{aligned}$$

In a similar fashion we also show that $C_k(\varphi_-, \infty) = 0$ for all $k \in \mathbb{N}_0$. □

This proposition permits the exact computation of the critical groups of φ at the two constant sign solutions $u_0 \in \mathrm{int}\,\widehat{C}_+$ and $v_0 \in -\mathrm{int}\,\widehat{C}_+$ (see Proposition 17.5.5).

Proposition 17.5.8 *If hypotheses* $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_2$ *hold and* $K_\varphi = \{0, u_0, v_0\}$, *then* $C_k(\varphi, u_0) = C_k(\varphi, v_0) = \delta_{k,1}\mathbb{Z}$ *for all* $k \in \mathbb{N}_0$.

Proof We have $K_{\varphi_+} = \{0, u_0\}$. Let $\tau < 0 < \lambda < m_\rho^+$ (see (17.92)) and consider the following triple of sets

$$\varphi_+^\tau \subseteq \varphi_+^\lambda \subseteq W^{1,p}(\Omega).$$

We consider the corresponding long exact sequence of singular homology groups (see Motreanu et al. [52, p. 151]). So, we have

$$\cdots \to H_k(W^{1,p}(\Omega), \varphi_+^\tau) \xrightarrow{i_*} H_k(W^{1,p}(\Omega), \varphi_+^\lambda) \xrightarrow{\widehat{\partial}_*} H_{k-1}(\varphi_+^\lambda, \varphi_+^\tau) \to \cdots . \tag{17.113}$$

From (17.113) we have

$$\begin{aligned} \operatorname{rank} H_k(W^{1,p}(\Omega), \varphi_+^\lambda) &= \operatorname{rank} \ \operatorname{im} \widehat{\partial}_* + \operatorname{rank} \ \ker \widehat{\partial}_* \\ &= \operatorname{rank} \ \operatorname{im} \widehat{\partial}_* + \operatorname{rank} \ \operatorname{im} i_* \ \text{((17.113) is exact).} \end{aligned} \tag{17.114}$$

Since $K_{\varphi_+} = \{0, u_0\}$ and $\tau < 0 < \lambda < m_\rho^+ \le \varphi_+(u_0)$, we have

$$\begin{aligned} & H_k(W^{1,p}(\Omega), \varphi_+^\tau) = C_k(\varphi_+, \infty) = 0 \text{ for all } k \in \mathbb{N}_0 \text{ (see Proposition 17.5.7)}, \\ \Rightarrow \quad & \operatorname{im} i_* = \{0\} \text{ (see (17.113))}. \end{aligned} \tag{17.115}$$

Also since $0 < \lambda < m_\rho^+ \le \varphi_+(u_0)$, we have

$$H_k(W^{1,p}(\Omega), \varphi_+^\tau) = C_k(\varphi_+, u_0) \text{ for all } k \in \mathbb{N}_0. \tag{17.116}$$

We know that $u_0 \in K_{\varphi_+}$ is of mountain pass type (see the proof of Proposition 17.5.5). Therefore $C_1(\varphi_+, u_0) \neq 0$ (see Motreanu et al. [52, p. 176]). Also since $\tau < 0 < \lambda < m_\rho^+ \le \varphi_+(u_0)$ we have

$$H_{k-1}(\varphi_+^\lambda, \varphi_+^\tau) = C_{k-1}(\varphi_+, 0) = \delta_{k-1,0}\,\mathbb{Z} = \delta_{k,1}\,\mathbb{Z} \text{ for all } k \in \mathbb{N}_0. \tag{17.117}$$

In (17.114) note that for $k \ge 2$ all terms are trivial. Using (17.115), (17.116), (17.117) we obtain

$$\begin{aligned} & \operatorname{rank} C_1(\varphi_+, u_0) \le 1, \\ \Rightarrow \quad & \operatorname{rank} C_1(\varphi_+, u_0) = 1, \\ \Rightarrow \quad & C_k(\varphi_+, u_0) = \delta_{k,1}\,\mathbb{Z} \text{ for all } k \in \mathbb{N}_0. \end{aligned} \tag{17.118}$$

Consider the homotopy $h : [0, 1] \times W^{1,p}(\Omega) \to W^{1,p}(\Omega)$ defined by

$$h(t, u) = (1 - t)\varphi(u) + t\varphi_+(u) \text{ for all } (t, u) \in [0, 1] \times W^{1,p}(\Omega).$$

Suppose we could find two sequences $\{t_n\}_{n\geq 1} \subseteq [0, 1]$ and $\{u_n\}_{n\geq 1} \subseteq W^{1,p}(\Omega)$ such that

$$t_n \to \widehat{t} \text{ in [0,1]}, u_n \to u_0 \text{ in } W^{1,p}(\Omega) \text{ and } h'_u(t_n, u_n) = 0 \text{ for all } n \in \mathbb{N}. \tag{17.119}$$

From (17.119) we have

$$\begin{aligned}
&\langle A(u_n), h\rangle + \int_\Omega \xi(z)|u_n|^{p-2}u_n h dz + \int_{\partial\Omega} \beta(z)|u_n|^{p-2}u_n h d\sigma \\
&\quad - t_n\eta \int_\Omega (u_n^-)^{p-1} h dz \\
&= \int_\Omega f\left(z, u_n^+\right) h dz + (1 - t_n) \int_\Omega f(z, -u_n^-) h dz \text{ for all } h \in W^{1,p}(\Omega), \text{ all } n \in \mathbb{N}.
\end{aligned}$$

It follows that

$$\begin{cases} -\operatorname{div} a(\nabla u_n) + \xi(z)|u_n|^{p-2}u_n = f\left(z, u_n^+\right) + (1 - t_n) f(z, -u_n^-) + \eta(u_n^-)^{p-1} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for a.a. } z \in \Omega, \\ \dfrac{\partial u_n}{\partial n_a} + \beta(z)|u_n|^{p-2}u_n = 0 \text{ on } \partial\Omega. \end{cases} \tag{17.120}$$

Then using Propositions 17.2.3 and 17.2.4, from (17.120) we see that we can find $\alpha \in (0, 1)$ and $c_{26} > 0$ such that

$$u_n \in C^{1,\alpha}(\overline{\Omega}) \text{ and } \|u_n\|_{C^{1,\alpha}(\overline{\Omega})} \leq c_{26} \text{ for all } n \in \mathbb{N}.$$

Recalling that $C^{1,\alpha}(\overline{\Omega})$ is embedded compactly in $C^1(\overline{\Omega})$ and using (17.119), we infer that $u_n \to u_0$ in $C^1(\overline{\Omega})$. But $u_0 \in \operatorname{int} \widehat{C}_+$. Hence $u_n \in \operatorname{int} \widehat{C}_+$ for all $n \geq n_0$ and so from (17.120) it follows that $\{u_n\}_{n\geq 1} \subseteq K_\varphi$, a contradiction. Therefore (17.119) cannot be true and from the homotopy invariance of critical groups (see Gasiński and Papageorgiou [34], p.836) we have

$$\begin{aligned}
&C_k(\varphi, u_0) = C_k(\varphi_+, u_0) \text{ for all } k \in \mathbb{N}_0, \\
\Rightarrow\quad &C_k(\varphi, u_0) = \delta_{k,1}\mathbb{Z} \text{ for all } k \in \mathbb{N}_0 \text{ (see (17.118))}.
\end{aligned}$$

Similarly using this time φ_-, we show that $C_k(\varphi, v_0) = C_k(\varphi_-, v_0) = \delta_{k,1}\mathbb{Z}$ for all $k \in \mathbb{N}_0$. □

Now we are ready to produce a third nontrivial solution and have the full multiplicity result for problem (17.68) (three solutions theorem).

Theorem 17.5.9 *If hypotheses $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_2$ hold, then problem* (17.68) *has at least three nontrivial solutions $u_0 \in \operatorname{int} \widehat{C}_+$, $v_0 \in -\operatorname{int} \widehat{C}_+$ and $y_0 \in C^1(\overline{\Omega}) \setminus \{0\}$.*

Proof From Proposition 17.5.5, we already have two nontrivial constant sign solutions $u_0 \in \operatorname{int} \widehat{C}_+$ and $v_0 \in -\operatorname{int} \widehat{C}_+$. From Proposition 17.5.8 we know that

$$C_k(\varphi, u_0) = C_k(\varphi, v_0) = \delta_{k,1}\mathbb{Z} \text{ for all } k \in \mathbb{N}_0. \tag{17.121}$$

From Proposition 17.5.4 we know that $u = 0$ is a local minimizer of φ. Hence

$$C_k(\varphi, 0) = \delta_{k,0}\mathbb{Z} \text{ for all } k \in \mathbb{N}_0. \tag{17.122}$$

Finally from Proposition 17.5.6, we have

$$C_k(\varphi, \infty) = 0 \text{ for all } k \in \mathbb{N}_0. \tag{17.123}$$

Suppose that $K_\varphi = \{0, u_0, v_0\}$. Then using (17.121), (17.122), (17.123) and the Morse relation with $t = -1$ (see (17.72)), we have

$$\begin{aligned} & (-1)^0 + 2(-1)^1 = 0, \\ \Rightarrow \quad & (-1)^1 = 0, \text{ a contradiction.} \end{aligned}$$

So, there exists $y_0 \in K_\varphi$, $y_0 \notin \{0, u_0, v_0\}$. Then y_0 is a solution of problem (17.68) and the nonlinear regularity theory (see Propositions 17.2.3 and 17.2.4) implies that $y_0 \in C^1(\overline{\Omega})$. □

17.6 Nodal Solutions

In this section we prove the existence of nodal solutions for the following nonlinear nonhomogeneous Robin problem.

$$\begin{cases} -\operatorname{div} a(\nabla u(z)) + \xi(z)|u(z)|^{p-2}u(z) = f(z, u(z)) \text{ in } \Omega, \\ \dfrac{\partial u}{\partial n_a} + \beta(z)|u|^{p-2}u = 0 \text{ on } \partial\Omega. \end{cases} \tag{17.124}$$

The reaction term $f(z, x)$ is a Carathéodory function which exhibits $(p-1)$-linear growth near $\pm\infty$ and a concave nonlinearity near the origin. Under such conditions we show that problem (17.124) admits at least one nodal solution.

The new conditions on the reaction term $f(z,x)$ are the following:

$H(f)_3$ $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function such that for a.a. $z \in \Omega$ $f(z,0) = 0$ and

(*i*) for every $\rho > 0$, there exists $a_\rho \in L^\infty(\Omega)$ such that $|f(z,x)| \le a_\rho(z)$ for a.a. $z \in \Omega$, all $|x| \le \rho$;

(*ii*) there exist $\vartheta \in L^\infty(\Omega)$ and $\gamma > 0$ such that

$$\vartheta(z) \le \frac{c_1}{p-1}\widehat{\lambda}_1(p,\xi_0,\beta_0) \text{ for a.a. } z \in \Omega, \quad \vartheta \not\equiv \frac{c_1}{p-1}\widehat{\lambda}_1(p,\xi_0,\beta_0),$$

$$-\gamma \le \liminf_{x\to\pm\infty} \frac{pF(z,x)}{|x|^p} \le \limsup_{x\to\pm\infty} \frac{pF(z,x)}{|x|^p} \le \vartheta(z) \quad \text{uniformly for a.a. } z \in \Omega$$

(recall $F(z,x) = \int_0^x f(z,s)ds$);

(*iii*) with $q \in (1,p]$ as in hypothesis $H(a)'(v)$ we have

$$\lim_{x\to 0} \frac{F(z,x)}{|x|^q} = +\infty \text{ uniformly for a.a. } z \in \Omega.$$

First we establish the existence of constant sign solutions for problem (17.124). So, let $S_\pm$ be the sets of positive (resp. negative) solutions of (17.124).

Proposition 17.6.1 *If hypotheses $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_3$ hold, then $S_\pm \neq \emptyset$ and $S_+ \subseteq \operatorname{int}\widehat{C}_+$, $S_- \subseteq -\operatorname{int}\widehat{C}_+$.*

Proof Let $\eta > \max\{\|\xi\|_\infty, \|\xi_0\|_\infty\}$ and consider the C^1-functional $\varphi_+ : W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\varphi_+(u) = \frac{1}{p}\widehat{\mu}(u) + \frac{\eta}{p}\|u^-\|_p^p - \int_\Omega F(z,u^+)dz \text{ for all } u \in W^{1,p}(\Omega).$$

Recall that $\widehat{\mu} : W^{1,p}(\Omega) \to \mathbb{R}$ is defined by

$$\widehat{\mu}(u) = \int_\Omega pG(\nabla u)dz + \frac{1}{p}\int_\Omega \xi(z)|u|^p dz + \frac{1}{p}\int_{\partial\Omega} \beta(z)|u|^p d\sigma \text{ for all } u \in W^{1,p}(\Omega).$$

Claim φ_+ is coercive.

Hypotheses $H(f)_3(i), (ii)$ imply that given $\varepsilon > 0$, we can find $c_{27} = c_{27}(\varepsilon) > 0$ such that

$$F(z,x) \le \frac{1}{p}[\vartheta(z) + \varepsilon]|x|^p + c_{27} \text{ for a.a. } z \in \Omega, \text{ all } x \in \mathbb{R}. \tag{17.125}$$

For all $u \in W^{1,p}(\Omega)$, we have

$$\begin{aligned}
\varphi_+(u) \geq & \frac{c_1}{p(p-1)} \left[\|\nabla u\|_p^p + \int_\Omega \xi_0(z)|u|^p dz + \int_{\partial\Omega} \beta_0(z)|u|^p d\sigma \right] \\
& - \frac{c_1}{p(p-1)} \int_\Omega \vartheta_0(z) \left(u^+\right)^p dz + \frac{\eta}{p}\|u^-\|_p^p - \frac{\varepsilon}{p}\|u\|^p - c_{28} \\
& \qquad \text{(with } \vartheta_0 = \frac{p-1}{c_1}\vartheta \text{ and some } c_{28} > 0 \text{ (see (17.125))} \\
= & \frac{c_1}{p(p-1)} \left[\|\nabla u^+\|_p^p + \int_\Omega \xi(z)(u^+)^p dz + \int_{\partial\Omega} \beta(z)(u^+)^p d\sigma \right. \\
& \left. - \int_\Omega \vartheta_0(z)(u^+)^p dz \right] + \frac{c_1}{p(p-1)} \left[\|\nabla u^-\|_p^p + \int_\Omega (\xi_0(z) + \eta)(u^-)^p dz \right. \\
& \left. + \int_{\partial\Omega} \beta_0(z)(u^-)^p d\sigma \right] - \frac{\varepsilon}{p}\|u\|^p - c_{28} \geq [c_{29} - \varepsilon]\|u^+\|^p \\
& + c_{30}\|u^-\|^p - c_{28} \text{ for some } c_{29}, c_{30} > 0 \text{ (see Proposition 17.4.6).}
\end{aligned}$$

Choosing $\varepsilon \in (0, c_{29})$, we conclude that

$$\begin{aligned}
& \varphi_+(u) \geq c_{31}\|u\|^p - c_{28} \text{ for all } u \in W^{1,p}(\Omega), \text{ some } c_{31} > 0, \\
\Rightarrow \quad & \varphi_+ \text{ is coercive.}
\end{aligned}$$

This proves the claim.

The Sobolev embedding theorem and the compactness of the trace map imply that φ_+ is sequentially weakly lower semicontinuous. So, by the Weierstrass-Tonelli theorem, we can find $u_0 \in W^{1,p}(\Omega)$ such that

$$\varphi_+(u_0) = \inf\left[\varphi_+(u) : u \in W^{1,p}(\Omega)\right]. \tag{17.126}$$

Hypothesis $H(f)_3(iii)$ implies that given any $\lambda > 0$, we can find $\delta \in (0, 1)$ such that

$$F(z, x) \geq \frac{\lambda}{q}|x|^q \text{ for a.a. } z \in \Omega, \text{ all } |x| \leq \delta. \tag{17.127}$$

Let $t \in (0, \delta]$. Then on account of (17.127) and since $\delta < 1$, $q < p$, we have

$$\begin{aligned}
\varphi_+(t) & \leq \frac{t^q}{q}\left[\|\xi\|_\infty|\Omega|_N + \|\beta\|_\infty\sigma(\partial\Omega)\right] - \frac{\lambda t^q}{q} \\
& = \frac{t^q}{q}\left[\|\xi\|_\infty|\Omega|_N + \|\beta\|_\infty\sigma(\partial\Omega) - \lambda\right].
\end{aligned}$$

Since $\lambda > 0$ is arbitrary, we choose $\lambda > \|\xi\|_\infty |\Omega|_N + \|\beta\|_\infty \sigma(\partial\Omega)$ and we have $\varphi_+(t) < 0$. Hence from (17.126) we have

$$\varphi_+(u_0) < 0 = \varphi_+(0),$$
$$\Rightarrow \quad u_0 \neq 0.$$

From (17.126) we have

$$\begin{aligned}
&\varphi_+'(u_0) = 0,\\
&\quad \Rightarrow \langle A(u_0), h\rangle + \int_\Omega \xi(z)|u_0|^{p-2}u_0 h dz + \int_{\partial\Omega} \beta(z)|u_0|^{p-2}u_0 h d\sigma\\
&\qquad - \eta \int_\Omega (u_0^-)^{p-1} h dz\\
&\quad = \int_\Omega f(z, u_0^+) h dz \text{ for all } h \in W^{1,p}(\Omega).
\end{aligned} \tag{17.128}$$

In (17.128) we choose $h = -u_0^- \in W^{1,p}(\Omega)$. Then

$$\frac{c_1}{p-1}\|\nabla u_0^-\|_p^p + \int_\Omega [\xi(z) + \eta](u_0^-)^p dz + \int_{\partial\Omega} \beta(z)(u_0^-)^p d\sigma \le 0,$$
$$\Rightarrow \quad u_0 \ge 0, \ u_0 \neq 0.$$

Then from (17.128) it follows that u_0 solves problem (17.124) and $u_0 \in C_+ \setminus \{0\}$ (nonlinear regularity). Let $\rho = \|u_0\|_\infty$. Hypotheses $H(f)_3$ imply that we can find $\widehat{\xi}_\rho > 0$ such that

$$f(z, x) + \widehat{\xi}_\rho x^{p-1} \ge 0 \text{ for a.a. } z \in \Omega, \text{ all } 0 \le x \le \rho.$$

Then we have

$$\operatorname{div} a(\nabla u_0(z)) \le \left[\|\xi\|_\infty + \widehat{\xi}_\rho\right] u_0(z)^{p-1} \text{ for a.a. } z \in \Omega,$$
$$\Rightarrow \quad u_0 \in \operatorname{int} \widehat{C}_+ \text{ (see Theorem 17.3.1)}.$$

So, we have proved that $S_+ \neq \emptyset$ and $S_+ \subseteq \operatorname{int} \widehat{C}_+$. Similarly we show that $S_- \neq \emptyset$ and $S_- \subseteq -\operatorname{int} \widehat{C}_+$. □

Hypotheses $H(f)_3$ imply that given any $\lambda > 0$, we can find $c_{32} = c_{32}(\lambda) > 0$ such that

$$f(z, x)x \ge \lambda |x|^q - c_{32}|x|^r \text{ for a.a. } z \in \Omega, \text{ all } x \in \mathbb{R},$$

with $r \in (p, p^*)$.

We consider the following auxiliary Robin problem

$$\begin{cases} -\operatorname{div} a(\nabla u(z)) + |\xi(z)||u(z)|^{p-2}u(z) = \lambda |u(z)|^{q-2}u(z) - c_{32}|u(z)|^{r-2}u(z) \text{ in } \Omega, \\ \dfrac{\partial u}{\partial n_a} + \beta(z)|u|^{p-2}u = 0 \text{ on } \partial\Omega. \end{cases} \tag{17.129}$$

Reasoning as in the proof of Proposition 17.4.14 and using hypothesis $H(a)'(v)$, we show that problem (17.129) admits a unique positive solution $u_* \in \operatorname{int} \widehat{C}_+$. Moreover, since (17.129) is odd $v_* = -u_* \in -\operatorname{int} \widehat{C}_+$ is the unique negative solution of (17.129). Then arguing as in the proof of Proposition 17.4.15, we show that

$$u_* \le u \text{ for all } u \in S_+, \quad v \le v_* \text{ for all } v \in S_-. \tag{17.130}$$

Using (17.130) and Lemma 3.10, p. 178, of Hu and Papageorgiou [43], we generate extremal constant sign solutions for problem (17.129) (see also Proposition 17.4.16).

Proposition 17.6.2 *If hypotheses $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_3$ hold, then problem (17.124) has a smallest positive solution $\widehat{u} \in S_+ \subseteq \operatorname{int} \widehat{C}_+$, that is, $\widehat{u} \le u$ for all $u \in S_+$ and a biggest negative solution $\widehat{v} \in S_- \subseteq -\operatorname{int} \widehat{C}_+$, that is, $v \le \widehat{v}$ for all $v \in S_-$.*

Now we focus on the order interval

$$[\widehat{v}, \widehat{u}] = \{y \in W^{1,p}(\Omega) : \widehat{v}(z) \le y(z) \le \widehat{u}(z) \text{ for a.a. } z \in \Omega\}$$

and we produce a nontrivial solution of (17.124) distinct from $\widehat{u} \in \operatorname{int} \widehat{C}_+$ and $\widehat{v} \in -\operatorname{int} \widehat{C}_+$. Then the extremality of $\widehat{u}$ and $\widehat{v}$ implies that this new solution is necessarily nodal.

From Sect. 17.3, we know that

$$a_0(t)t^2 - G_0(t) \ge \frac{c_1}{p} t^p \quad \text{for all } t > 0,$$

$$\Rightarrow \quad (a(y), y)_{\mathbb{R}^N} - G(y) \ge \frac{c_1}{p}|y|^p \text{ for all } y \in \mathbb{R}^N.$$

Proposition 17.6.3 *If hypotheses $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_3$ hold, then $C_k(\varphi, 0) = 0$ for all $k \in \mathbb{N}_0$.*

Proof Hypotheses $H(f)_3$ imply that given any $\lambda > 0$, we can find $c_{32} > 0$ such that

$$F(z, x) \ge \lambda\, |x|^q - c_{32}\, |x|^r \text{ for a.a. } z \in \Omega, \text{ all } x \in \mathbb{R}, \text{ with } r > p. \tag{17.131}$$

In addition hypothesis $H(a)'(v)$ and Corollary 17.2.2 imply that

$$G(y) \leq c_{33}\left[\,|y|^q + |y|^p\right] \text{ for some } c_{33} > 0, \text{ all } y \in \mathbb{R}^N. \tag{17.132}$$

For $u \in W^{1,p}(\Omega)$ and $t > 0$, we have

$$\varphi(tu) \leq c_{33}\left[t^q\|\nabla u\|_q^q + t^p\|\nabla u\|_p^p\right] + \frac{t^p}{p}\int_\Omega \xi(z)|u|^p dz + \frac{t^p}{p}\int_{\partial\Omega} \beta(z)|u|^p d\sigma$$
$$- \lambda t^q\|u\|_q^q + c_{32}\|u\|_r^r \quad \text{(see (17.131), (17.132))}.$$

Since $q \leq p < r$ and $\lambda > 0$ is arbitrary we can find $\tau^* = \tau^*(u) \in (0, 1)$ small such that

$$\varphi(tu) < 0 \text{ for all } t \in (0, \tau^*). \tag{17.133}$$

Let $u \in W^{1,p}(\Omega)$ with $0 < \|u\| \leq 1$ and $\varphi(u) = 0$. For $\eta > \|\xi\|_\infty$ we have

$$\begin{aligned}
\frac{d}{dt}\varphi(tu)\Big|_{t=1} &= \langle \varphi'(u), u\rangle \\
&= \int_\Omega [(a(\nabla u), \nabla u) - G(\nabla u)]\,dz + \left(1 - \frac{1}{p}\right)\int_\Omega [\xi(z) + \eta]|u|^p dz \\
&\quad + \int_\Omega [F(z, u) - f(z, u)u]dz - \left(1 - \frac{1}{p}\right)\eta\|u\|_p^p \\
&\geq c_{34}\|u\|^p - c_{35}\|u\|^r \text{ for some } c_{34}, c_{35} > 0
\end{aligned} \tag{17.134}$$

(see (17.131) and recall $\eta > \|\xi\|_\infty$).

Recall that $p < r$. So, from (17.134) it follows that we can find $\rho \in (0, 1)$ small such that

$$\frac{d}{dt}\varphi(tu)\Big|_{t=1} > 0 \text{ for all } u \in W^{1,p}(\Omega), 0 < \|u\| \leq \rho, \varphi(u) = 0. \tag{17.135}$$

Pick $u \in W^{1,p}(\Omega)$ such that $0 < \|u\| \leq \rho$, $\varphi(u) = 0$. We will show that

$$\varphi(tu) \leq 0 \text{ for all } t \in [0, 1]. \tag{17.136}$$

Arguing by contradiction, suppose that (17.136) is not true. We can find $t_0 \in (0, 1)$ such that $\varphi(t_0 u) > 0$. We set

$$t^* = \min\{t \in (t_0, 1] : \varphi(tu) = 0\} > t_0 > 0 \text{ (recall } \varphi(u) = 0).$$

Let $y = t^*u$. We have

$$0 < \|y\| \le \|u\| \le \rho \text{ and } \varphi(y) = 0.$$

Then (17.135) implies that

$$\frac{d}{dt}\varphi(ty)\Big|_{t=1} > 0. \tag{17.137}$$

Also we see that $\varphi(tu) > 0$ for all $t \in [t_0, t^*)$. Hence

$$\begin{aligned} &\varphi(y) = \varphi(t^*u) = 0 < \varphi(tu) \text{ for all } t \in [t_0, t^*), \\ \Rightarrow \quad &\frac{d}{dt}\varphi(ty)\Big|_{t=1} = t^*\frac{d}{dt}\varphi(tu)\Big|_{t=t^*} \le 0, \end{aligned}$$

which contradicts (17.137). Therefore (17.136) holds.

We can always choose $\rho \in (0, 1)$ small such that $K_\varphi \cap \overline{B}_\rho = \{0\}$ (here $\overline{B}_\rho = \{u \in W^{1,p}(\Omega) : \|u\| \le \rho\}$). Consider the deformation $h : [0, 1] \times (\varphi^\circ \cap \overline{B}_\rho) \to \varphi^\circ \cap \overline{B}_\rho$ defined by $h(t, u) = (1 - t)u$. On account of (17.136) this deformation is well-defined and shows that $\varphi^\circ \cap \overline{B}_\rho$ is contractible in itself.

Fix $u \in \overline{B}_\rho$ with $\varphi(u) > 0$. From (17.133), (17.135) and Bolzano's theorem, we see that there is a unique $t(u) \in (0, 1)$ such that $\varphi(t(u)u) = 0$. Then

$$\varphi(tu) < 0 \text{ if } t \in (0, t(u)), \quad \varphi(tu) > 0 \text{ if } t \in (t(u), 1]. \tag{17.138}$$

Consider the map $\widehat{k}_0 : \overline{B}_\rho \setminus \{0\} \to [0, 1]$ defined by

$$\widehat{k}_0(u) = \begin{cases} 1 & \text{if } u \in \overline{B}_\rho \setminus \{0\}, \varphi(u) \le 0, \\ t(u), & \text{if } u \in \overline{B}_\rho \setminus \{0\}, \varphi(u) > 0. \end{cases}$$

We can easily see using (17.138) that $\widehat{k}_0(\cdot)$ is continuous. Then consider the map $\widehat{\gamma}_0 : \overline{B}_\rho \setminus \{0\} \to (\varphi^\circ \cap \overline{B}_\rho) \setminus \{0\}$ defined by

$$\widehat{\gamma}_0(u) = \begin{cases} u & \text{if } u \in \overline{B}_\rho \setminus \{0\}, \varphi(u) \le 0, \\ \widehat{k}_0(u)u, & \text{if } u \in \overline{B}_\rho \setminus \{0\}, \varphi(u) > 0. \end{cases}$$

The continuity of $\widehat{k}_0(\cdot)$ implies the continuity of $\widehat{\gamma}_0(\cdot)$. Also, we have

$$\widehat{\gamma}_0\big|_{(\varphi^\circ \cap \overline{B}_\rho)\setminus\{0\}} = \mathrm{id}\big|_{(\varphi^\circ \cap \overline{B}_\rho)\setminus\{0\}}.$$

Therefore $(\varphi^\circ \cap \overline{B}_\rho) \setminus \{0\}$ is a retract of $\overline{B}_\rho \setminus \{0\}$. The infinite dimensionality of $W^{1,p}(\Omega)$ implies the contractibility of $\overline{B}_\rho \setminus \{0\}$. Hence $(\varphi^\circ \cap \overline{B}_\rho) \setminus \{0\}$ is contractible too (see Gasiński and Papageorgiou [34, p. 677]). Recall that $\varphi^\circ \cap \overline{B}_\rho$ is contractible. Therefore from Motreanu et al. [52, p. 147], we have

$$\begin{aligned} & H_k(\varphi^\circ \cap \overline{B}_\rho, (\varphi^\circ \cap \overline{B}_\rho) \setminus \{0\}) = 0 \text{ for all } k \in \mathbb{N}_0, \\ \Rightarrow \quad & C_k(\varphi, 0) = 0 \text{ for all } k \in \mathbb{N}_0. \end{aligned}$$

□

Using this proposition we can establish the existence of nodal solutions.

Theorem 17.6.4 *If hypotheses $H(a)'$, $H(\xi)$, $H(\beta)$, $H(f)_3$ hold, then problem (17.124) admits a nodal solution $\widehat{y} \in C^1(\overline{\Omega})$.*

Proof Let $\widehat{u} \in \operatorname{int} \widehat{C}_+$ and $\widehat{v} \in -\operatorname{int} \widehat{C}_+$ be the two extremal constant sign solutions of problem (17.124) produced in Proposition 17.6.2. We consider the Carathéodory function

$$\ell_0(z, x) = \begin{cases} f(z, \widehat{v}(z)) + \eta\, |\widehat{v}(z)|^{p-2}\widehat{v}(z) & \text{if } x < \widehat{v}(z), \\ f(z, x) + \eta\, |x|^{p-2} x & \text{if } \widehat{v}(z) \le x \le \widehat{u}(z), \\ f(z, \widehat{u}(z)) + \eta\, \widehat{u}(z)^{p-1} & \text{if } \widehat{u}(z) < x, \end{cases} \tag{17.139}$$

with $\eta > \|\xi\|_\infty$. We set $L_0(z, x) = \int_0^x \ell_0(z, s) ds$ and consider the C^1-functional $\varphi_0 : W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\varphi_0(u) = \frac{1}{p}\widehat{\mu}(u) + \frac{\eta}{p}\|u\|_p^p - \int_\Omega L_0(z, u)dz \text{ for all } u \in W^{1,p}(\Omega).$$

Evidently φ_0 is coercive (see (17.139) and recall that $\eta > \|\xi\|_\infty$). So, it satisfies the C-condition and

$$C_k(\varphi_0, \infty) = \delta_{k,0}\mathbb{Z} \text{ for all } k \in \mathbb{N}_0.$$

Also since $\widehat{u} \in \operatorname{int} \widehat{C}_+$, $\widehat{v} \in -\operatorname{int} \widehat{C}_+$, it follows that

$$\begin{aligned} & C_k(\varphi_0, 0) = C_k(\varphi, 0) \text{ for all } k \in \mathbb{N}_0, \\ \Rightarrow \quad & C_k(\varphi_0, 0) = 0 \text{ for all } k \in \mathbb{N}_0. \end{aligned} \tag{17.140}$$

Consider the positive and negative truncations of $\ell_0(z, \cdot)$, that is, the Carathéodory function $\ell_0^\pm(z, x) = \ell_0(z, \pm x^\pm)$ for all $(z, x) \in \Omega \times \mathbb{R}$. We set $L_0^\pm(z, x) = \int_0^x \ell_0^\pm(z, s) ds$ and consider the C^1-functional $\varphi_0^\pm : W^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\varphi_0^\pm(u) = \frac{1}{p}\widehat{\mu}(u) + \frac{\eta}{p}\|u\|_p^p - \int_\Omega L_0^\pm(z, u)dz \text{ for all } u \in W^{1,p}(\Omega).$$

Using (17.139) we can see that

$$K_{\varphi_0} \subseteq [\widehat{v}, \widehat{u}] \cap C^1(\overline{\Omega}), \quad K_{\varphi_0^+} \subseteq [0, \widehat{u}] \cap C_+, \quad K_{\varphi_0^-} \subseteq [\widehat{v}, 0] \cap (-C_+).$$

The extremality of $\widehat{u}$, $\widehat{v}$ implies that

$$K_{\varphi_0} \subseteq [\widehat{v}, \widehat{u}] \cap C^1(\overline{\Omega}), \quad K_{\varphi_0^+} = \{0, \widehat{u}\}, \quad K_{\varphi_0^-} = \{0, \widehat{v}\}. \tag{17.141}$$

Claim $\widehat{u}$, $\widehat{v}$ are local minimizers of φ_0.

The functional φ_0^+ is coercive and sequentially weakly lower semicontinuous. So, we can find $\widetilde{u} \in W^{1,p}(\Omega)$ such that

$$\varphi_0^+(\widetilde{u}) = \inf\Big[\varphi_0(u) : u \in W^{1,p}(\Omega)\Big]. \tag{17.142}$$

As before hypothesis $H(f)_3(iii)$ implies that

$$\begin{aligned} &\varphi_0^+(\widetilde{u}) < 0 = \varphi_0^+(0), \\ \Rightarrow \quad &\widetilde{u} \neq 0. \end{aligned}$$

Then from (17.141) and (17.142) it follows that $\widetilde{u} = \widehat{u} \in \operatorname{int} \widehat{C}_+$. Since $\varphi_0\big|_{C_+} = \varphi_0^+\big|_{C_+}$ it follows that $\widehat{u}$ is a local $C^1(\overline{\Omega})$-minimizer of φ_0^+. So, by Proposition 17.2.5 $\widehat{u} \in \operatorname{int} \widehat{C}_+$ is a local $W^{1,p}(\Omega)$-minimizer of φ_0. Similarly for $\widehat{v} \in -\operatorname{int} \widehat{C}_+$ using this time φ_0^-. This proves the claim.

Evidently we may assume that K_{φ_0} is finite (see (17.141)). Also without any loss of generality we can have $\varphi_0(\widehat{v}) \leq \varphi_0(\widehat{u})$. The claim implies that we can find $\rho \in (0, 1)$ small such that

$$\varphi_0(\widehat{v}) \leq \varphi_0(\widehat{u}) < \inf[\varphi_0(u) : \|u - \widehat{u}\| = \rho] = m_\rho^0 \tag{17.143}$$

(see Aizicovici et al. [1, proof of Proposition 29]). Recall that φ_0 satisfies the C-condition. This fact and (17.143) permit the use of the mountain pass theorem. So, we can find $\widehat{y} \in W^{1,p}(\Omega)$ such that

$$\widehat{y} \in K_{\varphi_0} \subseteq [\widehat{v}, \widehat{u}] \cap C^1(\overline{\Omega}) \text{ (see (17.141))}, \quad m_\rho^0 \leq \varphi_0(\widehat{y}). \tag{17.144}$$

From (17.143), (17.144) we have that $\widehat{y} \notin \{\widehat{v}, \widehat{u}\}$. Moreover, since $\widehat{y}$ is a critical point of φ of mountain pass type, we have

$$C_1(\varphi_0, \widehat{y}) \neq 0 \tag{17.145}$$

(see Motreanu et al. [52, p. 176]). Then comparing (17.140) and (17.145), we infer that $\widehat{y} \neq 0$. Hence

$$\begin{aligned} &\widehat{y} \in K_{\varphi_0} \subseteq [\widehat{v}, \widehat{u}] \cap C^1(\overline{\Omega}) \setminus \{0\}, \\ \Rightarrow \quad &\widehat{y} \in C^1(\overline{\Omega}) \text{ is a nodal solution of (17.124).} \end{aligned}$$

□

When $\xi \equiv 0$, $\beta > 0$, and $f(z, \cdot)$ is nondecreasing, we can generate a second nodal solution. So, the problem under consideration is the following:

$$\begin{cases} -\operatorname{div} a(\nabla u(z)) = f(z, u(z)) \text{ in } \Omega, \\ \dfrac{\partial u}{\partial n_a} + \beta(z)|u|^{p-2}u = 0 \text{ on } \partial\Omega. \end{cases} \tag{17.146}$$

The new hypotheses on the reaction term $f(z, x)$ are:

$H(f)_4$ $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function such that for a.a. $z \in \Omega$ $f(z, 0) = 0$, hypotheses $H(f)_4(i), (ii), (iii)$ are the same as the corresponding hypotheses $H(f)_3(i), (ii), (iii)$ and

(iv) for a.a. $z \in \Omega$ $f(z, \cdot)$ is nondecreasing.

Also we restrict further the boundary coefficient $\beta(\cdot)$.

$H(\beta)'$ $\beta \in C^{0,\alpha}(\partial\Omega)$ with $\alpha \in (0, 1)$, $\beta(z) > 0$ for all $z \in \partial\Omega$.

Theorem 17.6.5 *If the hypotheses $H(a)'$, $H(\beta)'$, $H(f)_4$ hold, then problem (17.146) has at least two nodal solutions $\widehat{y}, y_0 \in C^1(\overline{\Omega})$.*

Proof From Theorem 17.6.4 we already have a nodal solution $\widehat{y} \in [\widehat{v}, \widehat{u}] \cap C^1(\overline{\Omega})$. We have

$$\begin{aligned} &-\operatorname{div} a(\nabla\widehat{y}(z)) = f(z, \widehat{y}(z)) \leq f(z, \widehat{u}(z)) \leq -\operatorname{div} a(\nabla\widehat{u}(z)) \text{ for a.a. } z \in \Omega \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{(see hypothesis } H(f)_4(iv)), \\ \Rightarrow \quad &\widehat{u} - \widehat{y} \in \operatorname{int} C_+ \text{ (see Proposition 17.3.6).} \end{aligned}$$

Similarly we show that $\widehat{y} - \widehat{v} \in \operatorname{int} C_+$. Hence

$$\widehat{y} \in \operatorname{int}_{C^1(\overline{\Omega})}[\widehat{v}, \widehat{u}]. \tag{17.147}$$

On the other hand, from He et al. [40] we know that problem (17.146) has a nodal solution $y_0 \notin \operatorname{int}_{C^1(\overline{\Omega})}[\widehat{v}, \widehat{u}]$. Because of (17.147) we conclude that $\widehat{y} \neq y_0$. □

Remark 17.6.6 It is interesting to know if we can remove the stronger condition $H(\beta)'$ which excludes from consideration of Neumann problems. This can be done in the case of $(p, 2)$-equations.

17.7 Dirichlet $(p, 2)$-Equations

In this section we study the following nonlinear nonhomogeneous Dirichlet problem:

$$\begin{cases} -\Delta_p u(z) - \Delta u(z) = f(z, u(z)) \text{ in } \Omega, \\ u\big|_{\partial\Omega} = 0, \quad p > 2. \end{cases} \tag{17.148}$$

So, in this case

$$a(y) = |y|^{p-2}y + y \text{ for all } y \in \mathbb{R}^N,$$

and $q = 2$ (see hypothesis $H(a)'(v)$).

The hypotheses on the reaction term $f(z, x)$ are the following:

$H(f)_5$ $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a measurable function such that for a.a. $z \in \Omega$ $f(z, 0) = 0$, $f(z, \cdot) \in C^1(\mathbb{R})$ and

(i) $|f'_x(z, x)| \leq a(z)[1 + |x|^{r-2}]$ for a.a. $z \in \Omega$, all $x \in \mathbb{R}$, with $a \in L^\infty(\Omega)$, $p < r < p^*$;

(ii) there exist function $\vartheta \in L^\infty(\Omega)$ and $\gamma > 0$ such that

$$\vartheta(z) \leq \widehat{\lambda}_1(p) \text{ for a.a. } z \in \Omega, \quad \vartheta \not\equiv \widehat{\lambda}_1(p),$$

$$-\gamma \leq \liminf_{x \to \pm\infty} \frac{pF(z, x)}{|x|^p} \leq \limsup_{x \to \pm\infty} \frac{pF(z, x)}{|x|^p} \leq \vartheta(z) \text{ uniformly for a.a. } z \in \Omega;$$

(iii) there exist integer $m \geq 2$, $\widehat{\eta} \in L^\infty(\Omega)$ and $\delta_0 > 0$ such that

$$\widehat{\eta}(z) \leq \widehat{\lambda}_{m+1}(2) \text{ for a.a. } z \in \Omega, \quad \widehat{\eta} \not\equiv \widehat{\lambda}_{m+1}(2),$$

$$f(z, x)x \geq \widehat{\lambda}_m(2)x^2 \text{ for a.a. } z \in \Omega, \text{ all } |x| \leq \delta_0,$$

$$f'_x(z, 0) = \lim_{x \to 0} \frac{f(z, x)}{x} \leq \widehat{\eta}(z) \text{ uniformly for a.a. } z \in \Omega.$$

Remark 17.7.1 Now in contrast to $H(f)_3$, $f(z, \cdot)$ exhibits linear and not sublinear growth near zero (recall that in the present setting $q = 2$).

Reasoning as in Sect. 17.6, we can generate extremal constant sign solutions $\widehat{u} \in \operatorname{int} C^0_+$ and $\widehat{v} \in -\operatorname{int} C^0_+$. Then we truncate $f(z, \cdot)$ at $\{\widehat{v}(z), \widehat{u}(z)\}$. So, we introduce the Carathéodory function $\widehat{f}_0(z, x)$ defined by

$$\widehat{f}_0(z, x) = \begin{cases} f(z, \widehat{v}(z)) & \text{if } x < \widehat{v}(z), \\ f(z, x) & \text{if } \widehat{v}(z) \leq x \leq \widehat{u}(z), \\ f(z, \widehat{u}(z)) & \text{if } \widehat{u}(z) < x. \end{cases}$$

We set $\widehat{F}_0(z,x) = \int_0^x \widehat{f}_0(z,s)ds$ and consider the C^1-functional $\widehat{\varphi}_0 : W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\widehat{\varphi}_0(u) = \frac{1}{p}\|\nabla u\|_p^p + \frac{1}{2}\|\nabla u\|_2^2 - \int_\Omega \widehat{F}_0(z,u)dz \text{ for all } u \in W_0^{1,p}(\Omega).$$

Also, let $\widehat{\psi}_0 : H_0^1(\Omega) \to \mathbb{R}$ be the functional defined by

$$\widehat{\psi}_0(u) = \frac{1}{2}\|\nabla u\|_2^2 - \int_\Omega \widehat{F}_0(z,u)dz \text{ for all } u \in H_0^1(\Omega).$$

Evidently $\widehat{\psi}_0(\cdot)$ is C^2 in a neighborhood of the origin.

Proposition 17.7.2 *If hypothesis $H(f)_5(iii)$ holds, then $C_k(\widehat{\psi}_0, 0) = \delta_{k,d_m}\mathbb{Z}$ for all $k \in \mathbb{N}_0$ with $d_m = \dim \overline{H}_m \geq 2$.*

Proof We consider the orthogonal direct sum decomposition $H_0^1(\Omega) = \overline{H}_m \oplus \widehat{H}_m$ (see Sect. 17.4). The space $\overline{H}_m$ is finite dimensional. So, all norms are equivalent and we can find $\rho_1 \in (0,1)$ small such that

$$u \in \overline{H}_m, \quad \|u\|_{H_0^1(\Omega)} \leq \rho_1 \Rightarrow |u(z)| \leq \delta_0 \text{ for all } z \in \overline{\Omega}.$$

Then for $u \in \overline{H}_m$ with $\|u\|_{H_0^1(\Omega)} \leq \rho_1$, we have

$$\begin{aligned}\widehat{\psi}_0(u) &\leq \frac{1}{2}\|\nabla u\|_2^2 - \frac{\widehat{\lambda}_m(2)}{2}\|u\|_2^2 \text{ (see hypothesis } H(f)_5(iii))\\ &\leq 0 \text{ (see (17.20))}. \end{aligned} \tag{17.149}$$

On the other hand, from hypothesis $H(f)_5(iii)$ and the definition of $\widehat{f}_0$ we see that given $\varepsilon > 0$, we can find $c_{36} > 0$ such that

$$\widehat{F}_0(z,x) \leq \frac{1}{2}[\widehat{\eta}(z) + \varepsilon]x^2 + c_{36}|x|^r \text{ for a.a. } z \in \Omega, \text{ all } x \in \mathbb{R}, \text{ with } r > 2. \tag{17.150}$$

Then for $u \in \widehat{H}_m$ we have

$$\widehat{\psi}_0(u) \geq \frac{1}{2}\|\nabla u\|_2^2 - \frac{1}{2}\int_\Omega \widehat{\eta}(z)u^2dz - \varepsilon\|u\|^2 - c_{37}\|u\|^r$$

for some $c_{37} > 0$ (see (17.150)),

$$\Rightarrow \quad \widehat{\psi}_0(u) \geq [c_{38} - \varepsilon]\|u\|^2 - c_{37}\|u\|^r \text{ for some } c_{38} > 0 \text{ (see Proposition 17.4.7).}$$

Choosing $\varepsilon \in (0, c_{38})$ we have

$$\widehat{\psi}_0(u) \geq c_{39}\|u\|^2 - c_{37}\|u\|^r \text{ for some } c_{39} > 0, \text{ all } u \in \widehat{H}_m.$$

Since $r > 2$, we can find $\rho_2 \in (0, 1)$ small such that

$$\widehat{\psi}_0(u) > 0 \text{ for all } u \in \widehat{H}_m, 0 < \|u\| \leq \rho_2. \tag{17.151}$$

From (17.149) and (17.151) we infer that $\widehat{\psi}_0$ has a local linking at $u = 0$. So, invoking Proposition 2.3 of Su [73], we conclude that

$$C_k(\widehat{\psi}_0, 0) = \delta_{k,d_m}\mathbb{Z} \text{ for all } k \in \mathbb{N}_0, d_m = \dim \overline{H}_m \geq 2.$$

□

Using this proposition, we can prove the following multiplicity theorem for problem (17.148). We have sign-information for all the solutions we produce.

Theorem 17.7.3 *If hypotheses $H(f)_5$ hold, then problem* (17.148) *admits at least five nontrivial solutions $\widehat{u} \in \operatorname{int} C_+^0$, $\widehat{v} \in -\operatorname{int} C_+^0$ and $\widehat{y}$, y_0, $\widetilde{y} \in C_0^1(\overline{\Omega})$ all nodal solutions.*

Proof As we already mentioned, we have two constant sign solutions $\widehat{u} \in \operatorname{int} C_+^0$ and $\widehat{v} \in -\operatorname{int} C_+^0$, which we can assume to be extremal. Then using $\widehat{\varphi}_0$, Proposition 17.7.2 and the C^1-continuity of critical groups (see Gasiński and Papageorgiou [34, p. 836]), we obtain that if $\psi_0 = \widehat{\psi}_0\big|_{W_0^{1,p}(\Omega)}$, then

$$C_k(\psi_0, 0) = C_k(\widehat{\varphi}_0, 0) \text{ for all } k \in \mathbb{N}_0. \tag{17.152}$$

Since $W_0^{1,p}(\Omega)$ is dense in $H_0^1(\Omega)$, we have

$$\begin{aligned} & C_k(\psi_0, 0) = C_k(\widehat{\psi}_0, 0) \text{ for all } k \in \mathbb{N}_0, \\ \Rightarrow \quad & C_k(\psi_0, 0) = \delta_{k,d_m}\mathbb{Z} \text{ for all } k \in \mathbb{N}_0 \text{ (see Proposition 17.7.2)}, \\ \Rightarrow \quad & C_k(\widehat{\varphi}_0, 0) = \delta_{k,d_m}\mathbb{Z} \text{ for all } k \in \mathbb{N}_0 \text{ (see (17.152))}. \end{aligned} \tag{17.153}$$

Using (17.153) (recall that $d_m \geq 2$) as in the proof of Theorem 17.6.4, via the mountain pass theorem, we produce a solution $\widehat{y}$ such that

$$\widehat{y} \in [\widehat{v}, \widehat{u}] \cap C_0^1(\overline{\Omega}) \setminus \{0\}.$$

Hypotheses $H(f)_5$ imply that if $\rho = \max\{\|\widehat{v}\|_\infty, \|\widehat{u}\|_\infty\}$, then we can find $\widehat{\xi}_\rho > 0$ such that for a.a. $z \in \Omega$, the function

$$x \to f(z, x) + \widehat{\xi}_\rho |x|^{p-2}x$$

is nondecreasing on $[-\rho, \rho]$. We have

$$
\begin{aligned}
& -\Delta_p \widehat{y}(z) - \Delta \widehat{y}(z) + \widehat{\xi}_\rho |\widehat{y}(z)|^{p-2} \widehat{y}(z) \\
& = f(z, \widehat{y}(z)) + \widehat{\xi}_\rho |\widehat{y}(z)|^{p-2} \widehat{y}(z) \\
& \leq f(z, \widehat{u}(z)) + \widehat{\xi}_\rho \widehat{u}(z)^{p-1} \\
& = -\Delta_p \widehat{u}(z) - \Delta \widehat{u}(z) + \widehat{\xi}_\rho \widehat{u}(z)^{p-1} \text{ for a.a. } z \in \Omega.
\end{aligned} \tag{17.154}
$$

Recall that in the present setting $a(y) = |y|^{p-2}y + y$ for all $y \in \mathbb{R}^N$. Since $2 < p$, we have $a \in C^1(\mathbb{R}^N, \mathbb{R}^N)$ and we have

$$
\nabla a(y) = |y|^{p-2} \left[\text{id} + \frac{y \otimes y}{|y|^2} \right] + \text{id}.
$$

Therefore

$$
(\nabla a(y)\xi, \xi)_{\mathbb{R}^N} \geq |\xi|^2 \text{ for all } y, \xi \in \mathbb{R}^N.
$$

Hence the tangency principle of Pucci and Serrin [72, p. 35] implies that

$$
\widehat{y}(z) < \widehat{u}(z) \text{ for all } z \in \Omega. \tag{17.155}
$$

Since $\widehat{u} \in \operatorname{int} C_+^0$, from (17.154), (17.155) and Proposition 17.3.4, we infer that

$$
\widehat{u} - \widehat{y} \in \operatorname{int} C_+^0.
$$

Similarly we show that $\widehat{y} - \widehat{v} \in \operatorname{int} C_+^0$. Therefore

$$
\widehat{y} \in \operatorname{int}_{C_0^1(\overline{\Omega})} [\widehat{v}, \widehat{u}]. \tag{17.156}
$$

The solution $\widehat{y}$ is a critical point of $\widehat{\varphi}_0$ of mountain pass type. So, from Papageorgiou and Rădulescu [56], we have

$$
C_k(\widehat{\varphi}_0, \widehat{y}) = \delta_{k,1} \mathbb{Z} \text{ for all } k \in \mathbb{N}_0. \tag{17.157}
$$

Recall that $\widehat{u}, \widehat{v}$ are local minimizers of $\widehat{\varphi}_0$. Hence

$$
C_k(\widehat{\varphi}_0, \widehat{u}) = C_k(\widehat{\varphi}_0, \widehat{v}) = \delta_{k,1} \mathbb{Z} \text{ for all } k \in \mathbb{N}_0. \tag{17.158}
$$

Moreover, the coercivity of $\widehat{\varphi}_0$ implies that

$$
C_k(\widehat{\varphi}_0, \infty) = \delta_{k,0} \mathbb{Z} \text{ for all } k \in \mathbb{N}_0. \tag{17.159}
$$

Suppose that $K_{\widehat{\varphi}_0} = \{0, \widehat{u}, \widehat{v}, \widehat{y}\} \subseteq [\widehat{v}, \widehat{u}] \cap C_0^1(\overline{\Omega})$. From (17.153), (17.157), (17.158), (17.159), and the Morse relation with $t = -1$ (see (17.72), we have

$$(-1)^{d_m} + 2(-1)^0 + (-1)^1 = (-1)^0,$$
$$\Rightarrow \quad (-1)^{d_m} = 0, \text{ a contradiction.}$$

So, there exists $y_0 \in K_{\widehat{\varphi}_0} \subseteq [\widehat{v}, \widehat{u}] \cap C_0^1(\overline{\Omega})$, $y_0 \notin \{0, \widehat{u}, \widehat{v}, \widehat{y}\}$. Hence y_0 is a second nodal solution. Moreover, as for $\widehat{y}$ we show that

$$y_0 \in \operatorname{int}_{C_0^1(\overline{\Omega})}[\widehat{v}, \widehat{u}]. \tag{17.160}$$

Finally from Papageorgiou and Papalini [55], we can find a nodal solution $\widetilde{y} \in C_0^1(\overline{\Omega})$ such that

$$\widetilde{y} \notin \operatorname{int}_{C_0^1(\overline{\Omega})}[\widehat{v}, \widehat{u}].$$

Then from (17.156), (17.160), we have that $\widetilde{y} \notin \{\widehat{y}, y_0\}$. Therefore $\widetilde{y} \in C_0^1(\overline{\Omega})$ is the third nodal solution of (17.148). □

17.8 Remarks

Section 17.2 The nonlinear regularity theory was formulated by Lieberman [45]. The nonlinear maximum principle is discussed in the book of Pucci and Serrin [72]. Note that equations driven by the (p, q)-Laplace operator arise in problems of mathematical physics, see Cherfils and Il′yasov [18]. A good survey of such equations can be found in Marano and Mosconi [48]. The L^∞-regularity of the weak solutions (see Proposition 17.2.3) can be found in Guedda and Véron [38] (Dirichlet problems) and in Papageorgiou and Rădulescu [59] (Robin problems). The first result in the direction of Proposition 17.2.5 is due to Brezis and Nirenberg [15] with $X = H_0^1(\Omega)$ and $G(y) = \frac{1}{2}|y|^2$ for all $y \in \mathbb{R}^N$. In the general form proved in Proposition 17.2.5, can be found in Papageorgiou and Rădulescu [59]. A more general version of Proposition 17.2.7 can be found in Gasiński and Papageorgiou [32].

Section 17.3 Various weak and strong comparison principles can be found in Arcoya and Ruiz [8], Cuesta and Takáč [19], Damascelli [22], Fragnelli et al. [28], Gasiński and Papageorgiou [36], Lucia and Prashanth [47].

Section 17.4 The spectral theory of the p-Laplacian (with or without potential term) under various boundary conditions can be found in Anane [6], Anane and Tsouli [7], Gasiński and Papageorgiou [31, 34], Motreanu et al. [52], Mugnai and Papageorgiou [53], Papageorgiou and Rădulescu [57]. For the linear eigenvalue problem ($p = 2$), we refer to D'Aguì et al. [21]. Problem (17.26) in the particular case of the p-Laplacian (that is, $a(y) = |y|^{p-2}y$, $1 < p < +\infty$) can be thought as a perturbation of the classical eigenvalue problem. The semilinear case ($p = 2$) with Robin boundary condition can be found in Papageorgiou et al. [69]. For the Dirichlet p-Laplacian, we refer to Bonanno et al. [14]. Our work here subsumes both the aforementioned papers.

Section 17.5 More on critical groups and Morse theory can found in the books of Chang [16, 17] and Motreanu et al. [52]. Three solutions theorems for p-Laplacian equations with $(p - 1)$-superlinear reaction were proved by Aizicovici et al. [2], Bartsch and Liu [12], Filippakis et al. [27], Li and Yang [44], Liu [46], Sun [74]. For nonhomogeneous equations we mention Barletta and Papageorgiou [10], Fukagai and Narukawa [29], Papageorgiou and Rădulescu [58].

Section 17.6 Nodal solutions for nonhomogeneous boundary value problems were obtained by Aizicovici et al. [2–4], Gasiński and Papageorgiou [35], He et al. [40], He et al. [41], He et al. [39], He et al. [42], Papageorgiou and Rădulescu [57, 58, 60, 63], Papageorgiou et al. [68], Papageorgiou and Winkert [66, 67].

Section 17.7 Existence and multiplicity theorem for $(p, 2)$ and more generally (p, q)-equations can be found in Barile and Figueiredo [9], Barletta et al. [11], Benouhiba and Belyacine [13], Figueiredo [26], Gasiński and Papageorgiou [33, 35], Marano et al. [50, 51], Marano and Papageorgiou [49], Mugnai and Papageorgiou [54], Papageorgiou and Rădulescu [56, 61, 62], Papageorgiou et al. [68], Papageorgiou and Vetro [64], Papageorgiou et al. [70], Papageorgiou and Winkert [65], Pei and Zhang [71], Tanaka [75], Yang and Bai [77], Yin and Yang [78, 79].

References

1. S. Aizicovici, N.S. Papageorgiou, V. Staicu, *Degree Theory for Operators of Monotone Type and Nonlinear Elliptic Equations with Inequality Constraints*. Memoirs of the American Mathematical Society, vol. 196, no. 915 (American Mathematical Society, Providence, 2008), 70 pp.
2. S. Aizicovici, N.S. Papageorgiou, V. Staicu, Existence of multiple solutions with precise sign information for superlinear Neumann problems. Ann. Math. Pura Appl. **188**(4), 679–719 (2009)
3. S. Aizicovici, N.S. Papageorgiou, V. Staicu, Nodal solutions for $(p, 2)$-equations. Trans. Am. Math. Soc. **367**, 7343–7372 (2015)
4. S. Aizicovici, N.S. Papageorgiou, V. Staicu, Nonlinear, nonhomogeneous parametric Neumann problems. Topol. Methods Nonlinear Anal. **48**, 45–69 (2016)
5. W. Allegretto, Y.X. Huang, A Picone's identity for the p-Laplacian and applications. Nonlinear Anal. **32**, 819–830 (1998)
6. A. Anane, Etude des valeurs propres et de la résonance pour l'opérateur p-Laplacian. Ph.D. Thesis, Université Libre de Bruxelles, Belgium, 1987
7. A. Anane, N. Tsouli, On the second eigenvalue of the p-Laplacian, in *Nonlinear Partial Differential Equations (Fès, 1994)*, ed. by A. Benkirane, J.-P. Gossez, Pitman Research Notes in Mathematics Series, vol. 343 (Longman, Harlow, 1996), pp. 1–9

8. D. Arcoya, D. Ruiz, The Ambrosetti-Prodi problem for the p-Laplacian operator. Commun. Partial Differ. Equ. **31**, 849–865 (2006)
9. S. Barile, G.M. Figueiredo, Some classes of eigenvalues problems for generalized $p\&q$-Laplacian type operators on bounded domains. Nonlinear Anal. **119**, 457–468 (2015)
10. G. Barletta, N.S. Papageorgiou, A multiplicity theorem for p-superlinear Neumann problems with a nonhomogeneous differential operator. Adv. Nonlinear Stud. **14**, 895–913 (2014)
11. G. Barletta, G. D'Aguì, N.S. Papageorgiou, On resonant $(p, 2)$-equations. Minimax Theory Appl. **1**, 83–109 (2016)
12. T. Bartsch, Z. Liu, On a superlinear elliptic p-Laplacian equation. J. Differ. Equ. **198**, 149–175 (2004)
13. N. Benouhiba, Z. Belyacine, On the solutions of the (p, q)-Laplacian problem at resonance. Nonlinear Anal. **77**, 74–81 (2013)
14. G. Bonanno, P. Candito, R. Livrea, N.S. Papageorgiou, Existence, nonexistence and uniqueness of positive solutions for nonlinear eigenvalue problems. Commun. Pure Appl. Anal. **16**, 1169–1188 (2017)
15. H. Brezis, L. Nirenberg, H^1 versus C^1 local minimizers. C. R. Acad. Sci. Paris Sér. I Math. **317**, 465–472 (1993)
16. K.-C. Chang, *Infinite-Dimensional Morse Theory and Multiple Solution Problems*. Progress in Nonlinear Differential Equations and their Applications, vol. 6 (Birkhäuser, Boston, 1993)
17. K.-C. Chang, *Methods in Nonlinear Analysis*. Springer Monographs in Mathematics (Springer, Berlin, 2005)
18. L. Cherfils, Y. Il′yasov, On the stationary solutions of generalized reaction diffusion equations with $p\&q$-Laplacian. Commun. Pure Appl. Anal. **4**, 9–22 (2005)
19. M. Cuesta, P. Takáč, A strong comparison principle for positive solutions of degenerate elliptic equations. Differ. Integral Equ. **13**, 721–746 (2000)
20. M. Cuesta, D. de Figueiredo, J.-G. Gossez, The beginning of the Fučik spectrum for the p-Laplacian. J. Differ. Equ. **159**, 212–238 (1999)
21. G. D'Aguì, S.A. Marano, N.S. Papageorgiou, Multiple solutions to a Robin problem with indefinite weight and asymmetric reaction. J. Math. Anal. Appl. **433**, 1821–1845 (2016)
22. L. Damascelli, Comparison theorems for some quasilinear degenerate elliptic operators and applications to symmetry and monotonicity results. Ann. Inst. H. Poincaré Anal. Non Linéaire **15**, 493–516 (1998)
23. J.I. Díaz, J.E. Saá, Existence et unicité de solutions positives pour certaines équations elliptiques quasilinéaires. C. R. Acad. Sci. Paris Sér. I Math. **305**, 521–524 (1987)
24. J. Dugundji, *Topology* (Allyn and Bacon, Boston, 1966)
25. E.R. Fadell, P.H. Rabinowitz, Generalized cohomological index theories for Lie group actions with an application to bifurcation questions for Hamiltonian systems. Invent. Math. **45**, 139–174 (1978)
26. G.M. Figueiredo, Existence of positive solutions for a class of $p\&q$ elliptic problems with critical growth on $\mathbb{R}^N$. J. Math. Anal. Appl. **378**, 507–518 (2011)
27. M. Filippakis, A. Kristály, N.S. Papageorgiou, Existence of five nonzero solutions with exact sign for a p-Laplacian equation. Discrete Contin. Dyn. Syst. **24**, 405–440 (2009)
28. G. Fragnelli, D. Mugnai, N.S. Papageorgiou, Positive and nodal solutions for parametric nonlinear Robin problems with indefinite potential. Discrete Contin. Dyn. Syst. **36**, 6133–6166 (2016)
29. N. Fukagai, K. Narukawa, On the existence of multiple positive solutions of quasilinear elliptic eigenvalue problems. Ann. Mat. Pura Appl. **186**(4), 539–564 (2007)
30. J.P. García Azorero, I. Peral Alonso, J.J. Manfredi, Sobolev versus Hölder local minimizers and global multiplicity for some quasilinear elliptic equations. Commun. Contemp. Math. **2**, 385–404 (2000)
31. L. Gasiński, N.S. Papageorgiou, *Nonlinear Analysis*. Series of Mathematical Analysis and Applications, vol. 9 (Chapman and Hall/CRC Press, Boca Raton, 2006)
32. L. Gasiński, N.S. Papageorgiou, Existence and multiplicity of solutions for Neumann p-Laplacian-type equations. Adv. Nonlinear Stud. **8**, 843–870 (2008)

33. L. Gasiński, N.S. Papageorgiou, Multiplicity of positive solutions for eigenvalue problems of $(p, 2)$-equations. Bound. Value Probl. **2012**, 152 (2012)
34. L. Gasiński, N.S. Papageorgiou, *Exercises in Analysis. Part 2: Nonlinear Analysis* (Springer, Cham, 2016)
35. L. Gasiński, N.S. Papageorgiou, Nonlinear elliptic equations with a jumping reaction. J. Math. Anal. Appl. **443**, 1033–1070 (2016)
36. L. Gasiński, N.S. Papageorgiou, Positive solutions for the Robin p-Laplacian problem with competing nonlinearities. Adv. Calculus Variations **12**, 31–56 (2019)
37. A. Granas, J. Dugundji, *Fixed Point Theory*. Springer Monographs in Mathematics (Springer, New York, 2003)
38. M. Guedda, L. Véron, Quasilinear elliptic equations involving critical Sobolev exponents. Nonlinear Anal. **13**, 879–902 (1989)
39. T. He, H. Yan, Z. Sun, M. Zhang, On nodal solutions for nonlinear elliptic equations with a nonhomogeneous differential operator. Nonlinear Anal. **118**, 41–50 (2015)
40. T. He, Y. Huang, K. Liang, Y. Lei, Nodal solutions for noncoercive nonlinear Neumann problems with indefinite potential. Appl. Math. Lett. **71**, 67–73 (2017)
41. T. He, D. Wu, H. Sun, K. Liang, Sign-changing solutions for resonant Neumann problems. J. Math. Anal. Appl. **454**, 659–672 (2017)
42. T. He, Z. Yao, Z. Sun, Multiple and nodal solutions for parametric Neumann problems with nonhomogeneous differential operator and critical growth. J. Math. Anal. Appl. **449**, 1133–1151 (2017)
43. S. Hu, N.S. Papageorgiou, *Handbook of Multivalued Analysis. Vol. I. Theory*. Mathematics and its Applications, vol. 419 (Kluwer Academic Publishers, Dordrecht, 1997)
44. G. Li, C. Yang, The existence of a nontrivial solution to a nonlinear elliptic boundary value problem of p-Laplacian type without the Ambrosetti-Rabinowitz condition. Nonlinear Anal. **72**, 4602–4613 (2010)
45. G.M. Lieberman, On the natural generalization of the natural conditions of Ladyzhenskaya and Ural′tseva for elliptic equations. Commun. Partial Differ. Equ. **16**(2–3), 311–361 (1991)
46. S. Liu, Existence of solutions to a superlinear p-Laplacian equation. Electron. J. Differ. Equ. **2001**, 66 (2001)
47. M. Lucia, S. Prashanth, Strong comparison principle for solutions of quasilinear equations. Proc. Am. Math. Soc. **132**, 1005–1011 (2004)
48. S.A. Marano, S.J.N. Mosconi, Some recent results on the Dirichlet problem for (p, q)-Laplace equations. Discrete Contin. Dyn. Syst. Ser. S **11**, 279–291 (2018)
49. S.A. Marano, N.S. Papageorgiou, Constant-sign and nodal solutions of coercive (p, q)-Laplacian problems. Nonlinear Anal. **77**, 118–129 (2013)
50. S.A. Marano, S.J.N. Mosconi, N.S. Papageorgiou, Multiple solutions to (p, q)-Laplacian problems with resonant concave nonlinearity. Adv. Nonlinear Stud. **16**, 51–65 (2016)
51. S.A. Marano, S.J.N. Mosconi, N.S. Papageorgiou, On a (p, q)-Laplacian problem with parametric concave term and asymmetric perturbation. Rend. Lincei Mat Appl. **29**, 109–125 (2018)
52. D. Motreanu, V.V. Motreanu, N.S. Papageorgiou, *Topological and Variational Methods with Applications to Nonlinear Boundary Value Problems* (Springer, Berlin, 2014)
53. D. Mugnai, N.S. Papageorgiou, Resonant nonlinear Neumann problems with indefinite weight. Ann. Sc. Norm. Super. Pisa Cl. Sci. **11**, 729–788 (2012)
54. D. Mugnai, N.S. Papageorgiou, Wang's multiplicity result for superlinear (p, q)-equations without the Ambrosetti-Rabinowitz condition. Trans. Am. Math. Soc. **366**, 4919–4937 (2014)
55. N.S. Papageorgiou, F. Papalini, Seven solutions with sign information for sublinear equations with unbounded and indefinite potential and no symmetries. Israel J. Math. **201**, 761–796 (2014)
56. N.S. Papageorgiou, V.D. Rădulescu, Qualitative phenomena for some classes of quasilinear elliptic equations with multiple resonance. Appl. Math. Optim. **69**, 393–430 (2014)
57. N.S. Papageorgiou, V.D. Rădulescu, Multiple solutions with precise sign for nonlinear parametric Robin problems. J. Differ. Equ. **256**, 2449–2479 (2014)

58. N.S. Papageorgiou, V.D. Rădulescu, Coercive and noncoercive nonlinear Neumann problems with indefinite potential. Forum Math. **28**, 545–571 (2016)
59. N.S. Papageorgiou, V.D. Rădulescu, Nonlinear nonhomogeneous Robin problems with superlinear reaction term. Adv. Nonlinear Stud. **16**, 737–764 (2016)
60. N.S. Papageorgiou, V.D. Rădulescu, Infinitely many nodal solutions for nonlinear nonhomogeneous Robin problems. Adv. Nonlinear Stud. **16**, 287–299 (2016)
61. N.S. Papageorgiou, V.D. Rădulescu, Asymmetric, noncoercive, superlinear $(p, 2)$-equations. J. Convex Anal. **24**, 769–793 (2017)
62. N.S. Papageorgiou, V.D. Rădulescu, Noncoercive resonant $(p, 2)$-equations. Appl. Math. Optim. **76**, 621–639 (2017)
63. N.S. Papageorgiou, V.D. Rădulescu, Multiplicity theorems for nonlinear nonhomogeneous Robin problems. Rev. Mat. Iberoam. **33**, 251–289 (2017)
64. N.S. Papageorgiou, C. Vetro, Superlinear $(p(z), q(z))$-equations. Complex Var. Elliptic Equ. **64**, 8–25 (2019)
65. N.S. Papageorgiou, P. Winkert, Resonant $(p, 2)$-equations with concave terms. Appl. Anal. **94**, 342–360 (2015)
66. N.S. Papageorgiou, P. Winkert, Nonlinear nonhomogeneous Dirichlet equations involving a superlinear nonlinearity. Results Math. **70**, 31–79 (2016)
67. N.S. Papageorgiou, P. Winkert, Nonlinear Robin problems with a reaction of arbitrary growth. Ann. Mat. Pura Appl. **195**(4), 1207–1235 (2016)
68. N.S. Papageorgiou, V.D. Rădulescu, D.D. Repovš, On a class of parametric $(p, 2)$-equations. Appl. Math. Optim. **75**, 193–228 (2017)
69. N.S. Papageorgiou, V.D. Rădulescu, D.D. Repovš, Positive solutions for perturbations of the Robin eigenvalue problem plus an indefinite potential. Discrete Contin. Dyn. Syst. **37**, 2589–2618 (2017)
70. N.S. Papageorgiou, C. Vetro, F. Vetro, $(p, 2)$-equations with a crossing nonlinearity and concave terms. Appl. Math. Optim. https://doi.org/10.1007/s00245-018-9482-0
71. R. Pei, J. Zhang, Nontrivial solution for asymmetric $(p, 2)$-Laplacian Dirichlet problem. Bound. Value Probl. **2014**, 241 (2014)
72. P. Pucci, J. Serrin, *The Maximum Principle* (Birkhäuser Verlag, Basel, 2007)
73. J. Su, Semilinear elliptic boundary value problems with double resonance between two consecutive eigenvalues. Nonlinear Anal. **48**, 881–895 (2002)
74. M.-Z. Sun, Multiple solutions of a superlinear p-Laplacian equation without AR-condition. Appl. Anal. **89**, 325–336 (2010)
75. M. Tanaka, Generalized eigenvalue problems for (p, q)-Laplacian with indefinite weight. J. Math. Anal. Appl. **419**, 1181–1192 (2014)
76. J.L. Vázquez, A strong maximum principle for some quasilinear elliptic equations. Appl. Math. Optim. **12**, 191–202 (1984)
77. D. Yang, C. Bai, Nonlinear elliptic problem of 2-q-Laplacian type with asymmetric nonlinearities. Electron. J. Differ. Equ. **2014**, 170 (2014)
78. H. Yin, Z. Yang, A class of p-q-Laplacian type equation with concave-convex nonlinearities in bounded domain. J. Math. Anal. Appl. **382**, 843–855 (2011)
79. H. Yin, Z. Yang, Multiplicity of positive solutions to a p-q-Laplacian equation involving critical nonlinearity. Nonlinear Anal. **75**, 3021–3035 (2012)

Chapter 18
Summability of Double Sequences and Double Series Over Non-Archimedean Fields: A Survey

P. N. Natarajan and Hemen Dutta

Abstract In this chapter, K denotes a complete, non-trivially valued, non-Archimedean field. We introduce a new definition of convergence of a double sequence and a double series (Natarajan and Srinivasan, Ann Math Blaise Pascal 9:85–100, 2002), which seems to be most suitable in the non-Archimedean context. We study some of its properties. We then present a very brief survey of the results, proved so far, pertaining to the Nörlund, weighted mean, and $(M, \lambda_{m,n})$ (or Natarajan) methods of summability for double sequences. In this chapter, a Tauberian theorem for the Nörlund method for double series is presented.

Keywords Non-Archimedean field · Double sequence · Double series · 4-Dimensional infinite matrix · Conservative matrix · Regular matrix · Pringsheim · Silverman–Toeplitz theorem · Schur's theorem · Steinhaus theorem · Nörlund method · Weighted mean method · $(M, \lambda_{m,n})$ (or Natarajan) method · Tauberian theorem

18.1 Double Sequences and Double Series

At the outset, we suggest that the reader can refer to [1–3] for a study of double sequences and double series in the classical case. For a compact overview of basics of summability theory in the classical case, the reader can refer to [4].

P. N. Natarajan is a former Professor and Head of the Department of Mathematics of the Ramakrishna Mission Vivekananda College in Chennai, India

P. N. Natarajan (✉)
Independent Research Professional, Chennai, India

H. Dutta
Department of Mathematics, Gauhati University, Guwahati, India

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_18

Throughout this chapter, K denotes a complete, non-trivially valued, non-Archimedean field. Double sequences, double series, and 4-dimensional infinite matrices have entries in K. We now introduce a new definition of convergence of a double sequence in K (see [5]).

Definition 18.1.1 Let $\{x_{m,n}\}$ be a double sequence in K and $x \in K$. We say that

$$\lim_{m+n\to\infty} x_{m,n} = x,$$

if for every $\epsilon > 0$, the set

$$\{(m,n) \in \mathbb{N}^2 : |x_{m,n} - x| \geq \epsilon\}$$

is finite, where $\mathbb{N}$ denotes the set of all non-negative integers. In such a case, we say that the double sequence $\{x_{m,n}\}$ converges to x. Note that x is unique and we say that x is the limit of $\{x_{m,n}\}$.

Definition 18.1.2 Let $\{x_{m,n}\}$ be a double sequence in K and $s \in K$. We say that

$$\sum_{m,n=0}^{\infty,\infty} x_{m,n} = s,$$

if

$$\lim_{m+n\to\infty} s_{m,n} = s,$$

where

$$s_{m,n} = \sum_{k,\ell=0}^{m,n} x_{k,\ell}, \quad m,n = 0,1,2,\ldots.$$

We say that the double series $\sum_{m,n=0}^{\infty,\infty} x_{m,n}$ converges to s and s is the sum of $\sum_{m,n=0}^{\infty,\infty} x_{m,n}$.

Remark 18.1.3 If $\lim_{m+n\to\infty} x_{m,n} = x$, then the double sequence $\{x_{m,n}\}$ is bounded.

The following important results are easily proved.

Theorem 18.1.4 $\lim_{m+n\to\infty} x_{m,n} = x$
if and only if

(i) $\lim_{n\to\infty} x_{m,n} = x, \quad m = 0,1,2,\ldots;$
(ii) $\lim_{m\to\infty} x_{m,n} = x, \quad n = 0,1,2,\ldots;$ *and*

(iii) for each $\epsilon > 0$*, there exists* $N \in \mathbb{N}$ *such that*

$$|x_{m,n} - x| < \epsilon, \quad \textit{for all } m, n \geq N,$$

which we write as

$$\lim_{m,n\to\infty} x_{m,n} = x.$$

(This is well-known as Pringsheim's definition of convergence of a double sequence.)

Theorem 18.1.5 $\sum_{m,n=0}^{\infty,\infty} x_{m,n}$ *converges if and only if*

$$\lim_{m+n\to\infty} x_{m,n} = 0.$$

Remark 18.1.6 In the case of simple sequences $\{x_n\}$, $x_n \in K$, $n = 0, 1, 2, \ldots$, it is very well-known (see [6]) that the series $\sum_{n=0}^{\infty} x_n$ converges if and only if

$$\lim_{n\to\infty} x_n = 0.$$

So, Theorem 18.1.5 implies that Definition 18.1.1 is the most suitable definition of convergence of a double sequence in the non-Archimedean context.

18.2 Silverman–Toeplitz, Schur, and Steinhaus Theorems

Definition 18.2.1 Given a 4-dimensional infinite matrix $A = (a_{m,n,k,\ell})$, $a_{m,n,k,\ell} \in K$, $m, n, k, \ell = 0, 1, 2, \ldots$, and a double sequence $x = \{x_{k,\ell}\}$, $x_{k,\ell} \in K$, $k, \ell = 0, 1, 2, \ldots$, by the A-transform of $x = \{x_{k,\ell}\}$, we mean the double sequence $Ax = \{(Ax)_{m,n}\}$, where

$$(Ax)_{m,n} = \sum_{k,\ell=0}^{\infty,\infty} a_{m,n,k,\ell} x_{k,\ell}, \quad m, n = 0, 1, 2, \ldots,$$

it is being assumed that the double series on the right converge. If $\lim_{m+n\to\infty} (Ax)_{m,n} = s$, we say that $\{x_{k,\ell}\}$ is A-summable or summable A to s. If $\lim_{m+n\to\infty} (Ax)_{m,n} = s$, whenever $\lim_{k+\ell\to\infty} x_{k,\ell} = t$, we say that A is convergence preserving or conservative. If A is conservative and $s = t$, we say that A is regular.

Natarajan and Srinivasan [5] proved the following theorem which characterizes a regular 4-dimensional matrix in terms of its entries.

Theorem 18.2.2 (Silverman–Toeplitz Theorem) *The 4-dimensional matrix* $A = (a_{m,n,k,\ell})$ *is regular if and only if*

$$\lim_{m+n\to\infty} a_{m,n,k,\ell} = 0, \quad k, \ell = 0, 1, 2, \ldots; \tag{18.1}$$

$$\lim_{m+n\to\infty} \sum_{k,\ell=0}^{\infty,\infty} a_{m,n,k,\ell} = 1; \tag{18.2}$$

$$\lim_{m+n\to\infty} \sup_{k\geq 0} |a_{m,n,k,\ell}| = 0, \quad \ell = 0, 1, 2, \ldots; \tag{18.3}$$

$$\lim_{m+n\to\infty} \sup_{\ell\geq 0} |a_{m,n,k,\ell}| = 0, \quad k = 0, 1, 2, \ldots; \tag{18.4}$$

and

$$\sup_{m,n,k,\ell} |a_{m,n,k,\ell}| < \infty. \tag{18.5}$$

The following definitions are needed in the sequel (see [7]).

Definition 18.2.3 $A = (a_{m,n,k,\ell})$ is called a Schur matrix if $\{(Ax)_{m,n}\} \in c_{ds}$, whenever $x = \{x_{k,\ell}\} \in \ell_{ds}^{\infty}$, where $c_{ds}, \ell_{ds}^{\infty}$, respectively, denote the spaces of convergent and bounded double sequences.

Definition 18.2.4 The double sequence $\{x_{m,n}\}$ in K is called a Cauchy sequence if for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that the set

$$\{(m,n),(k,\ell) \in \mathbb{N}^2 : |x_{m,n} - x_{k,\ell}| \geq \epsilon, \quad m,n,k,\ell \geq N\}$$

is finite.

It is now easy to prove the following result.

Theorem 18.2.5 *The double sequence* $\{x_{m,n}\}$ *in* K *is Cauchy if and only if*

$$\lim_{m+n\to\infty} |x_{m+1,n} - x_{m,n}| = 0;$$

and

$$\lim_{m+n\to\infty} |x_{m,n+1} - x_{m,n}| = 0.$$

Definition 18.2.6 If every Cauchy double sequence of a non-Archimedean normed linear space X converges to an element of X, then X is said to be double sequence complete or ds-complete.

For $x = \{x_{m,n}\} \in \ell_{ds}^{\infty}$, define

$$\|x\| = \sup_{m,n} |x_{m,n}|. \tag{18.6}$$

One can easily prove that ℓ_{ds}^{∞} is a non-Archimedean normed linear space which is ds-complete. With the same definition of norm for elements of c_{ds}, c_{ds} is a closed subspace of ℓ_{ds}^{∞}.

In the rest of this section, we shall suppose that K is a non-trivially valued, non-Archimedean field which is ds-complete. The following result was proved by Natarajan [7].

Theorem 18.2.7 (Schur's Theorem) *$A = (a_{m,n,k,\ell})$ is a Schur matrix if and only if*

$$\lim_{k+\ell\to\infty} a_{m,n,k,\ell} = 0, \quad m, n = 0, 1, 2, \ldots; \tag{18.7}$$

$$\lim_{m+n\to\infty} \sup_{k,\ell} |a_{m+1,n,k,\ell} - a_{m,n,k,\ell}| = 0; \tag{18.8}$$

and

$$\lim_{m+n\to\infty} \sup_{k,\ell} |a_{m,n+1,k,\ell} - a_{m,n,k,\ell}| = 0. \tag{18.9}$$

Using Theorems 18.2.2 and 18.2.7, we can now deduce the following important result.

Theorem 18.2.8 (Steinhaus Theorem) *A 4-dimensional infinite matrix $A = (a_{m,n,k,\ell})$ cannot be both a regular and a Schur matrix, i.e., given a regular matrix $A = (a_{m,n,k,\ell})$, there exists a bounded, divergent double sequence, which is not A-summable.*

Proof If $A = (a_{m,n,k,\ell})$ is regular, then (18.1) and (18.2) hold. If A were a Schur matrix too, $\{a_{m,n,k,\ell}\}_{m,n=0}^{\infty,\infty}$ is uniformly Cauchy with respect to $k, \ell = 0, 1, 2, \ldots$. Since K is ds-complete, $\{a_{m,n,k,\ell}\}_{m,n=0}^{\infty,\infty}$ converges to 0 uniformly with respect to $k, \ell = 0, 1, 2, \ldots$ so that

$$\lim_{m+n\to\infty} \sum_{k,\ell=0}^{\infty,\infty} a_{m,n,k,\ell} = 0,$$

a contradiction of (18.2), completing the proof of the theorem. □

18.3 Characterization of 2-Dimensional Schur Matrices

In this section, we prove a characterization of 2-dimensional Schur matrices using Definition 18.1.1 (see [8, 9]). We shall now explain a notation. Let X, Y denote spaces of simple sequences $\{x_k\}$ in K. Given a 2-dimensional infinite matrix $A = (a_{nk})$ in K, we write $A \in (X, Y)$ if $\{(Ax)_n\} \in Y$, whenever $x = \{x_k\} \in X$, where

$$(Ax)_n = \sum_{k=0}^{\infty} a_{nk}x_k,\ n = 0, 1, 2, \ldots,$$

assuming that the series on the right converge. Let c_0, c, ℓ_∞, respectively, denote the spaces of null, convergent, and bounded sequences. We now have the following result.

Theorem 18.3.1 *The following statements are equivalent:*

(a) $A \in (\ell_\infty, c_0)$*;*

(b) (i)

$$\lim_{k\to\infty} a_{nk} = 0,\ n = 0, 1, 2, \ldots; \tag{18.10}$$

and

(ii)

$$\lim_{n\to\infty} \sup_{k\geq 0} |a_{nk}| = 0. \tag{18.11}$$

(c) (i)

$$\lim_{n\to\infty} a_{nk} = 0,\ k = 0, 1, 2, \ldots; \tag{18.12}$$

and

(ii)

$$\lim_{k\to\infty} \sup_{n\geq 0} |a_{nk}| = 0. \tag{18.13}$$

(d)

$$\lim_{n+k\to\infty} a_{nk} = 0. \tag{18.14}$$

Proof Natarajan proved that (a) and (b) are equivalent (see [10, p.422]). For the equivalence of (b) and (c), see [11, p.129]. We will prove that (c) and (d) are equivalent. It is clear that (c) implies (d), using Theorem 18.1.4. Conversely, let

(d) hold. Using Theorem 18.1.4 again, (18.10), (18.12) hold and

$$\lim_{n,k\to\infty} a_{nk} = 0. \tag{18.15}$$

Equation (18.10), along with (18.15), implies that (18.13) holds. Thus (c) holds, i.e., (d) implies (c), completing the proof of the theorem. □

It is now easy to prove the following.

Theorem 18.3.2 *The following statements are equivalent:*

(a) $A = (a_{nk}) \in (\ell_\infty, c)$;
(b) *(i) (18.10) holds;*
and
(ii)

$$\lim_{n\to\infty} \sup_{k\geq 0} |a_{n+1,k} - a_{nk}| = 0. \tag{18.16}$$

(c) *(i) (18.10) holds;*
(ii)

$$\lim_{n\to\infty} a_{nk} = \delta_k \text{ exists}, \ k = 0, 1, 2, \ldots; \tag{18.17}$$

and
(iii)

$$\lim_{k\to\infty} \sup_{n\geq 0} |a_{n+1,k} - a_{nk}| = 0. \tag{18.18}$$

(d) *(i) (18.10) holds;*
and
(ii)

$$\lim_{n+k\to\infty} (a_{n+1,k} - a_{nk}) = 0. \tag{18.19}$$

(e) *(i) (18.10) holds;*
(ii) (18.17) holds;
and
(iii)

$$\lim_{n,k\to\infty} (a_{n+1,k} - a_{nk}) = 0.$$

18.4 The Nörlund Method for Double Sequences

We now introduce the Nörlund method (or mean) for double sequences in K [5].

Definition 18.4.1 Let $p_{m,n} \in K, m, n = 0, 1, 2, \ldots$ with

$$|p_{ij}| < |p_{0,0}|, \ (i, j) \neq (0, 0), \ i, j = 0, 1, 2, \ldots.$$

Let

$$P_{m,n} = \sum_{i,j=0}^{m,n} p_{i,j}, \ m, n = 0, 1, 2, \ldots.$$

Given a double sequence $\{s_{m,n}\}$, we define

$$\begin{aligned} \sigma_{m,n} &= (N, p_{m,n})(\{s_{m,n}\}) \\ &= \frac{\sum_{i,j=0}^{m,n} p_{m-i,n-j} s_{i,j}}{P_{m,n}}, \ m, n = 0, 1, 2, \ldots. \end{aligned}$$

If $\lim_{m+n\to\infty} \sigma_{m,n} = \sigma$, we say that $\{s_{m,n}\}$ is $(N, p_{m,n})$ summable to σ, written as

$$s_{m,n} \to \sigma(N, p_{m,n}).$$

The summability method $(N, p_{m,n})$ is called a Nörlund method (or mean). Any double series $\sum_{m,n=0}^{\infty,\infty} x_{m,n}$ is said to be $(N, p_{m,n})$ summable to σ, if

$$s_{m,n} \to \sigma(N, p_{m,n}),$$

where

$$s_{m,n} = \sum_{i,j=0}^{m,n} x_{i,j}, \ m, n = 0, 1, 2, \ldots.$$

Using Theorem 18.2.2, it is easy to prove the following result.

Theorem 18.4.2 *The Nörlund method $(N, p_{m,n})$ is regular if and only if*

$$\lim_{m+n\to\infty} \sup_{0\le j\le n} |p_{m-i,n-j}| = 0, \ 0 \le i \le m; \tag{18.20}$$

and

$$\lim_{m+n\to\infty} \sup_{0\le i\le m} |p_{m-i,n-j}| = 0, \ 0 \le j \le n. \tag{18.21}$$

In the remaining part of the present section, we shall suppose that $(N, p_{m,n})$, $(N, q_{m,n})$ are regular Nörlund methods such that each row and each column of the 2-dimensional infinite matrices $(p_{m,n})$, $(q_{m,n})$ is a regular Nörlund method for simple sequences. Under this assumption, we can prove the following results (see [5] for details of proof).

Theorem 18.4.3 *Any two such regular Nörlund methods $(N, p_{m,n})$, $(N, q_{m,n})$ are consistent.*

Theorem 18.4.4

$$(N, p_{m,n}) \subseteq (N, q_{m,n}),$$

$$\text{i.e., whenever } s_{m,n} \to \sigma(N, p_{m,n}), s_{m,n} \to \sigma(N, q_{m,n}) \text{ too,}$$

if and only if

$$\lim_{m+n\to\infty} k_{m,n} = 0,$$

where $\{k_{m,n}\}$ is defined by:

$$k(x, y) = \sum_{m,n=0}^{\infty,\infty} k_{m,n} x^m y^n = \frac{q(x, y)}{p(x, y)},$$

$$p(x, y) = \sum_{m,n=0}^{\infty,\infty} p_{m,n} x^m y^n,$$

and

$$q(x, y) = \sum_{m,n=0}^{\infty,\infty} q_{m,n} x^m y^n.$$

Theorem 18.4.5 *The Nörlund methods $(N, p_{m,n})$, $(N, q_{m,n})$ are equivalent, i.e., $(N, p_{m,n}) \subseteq (N, q_{m,n})$ and vice versa if and only if*

$$\lim_{m+n\to\infty} k_{m,n} = 0$$

and

$$\lim_{m+n\to\infty} \ell_{m,n} = 0,$$

where $\{\ell_{m,n}\}$ *is defined by:*

$$\ell(x, y) = \sum_{m,n=0}^{\infty,\infty} \ell_{m,n} x^m y^n = \frac{p(x, y)}{q(x, y)}$$

and $\{k_{m,n}\}$ *is defined as in Theorem 18.4.4.*

Natarajan [12] is the motivation for proving the following Tauberian theorem.

Theorem 18.4.6 ([13]) *If* $\sum_{m,n=0}^{\infty,\infty} a_{m,n}$ *is* $(N, p_{m,n})$ *summable to* s*,* $(N, p_{m,n})$ *being regular and if*

$$a_{m,n} \to \ell^*, \ m + n \to \infty,$$

then $\sum_{m,n=0}^{\infty,\infty} a_{m,n}$ *converges to* s*.*

Proof We first note the following: Since $|p_{k,\ell}| < |p_{0,0}|$, $(k, \ell) \neq (0, 0)$, $k, \ell = 0, 1, 2, \ldots$, it follows that

$$p_{0,0} \neq 0.$$

Since the valuation of K is non-Archimedean,

$$|P_{m,n}| = |p_{0,0}|, \ m, n = 0, 1, 2, \ldots$$

(see [6, p. 8,Theorem 2.2]). Also, since $(N, p_{m,n})$ is regular,

$$\lim_{m+n\to\infty} a_{m,n,0,0} = 0,$$

using (18.1). So

$$\lim_{m+n\to\infty} |a_{m,n,0,0}| = 0,$$

$$\text{i.e.,} \lim_{m+n\to\infty} \left|\frac{p_{m,n}}{P_{m,n}}\right| = 0,$$

$$\text{i.e.,} \lim_{m+n\to\infty} |p_{m,n}| = 0, \ \text{since } |P_{m,n}| = |p_{0,0}|,$$

$$\text{i.e.,} \lim_{m+n\to\infty} p_{m,n} = 0.$$

So $\sum_{m,n=0}^{\infty,\infty} p_{m,n}$ converges, in view of Theorem 18.1.5. Consequently,

$$\lim_{m+n\to\infty} P_{m,n} \text{ exists (say) } = P.$$

Note that $P \neq 0$, since $|P_{m,n}| = |p_{0,0}|, m, n = 0, 1, 2, \ldots$, and $p_{0,0} \neq 0$. Let $\{t_{m,n}\}$ be the $(N, p_{m,n})$-transform of the double sequence $\{s_{m,n}\}$, where

$$s_{m,n} = \sum_{k,\ell=0}^{m,n} a_{k,\ell}, \; m, n = 0, 1, 2, \ldots.$$

Then,

$$t_{m,n} = \frac{\sum_{i,j=0}^{m,n} p_{m-i,n-j} s_{i,j}}{P_{m,n}}, \; m, n = 0, 1, 2, \ldots$$

and

$$\lim_{m+n\to\infty} t_{m,n} = s.$$

Thus,

$$\lim_{m+n\to\infty} \sum_{i,j=0}^{m,n} p_{m-i,n-j} s_{i,j} = sP.$$

Now,

$$\begin{aligned}
\sum_{i,j=0}^{m,n} p_{m-i,n-j} s_{i,j} &= \sum_{i,j=0}^{m,n} p_{m-i,n-j} \left(\sum_{k,\ell=0}^{i,j} a_{k,\ell} \right) \\
&= \sum_{i,j=0}^{m,n} a_{i,j} P_{m-i,n-j} \\
&= \sum_{i,j=0}^{m,n} (a_{i,j} - \ell^*)(P_{m-i,n-j} - P) \\
&\quad + \ell^* \sum_{i,j=0}^{m,n} P_{m-i,n-j}
\end{aligned}$$

$$
\begin{aligned}
&+ P \sum_{i,j=0}^{m,n} (a_{i,j} - \ell^*) \\
&= \sum_{i,j=0}^{m,n} (a_{i,j} - \ell^*)(P_{m-i,n-j} - P) \\
&+ \ell^* \sum_{i,j=0}^{m,n} P_{i,j} + P \sum_{i,j=0}^{m,n} (a_{i,j} - \ell^*). \qquad (18.22)
\end{aligned}
$$

We note that

$$\lim_{m+n\to\infty} \sum_{i,j=0}^{m,n} (a_{i,j} - \ell^*)(P_{m-i,n-j} - P) = 0,$$

since $\lim_{m+n\to\infty} a_{m,n} = \ell^*$ and $\lim_{m+n\to\infty} P_{m,n} = P$, using Theorem 2 of [14]. Taking limit as $m + n \to \infty$ in (18.22), we have

$$
\begin{aligned}
sP &= \lim_{m+n\to\infty} \left[\ell^* \sum_{i,j=0}^{m,n} P_{i,j} + P \left\{ \sum_{i,j=0}^{m,n} a_{i,j} - mn\ell^* \right\} \right] \\
&= \lim_{m+n\to\infty} \left[P s_{m,n} + \ell^* \left\{ \sum_{i,j=0}^{m,n} (P_{i,j} - P) \right\} \right] \\
&= P \lim_{m+n\to\infty} s_{m,n} + \ell^* \sum_{i,j=0}^{\infty,\infty} (P_{i,j} - P).
\end{aligned}
$$

Thus, $\lim_{m+n\to\infty} s_{m,n}$ exists and

$$\lim_{m+n\to\infty} s_{m,n} = \frac{1}{P} \left[sP - \ell^* \sum_{m,n=0}^{\infty,\infty} (P_{m,n} - P) \right].$$

In other words, $\sum_{m,n=0}^{\infty,\infty} a_{m,n}$ converges and so $\lim_{m+n\to\infty} a_{m,n} = 0$, proving that

$$\ell^* = 0.$$

It now follows that

$$\lim_{m+n\to\infty} s_{m,n} = s,$$

i.e., $\sum_{m,n=0}^{\infty,\infty} a_{m,n}$ converges to s, completing the proof of the theorem. □

18.5 Weighted Mean Method for Double Sequences

We now introduce weighted mean methods for double sequences and extend theorems dealing with weighted mean methods for simple sequences (for details, refer to [15]).

Definition 18.5.1 The $(\overline{N}, p_{m,n})$ method, called the weighted mean method, is defined by the 4-dimensional infinite matrix $(a_{m,n,k,\ell})$, $m, n, k, \ell = 0, 1, 2, \ldots,$ where

$$(a_{m,n,k,\ell}) = \begin{cases} \frac{p_{k,\ell}}{P_{m,n}}, & \text{if } k \le m,\ \ell \le n; \\ 0, & \text{if } k > m \text{ or } \ell > n, \end{cases}$$

$$P_{m,n} = \sum_{i,j=0}^{m,n} p_{i,j},\ m, n = 0, 1, 2, \ldots,$$

with the sequence $\{p_{m,n}\}$ of weights satisfying the conditions

$$p_{m,n} \neq 0,\ m, n = 0, 1, 2, \ldots, \tag{18.23}$$

for each fixed pair (i, j),

$$\begin{aligned} |p_{k,\ell}| \le |P_{i,j}|,\ & k = 0, 1, 2, \ldots, i; \\ & i = 0, 1, 2, \ldots; \\ & \ell = 0, 1, 2, \ldots, j; \\ & j = 0, 1, 2, \ldots. \end{aligned} \tag{18.24}$$

Remark 18.5.2 From (18.24), it is clear that for every fixed $i = 0, 1, 2, \ldots,$

$$\begin{aligned} |p_{i,\ell}| \le |P_{i,j}|,\ & \ell = 0, 1, 2, \ldots, j; \\ & j = 0, 1, 2, \ldots \end{aligned} \tag{18.25}$$

and for every fixed $j = 0, 1, 2, \ldots,$

$$|p_{k,j}| \le |P_{i,j}|, \;\; k = 0, 1, 2, \ldots, i; \quad i = 0, 1, 2, \ldots. \tag{18.26}$$

Note that (18.24) is equivalent to

$$\max_{\substack{0 \le k \le i \\ 0 \le \ell \le j}} |p_{k,\ell}| \le |P_{i,j}|, \;\; i, j = 0, 1, 2, \ldots. \tag{18.27}$$

Equation (18.25) is equivalent to

$$\max_{0 \le \ell \le j} |p_{i,\ell}| \le |P_{i,j}|, \;\; j = 0, 1, 2, \ldots, \tag{18.28}$$

while (18.26) is equivalent to

$$\max_{0 \le k \le i} |p_{k,j}| \le |P_{i,j}|, \;\; i = 0, 1, 2, \ldots. \tag{18.29}$$

Since the valuation of K is non-Archimedean,

$$|P_{i,j}| \le \max_{\substack{0 \le k \le i \\ 0 \le \ell \le j}} |p_{k,\ell}|, \;\; i, j = 0, 1, 2, \ldots. \tag{18.30}$$

Combining (18.27) and (18.30), we have, for every fixed pair (i, j),

$$|P_{i,j}| = \max_{\substack{0 \le k \le i \\ 0 \le \ell \le j}} |p_{k,\ell}|, \;\; i, j = 0, 1, 2, \ldots. \tag{18.31}$$

Using (18.31), we have

$$P_{m,n} \neq 0, \;\; m, n = 0, 1, 2, \ldots. \tag{18.32}$$

Remark 18.5.3 Equation (18.31) implies

$$|P_{m+1,n+1}| \ge |P_{m,n}|; \tag{18.33}$$

$$|P_{m,n+1}| \ge |P_{m,n}|; \tag{18.34}$$

and

$$|P_{m+1,n}| \ge |P_{m,n}|. \tag{18.35}$$

Proof Using (18.31), we have

$$\begin{aligned}|P_{m+1,n+1}| &= \max_{\substack{0\le k\le m+1\\0\le \ell\le n+1}} |p_{k,\ell}| \\ &= \max\left[\max_{\substack{0\le k\le m\\0\le \ell\le n}} |p_{k,\ell}|, |p_{m,n+1}|, \right. \\ &\qquad \left. |p_{m+1,n}|, |p_{m+1,n+1}|\right] \\ &= \max\left[|P_{m,n}|, |p_{m,n+1}|, |p_{m+1,n}|, \right. \\ &\qquad \left. |p_{m+1,n+1}|\right] \\ &\ge |P_{m,n}|.\end{aligned}$$

In a similar fashion, we can prove that (18.31) implies (18.34) and (18.35). □

Using Theorem 18.2.2, we can prove the following result.

Theorem 18.5.4 *The weighted mean method* $(\overline{N}, p_{m,n})$ *is regular if and only if*

$$\lim_{m+n\to\infty} |P_{m,n}| = \infty; \tag{18.36}$$

$$\lim_{m+n\to\infty} \frac{\max_{0\le k\le m} |p_{k,\ell}|}{|P_{m,n}|} = 0,\ \ \ell = 0, 1, 2, \ldots; \tag{18.37}$$

and

$$\lim_{m+n\to\infty} \frac{\max_{0\le \ell\le n} |p_{k,\ell}|}{|P_{m,n}|} = 0,\ \ k = 0, 1, 2, \ldots. \tag{18.38}$$

The next result puts a limitation on the $(\overline{N}, p_{m,n})$ summability of a double sequence $\{s_{m,n}\}$ (see [15]).

Theorem 18.5.5 (Limitation Theorem) *If* $\{s_{m,n}\}$ *is* $(\overline{N}, p_{m,n})$ *summable to s, then*

$$s_{m,n} - s = o\left(\frac{P_{m,n}}{p_{m,n}}\right),\ \ m+n\to\infty, \tag{18.39}$$

in the sense that

$$\frac{p_{m,n}}{P_{m,n}}(s_{m,n} - s) \to 0,\ \ m+n\to\infty.$$

Proof Let $\{t_{m,n}\}$ be the $(\overline{N}, p_{m,n})$-transform of $\{s_{m,n}\}$. Then,

$$\begin{aligned}
&\left|\frac{p_{m,n}}{P_{m,n}}(s_{m,n}-s)\right| \\
&= \left|\frac{\begin{array}{r}(P_{m,n}t_{m,n}-P_{m,n-1}t_{m,n-1}-P_{m-1,n}t_{m-1,n}\\ +P_{m-1,n-1}t_{m-1,n-1})\\ -(P_{m,n}-P_{m,n-1}-P_{m-1,n}+P_{m-1,n-1})s\end{array}}{P_{m,n}}\right| \\
&= \left|\begin{array}{r}(t_{m,n}-s)-\frac{P_{m,n-1}}{P_{m,n}}(t_{m,n-1}-s)-\frac{P_{m-1,n}}{P_{m,n}}(t_{m-1,n}-s)\\ +\frac{P_{m-1,n-1}}{P_{m,n}}(t_{m-1,n-1}-s)\end{array}\right| \\
&\le \max\Bigg[|t_{m,n}-s|, \left|\frac{P_{m,n-1}}{P_{m,n}}\right| |t_{m,n-1}-s|, \\
&\qquad \left|\frac{P_{m-1,n}}{P_{m,n}}\right| |t_{m-1,n}-s|, \\
&\qquad \left|\frac{P_{m-1,n-1}}{P_{m,n}}\right| |t_{m-1,n-1}-s|\Bigg] \\
&\le \max\left[|t_{m,n}-s|, |t_{m,n-1}-s|, |t_{m-1,n}-s|, |t_{m-1,n-1}-s|\right]
\end{aligned}$$

in view of Remark 18.5.3. Since $\lim_{m+n\to\infty} t_{m,n} = s$, it follows that

$$\lim_{m+n\to\infty}\left|\frac{p_{m,n}}{P_{m,n}}(s_{m,n}-s)\right| = 0,$$

$$\text{i.e., } s_{m,n}-s = o\left(\frac{P_{m,n}}{p_{m,n}}\right),\ m+n\to\infty,$$

which completes the proof of the theorem. □

We now list a few inclusion theorems involving weighted mean methods for double sequences (see [11, 15] for details).

Theorem 18.5.6 (Comparison Theorem for Two Weighted Mean Methods for Double Sequences) *Let* $(\overline{N}, p_{m,n})$, $(\overline{N}, q_{m,n})$ *be two weighted mean methods such that*

$$q_{m,n} = O(p_{m,n}),\ m+n\to\infty \tag{18.40}$$

in the sense that there exists $M > 0$ *such that*

$$\left|\frac{q_{m,n}}{p_{m,n}}\right| \leq M, \; m, n = 0, 1, 2, \ldots$$

and

$$P_{m,n} = o(Q_{m,n}), \; m + n \to \infty \tag{18.41}$$

in the sense that

$$\left|\frac{P_{m,n}}{Q_{m,n}}\right| \to 0, \; m + n \to \infty,$$

where $P_{m,n} = \sum_{k,\ell=0}^{m,n} p_{k,\ell}$*,* $Q_{m,n} = \sum_{k,\ell=0}^{m,n} q_{k,\ell}$*,* $m, n = 0, 1, 2, \ldots$.
Then,

$$\begin{aligned}&(\overline{N}, p_{m,n}) \subseteq (\overline{N}, q_{m,n}),\\ &\text{i.e., } s_{m,n} \to s(\overline{N}, p_{m,n}) \text{ implies that}\\ &\quad s_{m,n} \to s(\overline{N}, q_{m,n}).\end{aligned}$$

Theorem 18.5.7 (Comparison Theorem for a $(\overline{N}, p_{m,n})$ Method and a Regular Matrix Method) *Let* $(\overline{N}, p_{m,n})$ *be a weighted mean method and* $A = (a_{m,n,k,\ell})$ *be a regular matrix. If*

$$\lim_{k+\ell\to\infty} \frac{a_{m,n,k,\ell}}{p_{k,\ell}} P_{k,\ell} = 0, \; m, n = 0, 1, 2, \ldots \tag{18.42}$$

and

$$P_{m,n} = O(p_{m,n}), \; m + n \to \infty, \tag{18.43}$$

$$\text{i.e., } \left|\frac{P_{m,n}}{p_{m,n}}\right| \leq M, \; M > 0, \; m, n = 0, 1, 2, \ldots,$$

then

$$(\overline{N}, p_{m,n}) \subseteq A.$$

Theorem 18.5.8 (Comparison Theorem for a $(\overline{N}, p_{m,n})$ Method and a Regular Matrix Method) *Let* $(\overline{N}, p_{m,n})$ *be a weighted mean method and* $A = (a_{m,n,k,\ell})$ *be* *a regular matrix. If*

$$\lim_{k+\ell\to\infty} a_{m,n,k,\ell}\frac{P_{k,\ell}}{p_{k,\ell}} = 0,\ m,n = 0,1,2,\ldots; \tag{18.44}$$

$$\sup_{m,n,k,\ell}\left|a_{m,n,k,\ell}\frac{P_{k,\ell}}{p_{k,\ell}}\right| < \infty; \tag{18.45}$$

$$\lim_{m+n\to\infty}\sup_{k\geq 0}\left|a_{m,n,k,\ell}\frac{P_{k,\ell}}{p_{k,\ell}}\right| = 0,\ \ell = 0,1,2,\ldots; \tag{18.46}$$

and

$$\lim_{m+n\to\infty}\sup_{\ell\geq 0}\left|a_{m,n,k,\ell}\frac{P_{k,\ell}}{p_{k,\ell}}\right| = 0,\ k = 0,1,2,\ldots, \tag{18.47}$$

then

$$(\overline{N}, p_{m,n}) \subseteq A.$$

18.6 $(M, \lambda_{m,n})$ Method (or Natarajan Method) for Double Sequences

In an attempt to generalize the Nörlund method, the (M, λ_n) method for simple sequences in K was introduced earlier by Natarajan [16]. We now recall the definition.

Definition 18.6.1 Let $\{\lambda_n\}$ be a sequence in K such that $\lim_{n\to\infty}\lambda_n = 0$. The (M, λ_n) method is defined by the 2-dimensional infinite matrix (a_{nk}), where

$$a_{nk} = \begin{cases}\lambda_{n-k}, & \text{if } k \leq n;\\ 0, & \text{if } k > n.\end{cases}$$

Remark 18.6.2 Note that the (M, λ_n) method is a non-trivial method. In this context, we also note that the (M, λ_n) method reduces to the Y-method of Srinivasan [17], when $K = \mathbb{Q}_p$, the p-adic field for a prime p, $\lambda_0 = \lambda_1 = \frac{1}{2}$ and $\lambda_n = 0, n \geq 2$. More generally, the (M, λ_n) method reduces to the Y-method when K is a complete, non-trivially valued, non-Archimedean field of characteristic zero, $\lambda_0 = \lambda_1 = \frac{1}{2}$ and $\lambda_n = 0, n \geq 2$.

We now introduce the $(M, \lambda_{m,n})$ method for double sequences and list some of the results (see [18]).

Definition 18.6.3 Let $\{\lambda_{m,n}\}$ be a double sequence in K such that

$$\lim_{m+n\to\infty} \lambda_{m,n} = 0.$$

The method $(M, \lambda_{m,n})$ is defined by the 4-dimensional infinite matrix $(a_{m,n,k,\ell})$, where

$$(a_{m,n,k,\ell}) = \begin{cases} \lambda_{m-k,n-\ell}, & \text{if } k \le m \text{ and } \ell \le n; \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 18.6.4 *The method $(M, \lambda_{m,n})$ is regular if and only if*

$$\sum_{m,n=0}^{\infty\infty} \lambda_{m,n} = 1.$$

For Theorem 18.6.5, let $(M, \lambda_{m,n})$, $(M, \mu_{m,n})$ be regular methods such that each row and each column of the 2-dimensional infinite matrices $(\lambda_{m,n})$, $(\mu_{m,n})$ is a regular Natarajan method for simple sequences.

Theorem 18.6.5 *Any two such methods $(M, \lambda_{m,n})$, $(M, \mu_{m,n})$ are consistent.*

Theorem 18.6.6 *If $(M, \lambda_{m,n})$, $(M, \mu_{m,n})$ are regular, then*

$$(M, \lambda_{m,n}) \subseteq (M, \mu_{m,n})$$

if and only if

$$\lim_{m+n\to\infty} k_{m,n} = 0 \text{ and } \sum_{m,n=0}^{\infty,\infty} k_{m,n} = 1,$$

where $\{k_{m,n}\}$ is defined by

$$k(x, y) = \sum_{m,n=0}^{\infty,\infty} k_{m,n} x^m y^n = \frac{\mu(x, y)}{\lambda(x, y)},$$

$$\lambda(x, y) = \sum_{m,n=0}^{\infty,\infty} \lambda_{m,n} x^m y^n,$$

$$\mu(x, y) = \sum_{m,n=0}^{\infty,\infty} \mu_{m,n} x^m y^n.$$

Theorem 18.6.7 *The regular methods $(M, \lambda_{m,n})$, $(M, \mu_{m,n})$ are equivalent if and only if*

$$\lim_{m+n\to\infty} k_{m,n} = 0, \sum_{m,n=0}^{\infty,\infty} k_{m,n} = 1;$$

and

$$\lim_{m+n\to\infty} h_{m,n} = 0, \sum_{m,n=0}^{\infty,\infty} h_{m,n} = 1,$$

where $\{h_{m,n}\}$ is defined by

$$h(x, y) = \sum_{m,n=0}^{\infty,\infty} h_{m,n} x^m y^n = \frac{\lambda(x, y)}{\mu(x, y)}$$

and $\{k_{m,n}\}$ is defined as in Theorem 18.6.6.

We now record some results on the Cauchy multiplication of $(M, \lambda_{m,n})$-summable double sequences and double series (see [19]).

Theorem 18.6.8 *If $\lim_{m+n\to\infty} a_{m,n} = 0$ and $\{b_{m,n}\}$ is $(M, \lambda_{m,n})$ summable to B, then $\{c_{m,n}\}$ is $(M, \lambda_{m,n})$ summable to AB, where*

$$c_{m,n} = \sum_{k,\ell=0}^{m,n} a_{m-k,n-\ell} b_{k,\ell}, \ m, n = 0, 1, 2, \ldots$$

and

$$\sum_{m,n=0}^{\infty,\infty} a_{m,n} = A.$$

Theorem 18.6.9 *If $\lim_{m+n\to\infty} a_{m,n} = 0$ and $\sum_{m,n=0}^{\infty,\infty} b_{m,n}$ is $(M, \lambda_{m,n})$ summable to B, then $\sum_{m,n=0}^{\infty,\infty} c_{m,n}$ is $(M, \lambda_{m,n})$ summable to AB, where*

$$c_{m,n} = \sum_{k,\ell=0}^{m,n} a_{m-k,n-\ell} b_{k,\ell}, \ m, n = 0, 1, 2, \ldots$$

and

$$\sum_{m,n=0}^{\infty,\infty} a_{m,n} = A.$$

Theorem 18.6.10 *If* $\sum_{m,n=0}^{\infty,\infty} a_{m,n}$ *is* $(M, \lambda_{m,n})$ *summable to* A*,* $\sum_{m,n=0}^{\infty,\infty} b_{m,n}$ *is* $(M, \mu_{m,n})$ *summable to* B*, then* $\sum_{m,n=0}^{\infty,\infty} c_{m,n}$ *is* $(M, \gamma_{m,n})$ *summable to* AB*, where*

$$c_{m,n} = \sum_{k,\ell=0}^{m,n} a_{m-k,n-\ell} b_{k,\ell},\ m, n = 0, 1, 2, \ldots$$

and

$$\gamma_{m,n} = \sum_{k,\ell=0}^{m,n} \lambda_{m-k,n-\ell} \mu_{k,\ell},\ m, n = 0, 1, 2, \ldots.$$

We have the following result too.

Theorem 18.6.11 *Let* $(M, \lambda_{m,n})$*,* $(M, \mu_{m,n})$ *be regular methods. Then,* $(M, \lambda_{m,n})$ $(M, \mu_{m,n})$ *is regular too, where, we define, for* $x = \{x_{m,n}\}$*,*

$$((M, \lambda_{m,n})(M, \mu_{m,n}))(x) = (M, \lambda_{m,n})((M, \mu_{m,n})(x)).$$

References

1. M. Zeltser, On conservative matrix methods for double sequence spaces. Acta Math. Hungar. **95**, 225–242 (2002)
2. B.V. Limaye, M. Zeltser, On the Pringsheim convergence of double series. Proc. Est. Acad. Sci. Phys. Math. **58**(2), 108–121 (2009)
3. M. Mursaleen, S.A. Mohiuddine, *Convergence Methods for Double Sequences and Applications* (Springer, Berlin, 2014)
4. A. Aasma, H. Dutta, P.N. Natarajan, *An Introductory Course in Summability Theory* (Wiley, New York, 2017)
5. P.N. Natarajan, V. Srinivasan, Silverman–Toeplitz theorem for double sequences and series and its application to Nörlund means in non-Archimedean fields. Ann. Math. Blaise Pascal **9**, 85–100 (2002)
6. G. Bachman, *Introduction to p-Adic Numbers and Valuation Theory* (Academic Press, London, 1964)
7. P.N. Natarajan, The Schur and Steinhaus theorems for 4-dimensional matrices in ultrametric fields. Comment. Math. Prace Mat. **51**, 203–209 (2011)

8. P.N. Natarajan, Some characterizations of Schur matrices in ultrametric fields. Comment. Math. Prace Mat. **52**, 137–142 (2012)
9. P.N. Natarajan, An addendum to the paper "Some characterizations of Schur matrices in ultrametric fields. Comment. Math. Prace Mat. **53**, 81–82 (2013)
10. P.N. Natarajan, The Steinhaus theorem for Toeplitz matrices in non-Archimedean fields. Comment. Math. Prace Mat. **20**, 417–422 (1978)
11. P.N. Natarajan, *An Introduction to Ultrametric Summability Theory*, 2nd edn. (Springer, Berlin, 2015)
12. P.N. Natarajan, *Some Tauberian Theorems in Non-Archimedean Fields, p-Adic Functional Analysis*. Lecture Notes in Pure and Applied Mathematics, vol. 192 (Marcel Dekker, New York, 1997), pp. 297–303
13. P.N. Natarajan, A Tauberian theorem for the Nörlund means for double series in ultrametric fields. Indian J. Math. Proc. Sixth Dr. George Bachman Memorial Conf. **57**, 33–38 (2015)
14. P.N. Natarajan, S. Sakthivel, Multiplication of double series and convolution of double infinite matrices in non-Archimedean fields. Indian J. Math. **50**, 125–133 (2008)
15. P.N. Natarajan, S. Sakthivel, Weighted means for double sequences in non-Archimedean fields. Indian J. Math. **48**, 201–220 (2006)
16. P.N. Natarajan, Product theorems for certain summability methods in non-Archimedean fields. Ann. Math. Blaise Pascal **10**, 261–267 (2003)
17. V.K. Srinivasan, On certain summation processes in the p-adic field. Indag. Math. **27**, 319–325 (1965)
18. P.N. Natarajan, Natarajan summability method for double sequences and double series. Adv. Dev. Math. Sci. **6**, 9–17 (2014)
19. P.N. Natarajan, Some properties of the Natarajan method of summability for double sequences in ultrametric fields. Int. J. Phys. Math. Sci. **6**, 22–27 (2016)

Chapter 19
On Approximate Solutions of Linear and Nonlinear Singular Integral Equations

Nizami Mustafa and Veysel Nezir

Abstract Singular integral equation theory has broad applications to theoretical and practical investigations in mathematics, mathematical physics, hydrodynamic and elasticity theory. This fact motivated many researchers to work on this field and their studies have showed that finding approximate solutions of linear and nonlinear singular integral equations in Banach spaces provides many applications even if their definite solutions cannot be found or if there are difficulties in finding them. Thus, the central theme of the recent studies is to develop effective approximate solution methods for the linear and nonlinear singular integral equations in Banach spaces. This chapter has been devoted to investigating approximate solutions of linear and nonlinear singular integral equations in Banach spaces using technical methods such as collocation method, quadrature method, Newton–Kantorovich method, monotonic operators method, and fixed point theory depending on the type of the equations. We provide sufficient conditions for the convergence of these methods and investigate some properties.

19.1 Introduction

The theory of nonlinear singular integral equations (NLSIEs) has developed significant importance over the last few years as many engineering problems of applied mechanics and applied mathematics are reduced to the solution of such types of nonlinear equations.

It is well known that the solutions to a host of familiar problems of mathematical physics, such as elasticity, plasticity, thermo-elasticity, and fluid mechanics have been reduced to solving equations of the NLSIE type; besides, the application area of NLSIEs is outstanding in connection to the theories of elasticity, viscoelasticity,

N. Mustafa (✉) · V. Nezir
Department of Mathematics, Faculty of Science and Letters, Kafkas University, Kars, Turkey

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_19

thermo-elasticity, hydrodynamics, fluid mechanics, and many other fields outside mathematical physics [42, 44–47].

Recent investigations on this topic have observed that, for many nonlinear differential equation systems, the solutions of the Dirichlet boundary-value problems which have partial derivatives and are defined in a region can be reduced to solving equations of the NLSIE type [9, 15, 56, 86]. The solution of the seismic wave equation—of great importance in elastodynamics—is investigated by reducing it to the solution of NLSIE by using Hilbert transformation [9].

Many problems of applied mechanics are to be reduced to the solution of an NLSIE. This approach involves a nonlinear generalization of the linear singular integral equations of the finite-part type and the multidimensional form, which have been investigated by Ladopoulos [44, 45], Bojarskii [10], Lackau [41], Monahov [61]. Moreover, a nonlinear integro-differential equation analysis was proposed by Ladopoulos [62], with applications to some basic problems of orthotropic shallow spherical shell stress analysis. Beyond these applications, Ladopoulos [47], Tutshke [90] have examined the existence and the uniqueness of the solution of the NLSIEs defined in Banach spaces while investigating the application of such types of equations in two-dimensional fluid mechanics.

As it is known, the analytical solutions of NLSIEs can only be found in certain special cases. In the absence of analytical solutions, these types of equations are usually solved by approximation methods. From this point of view, it is important to know how to solve the NLSIEs with approximation methods. Over the past few years, there have been many studies of the approximate solutions of NLSIEs [1, 5, 8, 11, 16, 17, 19–21, 26, 28, 29, 32, 37, 43, 48–55, 59, 60, 68, 69, 74–78, 80–82, 84, 85, 87, 88, 91, 94–97].

Note that this chapter is formed by the works [69–73, 75].

19.2 Newton–Kantorovich Method for Two-Dimensional Nonlinear Singular Integral Equations

In this section, we investigate the following two-dimensional NLSIE and apply Newton–Kantorovich method to find its approximate solutions.

$$B(\varphi)(z) \equiv F(z, \varphi(z), T_G f(\cdot, \varphi(\cdot))(z), \Pi_G g(\cdot, \varphi(\cdot))(z)) = 0, \quad z \in G, \tag{19.1}$$

where $f, g : D_0 \to \mathbb{C}$, $F : G \to \mathbb{C}$ are known continuous functions in their domains of definition, $\varphi(z)$ is an unknown function, $D_0 = \{(z, \varphi) : z \in \bar{G} = \partial G \cup \overset{\circ}{G}, \ \varphi \in \mathbb{C}\} = \bar{G} \times \mathbb{C}$, ∂G denotes the boundary of the region $G \subset \mathbb{C}$, $\mathbb{C}$ is the complex plane, $\overset{\circ}{G}$ is a set of interior points of the region G, while its closure $\bar{G} = \partial G \bigcup \overset{\circ}{G}$ and $D = \{(z, \varphi, v, w) : z \in \bar{G}, \ \ \varphi, v, w \in \mathbb{C}\} = \bar{G} \times \mathbb{C}^3$.

Moreover,

$$T_G h(\cdot)(z) = \frac{-1}{\pi} \iint_G \frac{h(\zeta)}{\zeta - z} d\xi d\eta, \quad \Pi_G h(\cdot)(z) = \frac{-1}{\pi} \iint_G \frac{h(\zeta)}{(\zeta - z)^2} d\xi d\eta.$$

The existence and uniqueness of the solution of Eq. (19.1) was proved by Mustafa and Ardil [74]. In order to derive the approximate solution of Eq. (19.1), we show that the nonlinear operator $B(\varphi)$ defined by Eq. (19.1) is the Freshet differentiable operator. Furthermore, the Freshet derivative of nonlinear operator $B(\varphi)$ is calculated and sufficient conditions for the convergence of the Newton–Kantorovich method for the approximate solution of Eq. (19.1) are given.

Throughout Sect. 19.2, if the opposite is not indicated, the set $G \subset \mathbb{C}$ is considered as a bounded and simple connected region in the complex plane as it has already been stated in the introduction section.

If, for every $z_1, z_2 \in \bar{G}$ there exist $H > 0$ and $\alpha \in (0, 1)$ numbers such that

$$|\varphi(z_1) - \varphi(z_2)| \leq H \cdot |z_1 - z_2|^\alpha$$

then it is said that the function $\varphi : \bar{G} \to \mathbb{C}$ satisfies the Holder condition on the set $\bar{G}$ with exponent α. The symbol $H_\alpha(\bar{G})$ will denote the set of all functions that satisfies Holder condition on the set $\bar{G}$ with exponent α.

It is well known that the vector space $(H_\alpha(\bar{G}); \|\cdot\|_\alpha)$ is a Banach space with the norm

$$\|\varphi\|_\alpha = \|\varphi\|_{H_\alpha(\bar{G})} \equiv \|\varphi\|_\infty + H(\varphi, \alpha; \bar{G}).$$

Here, the sup norm satisfies $\|\varphi\|_\infty = \max\left\{|\varphi(z)| : z \in \bar{G}\right\}$ and defines

$$H(\varphi, \alpha; \bar{G}) = \sup\left\{\frac{|\varphi(z_1) - \varphi(z_2)|}{|z_1 - z_2|^\alpha} : z_1, z_2 \in \bar{G}, \quad z_1 \neq z_2\right\}.$$

Furthermore, for every $z_k \in \bar{G}$ and $(z_k, \varphi_k) \in D_0$, $(z_k, \varphi_k, v_k, w_k) \in D$ for $k = 1, 2$, suppose that the scalar $\alpha \in (0, 1)$ and the positive numbers $m_1, m_2, n_1, n_2, l_1, l_2, l_3, l_4$ exist such that the following inequalities are satisfied:

$$|f(z_1, \varphi_1) - f(z_2, \varphi_2)| \leq m_1 \cdot |z_1 - z_2|^\alpha + m_2 \cdot |\varphi_1 - \varphi_2|, \tag{19.2}$$

$$|g(z_1, \varphi_1) - g(z_2, \varphi_2)| \leq n_1 \cdot |z_1 - z_2|^\alpha + n_2 \cdot |\varphi_1 - \varphi_2|, \tag{19.3}$$

$$\begin{aligned} |F(z_1, \varphi_1, v_1, w_1) - F(z_2, \varphi_2, v_2, w_2)| &\leq l_1 \cdot |z_1 - z_2|^\alpha + l_2 \cdot |\varphi_1 - \varphi_2| \\ &+ l_3 \cdot |v_1 - v_2| + l_4 \cdot |w_1 - w_2|. \end{aligned} \tag{19.4}$$

The symbols $H_{\alpha,1}(m_1, m_2; D_0)$, $H_{\alpha,1}(n_1, n_2; D_0)$, and $H_{\alpha,1,1,1}(l_1, l_2, l_3, l_4; D)$ denote the sets of functions that satisfy the conditions (19.2), (19.3), and (19.4) respectively.

The following definition is well known in the literature.

Definition 19.1 ([19]) Let A be an nonlinear operator defined on a set E in a Banach space X. Recall that A is said to be Freshet-differentiable at a point $x_0 \in E$ if there exists a bounded linear operator B such that

$$\lim_{\|h\|\to 0} \frac{\|A(x_0+h) - Ax_0 - Bh\|}{\|h\|} = 0.$$

Operator B is called the Freshet derivative of operator A at point x_0, denoted by $A'(x_0)$.

In our study we will prove that the following nonlinear operator is Freshet differentiable:

$$B(\varphi)(z) \equiv F(z, \varphi(z), T_G f(\cdot, \varphi(\cdot))(z), \Pi_G g(\cdot, \varphi(\cdot))(z)), \quad z \in G. \tag{19.5}$$

Let $B'(\varphi)$ be the Freshet differential of nonlinear operator $B(\varphi)$. Assume that there exists a solution for the linear equation

$$B'(\varphi)h(z) = \phi(z), \quad z \in G, \tag{19.6}$$

for every $\phi \in H_\alpha(\bar{G}), 0 < \alpha < 1$. This means that the existence of the bounded linear inverse operator $\left[B'(\varphi)\right]^{-1}$ is assumed.

19.2.1 Newton–Kantorovich Method for Eq. (19.1)

Let X and Y be Banach spaces and $L(X, Y)$ denote the linear operator spaces from X to Y. If $kerA$ and $CokerA = Y/ImA$ are finite-dimensional, $A \in L(X, Y)$ is called a Fredholm operator. The index of operator A is defined by $\kappa = indA = \dim\ker A - \dim Co\ker A$. The family of the Fredholm transformations from X to Y with index κ is denoted by $\phi_\kappa(X, Y)$.

Let $U \subset X$ be an open set and $h : U \to Y$ be a transformation. If $h'(\varphi) \in \phi_\kappa(X, Y)$ for every $\varphi \in X$,, then the transformation $h : U \to Y$ is called a Fredholm operator with index κ from class C'. Here, $h'(\varphi)$ is a Freshet differential of operator $h : U \to Y$.

In this study the family of Fredholm transformations C' from U to Y with index κ is denoted by $\phi_\kappa C'(U, Y)$.

Now the following theorem about the existence of a Freshet derivative for nonlinear operator $B(\varphi)$ is presented.

Theorem 19.1 *Let functions* $F(t,u,v,w)$, $f(t,u)$, $g(t,u)$ *and derivatives* $F_u^{'}$, $F_v^{'}$, $F_w^{'}$, $F_{u^2}^{''}$, $F_{v^2}^{''}$, $F_{w^2}^{''}$, $F_{uv}^{''}$, $F_{uw}^{''}$, $F_{vw}^{''}$ *and* $f_u^{'}$, $f_{u^2}^{''}$, $g_u^{'}$, $g_{u^2}^{''}$ *be of class* $H_{\alpha,1,1,1}(l_1,l_2,l_3,l_4;D)$, $H_{\alpha,1}(m_1,m_2;D_0)$ *and* $H_{\alpha,1}(n_1,n_2;D_0)$, $0<\beta<\alpha\le 1$*, respectively. Then nonlinear operator* $B(\varphi)$ *defined by (19.5) is Freshet differentiable for every* $\varphi\in H_\beta(\bar{G})$*, and derivative can be written:*

$$\begin{aligned} B'(\varphi)h(z) = {} & F_u^{'}(z,\varphi(z),T_G f(\tau,\varphi(\tau))(z),\Pi_G g(\tau,\varphi(\tau))(z))\cdot h(z)\\ & + F_v^{'}(z,\varphi(z),T_G f(\tau,\varphi(\tau))(z),\Pi_G g(\tau,\varphi(\tau))(z))\\ & \cdot T_G(f_u^{'}(\tau,\varphi(\tau))h(\tau)) + F_W^{'}(z,\varphi(z),T_G f(\tau,\varphi(\tau))(z),\\ & \quad \Pi_G g(\tau,\varphi(\tau))(z))\cdot \Pi_G(g_u^{'}(\tau,\varphi(\tau))h(\tau)). \end{aligned} \tag{19.7}$$

Furthermore, Freshet derivative $B'(\varphi)$ *on the ball* $U(\varphi_0,r)=\left\{\varphi\in H_\beta(\bar{G}):\ \ \|\varphi_0-\varphi\|\le r\right\}$ *provides the following Lipchitz condition:*

$$\left\|B'(\varphi_1)-B'(\varphi_2)\right\|\le L\cdot\|\varphi_1-\varphi_2\|,\quad \varphi_1,\varphi_2\in U(\varphi_0,r), \tag{19.8}$$

where L *is a constant and depends on functions* F, f, g *and* r, $\varphi_0\in H_\beta(\bar{G})$.

Proof Firstly, to prove that the equality (19.7) is correct, let functions $\varphi, h\in H_\beta(\bar{G})$ and $\beta\in(0,1)$ be given.

Now

$$\begin{aligned} B(\varphi+h)-B(\varphi) = {} & F(z,(\varphi+h)(z),T_G f(z,(\varphi+h)(\tau))(z),\Pi_G g(z,(\varphi+h)(\tau))(z))\\ & - F(z,\varphi(z),T_G f(z,\varphi(\tau))(z),\Pi_G g(z,\varphi(\tau))(z)). \end{aligned}$$

Then

$$\begin{aligned} B(\varphi+h)-B(\varphi) = {} & [F(z,(\varphi+h)(z),T_G f(z,(\varphi+h)(\tau))(z),\Pi_G g(z,(\varphi+h)(\tau))(z))\\ & - F(z,\varphi(z),T_G f(z,(\varphi+h)(\tau))(z),\Pi_G g(z,(\varphi+h)(\tau))(z))]\\ & + [F(z,\varphi(z),T_G f(z,(\varphi+h)(\tau))(z),\Pi_G g(z,(\varphi+h)(\tau))(z))\\ & - F(z,\varphi(z),T_G f(z,\varphi(\tau))(z),\Pi_G g(z,(\varphi+h)(\tau))(z))]\\ & + [F(z,\varphi(z),T_G f(z,\varphi(\tau))(z),\Pi_G g(z,(\varphi+h)(\tau))(z))\\ & - F(z,\varphi(z),T_G f(z,\varphi(\tau))(z),\Pi_G g(z,\varphi(\tau))(z))]. \end{aligned}$$

From the assumptions of the theorem, the following can be written:

$$\begin{aligned}B(\varphi+h)-B(\varphi) &= F_u^{'}(z,\varphi(z),T_G f(\tau,\varphi(\tau))(z),\Pi_G g(\tau,\varphi(\tau))(z))\cdot h(z)\\ &\quad + F_v^{'}(z,\varphi(z),T_G f(\tau,\varphi(\tau))(z),\Pi_G g(\tau,\varphi(\tau))(z))\\ &\quad \cdot T_G(f_u^{'}(\tau,\varphi(\tau))\cdot h(\tau)) + F_w^{'}(z,\varphi(z),T_G f(\tau,\varphi(\tau))(z),\\ &\quad \Pi_G g(\tau,\varphi(\tau))(z))\cdot \Pi_G(g_u^{'}(\tau,\varphi(\tau))\cdot h(\tau))+\omega(\varphi,h)(z).\end{aligned} \tag{19.9}$$

Here,

$$\omega(\varphi,h)(z)=\omega_1(\varphi,h)(z)+\omega_2(\varphi,h)(z)+\omega_3(\varphi,h)(z), \tag{19.10}$$

$$\omega_1(\varphi,h)(z)=\int_0^1\left[\begin{matrix}F_u^{'}\begin{pmatrix}z,(\varphi+\theta\cdot h)(z),\\ T_G f(\tau,(\varphi+\theta\cdot h)(\tau))(z),\\ \Pi_G g(\tau,(\varphi+\theta\cdot h)(z))\end{pmatrix}\\ -F_u^{'}(z,\varphi(z),T_G f(\tau,\varphi(\tau))(z)),\Pi_G g(\tau,\varphi(z)))\end{matrix}\right]\cdot h(z)d\theta, \tag{19.11}$$

$$\omega_2(\varphi,h)(z)=\int_0^1\left[\begin{matrix}F_v^{'}\begin{pmatrix}z,(\varphi+\theta\cdot h)(z),\\ T_G f(\tau,(\varphi+\theta\cdot h)(\tau))(z),\\ \Pi_G g(\tau,(\varphi+\theta\cdot h)(z))\end{pmatrix}\\ -F_v^{'}\begin{pmatrix}z,\varphi(z),\\ T_G f(\tau,\varphi(\tau))(z),\\ \Pi_G g(\tau,\varphi(z))\end{pmatrix}\end{matrix}\right]\cdot T_G(f_u^{'}(\tau,\varphi(\tau))(z)h(z)d\theta, \tag{19.12}$$

$$\omega_3(\varphi,h)(z)=\int_0^1\left[\begin{matrix}F_w^{'}\begin{pmatrix}z,\varphi(z),\\ T_G f(\tau,\varphi(\tau))(z),\\ \Pi_G g(\tau,(\varphi+\theta\cdot h)(z))\end{pmatrix}\\ -F_v^{'}\begin{pmatrix}z,\varphi(z),\\ T_G f(\tau,\varphi(\tau))(z),\\ \Pi_G g(\tau,\varphi(z))\end{pmatrix}\end{matrix}\right]\cdot \Pi_G(g_u^{'}(\tau,\varphi(\tau))(z)h(z)d\theta. \tag{19.13}$$

Let $D_r=\{(z,\varphi,v,w):\ \ z\in\bar{G},\ \|\varphi-\varphi_0(\tau)\|\le r,\ v,w\in\mathbb{C}\},\ r>0$.

From the assumptions of the theorem and properties of operators T_G and Π_G, it can be seen that derivative $F_u^{'}(z,u,v,w)$ is uniformly continuous on D_r. Therefore, for any $\varepsilon > 0$, the following evaluation can be written:

$$\|\omega_1(\varphi, h)\|_\infty \leq \max\left\{ \int_0^1 \left| \begin{array}{l} F_u^{'}\begin{pmatrix} z, (\varphi+\theta\cdot h)(z), \\ T_G f(\tau, (\varphi+\theta\cdot h)(\tau))(z), \\ \Pi_G g(\tau, (\varphi+\theta\cdot h)(z)) \end{pmatrix} \\ -F_u^{'}\begin{pmatrix} z, \varphi(z), \\ T_G f(\tau, \varphi(\tau))(z), \\ \Pi_G g(\tau, \varphi(z)) \end{pmatrix} \end{array} \right| \cdot |h(z)|\, d\theta d\theta : z \in \bar{G} \right\}$$

$$\leq c_1 \cdot \varepsilon \cdot \|h\|_\infty .$$

It follows that

$$\lim_{\|h\|\to 0} \frac{\|\omega_1(\varphi, h)\|}{\|h\|} = 0. \tag{19.14}$$

The following limits can be proved in a manner similar to (19.14)

$$\lim_{\|h\|\to 0} \frac{\|\omega_2(\varphi, h)\|}{\|h\|} = 0 \text{ and } \lim_{\|h\|\to 0} \frac{\|\omega_3(\varphi, h)\|}{\|h\|} = 0. \tag{19.15}$$

From (19.14) and (19.15), the following is obtained:

$$\lim_{\|h\|\to 0} \frac{\|\omega(\varphi, h)\|}{\|h\|} = 0.$$

Hence by the definition of the Freshet derivative, the veracity of the equality (19.7) is proved.

Now to prove that Freshet derivative $B'(\varphi)$ provides Lipchitz condition (19.8) on ball $U(\varphi_0, r)$,
let $\varphi_1, \varphi_2 \in U(\varphi_0, r)$. Then $(B'(\varphi_1) - B'(\varphi_2))h(z)$ is obtained as

$$\begin{aligned} (B'(\varphi_1) - B'(\varphi_2))h(z) &= [F_u^{'}(z, \varphi_1(z), T_G f(\tau, \varphi_1(\tau))(z), \Pi_G g(\tau, \varphi_1(\tau))(z)) \\ &- F_u^{'}(z, \varphi_2(z), T_G f(\tau, \varphi_2(\tau))(z), \Pi_G g(\tau, \varphi_2(\tau))(z))] \cdot h(z) \\ &+ [F_v^{'}(z, \varphi_1(z), T_G f(\tau, \varphi_1(\tau))(z), \Pi_G g(\tau, \varphi_1(\tau))(z)) \cdot T_g(f_u^{'}(\xi, \varphi_1(\xi))(\tau)h(\tau)) \\ &- F_v^{'}(z, \varphi_2(z), T_G f(\tau, \varphi_2(\tau))(z), \Pi_G g(\tau, \varphi_2(\tau))(z)) \cdot T_g(f_u^{'}(\xi, \varphi_2(\xi))(\tau)h(\tau))] \\ &+ [F_w^{'}(z, \varphi_1(z), T_G f(\tau, \varphi_1(\tau))(z), \Pi_G g(\tau, \varphi_1(\tau))(z)) \cdot \Pi_g(g_u^{'}(\xi, \varphi_1(\xi))(\tau)h(\tau)) \\ &- F_v^{'}(z, \varphi_2(z), T_G f(\tau, \varphi_2(\tau))(z), \Pi_G g(\tau, \varphi_2(\tau))(z)) \cdot \Pi_g(g_u^{'}(\xi, \varphi_2(\xi))(\tau)h(\tau))]. \end{aligned} \tag{19.16}$$

From the assumptions of the theorem and properties of operators T_G and Π_G, it is seen that

$$\left\| \begin{array}{l} F_u^{'}(\cdot, \varphi_1(\cdot), T_G f(\cdot, \varphi_1(\cdot))(\cdot), \Pi_G g(\cdot, \varphi_1(\cdot))(\cdot)) \\ -F_u^{'}(\cdot, \varphi_2(\cdot), T_G f(\cdot, \varphi_2(\cdot))(\cdot), \Pi_G g(\cdot, \varphi_2(\cdot))(\cdot)) \end{array} \right\| \le L_1 \cdot \|\varphi_1 - \varphi_2\| . \tag{19.17}$$

Similar evaluations can be proved for the second and third terms of difference (19.16). From these evaluations, it is seen that condition (19.8) is true.

Thus, the proof of Theorem 19.1 is complete.

The Freshet derivative in the form of a linear singular integral operator can be written as follows:

$$B'(\varphi)h(z) = a(\varphi, z) \cdot h(z) + b(\varphi, z) \cdot T_G(f_u^{'}(\tau, \varphi(\tau))h(\tau)) \tag{19.18}$$
$$+c(\varphi, z) \cdot \Pi_G(g_u^{'}(\tau, \varphi(\tau))h(\tau)).$$

Here,

$$a(\varphi, z) = F_u^{'}(z, \varphi(z), T_G f(\tau, \varphi(\tau))(z), \Pi_G g(\tau, \varphi(\tau))(z)),$$

$$b(\varphi, z) = F_v^{'}(z, \varphi(z), T_G f(\tau, \varphi(\tau))(z), \Pi_G g(\tau, \varphi(\tau))(z)),$$

$$c(\varphi, z) = F_W^{'}(z, \varphi(z), T_G f(\tau, \varphi(\tau))(z), \Pi_G g(\tau, \varphi(\tau))(z)).$$

The existence of the only zero solution of the equation below in space $H_\beta(\bar{G})$ is assumed:

$$B'(\varphi)h(z) = 0. \tag{19.19}$$

In this case the following equation is the unique solution for every $\phi \in H_\beta(\bar{G})$:

$$B'(\varphi)h(z) = \phi(z). \tag{19.20}$$

Therefore, bounded linear inverse operator $\left[B'(\varphi_0)\right]^{-1} : H_\beta(\bar{G}) \to H_\beta(\bar{G})$ exists. As a result, the solution of Eq. (19.20) is given as follows:

$$h(z) = \left[B'(\varphi_0)\right]^{-1} \phi(z).$$

Now a theorem on convergence of the Newton–Kantorovich method for Eq. (19.1) is given.

Theorem 19.2 *Let the conditions of Theorem 19.1 be provided and* $\kappa = ind B'(\varphi_0) \geq 0$ *for a* $\varphi_0 \in H_\beta(\bar{G})$*. Furthermore, assume that homogeneous*

Eq. (19.19) *has only a trivial solution. Also, suppose that*

$$\left\| \left[B'(\varphi_0) \right]^{-1} \right\| \le m, \quad \left\| \left[B'(\varphi_0) \right]^{-1} \cdot B(\varphi_0) \right\| \le M.$$

If

$$\delta = LMm < \frac{1}{2} \quad \text{and} \qquad r \ge r_0 = \frac{1 - \sqrt{1 - 2\delta}}{\delta} \cdot M,$$

then the following equation has a unique solution φ^*, *which is in ball* $U(\varphi_0, r_0) = \left\{ \varphi \in H_\beta(\bar{G}) : \quad \|\varphi_0 - \varphi\| \le r_0 \right\}$

$$B(\varphi)(z) = 0, \quad z \in G. \tag{19.21}$$

Furthermore, the following successive approximations converge to the solution φ^* *of Eq.* (19.21) *in the ball* $U(\varphi_0, r_0)$

$$\varphi_{n+1}(z) = \varphi_n(z) - \left[B'(\varphi_0) \right]^{-1} \cdot B(\varphi_n), \quad n = 0, 1, \ldots. \tag{19.22}$$

Also, the convergence ratio is to be taken as the following:

$$\left\| \varphi^* - \varphi_n \right\| \le \frac{(1 - \sqrt{1 - 2\delta})^n}{\sqrt{1 - 2\delta}} \cdot M, \quad n = 0, 1, \ldots.$$

Proof Let $\varphi_0 \in H_\beta(\bar{G}), \quad r_0 = \frac{1 - \sqrt{1-2\delta}}{\delta} \cdot M$ and $U(\varphi_0, r_0) = \left\{ \varphi \in H_\beta(\bar{G}) \colon \|\varphi_0 - \varphi\| \le r_0 \right\}$. It must then be proved that the Newton–Kantorovich method is applied to the approximate solution of Eq. (19.21).

If

$$A(\varphi)(z) = \varphi(z) - \left[B'(\varphi_0) \right]^{-1} B(\varphi)(z),$$

then Eq. (19.21) can be written as

$$\varphi(z) = A(\varphi)(z). \tag{19.23}$$

In this case applying the Newton–Kantorovich method to the approximate solution of Eq. (19.21) is equivalent to applying the iteration method to the approximate solution of Eq. (19.23).

Now let us show that the iteration method to the approximate solution of Eq. (19.23) can be applied. To this end, it is sufficient to show that operator A satisfies the contraction mapping principle conditions.

Now it is to be shown that operator A maps from ball $U(\varphi_0, r_0)$ to itself and that it is a contraction mapping. It can be written that for every $\varphi \in U(\varphi_0, r_0)$,

$$\int_0^1 B'(\varphi_0 + \theta(\varphi - \varphi_0))(\varphi - \varphi_0)d\theta = \int_0^1 B'(\varphi_0 + \theta(\varphi - \varphi_0))d(\theta(\varphi - \varphi_0))$$
$$= \int_{\varphi_0}^{\varphi} B'(t)dt = B(\varphi) - B(\varphi_0).$$

Thus, the following equality is true:

$$B(\varphi) - B(\varphi_0) = \int_0^1 B'(\varphi_0 + \theta(\varphi - \varphi_0))(\varphi - \varphi_0)d\theta. \tag{19.24}$$

Let $\varphi_1, \varphi_2 \in U(\varphi_0, r_0)$. Using (19.24), it can be written that

$$B(\varphi_1)-B(\varphi_2)-B'(\varphi_2)(\varphi_1-\varphi_2) = \int_0^1 \left[B'(\varphi_2 + \theta(\varphi_1 - \varphi_2)) - B'(\varphi_2)\right] d\theta(\varphi_1-\varphi_2).$$

This gives the following:

$$\left\| B(\varphi_1) - B(\varphi_2) - B'(\varphi_2)(\varphi_1 - \varphi_2)\right\| \le L \int_0^1 \|\varphi_1 - \varphi_2\| \, \|\varphi_1 - \varphi_2\| \, \theta d\theta$$
$$= \frac{L}{2} \|\varphi_1 - \varphi_2\|^2 .$$

Therefore,

$$\left\| B(\varphi_1) - B(\varphi_2) - B'(\varphi_2)(\varphi_1 - \varphi_2)\right\| \le \frac{L}{2} \|\varphi_1 - \varphi_2\|^2 . \tag{19.25}$$

Now let us show that operator A maps from ball $U(\varphi_0, r_0)$ to itself. The following inequality is clear:

$$\begin{aligned}\|A(\varphi) - \varphi_0\| &\le \|A(\varphi) - A(\varphi_0)\| + \|A(\varphi_0) - \varphi_0\| \\ &= \left\| \left[B'(\varphi_0)\right]^{-1} \left[B(\varphi) - B(\varphi_0) - B'(\varphi_0)(\varphi - \varphi_0)\right] \right\| \\ &\quad + \left\| \left[B'(\varphi_0)\right]^{-1} B(\varphi_0) \right\| \\ &\le m \left\| B(\varphi) - B(\varphi_0) - B'(\varphi_0)(\varphi - \varphi_0) \right\| + M, \quad \varphi \in U(\varphi_0, r_0).\end{aligned}$$

Using (19.25), it is written that

$$\|A(\varphi) - \varphi_0\| \le \frac{1}{2} m L r_0^2 + M.$$

Furthermore, taking $mLr_0^2 - 2r_0 + 2M = 0$ from the previous inequality, it is obtained that

$$\|A(\varphi) - \varphi_0\| \leq r_0.$$

Hence $A(\varphi) \in U(\varphi_0, r_0)$.

Now let us show that operator A is a contraction mapping. Using (19.24) it can be written that

$$\begin{aligned} A(\varphi_1) - A(\varphi_2) &= \varphi_1 - \varphi_2 - \left[B'(\varphi_0)\right]^{-1} [B(\varphi_1) - B(\varphi_2)] \\ &= \left[B'(\varphi_0)\right]^{-1} \left[B'(\varphi_0)(\varphi_1 - \varphi_2) - B(\varphi_1) + B(\varphi_2)\right] \\ &= \left[B'(\varphi_0)\right]^{-1} \int_0^1 \left[B'(\varphi_0) - B'(\varphi_2 + \theta(\varphi_1 - \varphi_2))\right] d(\theta(\varphi_1 - \varphi_2)). \end{aligned}$$

Thus, it is obtained that

$$\|A(\varphi_1) - A(\varphi_2)\| \leq mLr_0 \|\varphi_1 - \varphi_2\|.$$

Now since $mLr_0 < 1$, operator A is a contraction mapping with coefficient $q = 1 - \sqrt{1 - 2\delta}$.

Therefore, according to the contraction mapping principle, Eq. (19.23) has a unique solution φ^* in ball $U(\varphi_0, r_0)$ and this solution is the limit of the following iteration:

$$\varphi_{n+1}(z) = A(\varphi_n)(z), \quad n = 0, 1, \ldots,$$

and

$$\left\|\varphi_n - \varphi^*\right\| \leq \frac{q^n}{1-q} \|\varphi_1 - \varphi_0\|.$$

Also, under the hypothesis of the theorem

$$\|\varphi_1 - \varphi_0\| = \|A(\varphi_0) - \varphi_0\| = \left\|\left[B'(\varphi_0)\right]^{-1} B(\varphi_0)\right\| \leq M.$$

Therefore, the proof of Theorem 19.2 is complete.

19.3 Some Integral Operators in Holder Space

In this section, some integral operators, which have broad applications in the theory of elementary particles and scattering, have been investigated in Holder space. We show that some important inequalities for the norm of these operators are also satisfied in Holder space.

Consider the following nonlinear singular integral equation

$$\varphi(t) = f(t)\left\{\varphi^2(t) + [\lambda - S\varphi(t) + \mu S_+\varphi(t)]^2\right\}, \ t \in [0, 1]. \tag{19.26}$$

Here,

$$S\varphi(t) = \frac{1}{\pi}\int_0^1 \frac{\varphi(\tau)}{\tau - t}d\tau \text{ and } S_+\varphi(t) = \frac{1}{\pi}\int_0^1 \frac{\varphi(\tau)}{\tau + t}d\tau. \tag{19.27}$$

Nonlinear singular integral Eq. (19.26) has crucial applications in the theory of elementary particles and scattering [12]. It is important to examine the type of such equations. In the investigation of the existence of the solution of Eq. (19.26) it is important to examine the operators (19.27).

In this section, we prove that the operators S and S_+ from (19.27) are bounded in Holder space. Moreover, for these operators, some important inequalities in the different norms are also given.

Firstly, we will introduce some necessary information required for the proof of main results.

As usual, throughout the work, $C[0, 1]$ is the set of continuous functions defined on $[0, 1]$ with maximum norm

$$\|f\|_\infty = \max\{|f(t)| : t \in [0, 1]\}.$$

Definition 19.2 ([13]) The function

$$\omega(\varphi, x) = \sup\{|\varphi(t_2) - \varphi(t_1)| : |t_2 - t_1| \le x\}, x \in [0, +\infty)$$

is called the modulus of continuity of the bounded function $\varphi : \mathbb{R} \to \mathbb{R}$.

Let us recall the properties of the modulus of continuity:

1. The modulus of continuity is a continuous function.
2. The modulus of continuity is a non-decreasing function.
3. For every $x_1, x_2 > 0$, $\omega(x_1 + x_2) \le \omega(x_1) + \omega(x_2)$.
4. $\omega(0) = 0$.
5. For every $x > 0$ and $\alpha > 0$, $\frac{\omega(x)}{x}$ and $\frac{\omega(x)}{x^\alpha}$ are decreasing functions.

Definition 19.3 ([58]) Assume $\varphi(t)$ is defined on $[0, 1]$. If

$$|\varphi(x_2) - \varphi(x_1)| \le K|x_2 - x_1|^\alpha, 0 < \alpha \le 1$$

for arbitrary points $x_1, x_2 \in [0, 1]$, where $K > 0$ and α are definite constants, then $\varphi(t)$ is said to satisfy the Holder condition of order α, or simply condition H_α, denoted by $\varphi \in H_\alpha[0, 1]$.

The functions in class H possess the following properties:

1. If $\varphi \in H[0,1]$, then $\varphi \in C[0,1]$, i.e., $H[0,1] \subset C[0,1]$.
2. If $\varphi \in H_\alpha[0,1]$ and $0 < \beta \le \alpha$, then $\varphi \in H_\beta[0,1]$ i.e., $H_\alpha[0,1] \subseteq H_\beta[0,1]$ if $0 < \beta \le \alpha$.
3. If both φ and $\psi \in H_\alpha[0,1]$, then so do $\varphi \pm \psi$, $\varphi.\psi$, φ/ψ $(\psi \neq 0$ on $[0,1])$.
4. If both φ and $\psi \in H_\alpha[0,1]$, then so do $\varphi + \psi$, $\varphi.\psi$ and

$$H(\varphi + \psi; \alpha) = H(\varphi; \alpha) + H(\psi; \alpha)$$

$$H(\varphi \cdot \psi; \alpha) \le H(\varphi; \alpha) \|\psi\|_\infty + \|\varphi\|_\infty H(\psi; \alpha).$$

Let $\overset{\circ}{C}[0,1] = \{f \in C[0,1] : f(0) = 0 = f(1)\}$ and $\overset{\circ}{H}_\alpha[0,1] = \{\varphi \in H_\alpha : \varphi(0) = 0 = \varphi(1)\}$.

The function spaces $H_\alpha[0,1]$ and $\overset{\circ}{H}_\alpha[0,1]$ are Banach spaces with norm

$$\|\varphi\|_\alpha = \max(\|\varphi\|_\infty, H(\varphi; \alpha)).$$

Here,

$$H(\varphi; \alpha) = \sup\left\{\frac{\omega(\varphi, x)}{x^\alpha} : 0 < x \le 1\right\}.$$

Furthermore, throughout the section, we denote $H_\alpha(\overset{\circ}{H}_\alpha)$ instead of $H_\alpha[0,1]$ $(\overset{\circ}{H}_\alpha[0,1])$, unless stated otherwise.

We denote the norm of function $\varphi \in \overset{\circ}{H}_\alpha$ by

$$\|\varphi\|_{\alpha,0} = \max(\|\varphi\|_\infty, H(\varphi; \alpha)).$$

The norm of a bounded linear operator $\Im : \overset{\circ}{H}_\alpha \to H_\alpha$ is defined as follows [38]:

$$\|\Im\|_\alpha = \|\Im\|_{\overset{\circ}{H}_\alpha \to H_\alpha} = \sup_{\varphi \neq 0}\left\{\frac{\|\Im\varphi\|_\alpha}{\|\varphi\|_{\alpha,0}} : \varphi \in \overset{\circ}{H}_\alpha\right\}$$

$$\|\Im\|_\infty = \|\Im\|_{\overset{\circ}{H}_\alpha \to C[0,1]} = \sup_{\varphi \neq 0}\left\{\frac{\|\Im\varphi\|_\infty}{\|\varphi\|_{\alpha,0}} : \varphi \in \overset{\circ}{H}_\alpha\right\}.$$

Let

$$J_0 = \left\{\varphi \in C[0,1] : \int_0^1 \frac{\omega(\varphi, \xi)}{\xi} < +\infty\right\}$$

and

$$Z(\omega(\varphi,\cdot),t)=\int_0^t\frac{\omega(\varphi,\xi)}{\xi}d\xi+t\int_t^1\frac{\omega(\varphi,\xi)}{\xi^2}d\xi,t\in[0,1].$$

Then, the following is provided:

1. If $\varphi\in\overset{\circ}{H}_\alpha$, then $\varphi\in J_0$, i.e., $\overset{\circ}{H}_\alpha\subset J_0$.
2. $Z(\omega(\varphi,\cdot),t)$ is a non-decreasing function on [0, 1].

19.3.1 Some Properties of the Integral Operator (19.27)

In this section, we provide some properties of the operators $S:\overset{\circ}{H}_\alpha\to H_\alpha$ and $S_+:\overset{\circ}{H}_\alpha\to H_\alpha$, which are defined in formula (19.27).

Theorem 19.3 *Let the operators $S:\overset{\circ}{H}_\alpha\to H_\alpha$ and $S_+:\overset{\circ}{H}_\alpha\to H_\alpha$ be defined as in formula (19.27) and $\varphi\in J_0$. Then, for every $x\in(0,1]$,*

$$\omega(S\varphi,x)\le c_1Z(\omega(\varphi,\cdot),x),\tag{19.28}$$

$$\omega(S_+\varphi,x)\le c_2Z(\omega(\varphi,\cdot),x).\tag{19.29}$$

Here, $c_1=\frac{1}{\pi}(\frac{67}{6}+\ln 3)$, $c_2=\frac{2}{\pi}$ are definite constants.

Proof Let

$$\phi(t)=\begin{cases}\varphi(t) & \text{for } t\in[0,1],\\ 0 & \text{for } t\in[-1,2]/[0,1]\end{cases}$$

and

$$F\phi(t)=\frac{1}{\pi}\int_{-1}^2\frac{\phi(\tau)}{\tau-t}d\tau,\ t\in(-1,2).$$

The operator $F:\overset{\circ}{H}_\alpha\to H_\alpha$ can be written as

$$F\phi(t)=\frac{1}{\pi}\int_{-1}^2\frac{\phi(\tau)-\phi(t)}{\tau-t}d\tau+\frac{1}{\pi}\phi(t)\ln\frac{2-t}{1+t},\ t\in(-1,2).$$

Let $t_1, t_2 \in [0, 1]$, $0 \le t_1 < t_2 \le 1$ $(0 < t_2 - t_1 \le 1)$ and $\varepsilon = \frac{t_2 - t_1}{2}$. In that case, we write

$$\begin{aligned}
\pi\,(F\phi(t_2) - F\phi(t_1)) &= \int_{-1}^{2} \frac{\phi(\tau)-\phi(t_2)}{\tau - t_2} d\tau - \int_{-1}^{2} \frac{\phi(\tau)-\phi(t_1)}{\tau - t_1} d\tau \\
&\quad + \phi(t_2)\ln\frac{2-t_2}{1+t_2} - \phi(t_1)\ln\frac{2-t_1}{1+t_1} \\
&= \int_{t_1-\varepsilon}^{t_2-\varepsilon} \frac{\phi(\tau)-\phi(t_2)}{\tau - t_2} d\tau - \int_{t_1-\varepsilon}^{t_2-\varepsilon} \frac{\phi(\tau)-\phi(t_1)}{\tau - t_1} d\tau \\
&\quad + \int_{-1}^{t_1-\varepsilon} \left[\frac{\phi(\tau)-\phi(t_2)}{\tau - t_2} - \frac{\phi(\tau)-\phi(t_1)}{\tau - t_1}\right] d\tau \\
&\quad + \int_{t_2-\varepsilon}^{2} \left[\frac{\phi(\tau)-\phi(t_2)}{\tau - t_2} - \frac{\phi(\tau)-\phi(t_1)}{\tau - t_1}\right] d\tau \\
&\quad + \phi(t_2)\ln\frac{2-t_2}{1+t_2} - \phi(t_1)\ln\frac{2-t_1}{1+t_1}
\end{aligned}$$

Thus,

$$\begin{aligned}
\pi\,(F\phi(t_2) - F\phi(t_1)) &= \int_{t_1-\varepsilon}^{t_2-\varepsilon} \frac{\phi(\tau)-\phi(t_2)}{\tau - t_2} d\tau - \int_{t_1-\varepsilon}^{t_2-\varepsilon} \frac{\phi(\tau)-\phi(t_1)}{\tau - t_1} d\tau \\
&\quad + \int_{-1}^{t_1-\varepsilon} \left[\frac{\phi(\tau)-\phi(t_1)}{\tau - t_2} - \frac{\phi(\tau)-\phi(t_1)}{\tau - t_1} + \frac{\phi(t_1)-\phi(t_2)}{\tau - t_2}\right] d\tau \\
&\quad + \int_{t_2-\varepsilon}^{2} \left[\frac{\phi(\tau)-\phi(t_2)}{\tau - t_2} - \frac{\phi(\tau)-\phi(t_2)}{\tau - t_1} + \frac{\phi(t_1)-\phi(t_2)}{\tau - t_1}\right] d\tau \\
&\quad + \phi(t_2)\ln\frac{2-t_2}{1+t_2} - \phi(t_1)\ln\frac{2-t_1}{1+t_1} \\
&= \int_{t_1-\varepsilon}^{t_2-\varepsilon} \frac{\phi(\tau)-\phi(t_2)}{\tau - t_2} d\tau - \int_{t_1-\varepsilon}^{t_2-\varepsilon} \frac{\phi(\tau)-\phi(t_1)}{\tau - t_1} d\tau \\
&\quad + (t_2 - t_1) \int_{-1}^{t_1-\varepsilon} \frac{\phi(\tau)-\phi(t_1)}{(\tau - t_1)(\tau - t_2)} d\tau \\
&\quad + [\phi(t_1) - \phi(t_2)] \int_{-1}^{t_1-\varepsilon} \frac{d\tau}{\tau - t_2} \\
&\quad + (t_2 - t_1) \int_{t_2-\varepsilon}^{2} \frac{\phi(\tau)-\phi(t_2)}{(\tau - t_1)(\tau - t_2)} d\tau \\
&\quad + [\phi(t_1) - \phi(t_2)] \int_{t_2-\varepsilon}^{2} \frac{d\tau}{\tau - t_1}
\end{aligned}$$

$$
\begin{aligned}
&+\phi(t_2)\ln\frac{2-t_2}{1+t_2}-\phi(t_1)\ln\frac{2-t_1}{1+t_1}\\
=&\int_{t_1-\varepsilon}^{t_2-\varepsilon}\frac{\phi(\tau)-\phi(t_2)}{\tau-t_2}d\tau-\int_{t_1-\varepsilon}^{t_2-\varepsilon}\frac{\phi(\tau)-\phi(t_1)}{\tau-t_1}d\tau\\
&+(t_2-t_1)\int_{-1}^{t_1-\varepsilon}\frac{\phi(\tau)-\phi(t_1)}{(\tau-t_1)(\tau-t_2)}d\tau\\
&+(t_2-t_1)\int_{t_2-\varepsilon}^{2}\frac{\phi(\tau)-\phi(t_2)}{(\tau-t_1)(\tau-t_2)}d\tau\\
&+[\phi(t_1)-\phi(t_2)]\,[\ln|t_1-\varepsilon-t_2|-\ln(1+t_2)]\\
&+[\phi(t_1)-\phi(t_2)]\,[\ln(2-t_1)-\ln|t_2-\varepsilon-t_1|]\\
&+\phi(t_2)\ln\frac{2-t_2}{1+t_2}-\phi(t_1)\ln\frac{2-t_1}{1+t_1}.
\end{aligned}
$$

As a result of simple calculations, we write

$$
\begin{aligned}
&[\phi(t_1)-\phi(t_2)]\,[\ln|t_1-\varepsilon-t_2|-\ln(1+t_2)]+\phi(t_2)\ln\frac{2-t_2}{1+t_2}\\
&\quad+[\phi(t_1)-\phi(t_2)]\,[\ln(2-t_1)-\ln|t_2-\varepsilon-t_1|]-\phi(t_1)\ln\frac{2-t_1}{1+t_1}\\
&\quad=[\phi(t_1)-\phi(t_2)]\left[\ln\frac{3}{2}(t_2-t_1)-\ln(1+t_2)\right]-\phi(t_1)\,[\ln(2-t_1)-\ln(1+t_1)]\\
&\quad+[\phi(t_1)-\phi(t_2)]\left[\ln(2-t_1)-\ln\frac{t_2-t_1}{2}\right]+\phi(t_2)\,[\ln(2-t_2)-\ln(1+t_2)]\,.
\end{aligned}
$$

That is,

$$
\begin{aligned}
&[\phi(t_1)-\phi(t_2)]\,[\ln|t_1-\varepsilon-t_2|-\ln(1+t_2)]+\phi(t_2)\ln\frac{2-t_2}{1+t_2}\\
&\quad+[\phi(t_1)-\phi(t_2)]\,[\ln(2-t_1)-\ln|t_2-\varepsilon-t_1|]-\phi(t_1)\ln\frac{2-t_1}{1+t_1}\\
&=[\phi(t_1)-\phi(t_2)]\left[\ln 3+\ln\frac{t_2-t_1}{2}-\ln(1+t_2)\right]\\
&\quad+[\phi(t_1)-\phi(t_2)]\left[\ln(2-t_1)-\ln\frac{t_2-t_1}{2}\right]\\
&\quad+\phi(t_2)\,[\ln(2-t_2)-\ln(1+t_2)]-\phi(t_1)\,[\ln(2-t_1)-\ln(1+t_1)]\\
&=\phi(t_1)\ln\frac{1+t_1}{1+t_2}+\phi(t_2)\ln\frac{2-t_2}{2-t_1}+[\phi(t_1)-\phi(t_2)]\ln 3.
\end{aligned}
$$

Thus, we obtain

$$\pi\,(F\phi(t_2) - F\phi(t_1)) = \sum_{\nu=1}^{7} I_\nu.$$

Here,

$$\begin{aligned}
I_1 &= \int_{t_1-\varepsilon}^{t_2-\varepsilon} \frac{\phi(\tau)-\phi(t_2)}{\tau - t_2} d\tau,\\
I_2 &= -\int_{t_1-\varepsilon}^{t_2-\varepsilon} \frac{\phi(\tau)-\phi(t_1)}{\tau - t_1} d\tau,\\
I_3 &= (t_2 - t_1)\int_{-1}^{t_1-\varepsilon} \frac{\phi(\tau)-\phi(t_1)}{(\tau - t_1)(\tau - t_2)} d\tau,\\
I_4 &= (t_2 - t_1)\int_{t_2-\varepsilon}^{2} \frac{\phi(\tau)-\phi(t_2)}{(\tau - t_1)(\tau - t_2)} d\tau,\\
I_5 &= \phi(t_1)\ln\frac{1+t_1}{1+t_2},\\
I_6 &= \phi(t_2)\ln\frac{2-t_2}{2-t_1},\\
I_7 &= (\phi(t_1)-\phi(t_2))\ln 3.
\end{aligned}$$

Now we consider $|I_\nu|$, $\nu = 1, \ldots, 7$.

For I_1, we write

$$\begin{aligned}
|I_1| &\le \int_{t_1-\varepsilon}^{t_2-\varepsilon} \frac{\omega(\phi, |\tau - t_2|)}{|\tau - t_2|} d\tau = \int_{\varepsilon}^{\varepsilon+t_2-t_1} \frac{\omega(\phi,\xi)}{\xi} d\xi = \int_0^{t_2-t_1} \frac{\omega(\phi,\xi+\varepsilon)}{\xi+\varepsilon} d\xi\\
&\le \int_0^{t_2-t_1} \frac{\omega(\phi,\xi)}{\xi} d\xi.
\end{aligned}$$

Similarly,

$$\begin{aligned}
|I_2| &\le \int_{t_1-\varepsilon}^{t_2-\varepsilon} \frac{\omega(\phi, |\tau - t_1|)}{|\tau - t_1|} d\tau\\
&= \int_{t_1-\varepsilon}^{t_1} \frac{\omega(\phi, t_1-\tau)}{t_1-\tau} d\tau + \int_{t_1}^{t_2-\varepsilon} \frac{\omega(\phi, \tau - t_1)}{\tau - t_1} d\tau\\
&= 2\int_0^{\varepsilon} \frac{\omega(\phi,\xi)}{\xi} d\xi
\end{aligned}$$

$$= 2\int_0^{t_2-t_1} \frac{\omega(\phi, \xi/2)}{\xi} d\xi$$
$$\leq 2\int_0^{t_2-t_1} \frac{\omega(\phi, \xi)}{\xi} d\xi.$$

Also, we can write

$$|I_3| \leq (t_2-t_1)\int_{-1}^{t_1-\varepsilon} \frac{\omega(\phi, t_1-\tau)}{(t_1-\tau)(t_2-\tau)} d\tau$$
$$= (t_2-t_1)\int_{\varepsilon}^{1+t_1} \frac{\omega(\phi, \xi)}{\xi(\xi+t_2-t_1)} d\xi = \sum_{\nu=1}^{3} I_3^{(\nu)}.$$

Here,

$$I_3^{(1)} = (t_2-t_1)\int_{\varepsilon}^{t_2-t_1} \frac{\omega(\phi, \xi)}{\xi(\xi+t_2-t_1)} d\xi,$$

$$I_3^{(2)} = (t_2-t_1)\int_{t_2-t_1}^{1} \frac{\omega(\phi, \xi)}{\xi(\xi+t_2-t_1)} d\xi,$$

$$I_3^{(3)} = (t_2-t_1)\int_{1}^{1+t_1} \frac{\omega(\phi, \xi)}{\xi(\xi+t_2-t_1)} d\xi.$$

For $I_3^{(1)}$ and $I_3^{(2)}$, we write

$$I_3^{(1)} \leq \int_{\varepsilon}^{t_2-t_1} \frac{\omega(\phi, \xi)}{\xi} d\xi = \int_0^{\varepsilon} \frac{\omega(\phi, \xi+\varepsilon)}{\xi+\varepsilon} d\xi$$
$$\leq \int_0^{\varepsilon} \frac{\omega(\phi, \xi)}{\xi} d\xi = \int_0^{t_2-t_1} \frac{\omega(\phi, \xi/2)}{\xi} d\xi$$
$$\leq \int_0^{t_2-t_1} \frac{\omega(\phi, \xi)}{\xi} d\xi,$$

$$I_3^{(2)} \leq (t_2-t_1)\int_{t_2-t_1}^{1} \frac{\omega(\phi, \xi)}{\xi^2} d\xi.$$

For $I_3^{(3)}$, we have

$$I_3^{(3)} = (t_2 - t_1)\int_1^{1+t_1} \frac{\omega(\phi,\xi)}{\xi(\xi + t_2 - t_1)} d\xi = (t_2 - t_1)\int_0^{t_1} \frac{\omega(\phi,\xi+1)}{(\xi+1)(\xi+1+t_2-t_1)} d\xi$$
$$\leq (t_2 - t_1)\int_0^{t_1} \frac{\omega(\phi,\xi)}{\xi(\xi + t_2 - t_1)} d\xi.$$

Here, we need to consider two cases: $t_1 \leq t_2 - t_1$ and $t_1 > t_2 - t_1$.

Case 1 If $t_1 \leq t_2 - t_1$, then

$$\begin{aligned} I_3^{(3)} &\leq (t_2 - t_1)\int_0^{t_1} \frac{\omega(\phi,\xi)}{\xi(\xi + t_2 - t_1)} d\xi \\ &\leq (t_2 - t_1)\int_0^{t_2-t_1} \frac{\omega(\phi,\xi)}{\xi(\xi + t_2 - t_1)} d\xi \\ &\leq \int_0^{t_2-t_1} \frac{\omega(\phi,\xi)}{\xi} d\xi. \end{aligned}$$

Case 2 If $t_1 > t_2 - t_1$, then

$$\begin{aligned} I_3^{(3)} &\leq (t_2 - t_1)\int_0^{t_1} \frac{\omega(\phi,\xi)}{\xi(\xi + t_2 - t_1)} d\xi \\ &= (t_2 - t_1)\int_0^{t_2-t_1} \frac{\omega(\phi,\xi)}{\xi(\xi + t_2 - t_1)} d\xi + (t_2 - t_1)\int_{t_2-t_1}^{t_1} \frac{\omega(\phi,\xi)}{\xi(\xi + t_2 - t_1)} d\xi. \end{aligned}$$

Furthermore, since we can write

$$(t_2 - t_1)\int_0^{t_2-t_1} \frac{\omega(\phi,\xi)}{\xi(\xi + t_2 - t_1)} d\xi \leq \int_0^{t_2-t_1} \frac{\omega(\phi,\xi)}{\xi} d\xi,$$

$$\begin{aligned} (t_2 - t_1)\int_{t_2-t_1}^{t_1} \frac{\omega(\phi,\xi)}{\xi(\xi + t_2 - t_1)} d\xi &\leq (t_2 - t_1)\int_{t_2-t_1}^{t_1} \frac{\omega(\phi,\xi)}{\xi^2} d\xi \\ &\leq (t_2 - t_1)\int_{t_2-t_1}^{1} \frac{\omega(\phi,\xi)}{\xi^2} d\xi, \end{aligned}$$

we obtain

$$I_3^{(3)} \leq \int_0^{t_2-t_1} \frac{\omega(\phi,\xi)}{\xi} d\xi + (t_2 - t_1)\int_{t_2-t_1}^{1} \frac{\omega(\phi,\xi)}{\xi^2} d\xi.$$

Thus, considering estimates for $I_3^{(1)}$, $I_3^{(2)}$, and $I_3^{(3)}$, we obtain

$$|I_3| \le 2\left(\int_0^{t_2-t_1} \frac{\omega(\phi,\xi)}{\xi}d\xi + (t_2-t_1)\int_{t_2-t_1}^1 \frac{\omega(\phi,\xi)}{\xi^2}d\xi\right).$$

Now we consider I_4. We can write

$$|I_4| \le (t_2-t_1)\int_{t_2-\varepsilon}^2 \frac{\omega(\phi,|\tau-t_2|)}{|\tau-t_1|\,|\tau-t_2|}d\tau = \sum_{\nu=1}^3 J_4^{(\nu)}.$$

Here,

$$I_4^{(1)} = (t_2-t_1)\int_{t_2-\varepsilon}^{t_2} \frac{\omega(\phi,t_2-\tau)}{(t_2-\tau)(\tau-t_1)}d\tau,$$

$$I_4^{(2)} = (t_2-t_1)\int_{t_2}^{t_2+\varepsilon} \frac{\omega(\phi,\tau-t_2)}{(\tau-t_2)(\tau-t_1)}d\tau,$$

$$I_4^{(3)} = (t_2-t_1)\int_{t_2+\varepsilon}^{2} \frac{\omega(\phi,\tau-t_2)}{(\tau-t_2)(\tau-t_1)}d\tau.$$

For $I_4^{(1)}$, we obtain

$$\begin{aligned} I_4^{(1)} &= (t_2-t_1)\int_0^{\varepsilon} \frac{\omega(\phi,\xi)}{\xi(t_2-t_1-\xi)}d\xi = (t_2-t_1)\int_{\varepsilon}^{t_2-t_1} \frac{\omega(\phi,t_2-t_1-\xi)}{(t_2-t_1-\xi)\xi}d\xi \\ &\le 2\int_{\varepsilon}^{t_2-t_1} \frac{\omega(\phi,t_2-t_1-\xi)}{t_2-t_1-\xi}d\xi = 2\int_0^{\varepsilon} \frac{\omega(\phi,\xi)}{\xi}d\xi = 2\int_0^{t_2-t_1} \frac{\omega(\phi,\xi/2)}{\xi}d\xi \\ &\le 2\int_0^{t_2-t_1} \frac{\omega(\phi,\xi)}{\xi}d\xi. \end{aligned}$$

Similarly,

$$\begin{aligned} I_4^{(2)} &= (t_2-t_1)\int_0^{\varepsilon} \frac{\omega(\phi,\xi)}{\xi(\xi+t_2-t_1)}d\xi \\ &\le \int_0^{\varepsilon} \frac{\omega(\phi,\xi)}{\xi}d\xi = \int_0^{t_2-t_1} \frac{\omega(\phi,\xi/2)}{\xi}d\xi \\ &\le \int_0^{t_2-t_1} \frac{\omega(\phi,\xi)}{\xi}d\xi. \end{aligned}$$

For $I_4^{(3)}$, we write

$$I_4^{(3)} = (t_2 - t_1)\int_{t_2+\varepsilon}^{2} \frac{\omega(\phi, \tau - t_2)}{(\tau - t_2)(\tau - t_1)} d\tau$$
$$= (t_2 - t_1)\int_{\varepsilon}^{2-t_2} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi = \sum_{\nu=1}^{3} I_4^{(3,\nu)}.$$

Here,

$$I_4^{(3,1)} = (t_2 - t_1)\int_{\varepsilon}^{t_2-t_1} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi,$$
$$I_4^{(3,2)} = (t_2 - t_1)\int_{t_2-t_1}^{1} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi,$$

$$I_4^{(3,3)} = (t_2 - t_1)\int_{1}^{2-t_2} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi$$

For $I_4^{(3,1)}$ and $I_4^{(3,2)}$, we obtain the following estimates

$$I_4^{(3,1)} = (t_2 - t_1)\int_{\varepsilon}^{t_2-t_1} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi = (t_2 - t_1)\int_{0}^{\varepsilon} \frac{\omega(\phi, \xi + \varepsilon)}{(\xi + \varepsilon)(\xi + 3\varepsilon)} d\xi$$
$$\leq \frac{t_2 - t_1}{3\varepsilon}\int_{0}^{\varepsilon} \frac{\omega(\phi, \xi + \varepsilon)}{\xi + \varepsilon} d\xi = \frac{2}{3}\int_{0}^{\varepsilon} \frac{\omega(\phi, \xi + \varepsilon)}{\xi + \varepsilon} d\xi \leq \frac{2}{3}\int_{0}^{\varepsilon} \frac{\omega(\phi, \xi)}{\xi} d\xi$$
$$= \frac{2}{3}\int_{0}^{t_2-t_1} \frac{\omega(\phi, \xi/2)}{\xi} d\xi$$
$$\leq \frac{2}{3}\int_{0}^{t_2-t_1} \frac{\omega(\phi, \xi)}{\xi} d\xi,$$

$$I_4^{(3,2)} = (t_2 - t_1)\int_{t_2-t_1}^{1} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi \leq (t_2 - t_1)\int_{t_2-t_1}^{1} \frac{\omega(\phi, \xi)}{\xi^2} d\xi.$$

Now, we consider $I_4^{(3,3)}$. Firstly, we can write

$$I_4^{(3,3)} = (t_2 - t_1)\int_{1}^{2-t_2} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi = (t_2 - t_1)\int_{0}^{1-t_2} \frac{\omega(\phi, \xi + 1)}{(\xi + 1)(\xi + 1 + t_2 - t_1)} d\xi$$
$$\leq (t_2 - t_1)\int_{0}^{1-t_2} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi.$$

Here, we need to consider two cases: $1 - t_2 \leq t_2 - t_1$ and $1 - t_2 > t_2 - t_1$.

Case 1 If $1 - t_2 \leq t_2 - t_1,$ then

$$I_4^{(3,3)} \leq (t_2 - t_1) \int_0^{1-t_2} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi \leq (t_2 - t_1) \int_0^{t_2 - t_1} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi$$
$$\leq \int_0^{t_2 - t_1} \frac{\omega(\phi, \xi)}{\xi} d\xi.$$

Case 2 If $1 - t_2 > t_2 - t_1,$ then

$$I_4^{(3,3)} \leq (t_2 - t_1) \int_0^{t_2 - t_1} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi + (t_2 - t_1) \int_{t_2 - t_1}^{1-t_2} \frac{\omega(\phi, \xi)}{\xi(\xi + t_2 - t_1)} d\xi$$
$$\leq \int_0^{t_2 - t_1} \frac{\omega(\phi, \xi)}{\xi} d\xi + (t_2 - t_1) \int_{t_2 - t_1}^{1-t_2} \frac{\omega(\phi, \xi)}{\xi^2} d\xi$$
$$\leq \int_0^{t_2 - t_1} \frac{\omega(\phi, \xi)}{\xi} d\xi + (t_2 - t_1) \int_{t_2 - t_1}^{1} \frac{\omega(\phi, \xi)}{\xi^2} d\xi.$$

Thus, using the results for $I_4^{(3,\nu)}$, $\nu = 1, 2, 3,$ we obtain the following estimate:

$$I_4^{(3)} \leq \frac{5}{3} \int_0^{t_2 - t_1} \frac{\omega(\phi, \xi)}{\xi} d\xi + 2(t_2 - t_1) \int_{t_2 - t_1}^{1-t_2} \frac{\omega(\phi, \xi)}{\xi^2} d\xi.$$

Using the results for $I_4^{(\nu)}$, $\nu = 1, 2, 3,$ we obtain the following estimate:

$$|I_4| \leq \frac{14}{3} \int_0^{t_2 - t_1} \frac{\omega(\phi, \xi)}{\xi} d\xi + 2(t_2 - t_1) \int_{t_2 - t_1}^{1} \frac{\omega(\phi, \xi)}{\xi^2} d\xi.$$

Now, we consider $I_\nu,\ \nu = 5, 6, 7.$

For I_5, we write

$$|I_5| = |\phi(t_1)| \left(\ln(1 + t_2) - \ln(1 + t_1)\right).$$

For the right-hand side, we use the mean value theorem on the interval $[1 + t_1, 1 + t_2]$ and obtain the following:

$$|I_5| = |\phi(t_1)| \frac{t_2 - t_1}{2 + t_1 + \theta.(t_2 - t_1)} = |\phi(t_1) - \phi(-1)| \frac{t_2 - t_1}{2 + t_1 + \theta.(t_2 - t_1)},\ \theta \in [0, 1].$$

Thus, we obtain the following estimate:

$$|I_5| \leq \omega(\phi, 1+t_1)\frac{t_2-t_1}{2+t_1+\theta\cdot(t_2-t_1)} \leq \frac{\omega(\phi, t_2-t_1)}{t_2-t_1}(1+t_1)\frac{t_2-t_1}{2+t_1+\theta.(t_2-t_1)}$$
$$\leq \frac{2}{2+t_1+\theta.(t_2-t_1)}\omega(\phi, t_2-t_1) \leq \omega(\phi, t_2-t_1).$$

Now, we consider the term $I_6 = \phi(t_2)\ln\frac{2-t_2}{2-t_1}$.

We rewrite the difference $\ln(2-t_1) - \ln(2-t_2)$ as follows:

$$\ln(2-t_1) - \ln(2-t_2) = \ln(1+(1-t_1)) - \ln(1+(1-t_2)).$$

We set $1-t_1 = w_1,\ 1-t_2 = w_2$ and deduce that $w_1, w_2 \in [0, 1]$ and $0 \leq w_2 < w_1 \leq 1$. Thus, we can write

$$\ln(2-t_1) - \ln(2-t_2) = \ln(1+w_1) - \ln(1+w_2).$$

Therefore,

$$|I_6| = |\phi(t_2)|\,(\ln(1+w_1) - \ln(1+w_2)).$$

Then, for the function $\ln(1+w)$, similarly to the previous term, we apply the mean value theorem on the interval $[1+w_2, 1+w_1]$ when $\theta \in [0, 1]$.,

$$|I_6| = |\phi(t_2)|\,\frac{w_1-w_2}{2+w_2+\theta.(w_1-w_2)} = |\phi(t_2)-\phi(0)|\,\frac{w_1-w_2}{2+w_2+\theta.(w_1-w_2)}.$$

Hence,

$$|I_6| = |\phi(t_2)-\phi(0)|\,\frac{t_2-t_1}{2+w_2+\theta.(t_2-t_1)} \leq \frac{t_2-t_1}{2}\omega(\phi, t_2)$$
$$\leq \frac{t_2-t_1}{2}\frac{\omega(\phi, t_2-t_1)}{t_2-t_1}t_2 \leq \frac{\omega(\phi, t_2-t_1)}{2}.$$

Also for I_7, we write

$$|I_7| = |\phi(t_1)-\phi(t_2)|\ln 3 \leq \omega(\phi, t_2-t_1)\ln 3.$$

Hence, we obtain the following estimate:

$$|I_5| + |I_6| + |I_7| \leq \left(\frac{3}{2} + \ln 3\right)\omega(\phi, t_2-t_1).$$

Furthermore, since we can write

$$\omega(\phi, t_2 - t_1) = \frac{\omega(\phi, t_2 - t_1)}{t_2 - t_1}(t_2 - t_1) = \int_0^{t_2 - t_1} \frac{\omega(\phi, t_2 - t_1)}{t_2 - t_1} d\xi,$$

we obtain

$$|I_5| + |I_6| + |I_7| \leq \left(\frac{3}{2} + \ln 3\right) \int_0^{t_2 - t_1} \frac{\omega(\phi, \xi)}{\xi} d\xi.$$

However, since for every $x \in (0, 1]$, $\omega(\varphi, x) = \omega(\phi, x)$ and for every $t \in [0, 1]$, $F\phi(t) = S\varphi(t)$, using the previous observations, we can deduce that

$$|S\varphi(t_2) - S\varphi(t_1)| \leq c_1 Z(\omega(\varphi, \cdot), t_2 - t_1), c_1 = \frac{67}{6} + \ln 3. \tag{19.30}$$

Now, we consider the difference $S_+\varphi(t_2) - S_+\varphi(t_1)$, which we can write

$$S_+\varphi(t_2) - S_+\varphi(t_1) = \frac{1}{\pi} \int_0^{t_2 - t_1} \frac{\varphi(\tau)}{\tau + t_2} d\tau - \frac{1}{\pi} \int_0^{t_2 - t_1} \frac{\varphi(\tau)}{\tau + t_1} d\tau$$
$$+ \frac{t_1 - t_2}{\pi} \int_{t_2 - t_1}^1 \frac{\varphi(\tau)}{(\tau + t_1)(\tau + t_2)} d\tau.$$

Since $|\varphi(\tau)| = |\varphi(\tau) - \varphi(0)| \leq \omega(\phi, \tau)$, we obtain

$$\begin{aligned} |S_+\varphi(t_2) - S_+\varphi(t_1)| &\leq \frac{1}{\pi} \int_0^{t_2 - t_1} \frac{\omega(\varphi, \tau)}{\tau + t_2} d\tau + \frac{1}{\pi} \int_0^{t_2 - t_1} \frac{\omega(\varphi, \tau)}{\tau + t_1} d\tau \\ &\quad + \frac{t_2 - t_1}{\pi} \int_{t_2 - t_1}^1 \frac{\omega(\varphi, \tau)}{(\tau + t_1)(\tau + t_2)} d\tau \\ &\leq \frac{2}{\pi} \int_0^{t_2 - t_1} \frac{\omega(\varphi, \xi)}{\xi} d\xi + \frac{t_2 - t_1}{\pi} \int_{t_2 - t_1}^1 \frac{\omega(\varphi, \xi)}{\xi^2} d\xi \\ &\leq \frac{2}{\pi} Z(\omega(\varphi, \cdot), t_2 - t_1). \end{aligned}$$

Thus

$$|S_+\varphi(t_2) - S_+\varphi(t_1)| \leq c_1 Z(\omega(\varphi, \cdot), t_2 - t_1), c_1 = \frac{2}{\pi}. \tag{19.31}$$

Since the function $Z(\omega(\varphi, \cdot), t)$, $t \in [0, 1]$ is non-decreasing, from inequalities (19.30) and (19.31), we see that inequalities (19.28) and (19.29) exist.

Thus, the proof of Theorem 19.3 is complete.

Theorem 19.4 *Let operators* $S : \overset{\circ}{H}_\alpha \to H_\alpha$ *and* $S_+ : \overset{\circ}{H}_\alpha \to H_\alpha$ *be defined as in formula (19.27) and* $\varphi \in \overset{\circ}{H}_\alpha$*. Then,*

$$\|S\|_\alpha \le A(\alpha), \;\; \|S_+\|_\alpha \le B(\alpha), \tag{19.32}$$

$$\|S\|_\infty \le C(\alpha), \;\; \|S_+\|_\infty \le D(\alpha). \tag{19.33}$$

Here, the constants $A(\alpha),\ B(\alpha),\ C(\alpha),$ *and* $D(\alpha)$ *depend only on parameter* α*.*

Proof According to Theorem 19.3, for each $\varphi \in J_0$ and $x \in (0, 1]$ we write

$$\begin{aligned}
\omega(S\varphi, x) &\le c_1 Z(\omega(\varphi, \cdot), x) = c_1 \left(\int_0^x \frac{\omega(\varphi, \xi)}{\xi} d\xi + x \int_x^1 \frac{\omega(\varphi, \xi)}{\xi^2} d\xi \right) \\
&= c_1 \left(\int_0^x \frac{\omega(\varphi, \xi)}{\xi^\alpha} \xi^{\alpha-1} d\xi + x \int_x^1 \frac{\omega(\varphi, \xi)}{\xi^\alpha} \xi^{\alpha-2} d\xi \right) \\
&\le c_1 H(\varphi, \alpha) \left(\frac{x^\alpha}{\alpha} + x . \frac{x^{\alpha-1} - 1}{1-\alpha} \right) \\
&= c_1 H(\varphi, \alpha) \frac{x^\alpha (1 - \alpha x^{1-\alpha})}{\alpha(1-\alpha)} \\
&\le \frac{c_1}{\alpha(1-\alpha)} x^\alpha H(\varphi, \alpha).
\end{aligned}$$

Thus,

$$\omega(S\varphi, x) \le \frac{c_1}{\alpha(1-\alpha)} x^\alpha H(\varphi, \alpha).$$

It follows that

$$H(S\varphi, \alpha) \le \frac{c_1}{\alpha(1-\alpha)} H(\varphi, \alpha) \le \frac{c_1}{\alpha(1-\alpha)} \|\varphi\|_{\alpha,0} . \tag{19.34}$$

Similarly

$$\omega(S_+\varphi, x) \le \frac{c_2}{\alpha(1-\alpha)} x^\alpha H(\varphi, \alpha)$$

and

$$H(S_+\varphi, \alpha) \le \frac{c_2}{\alpha(1-\alpha)} H(\varphi, \alpha) \le \frac{c_2}{\alpha(1-\alpha)} \|\varphi\|_{\alpha,0} . \tag{19.35}$$

From the definition of the operator $F\phi(t)$, we write

$$
\begin{aligned}
\pi\,|F\phi(t)| &\le \int_{-1}^{2}\frac{\omega(\phi,|\tau-t|)}{|\tau-t|}d\tau+\omega(\phi,1+t)\ln 2\\
&=\int_{-1}^{t}\frac{\omega(\phi,t-\tau)}{t-\tau}d\tau+\int_{t}^{2}\frac{\omega(\phi,\tau-t)}{\tau-t}d\tau+\omega(\phi,1+t)\ln 2\\
&=\int_{0}^{1+t}\frac{\omega(\phi,\xi)}{\xi}d\xi+\int_{0}^{2-t}\frac{\omega(\phi,\xi)}{\xi}d\xi+\omega(\phi,1+t)\ln 2\\
&\le H(\varphi,\alpha)\left(\int_{0}^{1+t}\xi^{\alpha-1}d\xi+\int_{0}^{2-t}\xi^{\alpha-1}d\xi+(1+t)^{\alpha}\ln 2\right)\\
&=H(\varphi,\alpha)\left(\frac{(1+t)^{\alpha}+(2-t)^{\alpha}}{\alpha}+(1+t)^{\alpha}\ln 2\right)\\
&\le 2^{\alpha}\left(\frac{2}{\alpha}+\ln 2\right)H(\varphi,\alpha).
\end{aligned}
$$

According to this

$$
\|S\varphi\|_{\infty}\le\frac{2^{\alpha}}{\pi}\left(\frac{2}{\alpha}+\ln 2\right)H(\varphi,\alpha)\le\frac{2^{\alpha}}{\pi}\left(\frac{2}{\alpha}+\ln 2\right)\|\varphi\|_{\alpha,0}\,. \tag{19.36}
$$

Hence,

$$
\|S\|_{\infty}\le C(\alpha),\,C(\alpha)=\frac{2^{\alpha}}{\pi}\left(\frac{2}{\alpha}+\ln 2\right).
$$

From (19.34) and (19.36), we obtain

$$
\|S\varphi\|_{\alpha}\le A(\alpha)\,\|\varphi\|_{\alpha,0}\,,\,A(\alpha)=\max\left(\frac{c_1}{\alpha(1-\alpha)},\,C(\alpha)\right).
$$

Finally,

$$
\|S\|_{\alpha}\le A(\alpha).
$$

Now, we consider $\|S_{+}\varphi\|_{\infty}$.

Writing

$$
S_{+}\varphi(t)=\frac{1}{\pi}\int_{0}^{1}\frac{\varphi(\tau)}{\tau+t}d\tau=\frac{1}{\pi}\int_{0}^{1}\frac{\varphi(\tau)-\varphi(0)}{\tau+t}d\tau=\frac{1}{\pi}\int_{0}^{1}\frac{\varphi(\tau)-\varphi(0)}{\tau^{\alpha}}\frac{\tau^{\alpha}}{\tau+t}d\tau
$$

we obtain

$$|S_+\varphi(t)| \leq \frac{H(\varphi,\alpha)}{\pi}\int_0^1 \frac{\tau^\alpha}{\tau+t}d\tau \leq \frac{H(\varphi,\alpha)}{\pi}\int_0^1 \tau^{\alpha-1}d\tau \leq \frac{1}{\alpha\pi}\,\|\varphi\|_{\alpha,0}\,.$$

Thus,

$$\|S_+\varphi\|_\infty \leq \frac{1}{\alpha\pi}\,\|\varphi\|_{\alpha,0}\,. \tag{19.37}$$

Hence,

$$\|S_+\|_\infty \leq D(\alpha),\, D(\alpha) = \frac{1}{\alpha\pi}.$$

From (19.35) and (19.37), we obtain

$$\|S_+\varphi\|_\alpha \leq B(\alpha)\,\|\varphi\|_{\alpha,0}\,,\, B(\alpha) = \max\left(\frac{c_2}{\alpha(1-\alpha)},\, D(\alpha)\right).$$

Finally,

$$\|S_+\|_\alpha \leq B(\alpha).$$

Thus, the proof of Theorem 19.4 is complete.

19.3.2 Existence of the Solutions due to Banach Contraction Principle

Using Theorems 19.3 and 19.4, we can show that the operator

$$A\varphi(t) = f(t)\left\{\varphi^2(t) + [\lambda - S\varphi(t) + \mu\cdot S_+\varphi(t)]^2\right\},\;\; t\in[0,1]$$

is a contraction mapping. Furthermore, we can show that the operator A maps a closed sphere of space into itself. Thus, the conditions of the Banach contraction mapping principle are satisfied for the operator equation

$$\varphi(t) = A\varphi(t),\;\; t\in[0,1]\,;$$

in other words, this equation has a unique solution.

19.4 Fixed Point Theory and Approximate Solutions of Nonlinear Singular Integral Equations

In this section, the existence and uniqueness of the solution of a nonlinear singular integral equation with Cauchy kernel defined on a closed, simple smooth curve in the complex plane are researched in Holder space. In addition, an iteration method is given for the solution of the nonlinear singular integral equation. It is demonstrated that the iteration which was constructed for approximate solution converges to the definite solution. To obtain the existence of the solutions, fixed point property is used.

In nonlinear singular integral equation theory, the solution of an equation is approached from two directions. The first problem is to prove the existence and uniqueness of the solution of the equation; the second problem is to find the solution of the equation (if a solution exists). Sometimes, proving the existence of the solution of nonlinear singular integral equations is difficult, and it is important to find the solution. Furthermore, the solution of that equation (if it is exists) may not necessarily be obtained analytically. In such cases, approximate solutions are investigated.

Thus, in this section, the solution of a general nonlinear singular integral equation

$$F(t, \varphi(t), S_\gamma k(., \varphi(.))(t)) = f(t), \ t \in \gamma \tag{19.38}$$

is studied, where γ is a closed, simple smooth curve for the sets D_0 and $\tilde{D}$ given by $D_0 = \{(t, \varphi) : \ t \in \gamma, \ \varphi \in \mathbb{C}\}$ and $\tilde{D} = \{(t, \varphi, s) : \ t \in \gamma, \ \varphi, s \in \mathbb{C}\}$. When the functions $k : D_0 \to \mathbb{C}$, $F : \tilde{D} \to \mathbb{C}$ and $f : \gamma \to \mathbb{C}$ are given functions,

$$S_\gamma k(., \varphi(.))(t) = \frac{1}{\pi i} \int_\gamma \frac{k(\tau, \varphi(\tau))}{\tau - 1} d\tau, \ \ t \in \gamma$$

is a nonlinear singular integral equation with a Cauchy kernel respect to unknown function $\varphi : \gamma \to \mathbb{C}$.

In our study, Eq. (19.38) was investigated by using functional analysis methods and conditions were found for the existence of the unique solution of the equation in Holder space $(H_\alpha(\gamma); \| \cdot \|_\alpha)$. In addition, an iteration was constructed for approximate solution and was demonstrated to be convergent.

Throughout this section, we will assume that the curve $\gamma \subset \mathbb{C}$ is a closed, simple smooth curve and that $\ell = |\gamma$ denotes the length of the curve γ.

Definition 19.4 ([22, 67]) If there exists a constant $c > 0$ and $\alpha \in (0, 1]$ such that

$$|\varphi(t_1) - \varphi(t_2)| \leq c. |t_1 - t_2|^\alpha ,$$

for $\forall t_1, t_2 \in \gamma$, then it is said that the function $\varphi : \gamma \to \mathbb{C}$ satisfies the Holder condition with the exponent α on the curve γ.

$H_\alpha(\gamma)$ denotes the set of all the functions $\varphi : \gamma \to \mathbb{C}$ that satisfy the Holder condition with the exponent α on the curve γ. For $\varphi \in H_\alpha(\gamma)$, $\alpha \in (0, 1)$ the vector space $(H_\alpha(\gamma); \|\,.\,\|_\alpha)$ is a Banach space with norm

$$\|\varphi\|_\alpha = \|\varphi\|_{h_\alpha(\gamma)} \equiv \|\varphi\|_\infty + H(\varphi; \alpha),$$

where

$$\|\varphi\|_\infty = \max\left\{|\varphi(t)| : t \in \gamma\right\}, \; H(\varphi; \alpha) = \sup\left\{\frac{|\varphi(t_1) - \varphi(t_2)|}{|t_1 - t_2|^\alpha} : t_1, t_2 \in \gamma, t_1 \neq t_2\right\}$$

are as given. Let $\varphi \in L_p(\gamma) = \{\varphi : \gamma \to \mathbb{C} : \int_\gamma |\varphi(\tau)|^p \, |d\tau| < +\infty \}$, $p > 1$.

In this case, $\forall \varphi \in L_p(\gamma)$, the vector space$(L_p(\gamma); \|.\|_p)$ is a Banach space with norm

$$\|\varphi\|_p = \|\varphi\|_{L_p(\gamma)} \equiv \left(\int_\gamma |\varphi(\tau)|^p \, |d\tau|\right)^{1/p}$$

(see [14]).

Theorem 19.5 *Given two metric spaces* (X, ρ_1) *and* (X, ρ_2) *let the following assumptions hold:*

1. (X, ρ_1) *is a compact metric space;*
2. *Every convergent sequence in X with respect to the* ρ_1 *metric is also convergent with respect to another metric* ρ_2*;*
3. *The operator* $A : X \to X$ *is a contraction map with respect to the* ρ_2 *metric. That is, there exists a number* $0 \leq q_1 < 1$*, such that*

$$\rho_2(Ax, Ay) \leq q.\rho_2(x, y), \; \forall x, y \in X$$

In this case, the operator equation

$$x = Ax$$

has a unique solution $x_* \in X$*, and for any initial approximation* $x_0 \in X$ *the sequence* $(x_n) \subset X, n = 1, 2, \ldots$ *defined as*

$$x_n = Ax_{n-1}, n = 1, 2, \ldots$$

converges to the solution x_* *with a "velocity"*

$$\rho_2(x_n, x_*) \leq \frac{q^n}{1 - q}.\rho_2(x_0, x_1).$$

Proof Proof is well known but if one would like to see [24], can be good source.

Let us express Eq. (19.38) as

$$\varphi(t) = \Phi(t, \varphi(t), f(t), S_\gamma k(., \varphi(.))(t)), t \in \gamma \tag{19.39}$$

where

$\Phi(t, \varphi(t), f(t), S_\gamma k(., \varphi(.))(t)) = \varphi(t) - f(t) + F(t, \varphi(t), S_\gamma k(., \varphi(.))(t)),\ t \in \gamma$ is as given.

Equations (19.38) and (19.39) are equivalent. Therefore, we will henceforth investigate Eq. (19.39) instead of Eq. (19.38).

Definition 19.5 If $\forall t_1, t_2 \in \gamma$ and $\forall (t_k, \varphi_k) \in D_0, \forall (t_k, \varphi_k, f_k, s_k) \in D,\ k = 1, 2, \alpha \in (0, 1)$ there exist positive numbers $m_1, m_2, n_1, n_2, n_3, n_4$ which satisfy the following inequalities

$$|k(t_1, \varphi_1) - k(t_2, \varphi_2)| \le m_1 |t_1 - t_2|^\alpha + m_2 |\varphi_1 - \varphi_2| \tag{19.40}$$

and

$$\begin{aligned} |\Phi(t_1, \varphi_1, f_1, s_1) - \Phi(t_2, \varphi_2, f_2, s_2)| \le n_1 |t_1 - t_2|^\alpha + n_2 |\varphi_1 - \varphi_2| \\ + n_3 |f_1 - f_2| + n_4 |s_1 - s_2| \end{aligned} \tag{19.41}$$

where $D = \tilde{D} \times \mathbb{C} = \{(t, \varphi, f, s) : t \in \gamma, \varphi, f, s \in \mathbb{C}\}$ then we will say that the functions $k(t, \varphi)$and $\Phi(t, \varphi, f, s)$ are from classes $H_{\alpha,1}(m_1, m_2; D_0)$ and $H_{\alpha,1,1,1}$ $(n_1, n_2, n_3, n_4; D)$, respectively.

19.4.1 *On the Solution of Nonlinear Singular Integral Equations*

This section addresses the solution of Eq. (19.39). First, let us give some auxiliary lemmas.

Lemma 19.1 *If* $\varphi \in H_\alpha(\gamma)$, $\alpha \in (0, 1)$*, then for* $\forall p > 1$*and* $\forall \varepsilon \in (0, \ell]$ *the following inequality is true*

$$\|\varphi\|_\infty \le \varepsilon^\alpha \|\varphi\|_\alpha + \frac{\varepsilon^{-1/p}}{\sqrt[p]{2}} \|\varphi\|_p \tag{19.42}$$

Proof Let the function $\varphi \in H_\alpha(\gamma)$, $\alpha \in (0, 1)$ be given. For any $t, \tau \in \gamma$, let us denote the shortest arc length between the point's t and τ on the curve γ with $\rho(t, \tau)$. In this case, since γ is a simple curve for any point t on γ there exists $\varepsilon > 0$ such that $U(t; \varepsilon) = \{\tau \in \gamma : p(t, \tau) < \varepsilon\} \subset \gamma$.

Therefore, it can be written that

$$\varphi(t) = \frac{1}{2\varepsilon} \int_{U(t;\varepsilon)} \varphi(\tau) d\tau + \frac{1}{2\varepsilon} \int_{U(t;\varepsilon)} [\varphi(t) - \varphi(\tau)] \, d\tau, \ \ t \in \gamma.$$

Accordingly, if we apply the Holder inequality to the first integral on the right-hand side of the above equality, for any point t on γ. Then we have

$$\begin{aligned} |\varphi(t)| &\leq \frac{1}{2\varepsilon} \int_{U(t;\varepsilon)} |\varphi(\tau)| \, |d\tau| + \frac{1}{2\varepsilon} H(\varphi; \alpha) \int_{U(t;\varepsilon)} |t - \tau|^{\alpha} \, |d\tau| \\ &\leq \frac{1}{2\varepsilon} \left(\int_{U(t;\varepsilon)} |\varphi(\tau)|^{p} \, |d\tau| \right)^{1/p} . \left(\int_{U(t;\varepsilon)} |d\tau| \right)^{1-\frac{1}{p}} + \varepsilon^{\alpha} . H(\varphi; \alpha) \\ &\leq \frac{1}{(2\varepsilon)^{1/p}} . \, \|\varphi\|_{p} + \varepsilon^{\alpha} . \, \|\varphi\|_{\alpha} \, , t \in \gamma. \end{aligned}$$

Consequently,

$$|\varphi(t)| \leq \varepsilon^{\alpha} . \, \|\varphi\|_{\alpha} + \frac{1}{(2\varepsilon)^{1/p}} . \, \|\varphi\|_{p} \, .$$

From the last inequality the proof of inequality (19.42) is immediate.

Lemma 19.2 $\forall \varphi \in H_{\alpha}(\gamma)$, $\alpha \in (0, 1)$ *and* $p > 1$, *the following inequality is true:*

$$\|\varphi\|_{\infty} \leq M(\alpha, p) . \, \|\varphi\|_{\alpha}^{\frac{1}{1+\alpha p}} \, . \, \|\varphi\|_{p}^{\frac{\alpha p}{1+\alpha p}} \, . \tag{19.43}$$

Here,

$$M\left(\alpha, p\right) = \max\left\{M_{1}\left(\alpha, p\right), M_{2}\left(\alpha, p\right)\right\},$$

$$M_{1}\left(\alpha, p\right) = \left(n\left(\alpha, p\right)\right)^{\alpha} + \left(2.n\left(\alpha, p\right)\right)^{-1/p},$$

$$M_{2}\left(\alpha, p\right) = \frac{\sqrt[p]{2}}{\sqrt[p]{2} - 1} \left(n\left(\alpha, p\right)\right)^{\alpha}, n\left(\alpha, p\right) = \left(\alpha . p . \sqrt[p]{2}\right)^{\frac{-p}{1+\alpha p}}$$

Proof For convenience, if we take $A = \|\varphi\|_{\infty}$, $B = \|\varphi\|_{\alpha}$, $C = \|\varphi\|_{p}$, then from Lemma 19.1 for $\varepsilon \in (0, \ell]$, we can write

$$A \leq \varepsilon^{\alpha} . B + \frac{C}{2^{1/p}} \varepsilon^{-1/p} \tag{19.44}$$

$\forall\alpha \in (0, 1)$, $p > 1$ let us define the function$\theta = \theta(\varepsilon)$, $\theta : (0, +\infty) \to (0, +\infty)$ as below

$$\theta\,(\varepsilon) = B.\varepsilon^{\alpha} + \frac{C}{2^{1/p}}\varepsilon^{-1/p}.$$

It is easy to show that the function $\theta = \theta(\varepsilon)$ takes its smallest value at the point $\varepsilon_* = n(\alpha, p).(C/B)^{\frac{p}{1+\alpha p}}$ where $n(\alpha, p) = (\alpha.p.\sqrt[p]{2})^{\frac{-p}{1+\alpha p}}$.

Therefore, we have

$$\min\{\theta(\varepsilon) : \varepsilon \in (0, +\infty)\} = \theta(\varepsilon_*) = M_1(\alpha, p).B^{\frac{1}{1+\alpha p}}.C^{\frac{\alpha p}{1+\alpha p}}, \tag{19.45}$$

where $M_1(\alpha, p) = (n(\alpha, p))^{\alpha} + (2.n(\alpha, p))^{-1/p}$.

If $\varepsilon_* \in (0, \ell]$, then from Eqs. (19.44) and (19.45) we obtain

$$\|\varphi\|_{\infty} \le M_1(\alpha, p).\ \|\varphi\|_{\alpha}^{\frac{1}{1+\alpha p}} .\ \|\varphi\|_{p}^{\frac{\alpha p}{1+\alpha p}}. \tag{19.46}$$

If $\varepsilon_* > \ell$, and if we take into consideration that

$C = \|\varphi\|_p \le \|\varphi\|_{\infty} \cdot \left(\int_{\gamma} |d\tau|\right)^{1/p} = \ell^{1/p}.A,$

then from Eq. (19.44) we have

$$A \le B.\ell^{\alpha} + \frac{\ell^{-1/p}}{\sqrt[p]{2}}.\ell^{-1/p}.A.$$

Consequently, we have

$$A \le \frac{\sqrt[p]{2}}{\sqrt[p]{2} - 1}.B.\ell^{\alpha}.$$

Thus, for $\ell < \varepsilon_*$ we have

$$A \le \frac{\sqrt[p]{2}}{\sqrt[p]{2} - 1}.B.\varepsilon_*^{\alpha} = \frac{\sqrt[p]{2}}{\sqrt[p]{2} - 1}.\ (n(\alpha, p))^{\alpha}.B^{\frac{1}{1+\alpha p}}.C^{\frac{\alpha p}{1+\alpha p}}.$$

Finally, we obtain

$$\|\varphi\|_{\infty} \le M_2(\alpha, p).\ \|\varphi\|_{\alpha}^{\frac{1}{1+\alpha p}} .\ \|\varphi\|_{p}^{\frac{\alpha p}{1+\alpha p}}, \tag{19.47}$$

where $M_2(\alpha, p) = \frac{\sqrt[p]{2}}{\sqrt[p]{2}-1}.\ (n(\alpha, p))^{\alpha}$.

From inequalities (19.46) and (19.47), Lemma 19.2 is proven.

Lemma 19.3 *Let* $D_{0r} = \{(t, \varphi) \in D_0 : t \in \gamma, |\varphi| \le r\}$, $r > 0$, $B_{\alpha}(0; r) = \left\{\varphi \in H_{\alpha}(\gamma) : \|\varphi\|_{\alpha} \le r\right\}$, $\alpha \in (0, 1)$ *and* $k_1(t) = k(t, \varphi(t))$, $t \in \gamma$.

If $\varphi \in B_\alpha(0; r)$ *and* $k \in H_{\alpha,1}(m_1, m_2; D_{0r})$*, for the function* $k_1(t)$ *the following inequalities are hold*

$$|k_1(t)| \leq m_0 + m_2.r, t \in \gamma,$$

$$|k_1(t_1) - k_1(t_2)| \leq (m_1 + m_2.r) \,.\, |t_1 - t_2|^\alpha \,, t_1, t_2 \in \gamma. \tag{19.48}$$

Here $m_0 = \max\{|k(t, 0)| : t \in \gamma\}$.

Proof $\forall t, t_1, t_2 \in \gamma$ and $\alpha \in B_\alpha(0; r)$, $0 < \alpha < 1$, $r > 0$, in accordance with (19.40) it can be seen that

$$|k_1(t)| \leq |k(t, \varphi(t)) - k(t, 0)| + |k(t, 0)| \leq m_2 |\varphi(t)| + m_0 \leq m_0 + m_2.r$$

and

$$\begin{aligned} |k_1(t_1) - k_1(t_2)| &= |k(t_1, \varphi(t_1)) - k(t_2, \varphi(t_2))| \\ &\leq m_1.\, |t_1 - t_2|^\alpha + m_2.\, |\varphi(t_1) - \varphi(t_2)| \\ &\leq (m_1 + m_2.H(\varphi, \alpha; \gamma))\,.\, |t_1 - t_2|^\alpha \\ &\leq \left(m_1 + m_2.\, \|\varphi\|_\alpha\right).\, |t_1 - t_2|^\alpha \\ &\leq (m_1 + m_2.r)\,.\, |t_1 - t_2|^\alpha \,. \end{aligned}$$

Therefore, inequality (19.48) is proven.

Corollary 19.1 *If the assumptions of Lemma 19.3 are satisfied, then* $k_1 \in H_\alpha(\gamma)$ *and for the norm* $\|k_1\|_\alpha$ *the following evaluation is true*

$$\|k_1\|_\alpha \leq m_0 + m_1 + 2.m_2.r. \tag{19.49}$$

Now, for $k \in H_{\alpha,1}(m_1, m_2; D_0)$ and $\varphi \in H_\alpha(\gamma)$,$0 < \alpha < 1$, let us the define the function $\tilde{k}_1 : \gamma \to \mathbb{C}$ as

$$\tilde{k}_1(t) = S_\gamma k(., \varphi(.))(t), t \in \gamma. \tag{19.50}$$

For the bounded operator $S_\gamma : H_\alpha(\gamma) \to H_\alpha(\gamma)$, $\alpha \in (0, 1)$ (see [22, 24]) let the norm below is given

$$\left\|S_\gamma\right\|_\alpha = \left\|S_\gamma\right\|_{H_\alpha(\gamma)} \equiv \sup\left\{\left\|S_\gamma \varphi\right\|_\alpha : \varphi \in H_\alpha(\gamma), \|\varphi\|_\alpha \leq 1\right\}. \tag{19.51}$$

Lemma 19.4 *If* $k \in H_{\alpha,1}(m_1, m_2; D_{0r})$*and*$\varphi \in B_\alpha(0; r)$,$0 < \alpha < 1$*, then the function* $\tilde{k}_1(t)$ *that is defined by formula (19.50) is of the class* $H_\alpha(\gamma)$ *and the following evaluation for the norm* $\|k_1\|_\alpha$ *is true,*

$$\|k_1\|_\alpha \leq (m_0 + m_1 + 2.m_2.r).\left\|S_\gamma\right\|_\alpha = L_1. \tag{19.52}$$

Proof From Corollary 19.1, the definition of $\tilde{k}_1(t)$ and the property of the operator $S_\gamma : H_\alpha(\gamma) \to H_\alpha(\gamma)$,$\alpha \in (0, 1)$ (see [4, 10]), it is obvious that $\tilde{k}_1 \in H_\alpha(\gamma)$, $\alpha \in (0, 1)$.

From formula (19.50) and the boundedness of the operator $S_\gamma : H_\alpha(\gamma) \to H_\alpha(\gamma)$, $\alpha \in (0, 1)$ we have

$$\left\|\tilde{k}_1\right\|_\alpha \leq \|k(., \varphi(.))\|_\alpha . \left\|S_\gamma\right\|_\alpha .$$

From this inequality and the inequality (19.49), the lemma is proven.

Lemma 19.5 *$\forall r > 0$ and $\forall t \in \gamma$, let $D_r = \{(t, \varphi, f, s) : t \in \gamma,\ |\varphi| \leq r, |f| \leq r, |s| \leq r\}$ and*

$$\tilde{\Phi}(t) = \Phi(t, \varphi(t), f(t), (S_\gamma \circ k_1)(t)).$$

If $k \in H_{\alpha,1}(m_1, m_2; D_{0r})$, $\Phi \in H_{\alpha,1,1,1}(n_1, n_2, n_3, n_4; D_r)$ and $\varphi, f \in B_\alpha(0; r)$,$\alpha \in (0, 1)$, then for $\forall t, t_1, t_2 \in \gamma$ the following inequalities hold

$$\left|\tilde{\Phi}(t)\right| \leq n_0+n_2.r+n_3.L_1, \left|\tilde{\Phi}(t_1) - \tilde{\Phi}(t_2)\right| \leq n_1+(n_2+n_3).r+n_4.L_1) \left|t_1 - t_2\right|^\alpha .$$

Here, $n_0 = \|\Phi(., 0, 0, 0)\|_\infty$.

Proof $\forall \varphi, f \in B_\alpha(0; r)$,$\alpha \in (0, 1)$ and $\forall t, t_1, t_2 \in \gamma$, as required by inequality (19.41) and Lemma 19.4, it can be easily seen that

$$\begin{aligned}\left|\tilde{\Phi}(t)\right| &\leq \left|\Phi(t, \varphi(t), f(t), (S_\gamma \circ k_1)(t) - \Phi(t, 0, 0, 0)\right| + |\Phi(t, 0, 0, 0)| \\ &\leq n_2. |\varphi(t)| + n_3. |f(t)| + n_4. \left|(S_\gamma \circ k_1)(t)\right| + n_0 \\ &\leq n_0 + (n_2 + n_3).r + n_4.L_1\end{aligned}$$

and

$$\begin{aligned}\left|\tilde{\Phi}(t_1) - \tilde{\Phi}(t_2)\right| &\leq (n_1. |t_1 - t_2|^\alpha + n_2. |\varphi(t_1) - \varphi(t_2)| + n_3. |f(t_1) - f(t_2)| \\ &\quad + n_4. \left|(S_\gamma \circ k_1)(t_1) - (S_\gamma \circ k_1)(t_2)\right| \\ &\leq (n_1 + (n_2 + n_3).r + n_4.L_1). |t_1 - t_2|^\alpha .\end{aligned}$$

From Lemma 19.5, we can obtain the following conclusions:

Corollary 19.2 *If the assumptions of Lemma 19.5 are satisfied, then $\tilde{\Phi} \in H_\alpha(\gamma)$ and, for the norm $\left\|\tilde{\Phi}\right\|_\alpha$, the following inequality holds*

$$\left\|\tilde{\Phi}\right\|_\alpha \leq 2. \{\max(n_0, n_1) + (n_2 + n_3).r + n_4.L_1)\} = L.$$

Corollary 19.3 *If the assumptions of Lemma 19.5 are satisfied, then the operator A that is defined by the following equality*

$$A(\varphi)(t) = \Phi(t, \varphi(t), f(t), (S_\gamma \circ k_1)(t)), t \in \gamma \quad (19.53)$$

transforms the sphere $B_\alpha(0; r)$ to the sphere $B_\alpha(0; L)$.

Corollary 19.4 *If the assumptions of Lemma 19.5 are satisfied and, if $L \leq r$, then the operator A that is defined by formula (19.53) transforms the sphere $B_\alpha(0; r)$ to itself.*

Lemma 19.6 *The sphere $B_\alpha(0; r)$,$\alpha \in (0, 1)$ is a compact set in the space $(H_\alpha(\gamma); \|\varphi\|_\infty)$.*

Proof From the definition of the sphere $B_\alpha(0; r)$, $\forall\alpha \in (0, 1)$ and $\forall\varphi \in B_\alpha(0; r)$, we have $\|\varphi\|_\alpha \leq r$ and thus we have $\|\varphi\|_\infty \leq r$.

Consequently, it is obvious that the sphere $B_\alpha(0; r)$is a uniformly bounded set in the space $(H_\alpha(\gamma); \|\varphi\|_\infty)$.

Furthermore, for $\forall\varepsilon > 0$ if we take$\delta = (\varepsilon/r)^{1/\alpha}$, then for $\forall t_1, t_2 \in \gamma$ and $\forall\varphi \in B_\alpha(0; r)$ we have

$$|t_1 - t_2| < \delta \Rightarrow |\varphi(t_1) - \varphi(t_2)| \leq r. |t_1 - t_2|^\alpha < \varepsilon.$$

From here, it is seen that the elements of the sphere $B_\alpha(0; r)$ are continuous at the same degree. Therefore, by the Arzela–Ascoli compactness theorem the sphere $B_\alpha(0; r)$ is a compact set of the space $(H_\alpha(\gamma); \|\varphi\|_\infty)$.

From that point, we yield the conclusion.

From Lemma 19.6 we can state the following corollary:

Corollary 19.5 $\forall\alpha \in (0, 1)$*, the sphere $B_\alpha(0; r)$ is a complete subspace of the space $(H_\alpha(\gamma); \|\varphi\|_\infty)$.*

Now, let us define two transformations in the space $H_\alpha(\gamma)$. $\forall\varphi, \tilde{\varphi} \in H_\alpha(\gamma)$, $\alpha \in (0, 1)$,$p > 1$ let $d_\infty(\varphi, \tilde{\varphi}) = \|\varphi - \tilde{\varphi}\|_\infty$ and $d_p(\varphi, \tilde{\varphi}) = \|\varphi - \tilde{\varphi}\|_p$ be as given.

It is obvious that the transformations $d_\infty : H_\alpha(\gamma)\text{x}H_\alpha(\gamma) \to [0, +\infty)$ and $d_p : H_\alpha(\gamma)\text{x}H_\alpha(\gamma) \to [0, +\infty)$ are metrics in the space $H_\alpha(\gamma)$, $\alpha \in (0, 1)$ and therefore $(H_\alpha(\gamma); d_\infty)$ and $(H_\alpha(\gamma); d_p)$ are metric spaces.

Lemma 19.7 $\forall\alpha \in (0, 1)$ *and $p > 1$, convergence according to the metrics d_∞ and d_p is equivalent in the subspace $B_\alpha(0; r)$.*

Proof Let $\varphi_0, \varphi_n \in B_\alpha(0; r)$, $\alpha \in (0, 1)$, $n = 1, 2, \ldots$. In this case, since

$$d_p(\varphi_0, \varphi_n) = \|\varphi_0 - \varphi_n\|_p \leq \ell^{1/p}. \|\varphi_0 - \varphi_n\|_\infty = \ell^{1/p}.d_\infty(\varphi_0, \varphi_n)$$

is as given, it is obvious that

$$\lim_{n\to\infty} d_\infty(\varphi_0, \varphi_n) = 0 \Rightarrow \lim_{n\to\infty} d_p(\varphi_0, \varphi_n) = 0.$$

Let us see that the converse of this is also true. From inequality (19.43) for$\varphi_0, \varphi_n \in B_\alpha(0; r)$, we can write

$$\|\varphi_0 - \varphi_n\|_\infty \le M(\alpha, p). \|\varphi_0 - \varphi_n\|_\alpha^{\frac{1}{1+\alpha p}} . \|\varphi_0 - \varphi_n\|_p^{\frac{\alpha p}{1+\alpha p}}$$

or

$$d_\infty(\varphi_0, \varphi_n) \le (2r)^{\frac{1}{1+\alpha p}} .M(\alpha, p).(d_p(\varphi_0, \varphi_n))^{\frac{\alpha p}{1+\alpha p}} .$$

From here, we have

$$\lim_{n\to\infty} d_p(\varphi_0, \varphi_n) = 0 \Rightarrow \lim_{n\to\infty} d_\infty(\varphi_0, \varphi_n) = 0.$$

Therefore, lemma is proven.

Now, for the bounded operator $S_\gamma : L_p(\gamma) \to L_p(\gamma)$, $p \in (1, +\infty)$ let us assume that

$$\left\| S_\gamma \right\|_p = \left\| S_\gamma \right\|_{L_p(\gamma)} \equiv \sup \left\{ \left\| S_\gamma \varphi \right\|_p : \|\varphi\|_p \le 1 \right\} .$$

Lemma 19.8 *If* $k \in H_{\alpha,1}(m_1, m_2; D_{0r})$ *and* $\Phi \in H_{\alpha,1,1,1}(n_1, n_2, n_3, n_4; D_{0r})$, $\alpha \in (0, 1)$*, then the operator A, that is defined by formula (19.53), satisfies the following inequality:*

$$d_p\left(A(\varphi), A(\tilde{\varphi})\right) \le M_3(p).d_\infty(\varphi, \tilde{\varphi}),\ p > 1, \tag{19.54}$$

$\forall \varphi, \tilde{\varphi} \in B_\alpha(0; r)$. *Here,*$M_3(p) = (n_2 + n_4.m_2. \left\| S_\gamma \right\|_p).\ell^{1/p}$.

Proof $\forall t \in \gamma$ and $\forall \varphi, \tilde{\varphi} \in B_\alpha(0; r)$ from the assumption of the lemma we can obtain

$$\begin{aligned} |A(\varphi)(t) - A(\tilde{\varphi})(t)| &= \left| \begin{array}{l} \Phi(t, \varphi(t), f(t), S_\gamma k(., \varphi(.)(t)) \\ -\Phi(t, \tilde{\varphi}(t), f(t), S_\gamma k(., \tilde{\varphi}(.)(t)) \end{array} \right| \\ &\le n_2. |\varphi(t) - \tilde{\varphi}(t)| + n_4. \left| S_\gamma k(., \varphi(.)) - S_\gamma k(., \tilde{\varphi}(.)))(t) \right| . \end{aligned}$$

From the last inequality, according to the Minkowski inequality and the assumptions of the lemma we have

$$\begin{aligned} \left(\int_\gamma |A(\varphi)(t) - A(\tilde{\varphi})(t)|^p |dt| \right)^{1/p} &= \left(\int_\gamma \left[\begin{array}{l} n_2. |\varphi(t) - \tilde{\varphi}(t)| \\ +n_4. \left| \begin{array}{l} S_\gamma (k(., \varphi(.)) \\ -k(., \tilde{\varphi}(.)))(t) \end{array} \right| \end{array} \right]^p |dt| \right)^{1/p} \\ &\le n_2. \|\varphi - \tilde{\varphi}\|_p \end{aligned}$$

$$+n_4. \left\| S_\gamma \right\|_p . \|k(., \varphi(.)) - k(., \tilde{\varphi}(.))\|_p$$
$$\leq \left(n_2 + n_4.m_2. \left\| S_\gamma \right\|_p \right) . \|\varphi - \tilde{\varphi}\|_p .$$

So, we have

$$d_p\left(A(\varphi), A(\tilde{\varphi})\right) \leq (n_2 + n_4.m_2. \left\| S_\gamma \right\|_p). \|\varphi - \tilde{\varphi}\|_p .$$

$\forall p > 1$, since

$$\|\varphi - \tilde{\varphi}\|_p \leq \ell^{1/p}. \|\varphi - \tilde{\varphi}\|_\infty$$

from the last inequality we have

$$d_p\left(A(\varphi), A(\tilde{\varphi})\right) \leq (n_2 + n_4.m_2. \left\| S_\gamma \right\|_p).\ell^{1/p}.d_\infty(\varphi, \tilde{\varphi}).$$

Therefore, the lemma is proven.

Lemma 19.9 *If* $k \in H_{\alpha,1}(m_1, m_2; D_{0r})$ *and* $\Phi \in H_{\alpha,1,1,1}(n_1, n_2, n_3, n_4; D_{0r})$, $\alpha \in (0, 1)$*, then the operator* $A : B_\alpha(0; r) \to B_\alpha(0; r)$*,* $\alpha \in (0, 1)$ *is a continuous operator according to the metric* d_∞*.*

Proof $\forall \varphi_0, \varphi_n \in B_\alpha(0; r)$, $\alpha \in (0, 1)$, $n = 1, 2, \ldots$, in accordance with Lemma 19.2 we can write

$$\|A(\varphi_0) - A(\varphi_n)\|_\infty \leq M(\alpha, p). \|A(\varphi_0) - A(\varphi_n)\|_\alpha^{\frac{1}{1+\alpha p}} . \|A(\varphi_0) - A(\varphi_n)\|_p^{\frac{\alpha p}{1+\alpha p}} .$$

From here, in accordance with inequality (19.54) we obtain:

$$\|A(\varphi_0) - A(\varphi_n)\|_\infty \leq c.(d_\infty(\varphi_0, \varphi_n))^{\frac{\alpha p}{1+\alpha p}} .$$

Here, the constant c that is defined as
$c = c(\alpha, p) = M(\alpha, p). (M_3(p))^{\frac{\alpha p}{1+\alpha p}} . (2.r)^{\frac{1}{1+\alpha p}}$ is a positive number independent of n.

From the last inequality, if $\lim\limits_{n\to\infty} d_\infty(\varphi_0, \varphi_n) = 0$, then it is understood that $\lim\limits_{n\to\infty} d_\infty(A(\varphi_0), A(\varphi_n)) = 0$.

Now, let us present the following corollaries about the solution of the nonlinear singular integral equation (19.39).

Theorem 19.6 *If* $k \in H_{\alpha,1}(m_1, m_2; D_{0r})$*,* $\Phi \in H_{\alpha,1,1,1}(n_1, n_2, n_3, n_4; D_r)$*,* $\alpha \in (0, 1)$ *and* $L \leq r$*, then the nonlinear singular integral equation* (19.39) *has at least one solution at the sphere* $B_\alpha(0; r)$*.*

Proof In accordance with Lemma 19.6, the sphere $B_\alpha(0; r)$ is a compact set of the space $H_\alpha(\gamma)$, $\alpha \in (0, 1)$. Moreover, if $L \le r$, then from Corollary 19.4 the operator A that is defined by formula (19.53) transforms the convex, closed, and compact sphere $B_\alpha(0; r)$ to itself. Therefore, the operator A is compact in the space $H_\alpha(\gamma)$, $\alpha \in (0, 1)$. From another perspective, in accordance with Lemma 19.9, since the operator A is continuous on the sphere $B_\alpha(0; r)$, it is completely continuous on the sphere $B_\alpha(0; r)$. Thus, from well-known Schauder's fixed point principle the operator A has a fixed point in the sphere $B_\alpha(0; r)$. Consequently, Eq. (19.39) has a solution in the space $H_\alpha(\gamma)$, $\alpha \in (0, 1)$.

Therefore, the theorem is proven.

Theorem 19.6 concern the existence of the solution of the nonlinear singular integral equation (19.39).

Now, let us investigate the uniqueness of the solution of Eq. (19.39) and the problem of how an approximation to this solution may be found. The next theorem concerns this.

Theorem 19.7 *If* $k \in H_{\alpha,1}(m_1, m_2; D_{0r})$, $\Phi \in H_{\alpha,1,1,1}(n_1, n_2, n_3, n_4; D_{0r})$, $\alpha \in (0, 1)$, $L \le r$ *and* $\Delta = n_2 + n_4.m_2.\left\|S_\gamma\right\|_p < 1$, *then the nonlinear singular integral equation (19.39) has only one solution* $\varphi_* \in B_\alpha(0; r)$. *This solution for* $\varphi_0 \in B_\alpha(0; r)$*at any starting approximation can be found as a limit of the sequence* (φ_n), $n = 1, 2, \dots$ *whose terms are defined as,*

$$\varphi_n(t) = \Phi(t, \varphi_{n-1}(t), f(t), S_\gamma k(., \varphi_{n-1}(.))(t), t \in \gamma. \tag{19.55}$$

Furthermore, for the approximation φ_n *the following evaluation is hold*

$$d_p(\varphi_*, \varphi_n) \le \frac{\Delta^n}{1 - \Delta}.d_p(\varphi_0, \varphi_1), n = 1, 2, \dots \tag{19.56}$$

Proof We will prove that Eq. (19.39) has a unique solution by using Theorem 19.5. Let us show that Eq. (19.39) satisfies the assumptions of Theorem 19.5.

If we take $B_\alpha(0; r) = X$, $d_\infty = \rho_1$ and $d_p = \rho_2$, then according to Lemma 19.6, Corollary 19.5, and Lemma 19.7 it is seen that the first and second assumptions of Theorem 19.5 are hold.

Now, let us show that when $\Delta < 1$, the operator A which is defined by formula (19.53) is a contraction transformation on the space $B_\alpha(0; r)$,$\alpha \in (0, 1)$ according to the metric d_p,$p > 1$.

$\forall \varphi_1, \varphi_2 \in B_\alpha(0; r)$,$\alpha \in (0, 1)$ in accordance with Lemma 19.8 we can prove the following inequality

$$\|A(\varphi_1) - A(\varphi_2)\|_p \le \left(n_2 + n_4.m_2.\left\|S_\gamma\right\|_p\right).\|\varphi_1 - \varphi_2\|_p .$$

From here, for$\forall \varphi_1, \varphi_2 \in B_\alpha(0; r)$,$\alpha \in (0, 1)$ we have

$$d_p(A(\varphi_1), A(\varphi_2)) \le \Delta.d_p(\varphi_1, \varphi_2). \tag{19.57}$$

Thus, when $\Delta < 1$, the operator A is a contraction transformation on the $B_\alpha(0; r)$, $\alpha \in (0, 1)$ according to the metric d_p, $p > 1$. So, according to Theorem 19.5, the operator equation

$$\varphi = A\varphi$$

and consequently Eq. (19.39) has only one solution $\varphi_* \in B_\alpha(0; r)$.

Now, let us show that the solution $\varphi_* \in B_\alpha(0; r)$ is the limit of the sequence $(\varphi_n), n = 1, 2, \ldots$ whose terms are defined by Eq. (19.55) as

$$\varphi_n(t) = A\varphi_{n-1}(t), t \in \gamma, n = 1, 2, \ldots$$

for any starting approximation $\varphi_0 \in B_\alpha(0; r)$.

$\forall n = 1, 2, \ldots$ from inequality (19.57) we can write

$$d_p(\varphi_{n+1}, \varphi_n) \leq \Delta . d_p(\varphi_n, \varphi_{n-1}).$$

From this inequality, we obtain

$$d_p(\varphi_{n+1}, \varphi_n) \leq \Delta^n . d_p(\varphi_0, \varphi_1).$$

Similarly, for $\forall m, n$ $(m, n \in \mathbb{N})$ we can write

$$d_p(\varphi_{m+n}, \varphi_n) \leq \Delta^n . d_p(\varphi_0, \varphi_m).$$

Furthermore, since

$$d_p(\varphi_0, \varphi_m) \leq (\Delta^{m-1} + \Delta^{m-2} + \cdots + \Delta + 1). d_p(\varphi_0, \varphi_1)$$

we obtain

$$d_p(\varphi_n, \varphi_{n+m}) \leq \frac{1 - \Delta^m}{1 - \Delta} \Delta^n . d_p(\varphi_0, \varphi_1). \tag{19.58}$$

From inequality (19.58), it can be seen that the sequence $(\varphi_n), n = 1, 2, \ldots$ is a Cauchy sequence according to the metric d_p, $p > 1$. Because the space $(B_\alpha(0; r); d_p)$ is a complete metric space according to Corollary 19.5 and Lemma 19.7, there exists an element $\tilde{\varphi} \in B_\alpha(0; r)$ such that $\lim\limits_{n\to\infty} \varphi_n = \tilde{\varphi}$ or $\lim\limits_{n\to\infty} d_p(\varphi_n, \tilde{\varphi}) = 0$.

Now, let us show that $\tilde{\varphi} = \varphi_*$.

$\forall n = 1, 2, \ldots$, since

$$d_p(\varphi_{n+1}, A(\tilde{\varphi})) \le d_p(A(\varphi_n), A(\tilde{\varphi})) \le \Delta . d_p(\varphi_n, \tilde{\varphi})$$

and

$$\lim_{n\to\infty} d_p(\varphi_n, \tilde{\varphi}) = 0$$

we obtain

$$\lim_{n\to\infty} d_p(\varphi_{n+1}, A(\tilde{\varphi})) = 0.$$

Thus, from the inequality

$$d_p(\tilde{\varphi}, A(\tilde{\varphi})) \le d_p(\tilde{\varphi}, \varphi_{n+1}) + d_p(\varphi_{n+1}, A(\tilde{\varphi}))$$

we have $\tilde{\varphi} = A\tilde{\varphi}$.

From here, it is seen that $\tilde{\varphi} = \varphi_*$.

Thus, the unique solution of the operator equation

$$\varphi(t) = A(\varphi)(t), t \in \gamma$$

is the limit of the sequence $(\varphi_n) \subset B_\alpha(0; r), \alpha \in (0, 1), n = 1, 2, \ldots$ whose terms are defined as

$$\varphi_n(t) = A\varphi_{n-1}(t), t \in \gamma, n = 1, 2, \ldots.$$

Furthermore, since

$$\lim_{m\to\infty} \varphi_{n+m}(t) = \varphi_*(t), \lim_{m\to\infty} \Delta^m = 0$$

according to inequality (19.58) it can be seen that inequality (19.56) is hold.

From that point, we yield the conclusion.

From Theorem 19.7 and Lemma 19.7 we will reach the following conclusion:

Corollary 19.6 *The sequence* $(\varphi_n) \subset B_\alpha(0; r), \alpha \in (0, 1), n = 1, 2, \ldots$ *whose terms* $\forall \varphi_0 \in B_\alpha(0; r), \alpha \in (0, 1)$ *are defined as*

$$\varphi_n(t) = A\varphi_{n-1}(t), t \in \gamma, n = 1, 2, \ldots.$$

converges to the unique solution of the nonlinear Cauchy singular integral equation (19.39) according to the metric d_∞.

19.5 Nonlinear Singular Integro-Differential Equations

In this section, it has been shown with the help of monotonic operators that a nonlinear singular integro-differential equation has a unique solution.

In the theory of elasticity, in an elastic plane with a linear crack the solution of the crack broadening is reduced to the solution of

$$f(\varphi(x)) - \frac{\lambda}{\pi}\int_{-a}^{a}\frac{\varphi'(t)}{t-x}dt = \mu, \qquad |x| \leq a, \quad a \geq 1 \tag{19.59}$$

type singular integro-differential equation with Cauchy kernel by taking the attractive forces along the borders of the crack [3]. Here, μ is the force causing the broadening in the crack, $2a$ is the width of the crack(in most of the problems a is taken as $a = 1$), λ is an arbitrary positive parameter, π is the known constant in mathematics(in this study π is only used for convenience in computation), $f(u)$ is a known function which is continuous and positive valued in the domain, and finally $\varphi(x)$is the characterization of the broadening in the crack and it is an unknown function.

In fact, in Eq. (19.59) μ is also an unknown parameter. However, after founding the solution that satisfies some of the boundary conditions (see (19.60) formula below) of Eq. (19.59), μ can be found with the help of the $\varphi(x)$ function (see (19.62) formula below). For this reason we do not take μ as an unknown parameter.

Equation (19.59) is used in several problems of mathematical physics. Mathematical models of several physical phenomena include integro-differential equation with Cauchy kernel. Furthermore, for several applications in important fields such as fractal mechanics [7], elastic contact problems [30], radiation and molecular contact problems of physics [25], (19.59) type equations are being used.

The analytic solution of singular integral equation was given by Muskhelishvili [4]. Later, several different methods have been used to analytically solve the singular integral and integro-differential equation with singular kernel [6, 22, 23, 36, 89]. There are several studies on approximate solution of integral and integro-differential equations [18, 23, 39, 57, 66]. The approximate solution of (19.59) type equation, known as Prandtl equation in the literature (see [66, p. 109]), can be set up by the methods given in references [2, 23].

In reference [3], the solution of (19.59) type equation was investigated asymptotically in $C^{(1)}[-1, 1]$ space and the existence of two solutions besides zero solution was tried to be proven.

In fact, for applications in several fields, problems in which the solution are reduced to the solution of (19.59) equation (for example, broadening of the crack in elastic plane) only the existence of the zero solution of the problem is required.

In this section, we prove that there exist a zero solution of the (19.59) type nonlinear integro-differential equation with Cauchy kernel which satisfies the following conditions for every $\lambda > 0$

$$\begin{aligned} \varphi'(-a) &= \varphi'(a) = 0, \\ \varphi(-a) &= \varphi(a) = 0. \end{aligned} \tag{19.60}$$

or this, the problem of finding the solution of Eq. (19.59) satisfying the (19.60) conditions is reduced to the solution of nonlinear Fredholm integral equation whose solution corresponds to this problem. It is shown that Fredholm integral equation has a single solution.

Throughout the section, unless the opposite is stated, the problem of finding the solution that satisfies the conditions (19.60) of Eq. (19.59) is called as (19.59)–(19.60) problem.

With the help of the formula of the inverse transformation of singular integral equation with Cauchy kernel at finite interval (see [4, 83]) the problem (19.59)–(19.60) can be reduced to the problem of finding the solution of following singular integro-differential equation

$$\varphi'(x) = -\frac{\sqrt{a^2 - x^2}}{\pi\lambda} \int_{-a}^{a} \frac{f(\varphi(t))}{(t - x)\sqrt{a^2 - t^2}} dt, \qquad |x| \le a \tag{19.61}$$

that satisfies the following condition:

$$\frac{1}{\pi} \int_{-a}^{a} \frac{f(\varphi(t))}{\sqrt{a^2 - t^2}} dt = \mu. \tag{19.62}$$

If both sides of Eq. (19.61) are integrated over the interval $[-a, x]$ and if the conditions (19.60) are taken into account, then we have the following nonlinear Fredholm integral equation

$$\varphi(x) = \frac{1}{\pi\lambda} \int_{-a}^{a} K(x, t) f(\varphi(t)) dt, \qquad |x| \le a. \tag{19.63}$$

Here,

$$K(x, t) = \ln \frac{a. |t - x|}{a^2 - xt + \sqrt{(a^2 - x^2).(a^2 - t^2)}} - \frac{\left(\arcsin \frac{x}{a} + \frac{\pi}{2}\right) t - \sqrt{a^2 - x^2}}{\sqrt{a^2 - t^2}}.$$

From Eq. (19.63), if we take the conditions (19.60) into account, then we have

$$\int_{-a}^{a} \frac{t.f(\varphi(t))}{\sqrt{a^2 - t^2}} dt = 0 \tag{19.64}$$

Thus, we have reduced the problem (19.59)–(19.60) to the problem of finding the solution of the Fredholm integral equation (19.63) that satisfies the conditions (19.62) and (19.64).

Furthermore, the solution of Eq. (19.63) that satisfies the condition (19.64) is also a solution of the following Fredholm integral equation

$$\varphi(x) = \frac{1}{\pi\lambda}\int_{-a}^{a} M(x,t) f(\varphi(t))dt, \qquad |x| \le a \tag{19.65}$$

Here,

$$M(x,t) = \ln \frac{a.\,|t-x|}{a^2 - xt + \sqrt{(a^2-x^2).(a^2-t^2)}} + \frac{\sqrt{a^2-x^2}}{\sqrt{a^2-t^2}} \tag{19.66}$$

Every solution of Eq. (19.65) is also a solution of Eq. (19.63) that satisfies the condition (19.64).

Thus, the problem (19.59)–(19.60) now is reduced to the problem of finding the solution of Eq. (19.65) that satisfies the condition (19.62). Therefore, from now on we will work on the solution of Eq. (19.65).

Remark 19.1 If it is shown that Eq. (19.65) has a unique solution and if the parameter μ is found from the formula (19.62), then it is also shown that the problem (19.59)–(19.60) has a unique solution.

Definition 19.6 [8, see p. 256] Let X be a Banach space and K be a nonempty convex subset of X. If $\forall x \in K\setminus\{0\}$, $\beta.x \in K$ when $\beta \ge 0$ and $\beta.x \notin K$ when $\beta < 0$, then K is said to be a conic in X.

In the space X, it is obvious that all conics are partially ordered sets and the partial order relation "$\prec$" in the conic K can be defined as follows

$$\prec \quad : \quad \text{for } \ x, y \in X \qquad x \prec y \Leftrightarrow y - x \in K$$

The relation $\prec$ satisfies the properties of the relation $\le$ in real number set.

Definition 19.7 ([8, see p. 259]) Let X be a Banach space, K is a conic with $K \subset X$, and B is a nonlinear operator that is defined on X. If for every subset $M \subset X$ we have $B(M) \subset K$, then the operator B is said to be nonnegative operator on M and if $\forall x, y \in X$, $Bx \prec By$ when $x \prec y$, then B is said to be monotone operator.

The class of functions—defined on the interval $[-a, a]$—that satisfies Hölder condition with the exponential $\alpha \in (0, 1]$ is a Banach space with the norm $\|.\|_\alpha = \|.\|_\infty + H(.;\alpha)$ and we denote this space with $H_\alpha[-a, a]$. Here,

$$\|\varphi\|_\infty = \max\{|\varphi(x)| : x \in [-a, a]\},$$

$$H(\varphi;\alpha) = \sup\left\{\frac{|\varphi(x_1) - \varphi(x_2)|}{|x_1 - x_2|^\alpha} : x_1, x_2 \in [-a, a], \quad x_1 \ne x_2\right\}.$$

Further in the study, unless the opposite is stated, we will denote space $H_\alpha[-a, a]$ with H_α. The set of functions which takes the value of zero at the endpoints of the interval $[-a, a]$ is a subspace of H_α and we denote this subspace with $\underset{\alpha}{\overset{\circ}{H}}$. The subset of $\underset{\alpha}{\overset{\circ}{H}}$ which composed of functions that do not take nonnegative values is denoted by $\overset{\circ}{H}{}_\alpha^+$.

Let us define a norm $\|.\|_{\alpha,0} = H(.; \alpha)$ on the subspace $\overset{\circ}{H}_\alpha$. It is obvious that the norm $\|.\|_{\alpha,0}$ and the norm $\|.\|_\alpha$ are equivalent at the space $\overset{\circ}{H}_\alpha$. Thus, $(\overset{\circ}{H}_\alpha, \|.\|_{\alpha,0})$ is a Banach space.

In the space $\overset{\circ}{H}{}_\alpha^+$ let us denote the subset of even functions that are monotone increasing on the interval $[-a, 0]$ and monotone decreasing on the interval $[0, a]$ with K_α^+.

For every $\alpha \in \left(0, \frac{1}{2}\right)$, K_α^+ is a conic in the subspace $\underset{\alpha}{\overset{\circ}{H}}$. The set of continuous functions on the interval $[-a, a]$ is a normed space with the norm$\|.\|_\infty$. Let us denote this space with $(C[-a, a], \|.\|_\infty) = Y$.

The operator $A_\lambda : Y \to Y$ whose domain is the set $D(A_\lambda) = H_\alpha$ is defined as follows:

$$A_\lambda : \varphi \to \frac{1}{\pi\lambda}\int_{-a}^{a} M(x,t) f(\varphi(t))dt = A_\lambda\varphi(x), \qquad |x| \le a \tag{19.67}$$

Here, $M(x, t)$ is the function which is given in the Formula (19.66).

Lemma 19.10 *If the function $f : \S_+ \to \S_+$ is monotone decreasing, continuously differentiable and $\|f'\|_\infty < +\infty$, then for $\forall\alpha \in \left(0, \frac{1}{2}\right)$ and $\forall\lambda > 0$ we have $A_\lambda(K_\alpha^+) \subset K_\alpha^+$.*

Proof Let the assumptions of the lemma be satisfied. $\forall x \in [-a, a]$ let us take

$$F(\varphi)(x) = \frac{1}{\pi\lambda}\int_{-a}^{a} M(x,t) f(\varphi(t))dt, \qquad |x| \le a, \quad a \ge 1$$

as given above.

Firstly, let us see that for $\forall\varphi \in K_\alpha^+$ we have $F(\varphi) \in H_\alpha$. $\forall x \in [-a, a]$

$$F'(\varphi)(x) = -\frac{\sqrt{a^2 - x^2}}{\pi\lambda}\int_{-a}^{a} \frac{f(\varphi(t))}{(t-x)\sqrt{a^2 - t^2}}dt.$$

We can write the derivative function $F'(\varphi)$ as follows

$$F'(\varphi)(x) = -\frac{\sqrt{a^2 - x^2}}{\pi\lambda}\int_{-a}^{a} \frac{G(t)}{(t-x)\sqrt{a^2 - t^2}}dt.$$

Here, $G(t) = f(\varphi(t)) - f(0)$. It is obvious that $G \in \overset{\circ}{\underset{\alpha}{H}}$. Therefore, from the last statement of $F'(\varphi)$ it is also obvious that $F'(\varphi) \in \overset{\circ}{\underset{\alpha}{H}}$.

Let $\|S\|_{\alpha,0}$ be a norm of the following bounded linear operator

$$S\theta(x) = -\frac{\sqrt{a^2 - x^2}}{\pi\lambda} \int_{-a}^{a} \frac{\theta(t)}{(t-x)\sqrt{a^2-t^2}} dt : \overset{\circ}{\underset{\alpha}{H}} \to \overset{\circ}{\underset{\alpha}{H}}$$

then we obtain

$$\left\|F'(\varphi)\right\|_\infty \le \frac{1}{|\lambda|} \left\|f'\right\|_\infty \cdot \|\varphi\|_{\alpha,0} \cdot \|S\|_{\alpha,0} \qquad (19.68)$$

Thus, $F(\varphi) \in H_\alpha$. Furthermore, since $M(-a,t) = M(a,t) = 0$ we have $F(\varphi)(-a) = F(\varphi)(a)$. Therefore, $F \in \overset{\circ}{\underset{\alpha}{H}}$.

Now, let us show that for every $\varphi \in K_\alpha^+$ the function $F(\varphi)$ is an even function which does not take negative values, monotone increasing on the interval $[-a, 0]$ and monotone decreasing on the interval $[0, a]$.

For every $\varphi \in K_\alpha^+$ and $\forall x \in [-a, a]$ we have
$F(\varphi)(-x) = \frac{1}{\pi\lambda}\int_{-a}^{a} M(-x,t) f(\varphi(t))dt = \frac{1}{\pi\lambda}\int_{-a}^{a} M(-x,-t) f(\varphi(t))dt = \frac{1}{\pi\lambda}\int_{-a}^{a} M(x,t) f(\varphi(t))dt = F(\varphi)(x)$.
$\forall x \in [0, a]$ we can write

$$F'(\varphi)(x) = -\frac{2x\sqrt{a^2-x^2}}{\pi\lambda} \int_{-a}^{a} \frac{f(\varphi(t)) - f(\varphi(x))}{(t-x)(t+x)\sqrt{a^2-t^2}} dt. \qquad (19.69)$$

Thus, from the assumptions of the lemma and equality (19.69) for every $\varphi \in K_\alpha^+$ and $\forall x \in [0, a]$ we have $F'(\varphi)(x) < 0$; hence, it is seen that the function $F'(\varphi)(x)$ is monotone decreasing on the interval$[0, a]$. On the other hand, since the function $F(\varphi)(x)$ is an even function on the interval $[-a, a]$, $F(\varphi)(x)$ is monotone increasing on the interval $[-a, 0]$.

Furthermore, since $F(\varphi)(-a) = F(\varphi)(a) = 0$ for $\forall x \in [-a, a]$ it is seen that $F(\varphi)(x) \ge 0$.

Thus, it has also proven that for every $\varphi \in K_\alpha^+$ and $\forall\lambda > 0$ we have $F(\varphi) \in K_\alpha^+$. It is proven that $A_\lambda(K_\alpha^+) \subset K_\alpha^+$. With this Lemma 19.10 has proven.

Let us take $K_\alpha^- = \left\{\varphi \in \overset{\circ}{\underset{\alpha}{H}} : -\varphi \in K_\alpha^+\right\}$ and in the space $\overset{\circ}{\underset{\alpha}{H}}$ define the operator M_λ as below

$$M_\lambda\varphi(x) = \frac{1}{\pi\lambda}\int_{-a}^{a} M(x,t) f(\varphi(t))dt, \qquad |x| \le a \qquad (19.70)$$

Since it can be proven similarly to Lemma 19.10, the next lemma is given without proof.

Lemma 19.11 *$\forall\alpha \in \left(0, \frac{1}{2}\right)$ and $\forall\lambda > 0$ the operator M_λ satisfies the following relations*

$$M_\lambda(K_\alpha^-) \subset K_\alpha^+,$$
$$M_\lambda(K_\alpha^+) \subset K_\alpha^-.$$

Lemma 19.12 *Let the function $f : \S_+ \to \S_+$ be monotone decreasing, continuously differentiable, and $\|f'\|_\infty < +\infty$. If also the derivative function $f'(u)$ is monotone decreasing on $\S_+$, for $\forall\alpha \in \left(0, \frac{1}{2}\right)$ and $\forall\lambda > 0$, then the operator A_λ that is defined by the formula (19.67) is a monotone operator on the conic K_α^+.*

Proof Let the assumptions of the lemma be satisfied. For any $\varphi_1, \varphi_2 \in K_\alpha^+$ and $\varphi_1 \prec \varphi_2$. With the help of this function let us define the following function

$$\Phi(t) = f(\varphi_2(t)) - f(\varphi_1(t)), \qquad t \in [-a, a] \tag{19.71}$$

For every $t \in [-a, a]$ it is obvious that$\Phi(-t) = \Phi(t)$. Furthermore, from the assumptions of the lemma it can be seen that $\Phi(t) \le 0$.

For every $t \in [-a, a]$ and $\theta \in (0, 1)$ we can write

$$\Phi(t) = f'\left(\varphi_1(t) + \theta.[\varphi_2(t) - \varphi_1(t)]\right).[\varphi_2(t) - \varphi_1(t)] \tag{19.72}$$

For every $\theta \in (0, 1)$, $H(t) = \varphi_1(t) + \theta.[\varphi_2(t) - \varphi_1(t)]$ is a monotone decreasing function on the interval $[0, a]$. Thus, $f'\{\varphi_1(t) + \theta.[\varphi_2(t) - \varphi_1(t)]\}$ is a function of variable t that does not take positive values and it is monotone decreasing on the interval $[0, a]$. Moreover, since K_α^+ is a conic, for every $\varphi_1, \varphi_2 \in K_\alpha^+$ and $\varphi_1 \prec \varphi_2$ we have$\varphi_2 - \varphi_1 \in K_\alpha^+$. Therefore, $\varphi_2 - \varphi_1$ is a function that does not take negative values and it is monotone decreasing on the interval $[0, a]$.

Thus, $\Phi(t)$ is a monotone increasing function on the interval $[0, a]$. Besides, since $\Phi(t)$ is an even function on the interval $[-a, a]$, it is monotone increasing on the interval $[-a, 0]$. On the other hand, $\Phi(-a) = \Phi(a) = 0$. As a result, we have the conclusion that $-\Phi \in K_\alpha^+$ and consequently $\Phi \in K_\alpha^-$. Hence, according to Lemma 19.11, we have $M_\lambda(\Phi) \in K_\alpha^+$. On the other hand, since $M_\lambda(\Phi) = A_\lambda(\varphi_2) - A_\lambda(\varphi_1)$, we have $A_\lambda(\varphi_2) - A_\lambda(\varphi_1) \in K_\alpha^+$ and as a result, we have $A_\lambda(\varphi_1) \prec A_\lambda(\varphi_2)$.

As a consequence, for every $\varphi_1, \varphi_2 \in K_\alpha^+$ and $\varphi_1 \prec \varphi_2$ we obtain that $A_\lambda(\varphi_1) \prec A_\lambda(\varphi_2)$ and this is the evidence of the fact that the operator A_λ is monotone on the conic K_α^+. With this the lemma is proven.

The following lemma can be proven using similar arguments in the proof of Lemma 19.12.

Lemma 19.13 *Let the function $f : \S_+ \to \S_+$ be monotone decreasing, continuously differentiable, and $\|f'\|_\infty < +\infty$. If also the derivative function $f'(u)$ is monotone decreasing for $\forall\alpha \in \left(0, \frac{1}{2}\right)$, $\forall\varphi \in K_\alpha^+$ and $\forall\lambda > 0$, then the following two relations are true*

$$A_\lambda(p.\varphi) \prec p.A_\lambda(\varphi), \quad \text{if} \quad p \in (0, 1),$$
$$p.A_\lambda(\varphi) \prec A_\lambda(p.\varphi), \quad \text{if} \quad p \in [1, +\infty)\,.$$

19.5.1 On the Solution of Eq. (19.65)

In this section we will give the proof of a theorem which is about the solution of Eq. (19.65).

Theorem 19.8 *If the function $f : \S_+ \to \S_+$ is monotone decreasing, continuously differentiable, $\|f'\|_\infty < +\infty$ and the derivative function $f'(u)$ is monotone decreasing, then for $\forall\alpha \in \left(0, \frac{1}{2}\right)$ and $\forall\lambda > 0$ Eq. (19.65) has a unique zero solution in the conic K_α^+.*

Proof Let the assumptions of the theorem be satisfied. From the following formula it can be concluded that $\forall\lambda > 0$, the zero function satisfies Eq. (19.65)

$$\int M(x,t)dt = (t-x)\ln \frac{a.\,|t-x|}{a^2 - xt + \sqrt{(a^2-x^2)(a^2-t^2)}}.$$

Now, let us show that $\forall\lambda > 0$ the function $\varphi(x) = 0$ is the unique solution of Eq. (19.65). We will show this by using contradiction. Let us assume that for one $\lambda = \lambda_0 > 0$ there exists a nonzero $\varphi_0 \in K_\alpha^+$ solution of Eq. (19.65).

If we write Eq. (19.65)

$$\varphi = A_\lambda(\varphi)$$

like as an operator equation and take $\nu = \frac{\lambda_0}{2}$, then the following equality is satisfied

$$2.\varphi_0 = A_\nu(\varphi_0). \tag{19.73}$$

If we multiply both sides of the above equality with 2 and take into account the second relation of the Lemma 19.13, then we have

$$2^2.\varphi_0 \prec A_\nu^2(\varphi_0). \tag{19.74}$$

This relation can be written in inductive form as given below

$$2^n.\varphi_0 \prec A_\nu^n(\varphi_0), \quad n = 1, 2, \ldots \tag{19.75}$$

From relation (19.75) we obtain

$$\lim_{n\to\infty} A_\nu^n(\varphi_0)(0) = +\infty. \tag{19.76}$$

On the other hand, from the following equality

$$A_\nu(\varphi_0)(0) = \frac{4}{\pi\lambda_0}\int_0^a \left[\ln\frac{a.t}{a+\sqrt{a^2-t^2}} + \frac{a}{\sqrt{a^2-t^2}}\right] f(A_\nu(\varphi_0)(t))dt,$$

we have

$$\left|A_\nu(\varphi_0)(0)\right| \le \frac{4a}{\lambda_0}.f(0). \tag{19.77}$$

Since

$$\begin{aligned} A_\nu^2(\varphi_0)(0) &= A_\nu\left(A_\nu(\varphi_0)(0)\right) \\ &= \frac{4}{\pi\lambda_0}\int_0^a \left[\ln\frac{a.t}{a+\sqrt{a^2-t^2}} + \frac{a}{\sqrt{a^2-t^2}}\right] f(A_\nu(\varphi_0)(t))dt, \end{aligned}$$

and for every $t \in [-a, a]$ we have

$$f(A_\nu(\varphi_0)(t) \le f(0),$$

then in this case we obtain

$$A_\nu^2(\varphi_0)(0) \le \frac{4a}{\lambda_0}.f(0).$$

In inductive form it can be easily seen that

$$A_\nu^n(\varphi_0)(0) \le \frac{4a}{\lambda_0}.f(0), \qquad n = 1, 2, \ldots$$

Therefore, $\left(A_\nu^n(\varphi_0)(0)\right)$, $n = 1, 2, \ldots$ is a bounded sequence. Thus, what obtained in the formula (19.76) is a contradiction. This contradiction proves that Eq. (19.65) has a unique zero solution. With this the theorem is proven.

From this theorem we have the following conclusion.

Corollary 19.7 *There is a unique zero solution of the nonlinear singular integro-differential equation (19.59) which satisfies the conditions (19.60). For the parameter μ, it can be found that $\mu = f(0)$ from Eq.* (19.62)*.*

19.6 The Collocation Method for the Solution of Boundary Integral Equations

This section is devoted to investigating the approximate solutions to a class of boundary integral equations over a closed, bounded, and smooth surface found via the collocation method. The section provides sufficient conditions for the convergence of the method in the space of continuous functions.

Consider the following singular integral equation:

$$\varphi(x) + \int_S \frac{\phi(x,y)}{|x-y|^2}\varphi(y)d\sigma_y = f(x), x \in S, \tag{19.78}$$

where $\phi(x, y)$ and $f(x)$ are known continuous functions in their domains of definition, $\varphi(x)$ is an unknown function, S is a closed, bounded, and smooth surface in $\mathbb{R}^3$ such that for every $x \in S$, $\phi(x, x) = 0$, and $|x - y|$ denotes the distance between the points x and y.

Many dispersion and radiation problems are related to finding the solution of Helmholtz's equation in an exterior region defined by an equation of the following form:

$$\Delta u + k^2.u = 0, Imk \geq 0.$$

Although these problems are typically solved approximately using the finite element method or the finite difference method, there are well-known difficulties in applying these methods in the general case. These difficulties have been overcome by using the integral equation method given by Jones instead (see [12, 27, 33, 34, 92]).

Analytical solutions of equations of the form (19.78) can only be found in certain special cases. In the absence of analytical solutions, these kinds of equations are usually solved by approximate methods. In addition, approximate solutions are often sufficient for a wide range of applications of problems of the form (19.78). From this point of view, it is important to know how to solve the boundary integral equations with approximate methods.

This section is an analysis of the collocation method for solving the boundary integral equation (19.78) over a closed, bounded, and smooth surface in $\mathbb{R}^3$. It gives sufficient conditions for the convergence of this method in the space of continuous functions.

We consider with the integral equation stated on a two-dimensional closed, bounded, and smooth surface in $\mathbb{R}^3$, where the kernel of the corresponding integral operator is a function of the form $\phi(x, y)/ |x - y|^2$ with continuous $\phi(x, y)$.

In this study, it is shown that if the modulus of continuity of $\phi(x, y)$satisfies some minor restrictions, then, in spite of the presence of a high level of singularity along the diagonal, the integral operator in question is well defined and compact in the space of continuous functions. In this case, the Fredholm theorem applies, so the integral equation has a solution for any continuous right-hand side if and only if the corresponding homogeneous equation does not have a nontrivial solution.

Then, in the second subsection, we prove the existence and uniqueness of the solution of the class of boundary integral equations under consideration.

In Sect. 19.6, we show the convergence of the collocation method applied to the boundary integral equation (19.78).

Firstly, we will introduce some necessary information for proving the main results.

Let $\ell = diam(S) = \sup\{|x-y| : x, y \in S\}$**,** where S is a closed, bounded, and smooth surface in $\mathbb{R}^3$. We denote the complex numbers by $\mathbb{C}$, the natural numbers by $\mathbb{N}$, the real numbers by $\mathbb{R}$, the nonnegative real numbers by $\mathbb{R}_+$, and the radius of the standard sphere associated to the surface S by d (see [40]). The symbol $C(X)$ denotes the space of continuous functions on X. We denote by $\|u\| = \max\{|u_i| : i = 1, 2, \ldots, n\}$ the norm of the vectors $(u_1, \ldots, u_n)$, $u_i \in \mathbb{C}$, $i = 1, 2, \ldots, n$; $n \in \mathbb{N}$, and the normed space is denoted $\mathbb{C}^n$.

Throughout this section, the numbers c_i, $i = 1, 2, \ldots$ will denote positive real numbers.

Definition 19.8 The function $\omega_\varphi : (0, \delta] \to \mathbb{R}_+$, which is defined by the following formula

$$\omega_\varphi(\delta) = \delta . \sup\left\{\bar{\omega}_\varphi(\tau).\tau^{-1} : \tau \geq \delta\right\}, \delta \in (0, \ell],$$

is the modulus of continuity of the function $\varphi \in C(S)$.

Here, $\bar{\omega}_\varphi(\delta) = \max\{|\varphi(x) - \varphi(y)| : |x-y| \leq \delta, x, y \in S\}, \delta \in (0, \ell]$.

Definition 19.9 Let us define the following functions for $\phi \in C(S \times S)$ and $\delta \in (0, \ell]$:

$$\bar{\omega}^*_\phi(\delta) = \sup\{|\phi(x, y)| : |x-y| \leq \delta, x, y \in S\},$$

$$\omega^*_\phi(\delta) = \delta . \sup\left\{\bar{\omega}^*_\phi(\tau).\tau^{-1} : \tau \geq \delta\right\},$$

$$\omega^{1,0}_\phi(\delta) = \sup\left\{\max_{y \in S} |\phi(x_1, y) - \phi(x_2, y)| : |x_1 - x_2| \leq \delta, x_1, x_2 \in S\right\},$$

$$\omega^{0,1}_\phi(\delta) = \sup\left\{\max_{x \in S} |\phi(x, y_1) - \phi(x, y_2)| : |y_1 - y_2| \leq \delta, y_1, y_2 \in S\right\}.$$

Remark 19.2 If a function $g : \mathbb{R} \to \mathbb{R}$ is increasing or decreasing in its domain, then we denote it by $g \uparrow$ or $g \downarrow$, respectively.

Definition 19.10 We define the following sets of functions $g : (0, \ell] \to \mathbb{R}_+$:

$$E_1(0, \ell] = \left\{g : g \geq 0, g \uparrow, g(\delta).\delta^{-1} \downarrow, g(\delta_1 + \delta_2) \leq g(\delta_1) + g(\delta_2), \lim_{\delta \to 0} g(\delta)) = 0\right\},$$

$$E_2(0, \ell] = \left\{g : g \geq 0, g \uparrow, \lim_{\delta \to 0} g(\delta) = 0\right\}.$$

Proposition 19.1 *For $\varphi \in C(S)$ and $\phi \in C(S \times S)$, we have $\omega_\varphi, \omega_\phi^* \in E_1(0, \ell]$ and $\omega_\phi^{1,0}, \omega_\phi^{0,1} \in E_2(0, \ell]$.*

Definition 19.11 ([40]) If the surface components $S_i \subset S, i = 1, 2, \ldots, n; n \in \mathbb{N}$ satisfy the following conditions, then S_i's will be called the "regular" elementary regions of the surface S :

(1) For every $i = 1, 2, \ldots, n$, S_i is closed and the interior $\overset{\circ}{S_i}$ is not empty. Furthermore,
$mes(S_i) = mes(\overset{\circ}{S_i})$ and for every $i, j = 1, 2, \ldots n, i \neq j$ $S_i \bigcap S_j = \emptyset$. Here, $mes(X)$ is the surface area of X;
(2) $S = \bigcup_{i=1}^{n} S_i$ and the S_i's are connected and have continuous boundaries;
(3) For every $i = 1, 2, \ldots, n$, $\exists x_i \in S_i$ such that
  (3.1) for all $r_i = \inf\{|x_i - x| : x \in \partial S_i\}$, $R_i = \sup\{|x_i - x| : x \in \partial S_i\}$ we have $r_i \approx R_i$, where ∂X is the boundary of X;
  (3.2) for every $i = 1, 2, \ldots, n$, we have $R_i \leq d/2$;
  (3.3) for every $i, j = 1, 2, \ldots, n$, we have $r_i \approx r_j$;
(4) $h = \max\{h_i : i = 1, 2, \ldots, n\}$, $h_i = \sup\{|x - y| : x, y \in S_i\} = diam(S_i)$ and $h \to 0$.

Definition 19.12 The points $x_i \in S_i, i = 1, 2, \ldots, n$ that satisfy conditions 3.1–3.3 of Definition 19.11 are called the "support points" of S_i (see [40]).

Remark 19.3 Actually, the number n (and consequently, the numbers r_i and R_i) in Definition 19.11 depends on h. Therefore, we will sometimes denote it by $n(h)$instead of n (for $r_i$and R_i we will use $r_i(h)$ and$R_i(h)$, respectively). It is also clear that as $h \to 0$, then $n(h) \to +\infty$. For every bounded surface, it can be seen that $r(h) \approx R(h)$. Here, $r(h) = \min\{r_i(h) : i = 1, 2, \ldots n\}$, and $R(h) = \max\{R_i(h) : i = 1, 2, \ldots, n\}$.

Definition 19.13 ([93]) If

(1) for every $u \in C(S)$ when $h \to 0$, we have $\left\|q^{n(h)}u\right\| \to \|u\|$ and
(2) for $u, u^1 \in C(S)$ and $a, a^1 \in \mathbb{C}$ when $h \to 0$, we have

$$\left\|q^{n(h)}(a.u + a^1.u^1) - (a.q^{n(h)}u + a^1.q^{n(h)}u^1)\right\| \to 0,$$

then the system $G = \left\{q^{n(h)}\right\}$, $n(h) = 1, 2, \ldots$ of operators $q^{n(h)} : C(S) \to \mathbb{C}^{n(h)}$ is called a **"**connective system" for the space $C(S)$ and $\mathbb{C}^{n(h)}$.

Definition 19.14 ([93]) If $\left\|u^{n(h)} - q^{n(h)}u\right\| \to 0$ as $h \to 0$, then we say that the sequence $\left\{u^{n(h)}\right\}$ of elements $\mathbb{C}^{n(h)}$, $n(h) = 0, 1, 2, \ldots$ is G-converging to the $u \in C(S)$, and denote this by the shorthand notation $u^{n(h)} \xrightarrow{G} u$.

Definition 19.15 ([93]) If every subsequence of the sequence $\left\{u^{n(h)}\right\}, u^{n(h)} \in \mathbb{C}^{n(h)}$, $n(h) = 1, 2, \ldots$ has a G-converging subsequence, then the sequence $\left\{u^{n(h)}\right\}$ is called G-compact.

Proposition 19.2 ([93]) *If $q^{n(h)} : C(S) \to \mathbb{C}^{n(h)}$, $n(h) = 1, 2, \ldots$ is linear and bounded, then the following two statements are equivalent:*

(1) The sequence $\left\{u^{n(h)}\right\}$ is G-compact.
(2) There exists a relatively compact sequence $\left\{u_{n(h)}\right\} \subset C(S)$ such that a $h \to 0$, we have $\left\|u^{n(h)} - q^{n(h)}u_{n(h)}\right\| \to 0$.

Definition 19.16 ([93]) If, as $h \to 0$, we have $u^{n(h)} \xrightarrow{G} u$ and at the same time $F^{n(h)}u^{n(h)} \xrightarrow{G} Fu$, then we will say that the operator sequence $F^{n(h)} \in L(\mathbb{C}^{n(h)}, \mathbb{C}^{n(h)})$ GG-converges to $F \in L(C(S), C(S))$ (where $L(X, Y)$ denotes the space of bounded linear operators from the Banach space X to the Banach space Y), and we denote this by the shorthand notation: $F^{n(h)} \xrightarrow{GG} F$. If the following two conditions are satisfied:

(1) $F^{n(h)} \xrightarrow{GG} F$,
(2) For every $u^{n(h)} \in \mathbb{C}^{n(h)}$, $\left\|u^{n(h)}\right\| \leq c_1$ the sequence $\left\{F^{n(h)}u^{n(h)}\right\}, n(h) = 1, 2, \ldots$ is G -compact, then the sequence $\left\{F^{n(h)}\right\}$ that converges to F is said to be GG-compact. For this, we use the shorthand notation $F^{n(h)} \xrightarrow{GG} F$-compact.

Proposition 19.3 ([93]) *If $F^{n(h)} \xrightarrow{GG} F$ is compact, then for every $n(h) \geq n_0$, there exists a number $n_0 \in \mathbb{N}$ such that $\dim Ker(I^{n(h)} + F^{n(h)}) \leq \dim Ker(I + F)$, where $I^{n(h)} : \mathbb{C}^{n(h)} \to \mathbb{C}^{n(h)}$ is the identity operator.*

Theorem 19.9 ([93]) *Let the following conditions hold:*

(1) $F^{n(h)} \xrightarrow{GG} F$ is compact;
(2) $Ker(I + F) = \{0\}$;
(3) For every $n(h) \geq n_0$, there exists a number $n_0 \in \mathbb{N}$ such that $I^{n(h)} + F^{n(h)}$ is a Fredholm operator with zero indexes;
(4) The sequence $\left\{v^{n(h)}\right\} \subset \mathbb{C}^{n(h)}$, $n(h) = 1, 2, \ldots$ G-converges to $v \in C(S)$.

In this case, the operator equations $(I + F)u = v$ and $(I^{n(h)} + F^{n(h)})u^{n(h)} = v^{n(h)}$ for $n(h) \geq n_0$ have only the unique solutions $u_* \in C(S)$ and $u_*^{n(h)} \in \mathbb{C}^{n(h)}$, respectively, and $u_*^{n(h)} \to u_*$ as $h \to 0$. We have the following estimate for the velocity of convergence:

$$c_2.\left\|\left(I^{n(h)} + F^{n(h)}\right)q^{n(h)}u_* - v^{n(h)}\right\| \leq \left\|u_*^{n(h)} - q^{n(h)}u_*\right\|$$
$$\leq c_3.\left\|\left(I^{n(h)} + F^{n(h)}\right)q^{n(h)}u_* - v^{n(h)}\right\|,$$

where $c_2 = 1/\sup\left\{\left\|I^{n(h)} + F^{n(h)}\right\| : n(h) \geq n_0\right\}$ *and* $c_3 = \sup\left\{\left\| \left(I^{n(h)}+F^{n(h)}\right)^{-1} \right\| : n(h) \geq n_0\right\}$.

19.6.1 On the Existence and Uniqueness of the Solution

Before proving the existence and uniqueness of the solution of the singular integral equation (19.78), we give the following lemma:

Lemma 19.14 *If (1)* $\int_0^\ell \omega_\phi^*(\tau).\tau^{-1}d\tau < +\infty$ *and (2) as* $\delta \to 0$ *we have* $\omega_\phi^{1,0}(\delta) = o(\ln^{-1}\delta)$*, then, for* $\phi \in C(S \times S)$*,*

$$(K\varphi)(x) = \int_S \frac{\phi(x, y)}{|x - y|^2}\varphi(y)d\sigma_y, x \in S \tag{19.79}$$

is a compact operator in $C(S)$.

Proof For $\varphi \in C(S), \phi \in C(S \times S)$ and for every$x \in S$, the following inequality can be easily proven:

$$\left|\int_S \frac{\phi(x, y)}{|x - y|^2}\varphi(y)d\sigma_y\right| \leq c_4. \|\varphi\| . \left\{1 + \int_0^d \frac{\omega_\phi^*(\tau)}{\tau}d\tau\right\}. \tag{19.80}$$

From this inequality, according to the first assumption of the lemma, we can see that the integral on the right-hand side of (19.79) is converging.

Now, let us show that the operator K that is defined by (19.79) is continuous in $C(S)$.

Let $x_1, x_2 \in S, |x_1 - x_2| = \delta$ and $\delta \in (0, d/2]$. In that case, we can write

$$\begin{aligned}(K\varphi)(x_1) - (K\varphi)(x_2) &= \int_{S_1} \frac{\phi(x_1, y)}{|x_1 - y|^2}\varphi(y)d\sigma_y - \int_{S_1} \frac{\phi(x_2, y)}{|x_2 - y|^2}\varphi(y)d\sigma_y \\ &\quad + \int_{S_2}\left[\frac{\phi\left(x_{1,y}\right)}{|x_1 - y|^2} - \frac{\phi\left(x_2, y\right)}{|x_2 - y|^2}\varphi\left(y\right)d\sigma_y\right]\end{aligned}$$

where $S_1 = S_{\delta/2}(x_1)\bigcup S_{\delta/2}(x_2)$, $S_2 = S\backslash S_1$, $S_\delta(x) = \{y \in S : |x - y| \leq \delta\}$. Let us denote the integrals on the right-hand side of the above equality by I_1, I_2, I_3, respectively. It is obvious that

$$I_1 = \int_{S_1^1} \frac{\phi(x_1, y)}{|x_1 - y|^2}\varphi(y)d\sigma_y + \int_{S_1^2} \frac{\phi(x_1, y)}{|x_1 - y|^2}\varphi(y)d\sigma_y,$$

where$S_1^i = S_{\delta/2}(x_i)$, $i = 1, 2$. In a manner similar to the proof of Inequality (19.80), we can prove the following inequality:

$$\left|I_1^1\right| = \left|\int_{S_1^1} \frac{\phi(x_1, y)}{|x_1 - y|^2}\varphi(y)d\sigma_y\right| \le c_5. \|\varphi\| . \int_0^\delta \frac{\omega_\phi^*(\tau)}{\tau}d\tau. \tag{19.81}$$

Since $\delta/2 \le |x_1 - y| \le 3\delta/2$ for$y \in S_1^2$, we have

$$\left|I_1^2\right| = \left|\int_{S_1^2} \frac{\phi(x_1, y)}{|x_1 - y|^2}\varphi(y)d\sigma_y\right| \le \|\varphi\| . \int_{S_1^2} \frac{\omega_\phi^*(|x_1 - y|)}{|x_1 - y|}d\sigma_y \le c_6. \|\varphi\| .\omega_\phi^*(\delta). \tag{19.82}$$

From inequalities (19.81) and (19.82), it follows that

$$|I_1| \le c_7. \|\varphi\| . \left[\omega_\phi^*(\delta) + \int_0^\delta \frac{\omega_\phi^*(\tau)}{\tau}d\tau\right]. \tag{19.83}$$

The following inequality can be proved in a manner similar to Inequality (19.83):

$$|I_2| \le c_8. \|\varphi\| . \left[\omega_\phi^*(\delta) + \int_0^\delta \frac{\omega_\phi^*(\tau)}{\tau}d\tau\right]. \tag{19.84}$$

Now, let us evaluate the integral I_3. We can write

$$I_3 \,\mathrm{ker} = \int_{S_2} \frac{\phi(x_1, y) - \phi(x_2, y)}{|x_1 - y|^2}\varphi(y)d\sigma_y + \int_{S_2} \phi(x_2, y)\left[\frac{1}{|x_1 - y|^2} - \frac{1}{|x_2 - y|^2}\right]\varphi(y)d\sigma_y.$$

Let us denote the individual integrals that are on the right-hand side of the above equality with I_3^1, I_3^2, respectively. For I_3^1, we obtain

$$\left|I_3^1\right| \le \|\varphi\| .\omega_\phi^{1,0}(\delta). \int_{S_2} \frac{d\sigma_y}{|x_1 - y|^2} \le c_9. \|\varphi\| .\omega_\phi^{1,0}(\delta). |\ln \delta| . \tag{19.85}$$

Since $|x_1 - y|/3 \le |x_1 - y| \le 3. |x_2 - y|$ for $y \in S_2$, we have

$$\left|I_3^2\right| \le \|\varphi\| . \int_{S_2} \frac{\omega_\phi^*(|x_2 - y|)\,|x_1 - x_2|\,[|x_1 - y| + |x_2 - y|]}{|x_1 - y|^2\,|x_2 - y|^2}d\sigma_y \tag{19.86}$$

$$\le c_{10}. \|\varphi\| .\delta. \int_\delta^\ell \frac{\omega_\phi^*(\tau)}{\tau}d\tau.$$

From (19.85) and (19.86), we have

$$|I_3| \leq c_{11}. \|\varphi\| . \left[\omega_\phi^{1,0} (\delta) . |\ln \delta| + \delta. \int_\delta^\ell \frac{\omega_\phi^* (\tau)}{\tau} d\tau \right]. \tag{19.87}$$

From inequalities (19.83), (19.84), (19.87), we obtain

$$|(K\varphi)(x_1) - (K\varphi)(x_2)| \leq c_{12}. \|\varphi\| . \begin{bmatrix} \omega_\phi^* (\delta) + \omega_\phi^{1,0} (\delta) . |\ln \delta| \\ + \int_0^\delta \frac{\omega_\phi^*(\tau)}{\tau} d\tau + \delta. \int_\delta^\ell \frac{\omega_\phi^*(\tau)}{\tau} d\tau \end{bmatrix}$$

From this inequality, according to the assumptions of the lemma and Proposition 19.1, we can see that $K : C(S) \to C(S)$.

Now, we will prove that the operator K is compact. For any $\varphi \in C(S)$, let us define the following operators:

$$(G_n\varphi) = \int_S g_n (x, y) \varphi (y) d\sigma_y, x \in S,$$

where

$$g_n (x, y) = \begin{cases} 0 & , \quad |x-y| \leq \frac{1}{2n}, \\ \frac{[2n.|x-y|-1].\phi(x,y)}{|x-y|^2} & , \quad \frac{1}{2n} < |x-y| \leq \frac{1}{n}, \\ \frac{\phi(x,y)}{|x-y|^2} & , \quad \frac{1}{n} < |x-y|, \end{cases}$$

$$n = 1, 2, \ldots.$$

It is obvious that the operators $G_n : C(S) \to C(S), n = 1, 2, \ldots$ are compact. Furthermore, from the inequality

$$|(K\varphi)(x) - (G_n\varphi)(x)| \leq c_{13}. \|\varphi\| . \int_0^{1/n} \frac{\omega_\phi^* (\tau)}{\tau} d\tau$$

it can be seen that the compact operator sequence $\{G_n\}, n = 1, 2, \ldots$ is converging to the operator K. Therefore, the operator K is also compact (see [35, Theorem 1, p241]).

This completes the proof of the lemma.

We can write the integral equation (19.78) in the form of an operator equation as follows:

$$(I + K) \varphi (x) = f (x) . \tag{19.88}$$

Here, I is the identity operator of $C(S)$, and K is the operator that is defined by the Formula (19.79).

Now, we present a theorem on the existence and uniqueness of the solution to the operator equation (19.88).

Theorem 19.10 *If (1)* $\int_0^{\ell} \frac{\omega_\phi^*(\tau)}{\tau} d\tau < +\infty$*; (2) as* $\delta \to 0$*, we have* $\omega_\phi^{1,0}(\delta) = o\left(\ln^{-1}\delta\right)$*; and (3)* $Ker\,(I+K) = \{0\}$*, then, for* $f \in C(S)$ *and* $\phi \in C(S \times S)$*, the operator equation* (19.88) *has a unique solution in the space* $C(S)$*.*

Proof To prove the theorem, it is sufficient to show that the linear operator $I + K : C(S) \to C(S)$ is bounded and one to one. The operator K (according to the Lemma 19.14) is compact. Since every compact operator is bounded, the operator K is bounded. Thus, the operator $I + K$ is bounded. Furthermore, by condition (3) of the theorem, the operator $I + K$ is a one-to-one operator. Therefore, according to the Banach Theorem on the existence of a bounded inverse operator, the operator $I + K$ has a bounded inverse (see [35, Theorem 3, p. 225]).

This completes the proof of Theorem 19.10.

19.6.2 Collocation Method

Let S be a closed, bounded, and smooth surface in $\mathbb{R}^3$, $S_i \subset S, i = 1, 2, \ldots, n(h)$ be the "regular" elementary regions of S, and $x_i, i = 1, 2, \ldots, n(h)$ be the support points of the S_i's. We will use the following equality as an estimate of the integral on the right-hand side of (19.79) at the support points $x_i \in S_i$:

$$\left(K^{n(h)}\varphi\right)(x_i) = \sum_{j=1, j\neq i}^{n(h)} \frac{\phi\left(x_i, x_j\right)}{\left|x_i - x_j\right|^2} .\varphi\left(x_j\right).mes\left(S_j\right). \tag{19.89}$$

We will refer to Formula (19.89), as the cubature formula of the integral that is on the right-hand side of (19.79).

Now, we will give a theorem below, which can be proven in a manner similar to the proof of Theorem 2.2.2 in [31, p. 75], concerns the following difference:

$$R^{n(h)}(x_i) = (K\varphi)(x_i) - \left(K^{n(h)}\varphi\right)(x_i),\ i = 1, 2, \ldots, n(h).$$

Theorem 19.11 *If* S *is a closed, bounded, and smooth surface in* $\mathbb{R}^3$*, then the following estimate is true:*

$$\max\left\{\left|R^{n(h)}(x_i)\right| : i = 1, 2, \ldots, n(h)\right\}$$
$$\leq c_{14}.\left\{\omega_\varphi(R(h)) + \|\varphi\| \left[\begin{array}{l}\omega_\phi^{0,1}(R(h))\,|\ln R(h)| \\ + \int_0^{R(h)} \frac{\omega_\phi^*(\tau)}{\tau} d\tau + R(h). \int_{r(h)}^{R(h)} \frac{\omega_\phi^*(\tau)}{\tau^2} d\tau\end{array}\right]\right\}.$$

For every $u^{n(h)} \in \mathbb{C}^{n(h)}$,$n(h) = 1, 2, \ldots$ let us define the operators $K^{n(h)} : \mathbb{C}^{n(h)} \to \mathbb{C}^{n(h)}$ via the following formulas:

$$K^{n(h)}u^{n(h)} = \left(K_1^{n(h)}u^{n(h)}, K_2^{n(h)}u^{n(h)}, \ldots, K_{n(h)}^{n(h)}u^{n(h)}\right), \tag{19.90}$$

where

$$K_i^{n(h)}u^{n(h)} = \sum_{j=1, j\neq i}^{n(h)} \frac{\phi\left(x_i, x_j\right)}{\left|x_i - x_j\right|^2}.u_j.mes\left(S_j\right), i = 1, 2, \ldots, n(h).$$

For $u \in C(S)$, we will call the operator $p^{n(h)} : C(S) \to \mathbb{C}^{n(h)}$, defined by $p^{n(h)}u = \left(u_1, u_2, \ldots, u_{n(h)}\right) = u^{n(h)}, u_i = u\left(x_i\right), i = 1, 2, \ldots, n(h), n(h) = 1, 2, \ldots$, the simple drift operator. It is clear that $p^{n(h)} \in L(C(S), \mathbb{C}^{n(h)})$.

If we use the cubature formula (19.89) instead of the integral in Eq. (19.78) and substitute the u_i's (u_i's are the approximate values of $\varphi\left(x_i\right)$) for the $\varphi\left(x_i\right)$'s**,** then we obtain the following linear system of equations in the unknowns u_i, $i = 1, 2, \ldots, n(h)$:

$$u_i + K_i^{n(h)}u^{n(h)} = f_i, i = 1, 2, \ldots, n(h), \tag{19.91}$$

where $f_i = f\left(x_i\right)$. We can write system (19.91) as an operator equation in the following form by using the simple drift operator $p^{n(h)}$:

$$\left(I^{n(h)} + K^{n(h)}\right)u^{n(h)} = f^{n(h)}. \tag{19.92}$$

The following theorem posits the existence of a unique solution of the operator equation (19.92) and the convergence of this solution to the solution of the singular integral equation (19.78).

Theorem 19.12 *Let the following conditions be satisfied:*

(1) $\int_0^{\ell} \frac{\omega_\phi^*(\tau)}{\tau}d\tau < \infty$;
(2) as $\delta \to 0$, *we have* $\omega_\phi^{1,0}(\delta) = o\left(\ln^{-1}\delta\right)$ *and* $\omega_\phi^{0,1}(\delta) = o\left(\ln^{-1}\delta\right)$;
(3) $Ker\,(I + K) = \{0\}$.

In this case, Eqs. (19.78) and (19.92), for every $f \in C(S), \phi \in C(S \times S)$*, have only the unique solutions* $\varphi_* \in C(S)$ *and* $u_*^{n(h)} \in \mathbb{C}^{n(h)}$*, respectively, such that* $\exists n_0 \in, \forall n(h) \geq n_0$.

Furthermore, the following estimate is true:

$$c_{15}.\Delta_{n(h)} \leq \left\| u_*^{n(h)} - p^{n(h)}\varphi_* \right\| \leq c_{16}.\Delta_{n(h)}, \tag{19.93}$$

where $c_{15} = 1/\sup\left\{ \left\| I^{n(h)} + K^{n(h)} \right\| : n(h) \geq n_0 \right\}$,
$c_{16} = \sup\left\{ \left\| \left(I^{n(h)} + K^{n(h)} \right)^{-1} \right\| : n(h) \geq n_0 \right\}$,
$\Delta_{n(h)} = \max\left\{ \left| K_i^{n(h)} \left(p^{n(h)}\varphi_* \right) - (K\varphi_*)(x_i) \right| : i = 1, 2, \ldots, n(h) \right\}$.

Proof Let $P = \left\{ p^{n(h)} \right\}, n(h) = 1, 2, \ldots$ be a system of simple drift operators**.** From the definition of P, it is obvious that the system P is a connective system for $C(S)$ and $\mathrm{c}^{n(h)}, n(h) = 1, 2, \ldots$. From Lemma 19.14, the definition of $K^{n(h)}$, for $n(h) = 1, 2, \ldots$, and Theorem 19.11, we can see that $K^{n(h)} \xrightarrow{GG} K$.

Now, we show that $K^{n(h)} \xrightarrow{GG} K$ is compact. For every $u^{n(h)} \in \mathbb{C}^{n(h)}$, let us take

$$\left(K_{n(h)} u^{n(h)} \right)(x) = \sum_{j=1}^{n(h)} \left[\int_{S_j} \frac{\phi(x, y)}{|x - y|^2} d\sigma_y \right] .u_j, x \in S.$$

From the assumptions of the theorem (and using Lemma 19.14), it can be easily shown that the class of $\left\{ K_{n(h)} u^{n(h)} \right\}$ is uniformly bounded and equicontinuous. Thus, according to the Arzela–Ascoli Theorem, the sequence $\left\{ K_{n(h)} u^{n(h)} \right\}$ is relatively compact (see [35, p. 110, Theorem 4]).

Just as in the proof of Theorem 19.11, the following inequality can be proven:
$\left| K_i^{n(h)} u^{n(h)} - \left(K_{n(h)} u^{n(h)} \right)(x_i) \right| \leq$

$$\leq c_{17}. \left\{ \omega_\phi^{0,1}(R(h)) . |\ln R(h)| + \int_0^{R(h)} \frac{\omega_\phi^*(\tau)}{\tau} d\tau + R(h). \int_{R(h)}^{\ell} \frac{\omega_\phi^*(\tau)}{\tau^2} d\tau \right\} . \left\| u^{n(h)} \right\|.$$

From this inequality**,** it can be seen that $\left\| K^{n(h)} u^{n(h)} - p^{n(h)} \left(K_{n(h)} u^{n(h)} \right) \right\| \to 0$ as $h \to 0$. Thus, from Proposition 19.2 and Definition 19.16, $K^{n(h)} \xrightarrow{GG} K$ is compact. From the third assumption of the theorem and Proposition 19.3, $\exists n_0 \in$n such that, for every$n(h) \geq n_0$, the operator $I^{n(h)} + K^{n(h)}$ is a Fredholm operator with zero indexes.

Furthermore, from the definition of the system P, it is obvious that $f^{n(h)} \xrightarrow{G} f$. Therefore, Theorem 19.12 follows from Theorems 19.9 and 19.10.

Corollary 19.8 *If the assumptions of Theorem 19.12 are satisfied, then the solution of Eq.* (19.92) *G-converges to the solution of Eq.* (19.78)*.*

This corollary is obvious from Estimate (19.93) and Theorem 19.11.

19.7 On the Approximate Solution of Singular Integral Equations with Negative Index

This part is devoted to investigating a class of singular integral equations with a negative index on a closed, simple, and smooth curve. In this section, we propose the collocation method to solve negative index linear singular integral equations. In the section, sufficient conditions are given for the convergence of this method in Hölder space.

The purpose of this study is to examine the collocation method to identify an approximate solution of the singular integral equations in cases where the index is negative, on a closed simple smooth curve.

In this section, we investigate the following type of linear singular integral equation:

$$K\varphi(t) = K^0\varphi(t) + \lambda \cdot k\varphi(t) = f(t), t \in \gamma. \tag{19.94}$$

In this equation,

$$K^0\varphi(t) = a(t)\varphi(t) + b(t)S\varphi(t),\ \ S\varphi(t) = \frac{1}{\pi i}\int_\gamma \frac{\varphi(\tau)}{\tau - t}d\tau,$$

$$k\varphi(t) = \int_\gamma k(t,\tau)\varphi(\tau)d\tau, t \in \gamma,$$

where γ is a closed, simple, and smooth curve in the complex plane, the functions $a(t), b(t), f(t)$and $k(t,\tau)$are known functions in Hölder space, and $a^2(t) - b^2(t) \neq 0$ in γ, and λ is a complex parameter and $\varphi(t)$ is the unknown function.

In this section, the present studies about the approximate solution of singular integral equations of type (19.94) are improved in two ways, then the collocation method is applied for the approximate solution of singular integral equations with negative index defined on an Alper curve.

First, we introduce some concepts needed to prove our main results.

Then, we show the convergence of the collocation method applied to singular integral equation (19.94).

Now, we will introduce some necessary information for proving the main results.

We consider a closed, simple, and smooth curve γ with equation $t = t(s)$, $0 \leq s \leq \ell$ in the complex plane, where s is the arc length calculated from a fixed point and $\ell = |\gamma|$ is the length of the curve γ. The interior and exterior of the curve γ are denoted by γ^+ and γ^-, respectively. Let the origin $0 \in \gamma^+$.

Definition 19.17 ([22, 58, 67]) Let the function $\theta(s)$ be the slope angle of the curve γ at the point $t(s)$ and let $\omega(\theta, x)$ be the continuity modulo of this function. If the condition $\int_0^{d_1} x^{-1}\omega(\theta,x)\,|\ln x|\,dx < \infty$ is satisfied, then the curve γ is called an Alper curve.

We will denote the class of Alper curves with(A). From here on $d_1, d_2, \ldots$ will denote positive real numbers.

Remark 19.4 It is known that if the function $\theta(s)$ in Definition 19.17 satisfies the condition $|\theta(s_1) - \theta(s_2)| \leq d_2 \cdot |s_1 - s_2|^\alpha$, $0 < \alpha < 1$, $s_1, s_2 \in [0, \ell]$, then the curve γ is said to be a Lyapunov curve.

Based on Remark 19.4 and Definition 19.17, it is obvious that every Lyapunov curve is an Alper curve.

Let $C(\gamma)$ be the set of continuous functions, which are defined on the curve γ. For $0 < \alpha < 1$, let us take

$$H_\alpha(\gamma) = \left\{ f \in C(\gamma) : H(f;\alpha) = \sup \left\{ \frac{|f(t_1) - f(t_2)|}{|t_1 - t_2|^\alpha} : t_1, t_2 \in \gamma, t_1 \neq t_2 \right\} < \infty \right\}$$

as given above.

The set $H_\alpha(\gamma)$, $0 < \alpha < 1$ is a Banach space with the norm $\|f\|_\alpha = \|f\|_{C(\gamma)} + H(f;\alpha)$ (see [22, 58, 67]). Here, $\|f\|_{C(\gamma)} = \max\{|f(t)| : t \in \gamma\}$.

Let $H_\alpha^{(r)}(\gamma)$ be the set of functions whose derivative of order r is from the space$H_\alpha(\gamma)$. Here, r is a nonnegative integer.

Definition 19.18 ([22, 58, 67]) The integer number $\nu = \frac{1}{2\pi}\left[\arg(D(t)/C(t))\right]_\gamma$ is called the index of the singular integral equation (19.94) (or of the operator K). Here, $D(t) = a(t) - b(t)$ and$C(t) = a(t) + b(t)$.

In the linear singular integral equation theory the following equation

$$K^0\varphi(t) = f(t), t \in \gamma \tag{19.95}$$

is called the characteristic equation of the singular integral equation (19.94) (see [7, 18, 22]). The sufficient conditions for the solvability of the singular integral equation (19.94) can be derived from the existence of the solution of the characteristic equation.

Let us state the necessary assumptions.

Theorem 19.13 ([22, 58, 67]) *Let $a, b, f \in H_\alpha(\gamma)$, $0 < \alpha < 1$, and ν is the index of the singular integral equation (19.94), and γ is a closed simple smooth curve.*

1. If $\nu > 0$ and the condition

$$\int_\gamma \tau^{p-1}\varphi(\tau)d\tau = 0, p = 1, 2, \ldots, \nu \tag{19.96}$$

is satisfied, then the characteristic equation (19.95) *has the unique solution* $\varphi(t) = Rf(t)$ *in the space* $H_\alpha(\gamma), 0 < \alpha < 1$. *Here,*

$$Rf(t) = \frac{Z(t)}{D(t)C(t)}\bar{K}^0(f(t)/Z(t)), \;\; \bar{K}^0(f(t)/Z(t)) = a(t)\frac{f(t)}{Z(t)} - b(t)S(f(t)/Z(t))$$

$$Z(t) = C(t)\psi^+(t) = t^{-\nu}D(t)\psi^-(t), \psi^{\pm}(t) = \exp[\Gamma^{\pm}(t)],$$

$$\Gamma(z) = \frac{1}{2\pi i}\int_\gamma \frac{\ln\left[\tau^{-\nu}G(\tau)\right]}{\tau - z}d\tau, \;\; z \notin \gamma, \;\; G(t) = D(t)/C(t),$$

$$\Gamma^{\pm}(t) = \Gamma(t) \pm \frac{1}{2}\ln[t^{-\nu}\cdot G(t)], t \in \gamma. \tag{19.97}$$

1. *If* $\nu = 0$, *then the characteristic equation* (19.95) *has the unique solution* $\varphi(t) = Rf(t)$ *in* $H_\alpha(\gamma), 0 < \alpha < 1$.
2. *If* $\nu < 0$, *then the existence of the unique solution* $\varphi(t) = Rf(t)$ *in the space* $H_\alpha(\gamma), 0 < \alpha < 1$ *depends upon meeting the following condition:*

$$\int_\gamma \frac{f(\tau)}{Z(\tau)}\tau^{p-1}d\tau = 0, \; p = 1, 2, \ldots, -\nu \tag{19.98}$$

As always, we will denote the complex numbers by $\mathbb{C}$.

Theorem 19.14 ([63]) *Let* γ *be a close simple smooth curve,* $a, b, f \in H_\alpha(\gamma), 0 < \alpha < 1$, *and suppose that the index of the singular integral equation (19.94) is* $\nu < 0$. *In this case, then the following equation has the unique solution* $x(t) = (\varphi, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{-\nu})$

$$K^0_\varepsilon x(t) \equiv K^0\varphi(t) + \sum_{k=1}^{-\nu}\varepsilon_k h_k(t) = f(t), t \in \gamma, \tag{19.99}$$

in the space $X = \{x = (\varphi, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{-\nu}) : \varphi \in H_\alpha(\gamma), \varepsilon_k \in \mathbb{C}, \; k = 1, 2, \ldots, -\nu\}$. *Here, the functions* $h_k(t) = b(t)\cdot t^{k-1}$, $k = 1, 2, \ldots, -\nu$ *are linearly independent solutions of the equation* $Rh(t) = 0$.

Now suppose that the index of the singular integral equation (19.94) is $\nu < 0$. In this case, the singular integral equation (19.94) together with the following conditions

$$L_p(f, k, \varphi) \equiv \int_\gamma \frac{\tau^{p-1}}{Z(\tau)}\left[f(\tau) - \lambda\int_\gamma k(\tau, \xi)\varphi(\xi)d\xi\right]d\tau = 0, \;\; p = 1, 2, \ldots, -\nu \tag{19.100}$$

is equivalent to the following Fredholm integral equation

$$F\varphi(t) \equiv \varphi(t) + \lambda\int_\gamma F(t, \tau)\varphi(\tau)d\tau = Rf(t), \tag{19.101}$$

in the subspace $\bar{H}_\alpha(\gamma) = \left\{ f \in H_\alpha(\gamma) : L_p(f, k, \varphi) = 0,\, p = 1, 2, \ldots, -\nu \right\}$, $0 < \alpha < 1$.

Here,

$$F(t,\tau) = \frac{a(t)k(t,\tau)}{D(t)C(t)} - \frac{b(t)Z(t)}{D(t)C(t)} \frac{1}{\pi i} \int_\gamma \frac{k(\xi,\tau)}{Z(\xi)} \frac{d\xi}{\xi - t}.$$

Really, if (19.94) has a solution φ then, due to (19.98) this solution automatically satisfies (19.100). Then, by (19.97) it follows that φ is a solution of (19.101). Conversely, if $\varphi \in \bar{H}_\alpha(\gamma)$, then φ is a solution of (19.94). Hence, problem (19.94) is equivalent to problem (19.100)–(19.101).

Moreover, the equation

$$K_\varepsilon x(t) \equiv K\varphi(t) + \sum_{k=1}^{-\nu} \varepsilon_k h_k(t) = f(t),\; x(t) = (\varphi(t), \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{-\nu}) \tag{19.102}$$

is equivalent to the Fredholm integral equation (19.101) together with the following conditions

$$\sum_{k=1}^{-\nu} \varepsilon_k \int_\gamma \frac{h_k(\tau)}{Z(\tau)} \tau^{p-1} d\tau = L_p(f, k, \varphi),\; p = 1, 2, \ldots, -\nu \tag{19.103}$$

in the subspace

$\bar{X} = \left\{ x \in X : L_p(f, k, \varphi) = \sum_{k=1}^{-\nu} \varepsilon_k \int_\gamma \frac{h_k(\tau)}{Z(\tau)} \tau^{p-1} d\tau \;,\, p = 1, 2, \ldots, -\nu \right\}$.

Really, if $x(t) = (\varphi(t), \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{-\nu})$ is a solution of (19.102), then (19.103) is automatically satisfied. Conversely, if $x(t) = (\varphi(t), \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{-\nu})$ is a solution of (19.101), then (19.103) is the solution of (19.102).

Remark 19.5 We call Eq. (19.102) the "regularization" of the singular integral equation (19.94). We also want to indicate that previously in the studies of V. V. Ivanov (see [26]) the idea of presenting the unknowns $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{-\nu}$(regularized parameter) was given. Afterwards this idea was used by B. I. Musaev (see [63–65]).

We now provide certain pertinent information regarding the Fredholm integral equation theory. Let us take the following homogeneous Fredholm integral equation in the space $H_\beta(\gamma)$, $0 < \beta < \alpha < 1$

$$F\varphi(t) \equiv \varphi(t) + \lambda \int_\gamma F(t,\tau)\varphi(\tau) d\tau = 0. \tag{19.104}$$

Definition 19.19 ([22, 58]) If, for a value of theλ parameter, there exists a non-zero solution of the homogeneous Fredholm integral equation (19.104), then we will call this value of the parameter an eigenvalue of the kernel $F(t,\tau)$ (or, equivalently, of Eq. (19.104)).

Theorem 19.15 ([21, 96]) *If the parameter λ is not an eigenvalue, or, equivalently, if the homogeneous Fredholm integral equation (19.104) has only the zero solution, then the non-homogeneous Fredholm integral equation (19.101) has only one solution for every $f, k \in H_\alpha(\gamma)$, $0 < \alpha < 1$.*

In this study, we assume that the homogeneous equation (19.104) has only the zero solution. In this case, according to Theorems 19.13 and 19.14, Eq. (19.102) has the unique solution $x(t) = (\varphi, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{-\nu})$ for every $f \in H_\alpha(\gamma)$, $0 < \alpha < 1$. Here, the function $\varphi(t) == Rf(t) - \int_\gamma \bar{F}(t,\tau)Rf(\tau)d\tau$ is the solution of the non-homogeneous Fredholm integral equation (19.101). $\bar{F}(t,\tau)$ is the resolvent kernel of Eq. (19.101) and can be clearly expressed with the help of the function $F(t,\tau)$ (see [21]). The components ε_k, $k = 1, 2, \ldots, -\nu$ are found from Eq. (19.103).

We will denote the set of natural numbers by $\mathbb{N}$. For every function $f \in C(\gamma)$ and $n \in \mathbb{N}$ let us define the Lagrange interpolation polynomial using the following formula (see [79]).

$$U_n f(t) \equiv U_n(f,t) = \sum_{j=1}^{2n} f(t_j) l_j(t), t \in \gamma. \tag{19.105}$$

Here,

$$l_j(t) = \prod_{k=0,k\neq j}^{2n} \frac{t - t_k}{t_j - t_k} \left(\frac{t_j}{t}\right)^n, \ t \in \gamma, \ t_j = \phi(w_j),$$

$$w_j = \exp\left[\frac{2\pi i}{2n+1}(j-k)\right], \ i^2 = -1, j = 0, 1, \ldots, 2n \tag{19.106}$$

and $z = \phi(w)$ is a conform transformation which satisfies the conditions $\phi(\infty) = \infty$, $\phi'(\infty) > 0$ and transforms the region outside the unit circle centred at the origin to the region γ^-.

Lemma 19.15 (See [80, Corollary 1.2.1]) *Let $\gamma \in (A)$, and $f \in H_\alpha^{(r)}(\gamma)$, $0 < \beta < \alpha < 1$ ($0 \le r$- is an integer). In this case, for every $n \in \mathbb{N}$, we have*

$$\|f - U_n f\|_\beta \le (d_3 + d_4 \ln n)\, H(f,\alpha) n^{\beta-\alpha-r}. \tag{19.107}$$

The following information is derived from the works of Gabdulkhaev [19–21].

Let X and Y be normed spaces and let X_n and Y_n be respective, subspaces of X and Y with finite dimension n. Let us take the following operator equations:

$$Kx = y (x \in X, y \in Y), \tag{19.108}$$

$$K_n x_n = y_n (x_n \in X_n, y_n \in Y_n). \tag{19.109}$$

Here, K and K_n are bounded linear operators from X to Y and from X_n toY_n, respectively.

Theorem 19.16 (See [20, Theorem 7]) *Let the following conditions be satisfied:*

1. *The operator $K : X \to Y$ has a bounded inverse;*
2. *When $n \to \infty$ then $\|K_n - K\| \to 0$;*
3. $\dim X_n = \dim Y_n < \infty \ (n = 1, 2, \ldots)$
   Then, for every $n \in \mathbb{N}$ that satisfies the condition $p_n \equiv \left\|K^{-1}\right\| . \|K_n - K\| < 1$, Eq. (19.109) has only one solution:$x_n \in X_n$. Furthermore,

$$\|x_n\| \leq \left\|K_n^{-1}\right\| . \|y_n\| , \left\|K_n^{-1}\right\| \leq \left\|K^{-1}\right\| (1 - p_n)^{-1} .$$

   If the condition
4. *when $n \to \infty$ then $\|y_n - y\| \to 0$*
   is also satisfied, then the solution of Eq. (19.109) convergences to the solution $x \in X$ of Eq. (19.108). In this case, the following is true:

$$\|x_n - x\| \leq (\|y_n - y\| + p_n \|y\|) \left\|K^{-1}\right\| (1 - p_n)^{-1} .$$

19.7.1 Collocation Method

Let us take

$$X = \left\{x = (\varphi, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{-\nu}) : \varphi \in H_\beta(\gamma), \varepsilon_k \in \mathbb{C},\ k = 1, 2, \ldots, -\nu\right\}$$

with norm $\|x\|_X = \|\varphi\|_\beta + \sum_{k=1}^{-\nu} |\varepsilon_k|$,$0 < \beta < \alpha < 1$.

We can write Eq. (19.102) in the space X, as following linear operator equation

$$C(t) \cdot P\varphi(t) + D(t) \cdot Q\varphi(t) + \lambda k\varphi(t) + \sum_{k=1}^{-\nu} \varepsilon_k h_k(t) = f(t), \qquad (19.110)$$

where $P = \frac{1}{2}(I + S)$, $Q = \frac{1}{2}(I - S)$ are projection operators, I is the identity operator on $H_\alpha(\gamma)$ and S is the linear singular integral operator with Cauchy kernel that is defined by (19.94).

We will seek the approximate solution of Eq. (19.102) in the form $x_{n-\nu} = (\varphi_{n-\nu}, \varepsilon_{1,n}, \varepsilon_{2,n}, \ldots, \varepsilon_{-\nu,n})$. Here, $\varphi_{n-\nu} = \varphi^+_{n-\nu} - \varphi^-_n$,$\varphi^+_{n-\nu}(t) = \sum_{k=0}^{n-\nu} \alpha_k t^k$,$\varphi^-_n(t) = -\sum_{k=-n}^{-1} \alpha_k t^k$.We will find the $\alpha_{-n}, \ldots, \alpha_{n-\nu}, \varepsilon_{1,n}, \ldots, \varepsilon_{-\nu,n}$ unknown values from the following linear equation system:

$$C_j \cdot \sum_{k=0}^{n-\nu} \alpha_k t_j^k + D_j \cdot \sum_{k=-n}^{-1} \alpha_k t_j^k + \lambda \int_\gamma k(t_j, \tau)\varphi_{n-\nu}(\tau)d\tau + \sum_{k=1}^{-\nu} \varepsilon_k h_k(t_j) = f_j. \qquad (19.111)$$

Here, C_j, D_j, $k(t_j, \tau)$, and f_j are the respective values of the functions $C(t)$, $D(t)$, $k(t, \tau)$ and $f(t)$ at the points $t = t_j$. The points t_j, $j = 0, 1, \ldots, 2(n-\nu)$ are the collocation points that are defined by (19.106).

Let us denote the $(2(n-\nu)+1)$-dimensional subspace of X using $X_{n-\nu} = \{x_{n-\nu} : x_{n-\nu} = (\varphi_{n-\nu}, \varepsilon_{1,n}, \varepsilon_{2,n}, \ldots, \varepsilon_{-\nu,n})\}$.

In the space X, we can write Eq. (19.102) in the form of a linear operator equation:

$$\bar{K}_\nu x \equiv \psi^- P\varphi + t^\nu \psi^+ Q\varphi + \lambda d.k\varphi + d \cdot \sum_{k=1}^{-\nu} \varepsilon_k h_k = g, \tag{19.112}$$

where the functions $\psi^\pm$ are the functions that are defined in formulas (19.97) and where $x = (\varphi, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{-\nu}) \in X$, $d = \psi^-/C$, $g = d \cdot f$.

Now we offer the following theorem about the existence of the approximate solution of Eq. (19.102) and the convergence of the collocation method.

Theorem 19.17 *Let $\gamma \in (A)$. The functions $a(t), b(t)$, $f(t)$, and $k(t, \tau)$ (for every two variable) are of the class $H_\alpha^{(r)}(\gamma)$, $0 < \alpha < 1$ ($0 \le r$- is an integer). For every $t \in \gamma$, $a^2(t) - b^2(t) \neq 0$ and index $\nu < 0$.*

If the homogeneous Fredholm integral equation (19.104) has only the zero solution, then, for every $n > n_0 = \min\{n \in \mathbb{N}: \delta_{n,\nu} \equiv (d_5 + d_6 \ln(n-\nu))$ $(n-\nu)^{\beta-\alpha-r} \cdot \left\| K_\nu^{-1} \right\| < 1\}$, $0 < \beta < \alpha < 1$ the linear equation system (19.111) has the unique solution $\left(\alpha_{-n}^*, \ldots, \alpha_{n-\nu}^*, \varepsilon_{1,n}^*, \ldots, \varepsilon_{-\nu,n}^*\right)$ in the space $X_{n-\nu}$. The approximate solution $x_{n-\nu}^* = \left(\varphi_{n-\nu}^*, \varepsilon_{1,n}^*, \ldots, \varepsilon_{-\nu,n}^*\right)$ of Eq. (19.102) converges to the unique solution $x^* = \left(\varphi^*, \varepsilon_1^*, \ldots, \varepsilon_{-\nu}^*\right)$, and the following estimate is correct:

$$\left\| x_{n-\nu}^* - x^* \right\|_X \le (d_7 + d_8 \ln(n-\nu))(n-\nu)^{\beta-\alpha-r}, 0 < \beta < \alpha \tag{19.113}$$

where $\varphi_{n-\nu}^*(t) = \sum_{k=-n}^{n-\nu} \alpha_k^* t^k$.

Proof Let the index of the singular integral equation (19.102) be $\nu < 0$. Let us take the collocation points $t_j \in \gamma$, $j = 0, 1, \ldots, 2(n-\nu)$ as the points that are defined in formula (19.106), and the operator $U_{n-\nu}$ as a Lagrange interpolation operator of the degree $n-\nu$ that is defined in formula (19.105). Let us write the linear equations system (19.111) in the space $X_{n-\nu}$ as the following linear operator equation:

$$\bar{K}_{n,\nu} x_{n-\nu} \equiv U_{n-\nu}\left(\psi^- P\varphi_{n-\nu} + t^\nu \psi^+ Q\varphi_{n-\nu} + \lambda d.k\varphi_{n-\nu}\right)$$
$$+ U_{n-\nu}\left(d \cdot \sum_{k=1}^{-\nu} \varepsilon_{k,n} h_k\right) = U_{n-\nu} g \equiv g_{n-\nu}. \tag{19.114}$$

From (19.112) and (19.114), we can write:

$$\bar{K}_{n,\nu}x_{n-\nu} - \bar{K}_{\nu}x_{n-\nu} = \left(\psi_{n-\nu}^{-} - \psi^{-}\right) \cdot P\varphi_{n-\nu} + \left(\psi_{n-\nu}^{+} - \psi^{+}\right) \cdot t^{\nu} Q\varphi_{n-\nu}$$
$$- \psi_{n-\nu}^{-} \cdot P\varphi_{n-\nu} - t^{\nu}\psi_{n-\nu}^{+} \cdot Q\varphi_{n-\nu} + U_{n-\nu}\left(\psi^{-} \cdot P\varphi_{n-\nu}\right.$$
$$\left. + t^{\nu}\psi_{n-\nu}^{+} \cdot Q\varphi_{n-\nu}\right) - \lambda \cdot \left(U_{n-\nu}\left(d \cdot k\varphi_{n-\nu}\right) - d \cdot k\varphi_{n-\nu}\right)$$
$$- \sum_{k=1}^{-\nu} \varepsilon_{k,n}\left(U_{n-\nu}\left(d \cdot h_k\right) - d \cdot h_k\right). \tag{19.115}$$

Here, $\psi_{n-\nu} = \psi_{n-\nu}^{+} - \psi_{n-\nu}^{-}$ ($\psi_{n-\nu}^{+}(t) = \sum_{k=0}^{n-\nu} \beta_k t^k$, $\psi_{n-\nu}^{-}(t) = \sum_{k=-n+\nu}^{-1} \beta_k t^k$) is the best approximation to the function $\psi = \psi^{+} - \psi^{-}$ with rational polynomial whose degree does not exceed $n-\nu$. Since $\psi_{n-\nu}^{-} \cdot P\varphi_{n-\nu} + t^{\nu}\psi_{n-\nu}^{+} \cdot Q\varphi_{n-\nu}$ is a rational polynomial whose degree does not exceed $n-\nu$ and $U_{n-\nu}(\psi_{n-\nu}^{-} \cdot P\varphi_{n-\nu} + t^{\nu}\psi_{n-\nu}^{+} \cdot Q\varphi_{n-\nu}) = \psi_{n-\nu}^{-} \cdot P\varphi_{n-\nu} + t^{\nu}\psi_{n-\nu}^{+} \cdot Q\varphi_{n-\nu}$, we can write (19.115) as follows:

$$\bar{K}_{n,\nu}x_{n-\nu} - \bar{K}_{\nu}x_{n-\nu} = \left(I - U_{n-\nu}\right)\left[\left(\psi_{n-\nu}^{-} - \psi^{-}\right) P\varphi_{n-\nu} + t^{\nu}\right.$$
$$\left. \cdot \left(\psi_{n-\nu}^{+} - \psi^{+}\right) Q\varphi_{n-\nu}\right] - \lambda \cdot \left[U_{n-\nu}\left(d \cdot k\varphi_{n-\nu}\right) - d \cdot k\varphi_{n-\nu}\right]$$
$$- \sum_{k=1}^{-\nu} \varepsilon_{k,n}\left[U_{n-\nu}\left(d \cdot h_k\right) - d \cdot h_k\right]. \tag{19.116}$$

From the boundedness of the operators P and Q in Hölder space (see [22, 58, 67]) and the following estimates (see [80, Corollary 1.1.5]):

$$\left\|\psi_{n-\nu}^{+} - \psi^{+}\right\|_{\beta} \leq d_9\left(n-\nu\right)^{\beta-\alpha-r}, \left\|\psi_{n-\nu}^{-} - \psi^{-}\right\|_{\beta} \leq d_{10}\left(n-\nu\right)^{\beta-\alpha-r},$$

we have

$$\left\|\left(\psi_{n-\nu}^{-} - \psi^{-}\right) P\varphi_{n-\nu} + t^{\nu} \cdot \left(\psi_{n-\nu}^{+} - \psi^{+}\right) Q\varphi_{n-\nu}\right\|_{\beta} \leq d_{11}\left(n-\nu\right)^{\beta-\alpha-r} \cdot \left\|\varphi_{n-\nu}\right\|_{\beta}. \tag{19.117}$$

In this case, from the following inequality (see [79, Lemma 2.1])

$$\left\|U_{n-\nu}\right\|_{\beta} \leq d_{12} + d_{13}\ln\left(n-\nu\right)$$

and, from (19.117), it is clear that the following estimate is correct:

$$\left\|\left(I - U_{n-\nu}\right)\left[\left(\psi_{n-\nu}^{-} - \psi^{-}\right) P\varphi_{n-\nu} + t^{\nu} \cdot \left(\psi_{n-\nu}^{+} - \psi^{+}\right) Q\varphi_{n-\nu}\right]\right\|_{\beta}$$
$$\leq \left(d_{14} + d_{15}\ln\left(n-\nu\right)\right)\left(n-\nu\right)^{\beta-\alpha-r} \cdot \left\|\varphi_{n-\nu}\right\|_{\beta}. \tag{19.118}$$

From the Lemma 19.15 we can write

$$\|U_{n-\nu}(d.k\varphi_{n-\nu}) - d.k\varphi_{n-\nu}\|_\beta \leq (d_{16} + d_{17}\ln(n-\nu))(n-\nu)^{\beta-\alpha-r} . \|\varphi_{n-\nu}\|_\beta . \tag{19.119}$$

In the evaluation of the last term of equality (19.116), the following estimate is evident:

$$\left\| \sum_{k=1}^{-\nu} \varepsilon_{k,n} \left[U_{n-\nu}\left(d \cdot h_k\right) - d \cdot h_k \right] \right\|_\beta \leq \sum_{k=11}^{-\nu} \left|\varepsilon_{k,n}\right| \left\| U_{n-\nu}\left(d \cdot h_k\right) - d \cdot h_k \right\|_\beta \tag{19.120}$$

Thus, from the evaluations (19.118)–(19.120) we obtain the following estimate of the difference $\bar{K}_{n,\nu}x_{n-\nu} - \bar{K}_\nu x_{n-\nu}$:

$$\left\| \bar{K}_{n,\nu}x_{n-\nu} - \bar{K}_\nu x_{n-\nu} \right\|_\beta \leq \left(d_{18} + d_{19}\ln\left(n-\nu\right)\right)\left(n-\nu\right)^{\beta-\alpha-r} \cdot \|x_{n-\nu}\|_X . \tag{19.121}$$

Besides, according to Lemma 19.15, the following evaluation is true:

$$\|g_{n-\nu} - g\|_\beta \leq \left(d_{20} + d_{21}\ln\left(n-\nu\right)\right)\left(n-\nu\right)^{\beta-\alpha-r} . \tag{19.122}$$

From the assumptions of the theorem, the operator equation (19.112) has only one solution. Consequently, exists the bounded linear inverse operator:$\bar{K}_\nu^{-1} : H_\beta(\gamma) \to X$,$0 < \beta < \alpha$. Thus, from the estimates (19.121) and (19.122) (according to the Theorem 19.16), there exists a unique solution $x^*_{n-\nu} \in X_{n-\nu}$ for operator equation (19.114) that satisfies the following condition for every $n \in \mathbb{N}$:

$$n > n_0 = \min\left\{n \in \mathbb{N} : \delta_{n,\nu} \equiv \left\| \bar{K}_\nu^{-1} \right\| \left(d_{22} + d_{23}\ln\left(n-\nu\right)\right)\left(n-\nu\right)^{\beta-\alpha-r} < 1 \right\}.$$

In this case, for the $x^* \in X$ to be the solution of operator equation (19.112), then

$$\left\| x^*_{n-\nu} - x^* \right\|_X \leq \left(d_{24} + d_{25}\ln\left(n-\nu\right)\right)\left(n-\nu\right)^{\beta-\alpha-r} . \tag{19.123}$$

Therefore, based on the conclusions that we have obtained so far, we can state that linear equation system (19.111) has a unique solution. Therefore, Eq. (19.102) can be solved approximately. Furthermore, the solution of Eq. (19.114) is the approximate solution of Eq. (19.102). Besides, the unique solution of Eq. (19.112) is the unique solution of Eq. (19.102).

Theorem 19.17 is thus completely proved.

From Theorem 19.17, proof of the following theorem is obvious.

Theorem 19.18 *Let us suppose that the parameter* λ *is not an eigenvalue of the homogeneous Fredholm integral equation (19.104) and let* $\nu < 0$. *Let the unique solution of Eq.* (19.102) *be* $x^* = \left(\varphi^*, \varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_{-\nu}^*\right)$ *and the approximate solution be* $x_{n-\nu}^* = \left(\varphi_{n-\nu}^*, \varepsilon_{1,n}^*, \varepsilon_{2,n}^*, \ldots, \varepsilon_{-\nu,n}^*\right)$. *The necessary and sufficient condition for the function* φ^* *to be the unique solution of the singular integral equation (19.94) is* $\lim\limits_{n\to\infty} \varepsilon_{k,n}^* = 0$ *for every* $k = 1, 2, \ldots, -\nu$.

19.7.2 Conclusions for Sect. 19.7

In Sect. 19.7, we apply the collocation method not to the singular integral equation (19.94), but to the "regularization" of Eq. (19.94). We proposed the collocation method in the context of the singular integral equation (19.102) and we derived sufficient conditions for the convergence of this method.

As evidenced by Theorem 19.18, if the vector $x^* = \left(\varphi^*, \varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_{-\nu}^*\right)$ is the unique solution of Eq. (19.102) and the vector $x_{n-\nu}^* = \left(\varphi_{n-\nu}^*, \varepsilon_{1,n}^*, \varepsilon_{2,n}^*, \ldots, \varepsilon_{-\nu,n}^*\right)$ is the approximate solution, and if condition (19.100) is satisfied for the function φ^*, then $\varepsilon_1^* = \varepsilon_2^* = \cdots = \varepsilon_{-\nu}^* = 0$. Therefore, the function φ^* is the unique solution of singular integral equation (19.94). In this case, for sufficiently large values of the natural number n, we can take the rational polynomial $\varphi_{n-\nu}^*$ as the approximate solution of the singular integral equation (19.94). Furthermore, from Theorem 19.17 the following estimate is true:

$$\left\|\varphi_{n-\nu}^* - \varphi^*\right\|_\beta \leq \left(d_{24} + d_{25} \ln\left(n-\nu\right)\right)\left(n-\nu\right)^{\beta-\alpha-r}, 0 < \beta < \alpha < 1.$$

References

1. F.A. Abdullayev, E.H. Khalilov, Ground of the collocation method for a class of boundary integral equations. Differ. Equ. **49**(1), 82–86 (2004)
2. V.M. Alexandrov, E.V. Kovalenko, *Problems with Mixed Boundary Conditions in Continuum Mechanics* (Nauka, Moscow 1986)
3. V.M. Alexandrov, I.I. Kudish, Asymptotic methods in Criffits problem. Appl. Math. Mech. **53**, 665–671 (1989)
4. V.M. Alexandrov, S.M. Mkitaryan, *Contact Problems for Bodies with Thin Coverings and Interlayers* (Nauka, Moscow, 1983)
5. S.M. Amer, On solution of non-linear singular integral equations with shift in generalized Holder space. Chaos, Solutions Fractals **12**, 1323–1334 (2001)
6. K.E. Atkinson, *A Survey of Numerical Methods for the Solution of Fredholm Integral Equations of the Second Kind* (SIAM, Philadelphia, 1976)
7. A.A. Badr, Integro-differential equations with Cauchy kernel. J. Comput. Appl. Math. **134**, 191199 (2001)
8. S.M. Belotserkhovskii, I. K. Lifanov, *Numerical Solutions of Singular Integral Equations* (Nauka, Moscow, 1985)

9. L. Bers, L. Nirenberg, On a representation for linear elliptic systems with discontinuous coefficients and its application. Convegno Internazionale sulle Equazioni Lineari alle Derivate Parziali (Trieste, 1954)
10. B.V. Bojarskii, Quasiconformal mappings and general structural properties of system of non-linear elliptic in the sense of Lavrentev. Sympos. Math. **18**, 485–499 (1976)
11. Y.G. Borisovich, V.G. Zvyagin, Non-linear Fredholm maps and the Leray-Schauder theory. Russ. Math. Surv. **32**(4), 1–54 (1977)
12. D.L. Colton, R. Kress, *Integral Equation Methods in Scattering Theory* (Wiley, New York 1983)
13. I.K. Daugavet, *Introduction to Approximation Theory of Functions* (Leningrad University, Leningrad, 1977) (in Russian)
14. G. David, Courbes corde-are et espaces de Hardy generalizes. Ann. Inst. Fourier **32**(3), 227–239 (1982)
15. R. Duduchava, An application of singular integral equations to some problems of elasticity. Integr. Equ. Oper. Theory **5**(1), 475–489 (1982)
16. R. Duduchava, S. Prösdorf, On the approximation of singular integral equations by equations with smooth kernels. Integr. Equ. Oper. Theory **21**(2), 224–237 (1995)
17. L.R. Duduchava, D. Mitrea, M. Mitrea, Differential operators and boundary value problems on hypersurfaces. Math. Nachr. **279**(9–10), 996–1023 (2006)
18. J. Frankel, A. Galerkin, Solution to a regularized Cauchy singular integro-differential equation. Q. Appl. Math. **L11**(2), 245–258 (1995)
19. B.G. Gabdulkhaev, Finite approximations of singular integrals, direct solution methods of singular integral and integro-differential equations. Itogi Nauki i Tekniki, VINITI AN SSSR, Math. Anal. **18**, 25–31 (1980)
20. B.G. Gabdulkhaev, *Optimal Approximation to Linear Problem* (Kazan University Publications, Kazan, 1980)
21. B.G. Gabdulkhaev, V.E. Gorlov, On the optimal algorithm of the approximate solutions of singular integral equations. Izv. Vuzov Math. **11**, 13–31 (1976)
22. F.D. Gakhov, *Boundary Value Problems*, English Edition (Pergamon Press, Oxford, 1966)
23. C.D. Green, *Integral Equation Methods* (Thomas Nelson, New York, 1969)
24. A.I. Gusseinov, K.S. Mukhtarov, *Introduction to the Theory of Nonlinear Singular Integral Equations* (Nauka, Moscow, 1980) (in Russian)
25. H. Hochstadt, *Integral Equations* (Wiley Interscience, New York, 1973)
26. V.V. Ivanov, *The Theory of Approximate Methods and its Application to the Numerical Solution of Singular Integral Equations* (Naukova Dumka, Kiev, 1968)
27. D.S. Jones, Integral equations for the exterior acoustic problem. Q. J. Mech. Appl. Math. **27**, 129–142 (1974)
28. P. Junghanns, K. Müller, A collocation method for non-linear Cauchy singular integral equations. J. Comput. Appl. Math. **115**, 283–300 (2000)
29. A.I. Kalandia, *Mathematical Methods of the Two-Dimensional Elastics* (Nauka, Moscow, 1973)
30. A.C. Kaya, F. Erdogan, On the solution of integral equation with strongly singular kernels. Q. Appl. Math. **45**, 105–122 (1987)
31. E.G. Khalilov, Approximate methods for the solution of surface integral equations, Ph.D. thesis, Baku, 1999 (in Russian)
32. E.H. Khalilov, On an approximate solution of a boundary integral equation of mixed problem for Helmholtz equation. Proc. Inst. Math. Mach. NASA **31**, 105–110 (2009)
33. R.E. Kleinman, G.F. Roach, Boundary integral equations for the three-dimensional Helmholtz equation. SIAM Rev. **16**, 214–236 (1974)
34. R.E. Kleinman, G.F. Roach, On modified Green functions in exterior problems for the Helmholtzs equation. Proc. R. Soc. Lond. A **383**, 313–333 (1982)
35. A.N. Kolmogorov, S.V. Fomin, *Elements of the Theory of Functions and Functional Analysis* (Nauka, Moscow, 1981) (in Russian)

36. M.A. Krasnosel'skii, P.P. Zabreyko, *Geometric Methods of Nonlinear Analysis* (Nauka, Moscow, 1975)
37. M.A. Krasnosel'skii, G.M. Vainikko, P.P. Zabreiko, Ya. B. Rutitskii, V. Ya. Stetsenko, *Approximate Solution of Operator Equations* (Wolters-Noordhoff Publishing, Groningen, 1972)
38. E. Kreyszig, *Introductory Functional Analysis with Applications* (Wiley, New York, 1978)
39. I.I. Kudish, Numerical solution methods of a one class nonlinear integral and integro-differential equations. J. Numer. Math. Math. Phys. **26**, 14931511 (1986)
40. Y.A. Kustov, B.I. Musaev, A Cubature Formula for Double Singular Integral and its Application. VINITI, No 4281 (1981) (in Russian)
41. E. Lackau, W. Tutschke, *Complex Analysis, Methods, Trends and Applications* (Pergamon Press, London, 1985)
42. E.G. Ladopoulos, On the numerical solution of the finite-part singular integral equations of the first and the second kind used in fracture mechanics. Comput. Methods Appl. Mech. Eng. **65**, 253–266 (1987)
43. E.G. Ladopoulos, Singular integral representation of three-dimensional plasticity problem. Theor. Appl. Fract. Mech. **8**, 205–211 (1987)
44. E.G. Ladopoulos, On a new integration rule with the Gegenbauer polynomials for singular integral equations used in the theory of elasticity. Ing. Arch. **58**, 35–46 (1988)
45. E.G. Ladopoulos, On the numerical evaluation of the general type of finite-part singular integrals and integral equations used in fracture mechanics. J. Eng. Fract. Mech. **31**, 315–337 (1988)
46. E.G. Ladoloulos, The general type of finite-part singular integrals and integral equations with logarithmic singularities used in fracture mechanics. Acta Mech. **75**, 275–285 (1988)
47. E.G. Ladopoulos, On the solution of the finite-part singular integro-differential equations used in two-dimensional aerodynamics. Arch. Mech. **41**, 925–936 (1989)
48. E.G. Ladopoulos, Singular integral operators method for two-dimensional plasticity problems. Comput. Struct. **33**, 859–865 (1989)
49. E.G. Ladopoulos, Non-linear integro-differential equations used in orthotropic shallow spherical shell analysis. Mech. Res. Commun. **18**, 111–119 (1991)
50. E.G. Ladopoulos, Relativistic elastic stress analysis for moving frames. Rev. Roum. Sci. Tech. Mech. Appl. **36**, 195–209 (1991)
51. E.G. Ladopoulos, Singular integral operators method for three-dimensional elastoplastic stress analysis. Comput. Struct. **38**, 1–8 (1991)
52. E.G. Ladopoulos, Singular integral operators method for two-dimensional elastoplastic stress analysis. Forsch. Ingenieurwes. **57**, 152–158 (1991)
53. E.G. Ladopoulos, New aspects for generalization of the Sokhotski-Plemelj formulae for the solution of finite-part singular integrals used in fracture mechanics. Int. J. Fract. **54**, 317–328 (1992)
54. E.G. Ladopoulos, Non-linear singular integral equations elastodynamics by using Hilbert transformations. J. Nonlinear Anal. Real World Appl. **6**, 531–536 (2005)
55. E.G. Ladopoulos, V.A. Zisis, Existence and uniqueness for non-linear singular integral equations used in fluid mechanics. Appl. Math. **42**, 345–367 (1997)
56. E.G. Ladopoulos, V.A. Zisis, Non-linear finite-part singular integral equations arising in two-dimensional fluid mechanics. J. Nonlinear Anal. **42**, 277–290 (2000)
57. P. Linz, *Analytical and Numerical Methods for Volterra Equations* (SIAM, Philadelphia, 1985)
58. J.K. Lu, *Boundary Value Problems for Analytic Functions* (World Scientific, Singapore, 1993)
59. K. Mamedov, N. Kosar, Continuity of the scattering function and the Levinson type formula of the boundary value problem. Int. J. Contemp. Math. Sci. **5**(4), 159–170 (2010)
60. K. Mamedov, H. Menken, On the inverse problem of the scattering theory for a boundary problem. Geom. Integrability Quantization **7**, 226–237 (2006)
61. V.N. Monahov, *Boundary Value Problems with Free Boundaries for Elliptic Systems* (Nauka, Novosibirsk, 1977)
62. A.S. Mshim Ba, W. Tutschke, *Functional-Analytic Methods in Complex Analysis and Applications to Partial Differential Equations* (World Scientific, Singapore, 1990)

63. B.I. Musaev, On approximate solution of the singular integral equations. AN Az. SSR, Institute of Physics Preprint No 17 (1986)
64. B.I. Musaev, On the approximate solution of the singular integral equations. Izv. AN Az. SSSR, Fizik-Teknik Science **5**, 1521 (1986)
65. B.I. Musaev, On the approximate solution of singular integral equations with negative index by Bubnov-Galerkin and collocation methods. Sov. Math. Dokl. **35**(2), 411–416 (1987)
66. N.I. Muskelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953)
67. N.I. Muskhelishvili, *Singular Integral Equations*, English Edition (Noordhoff, Groningen, 1968)
68. N. Mustafa, On the approximate solution of non-linear operator equations. Far East J. Appl. Math. **27**(1), 121–136 (2007)
69. N. Mustafa, Fixed point theory and approximate solutions of non-linear singular integral equations. Complex Variables Elliptic Equ. **53**(11), 1047–1058 (2008)
70. N. Mustafa, Non-linear singular integro-differential equations. Complex Variables Elliptic Equ. **53**(9), 879–886 (2008)
71. N. Mustafa, On the approximate solution of singular integral equations with negative index. Complex Variables **55**(7), 621–631 (2010)
72. N. Mustafa, Newton–Kantorovich method for two-dimensional non-linear singular integral equations. Maejo Int. J. Sci. Technol. **10**(1), 41 (2016)
73. N. Mustafa, Some integral operators and their properties. Kuwait J. Sci. **43**(4), 45–55 (2016)
74. N. Mustafa, C. Ardil, On the approximate solution of a non-linear singular integral equation. Int. J. Comput. Math. Sci. **3**(1), 1–7 (2009)
75. N. Mustafa, E.H. Khalilov, The collocation method for the solution of boundary integral equations. Appl. Anal. **88**(12), 1665–1675 (2009)
76. N. Mustafa, M.I. Yazar, On the approximate solution of a non-linear singular integral equation with Cauchy kernel. Far East J. Appl. Math. **27**(1), 101–119 (2007)
77. N.M. Mustafaev, *On the Approximate Solution of the Singular Integral Equation that is Defined on Closed Smooth Curve*. Singular Integral Operators, vol. 1 (AGU Publications, Baku, 1987), pp. 91–99
78. N. Mustafaev, Approximate solution of non-linear singular integral equations. VINITI (338-B88), 1–36 (1988)
79. N.M. Mustafaev, Error of the approximation to the singular integral equation that is defined on closed smooth curve. In Az. NIINTI (338-B88), 137 (1988)
80. N.M. Mustafaev, Approximate formulas for singular integrals and their application to the approximate solution of singular integral equations that are defined on closed smooth curve, Ph.D. thesis, AN Az. SSR, Institute of Math. and Mech., Baku., 1991
81. V.V. Panasyuk, M.P. Savruk, Z.T. Nazarchuk, *Singular Integral Equations Methods in Two-Dimensional Diffraction Problems* (Naukova Dumka, Kiev, 1984)
82. V.Z. Parton, P.I. Perlin, *Integral Equations of Elasticity Theory* (Nauka, Moscow, 1977)
83. G.Y. Popov, *Contact Problems for a Linearly Deformable Base* (Kiev, Odessa, 1982)
84. S. Prösdorf, B. Silberman, *Projektionsverfahren und die naherungsweise Losung Singularer* (Gleichungen, Leipziq, 1977)
85. S. Prösdorf, *Some Class Singular Integral Equations* (Mir, Moscow, 1979)
86. H. Reşidoğlu (Kh. amedov), On the planar problem of the theory elasticity. Works SSU **1**, 27–32 (2001)
87. M.H. Saleh, Basis of quadrature method for non-linear singular integral equations with Hilbert kernel in the space$H_{\varphi,k}$. In Az. NIINTI **279**, 1–40 (1984)
88. V.N. Seychuk, Direct methods of the solutions of singular integral equations that are defined on Lyapunov curve, Ph.D. thesis, Kishinev University, Kishinev, 1987
89. F.G. Trikomi, *Integral Equations* (Dover, New York, 1985)
90. W. Tutshke, Lözung nichtlinearer partieller differential-gleichungssysteme erster Ordnung in der Ebene cluch verwendung einer komlexen Normalform. Math. Nachr. **75**, 283–298 (1976)

91. W. Tzong-Mou, Solving the non-linear equations by the Mewton-homotopy continuation method with adjustable auxiliary homotopy functions. J. Appl. Math. Comput. **173**, 383–388 (2006)
92. F. Ursell, On the exterior problems of acoustics: II. Proc. Camb. Philol. Soc. **84**, 545–548 (1978)
93. G.M. Vaynicco, The regular convergence of operators and the approximate solution of equations, Moscow V sb. Math. Anal. (Itogi nauki i tekniki VINITI AN SSSR) **16**, 553 (1979) (in Russian)
94. I.N. Vekua, *Generalized Analytic Functions* (Pergamon Press, London, 1962)
95. V.A. Zolotaryevskii, On the Approximate solution of singular integral equations. Math. Res. Kishinev Shtiintsa **9**(3), 82–94 (1974)
96. V.A. Zolotaryevskii, V.N. Seychuk, The solution of the singular integral equation that is defined on Lyapunov curve by collocation method. Differ. Equ. **19**(6), 1056–1064 (1983)
97. A. Zygmund, A.P. Calderon, On the existence of singular integrals. Acta Math. **88**, 85–139 (1952)

Chapter 20
On Difference Double Sequences and Their Applications

L. Nayak and P. Baliarsingh

20.1 Introduction

Due to its numerous applications in the diverse fields of pure and applied sciences, recently the theory of difference single sequences has been attracted by several researchers. The applications of difference sequences become more apparent in linear algebra, approximation theory, and calculus in both classical and fractional cases. The idea of difference single sequence spaces based on order one has been introduced by Kızmaz [1] in the year 1981. Further, Et and Çolak [2] extended this idea to the case of an integral order in 1995. In order to stimulate its utility and applications, several extensions of this idea have been provided by many prominent authors. Quite recently, the notion of difference sequence spaces based on fractional order was provided by Baliarsingh [3] (see also [4–7]) and the idea was directly being used to study the fractional derivatives of certain functions and their geometrical interpretations.

As it is quite natural to extend the above idea to the case of double sequences, the primary aim of this work is to define certain related difference double sequence spaces and apply the idea in the study of approximations of partial derivatives based on both integer and fractional orders. Several prominent authors such as Gökhan and Çolak [8] studied certain paranormed double sequences and determined their alpha-, beta- , gamma-duals. Mursaleen [9] and Mursaleen and Edely [10] have studied the almost strong regularity of matrices for double sequences and Altay and Başar [11] have studied some double sequences involving sequence of partial sums of the double series. For more detail on the domain of difference sequence spaces and summability theory, one may refer to [12–28].

L. Nayak · P. Baliarsingh (✉)
Department of Mathematics, School of Applied Sciences, KIIT University, Bhubaneswar, India

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_20

20.2 Definitions

Let $\mathbb{N}_0$ be the set of all nonnegative integers. Then for a double sequence x of real or complex numbers, we write $x = (x_{m,n}),\ m, n \in \mathbb{N}_0$ whose elements are represented by an infinite two-dimensional matrix.

The double sequence $x = (x_{m,n})$ is said to be convergent in the *Pringsheim's* sense (or p-convergent) if for every $\epsilon > 0$, there exists a positive integer $N(\epsilon)$ and $L \in \mathbb{C}$, the complex field such that

$$|x_{m,n} - L| < \epsilon, \text{ for all } m, n \geq N_0(\epsilon).$$

More briefly, we write

$$p - \lim_{m,n\to\infty,\infty} x_{m,n} = L,$$

where m and n are tending to infinity independent of each other (see [12]). The limit L is called the double limit or *Pringsheim limit* of the sequence $x = (x_{m,n})$.

A double sequence $x = (x_{m,n})$ is called bounded if there exists a positive number M such that $|x_{m,n}| < M$ for all $m, n \in \mathbb{N}_0$, i.e.,

$$\|x\|_{(\infty,2)} = \sup_{m,n} |x_{m,n}| < \infty. \tag{20.1}$$

Let Ω be the space of all real double sequences of the form $(x_{m,n})$. By $\mathcal{M}_u, \mathcal{C}_p$, and $\mathcal{C}_{p0}$, we denote the spaces of all bounded, convergent, and null double sequences (in Pringsheim's sense), respectively. Then,

$$\mathcal{M}_u := \{x \in \Omega : \sup_{m,n} |x_{m,n}| < \infty\},$$

$$\mathcal{C}_p := \{x \in \Omega : p - \lim_{m,n} |x_{m,n} - l| = 0, \text{ for some } l \in \mathbb{C}\},$$

$$\mathcal{C}_{p0} := \{x \in \Omega : p - \lim_{m,n} |x_{m,n}| = 0\}.$$

In particular cases, it is known that $\mathcal{M}_u$ does not include $\mathcal{C}_p$; therefore, we write $\mathcal{C}_{bp}$ for the space of all double sequences which are bounded and convergent, and explicitly, we write $\mathcal{C}_{bp} = \mathcal{M}_u \cap \mathcal{C}_p$.

Let $A = (a^{ij}_{m,n}), m, n, i, j \in \mathbb{N}_0$, be a four-dimensional matrix and the A-transform of a double sequence $x = (x_{m,n})$ is denoted by $Ax = y$, provided the following double summation exists for all $i, j \in \mathbb{N}_0$:

$$(Ax)_{i,j} := y_{i,j} = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a^{ij}_{m,n} x_{m,n}.$$

A four-dimensional matrix $A = (a^{ij}_{m,n})$ is said to be conservative if $x \in \mathcal{C}_{bp}$ implies $Ax \in \mathcal{C}_p$ and we write $A \in (\mathcal{C}_{bp}, \mathcal{C}_p)$. Equivalently, A is conservative if and only if

$$p - \lim_{i,j} a^{ij}_{m,n} = z_{m,n} \quad \text{for each } m, n; \tag{20.2}$$

$$p - \lim_{i,j} \sum_m \sum_n a^{ij}_{m,n} = z; \tag{20.3}$$

$$p - \lim_{i,j} \sum_m |a^{ij}_{m,n}| = z_n \quad \text{for each } n; \tag{20.4}$$

$$p - \lim_{i,j} \sum_n |a^{ij}_{m,n}| = z_m \quad \text{for each } m; \tag{20.5}$$

$$p - \lim_{i,j} \sum_m \sum_n |a^{ij}_{m,n}| \text{ exists;} \tag{20.6}$$

$$\|A\| = \sup_{i,j} \sum_m \sum_n |a^{ij}_{m,n}| < \infty. \tag{20.7}$$

If A is conservative and $p - \lim Ax = p - \lim x$ for each $x \in \mathcal{C}_{bp}$, then A is called RH-regular (see [5, 6]). Also, it can be shown that A is RH-regular if and only if the conditions (20.2) with $z_{m,n} = 0$, (20.3) with $z = 1$, (20.4) and (20.5) with $z_n = z_m = 0$, and the conditions (20.6) and (20.7) hold.

The α-dual, $\beta(p)$-dual with respect to p-convergence, and γ-duals of the double sequence space X are, respectively, defined by

$$X^{\alpha} = \left\{ (a_{i,j}) \in \Omega : \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} |a_{i,j} x_{i,j}| < \infty, \text{ for all } (x_{m,n}) \in X \right\},$$

$$X^{\beta(p)} = \left\{ (a_{i,j}) \in \Omega : p - \lim \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{i,j} x_{i,j} \text{ exists, for all } (x_{m,n}) \in X \right\},$$

$$X^{\gamma} = \left\{ (a_{i,j}) \in \Omega : \sup_{m,n} \left| \sum_{i=0}^{m} \sum_{j=0}^{n} a_{i,j} x_{i,j} \right| < \infty, \text{ for all } (x_{m,n}) \in X \right\}.$$

For single sequences, X^{α}, X^{β}, and X^{γ} are called Köthe Toeplitz dual of X. For double sequences, -alpha and -gamma dual of X are unique, whereas -beta dual is different with respect to the type of convergence.

For a positive proper fraction $\bar{\alpha}$ and $x \in \Omega$, recently, the double difference sequence of fractional order $\bar{\alpha}$ has been defined by Baliarsingh [29] as

$$({}_2\Delta^{(\bar{\alpha})}x)_{m,n} = {}_2\Delta^{(\bar{\alpha})}x_{m,n} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{i+j}\frac{\Gamma(\bar{\alpha}+1)^2}{i!j!\Gamma(\bar{\alpha}-i+1)\Gamma(\bar{\alpha}-j+1)}x_{m-i,n-j}, \tag{20.8}$$

$$({}_2\Delta^{\bar{\alpha}}x)_{m,n} = {}_2\Delta^{\bar{\alpha}}x_{m,n} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{i+j}\frac{\Gamma(\bar{\alpha}+1)^2}{i!j!\Gamma(\bar{\alpha}-i+1)\Gamma(\bar{\alpha}-j+1)}x_{m+i,n+j}, \tag{20.9}$$

where $\Gamma(\bar{\alpha}+1)$ denotes *Euler gamma function* of a real number $\bar{\alpha}$, which can be defined by an improper integral

$$\Gamma(\bar{\alpha}) = \int_0^{\infty} e^{-t}t^{\bar{\alpha}-1}dt.$$

It is noted that $\bar{\alpha} \notin \{0, -1, -2, -3 \ldots\}$ and $\Gamma(\bar{\alpha}+1) = \bar{\alpha}\Gamma(\bar{\alpha})$.

The infinite series defined in (20.8) and (20.9) can be reduced to finite series if $\bar{\alpha}$ is a positive integer. An infinite series has no meaning unless it converges; therefore, throughout the text it is being presumed that the series (20.8) and (20.9) are convergent (in Pringsheim's sense), and $x_{i,j} = 0$ for any negative integers of i, j. It is well-known that convergent double series (in Pringsheim's sense) may not be bounded.

In particular, for $\bar{\alpha} = 1$, we have

$${}_2\Delta^1 x_{m,n} = x_{m,n} - x_{m,n+1} - x_{m+1,n} + x_{m+1,n+1}.$$

For any positive reals α, β and $h, k \in (0, 1]$, we define the generalized difference double sequence based on arbitrary orders as

$$({}_2\Delta_{h,k}^{(\alpha,\beta)}x)_{m,n} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-\alpha)_i(-\beta)_j}{i!j!h^{\alpha}k^{\beta}}x_{m-i,n-j};\ (m, n \in \mathbb{N}), \tag{20.10}$$

where $(\alpha)_k$ denotes the *Pochhammer* symbol or *shifted factorial* of a real number α which is being defined using familiar Euler gamma function as

$$(\alpha)_k = \begin{cases} 1, & (\alpha = 0 \ \text{ or } \ k = 0) \\ \dfrac{\Gamma(\alpha+k)}{\Gamma(\alpha)} = \alpha(\alpha+1)(\alpha+2)\ldots(\alpha+k-1), & (k \in \mathbb{N}). \end{cases}$$

It is noted that if $\alpha = \beta$, then the above difference sequence generalizes the sequence defined in (20.8). The related double difference operator ${}_2\Delta_{h,k}^{(\alpha,\beta)}$ includes the following known operators in different special cases:

Special cases of the operator ${}_2\Delta_{h,k}^{(\alpha,\beta)}$

α	β	h	k	Operators	cf.
1	1	1	1	${}_2\Delta^{(1)}$	[1, 30]
r	r	1	1	${}_2\Delta^{(r)}$	[31]
α	α	1	1	${}_2\Delta^{(\alpha)}$	[29]
1	0	1	1	Δ	[14]
m	0	1	1	$\Delta^{(m)}$	[2, 32]
α	0	1	1	$\Delta^{(\alpha)}$	[3]
α	0	h	1	$\Delta_h^{a,b,b}$	[7]

Now we discuss the convergence of the double sequence defined in (20.8) by providing the following numerical examples:

Example Suppose the double sequence $x = (x_{m,n})$ is defined by $x_{m,n} = m + n$, then it is trivial to calculate that

$$
\begin{aligned}
({}_2\Delta_{h,k}^{(\alpha,\beta)}x)_{m,n} &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-\alpha)_i(-\beta)_j}{i!j!h^\alpha k^\beta}(m-i+n-j)\\
&= (m+n)\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-\alpha)_i(-\beta)_j}{i!j!h^\alpha k^\beta}1-\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-\alpha)_i(-\beta)_j}{i!j!h^\alpha k^\beta}j\\
&\quad -\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-\alpha)_i(-\beta)_j}{i!j!h^\alpha k^\beta}i = \frac{(m+n)}{h^\alpha k^\beta}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-\alpha)_i(-\beta)_j}{i!j!}\\
&\quad +\frac{\beta}{h^\alpha k^\beta}\sum_{i=0}^{\infty}\sum_{j=1}^{\infty}\frac{(-\alpha)_i(-\beta+1)_j}{i!(j-1)!}\\
&\quad +\frac{\alpha}{h^\alpha k^\beta}\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}\frac{(-\alpha+1)_i(-\beta)_j}{(i-1)!j!}\\
&\quad \to 0, \text{ as } m,n\to\infty \text{ and} \alpha,\beta\geq 1.
\end{aligned}
$$

It has been observed that the difference double sequence ${}_2\Delta_{h,k}^{(\alpha,\beta)}x$ is convergent (in Pringsheim's sense) if the primary double sequence $x = (x_{m,n})$ is of order $(m+n)^\gamma$, where $\gamma \leq \max(1, \alpha+\beta)$. For details we have the following examples:

Example Consider a double sequence $x = (x_{m,n})$, defined by $x_{m,n} = (m+n)^3$, and choose $\alpha = 2$ and $\beta = 3$. Then it is clear that $3 = \gamma \le \max(1, \alpha + \beta) = 5$ and

$$\begin{aligned}({}_2\Delta_{h,k}^{(2,3)}x)_{m,n} &= \sum_{i=0}^{2}\sum_{j=0}^{3}\frac{(-2)_i(-3)_j}{i!j!h^2k^3}((m-i+n-j)^3)\\ &= \sum_{i=0}^{2}\frac{(-2)_i}{i!h^2k^3}((m-i+n)^3 - 3(m-i+n-1)^3\\ &\quad + 3(m-i+n-2)^3 - (m-i+n-3)^3\\ &= \sum_{i=0}^{2}\frac{(-2)_i}{i!h^2k^3}6 \to 0, \text{ as } m,n \to \infty.\end{aligned}$$

On the contrary, if we take $\alpha = 1/2$ and $\beta = 2$, then $3 = \gamma > \max(1, \alpha+\beta) = 5/2$ and observe that

$$\begin{aligned}({}_2\Delta_{h,k}^{(1/2,2)}x)_{m,n} &= \sum_{i=0}^{\infty}\sum_{j=0}^{2}\frac{(-1/2)_i(-2)_j}{i!j!h^{1/2}k^2}((m-i+n-j)^3)\\ &= \sum_{i=0}^{\infty}\frac{(-1/2)_i}{i!h^{1/2}k^2}((m-i+n)^3 - 2(m-i+n-1)^3\\ &\quad + (m-i+n-2)^3)\\ &= 6\sum_{i=0}^{\infty}\frac{(-1/2)_i}{i!h^{1/2}k^2}(m+n-i-1)\\ &= 6(m+n)\sum_{i=0}^{\infty}\frac{(-1/2)_i}{i!h^{1/2}k^2}1 - 6\sum_{i=0}^{\infty}\frac{(-1/2)_i}{i!h^{1/2}k^2}i - 6\sum_{i=0}^{\infty}\frac{(-1/2)_i}{i!h^{1/2}k^2}1\\ &= 3\sum_{i=1}^{\infty}\frac{(1/2)_i}{(i-1)!h^{1/2}k^2} \to \infty, \text{ as } m,n \to \infty.\end{aligned}$$

Let $\alpha = r_1$ and $\beta = r_2$ be two nonnegative integers. Then Eq. (20.10) can be rewritten as

$$\begin{aligned}({}_2\Delta_{h,k}^{(r_1,r_2)}x)_{m,n} &= \sum_{i=0}^{r_1}\sum_{j=0}^{r_2}\frac{(-r_1)_i(-r_2)_j}{i!j!h^{r_1}k^{r_2}}x_{m-i,n-j}\\ &= \frac{1}{h^{r_1}k^{r_2}}\sum_{i=0}^{r_1}\sum_{j=0}^{r_2}\binom{r_1}{i}\binom{r_2}{j}x_{m-i,n-j};\ (m,n \in \mathbb{N}).\end{aligned} \tag{20.11}$$

The double sequence based on inverse of double difference operator used in (20.10) is given by

$$({}_2\Delta_{h,k}^{(-\alpha,-\beta)}x)_{m,n} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(\alpha)_i(\beta)_j}{i!j!h^{-\alpha}k^{-\beta}}x_{m-i,n-j};\ (m,n\in\mathbb{N}). \tag{20.12}$$

Now, we quote some results related to the double difference operator ${}_2\Delta_{h,k}^{(\alpha,\beta)}$.

Theorem 20.2.1 *If α and β are two nonnegative reals and $h,k\to 1$, then the double difference operators ${}_2\Delta_{h,k}^{(\alpha,\beta)}:\Omega\to\Omega$ is a linear operator over the field $\mathbb{R}$ and $\|{}_2\Delta_{h,k}^{(\alpha,\beta)}\| = \frac{2^{\alpha+\beta}}{h^\alpha k^\beta}$, where $\|A\|$ is the supremum over ℓ_1 norms of the rows of the matrix A.*

Proof Proof is trivial, hence omitted. □

Theorem 20.2.2 *Let r_1, r_2, p_1, and p_2 be any positive integers and $x=(x_{mn})\in\mathcal{C}_{bp}$. Then*

(i) $({}_2\Delta_{h,k}^{(r_1,r_2)}\ {}_2\Delta_{h,k}^{(p_1,p_2)}x)_{m,n}=({}_2\Delta_{h,k}^{(r_1,r_2)}\ {}_2\Delta_{h,k}^{(p_1,p_2)}x)_{m,n} = ({}_2\Delta_{h,k}^{(r_1+p_1,r_1+p_2)}x)_{m,n}$,
(ii) $({}_2\Delta_{h,k}^{(r_1,r_2)}\ {}_2\Delta_{h,k}^{(-r_1,-r_2)}x)_{m,n}=({}_2\Delta_{h,k}^{(-r_1,-r_2)}\ {}_2\Delta_{h,k}^{(r_1,r_2)}x)_{m,n} = x_{m,n}$.

Proof Proof is a straightforward calculation, hence omitted. □

20.3 Related Difference Double Sequence Spaces

In the present section, using (20.11), we define the related difference double sequence spaces based on the difference operator ${}_2\Delta_{h,k}^{(r_1,r_2)}$. Let r_1, r_2 be any two positive integers and $h,k\to 1$. Then we define

$$\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)}) = \{(x_{m,n})\in\Omega : \sup_{m,n}|({}_2\Delta_{h,k}^{(r_1,r_2)}x)_{m,n}|<\infty\},$$

$$c^2({}_2\Delta_{h,k}^{(r_1,r_2)}) = \{(x_{m,n})\in\Omega : p-\lim_{m,n}|({}_2\Delta_{h,k}^{(r_1,r_2)}x)_{m,n}-l|=0,\ \text{for some } l\in\mathbb{R}\},$$

$$c_0^2({}_2\Delta_{h,k}^{(r_1,r_2)}) = \{(x_{m,n})\in\Omega : p-\lim_{m,n}|({}_2\Delta_{h,k}^{(r_1,r_2)}x)_{m,n})|=0\}$$

and

$$c_b^2({}_2\Delta_{h,k}^{(r_1,r_2)}) = c^2({}_2\Delta_{h,k}^{(r_1,r_2)})\cap\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)}).$$

It is observed that the above double sequence spaces are being derived by taking ${}_2\Delta_{h,k}^{(r_1,r_2)}$-transform of the double sequence $x = (x_{m,n})$, i.e.,

$$y_{m,n} = ({}_2\Delta_{h,k}^{(r_1,r_2)} x)_{m,n} = \sum_{i=m-r_1}^{m} \sum_{j=n-r_2}^{n} \frac{(-1)^{m+n-(i+j)}}{h^{r_1} k^{r_2}} \binom{r_1}{m-i} \binom{r_2}{n-j} x_{i,j}$$

$$= \sum_{i=m-r_1}^{m} \sum_{j=n-r_2}^{n} ((\delta_{h,k}^{r_1,r_2})_{m,n})_{j,i} x_{i,j},$$

where ${}_2\Delta_{h,k}^{(r_1,r_2)} = ((\delta_{h,k}^{r_1,r_2})_{m,n})_{j,i}$ represents a four-dimensional matrix, defined by

$$((\delta_{h,k}^{r_1,r_2})_{m,n})_{j,i} = \begin{cases} \frac{(-1)^{m+n-(i+j)}}{h^{r_1} k^{r_2}} \binom{r_1}{m-i} \binom{r_2}{n-j}, & (m - r_1 \le j \le m), (n - r_2 \le i \le n) \\ 0, & (\text{ otherwise}). \end{cases}$$

In fact, each element of the four-dimensional matrix $((\delta_{h,k}^{r_1,r_2})_{m,n})$ is being expressed by the corresponding two-dimensional matrix $(\delta_{h,k}^{r_1,r_2})_{m,n}$, $(m, n \in \mathbb{N}_0)$, namely

$$(\delta_{h,k}^{r_1,r_2})_{0,0} = \frac{1}{h^{r_1} k^{r_2}} \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, (\delta_{h,k}^{r_1,r_2})_{1,0} = \frac{1}{h^{r_1} k^{r_2}} \begin{pmatrix} -r_1 & 1 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

$$(\delta_{h,k}^{r_1,r_2})_{1,1} = \frac{1}{h^{r_1} k^{r_2}} \begin{pmatrix} r_1 r_2 & -r_2 & 0 & \dots \\ -r_1 & 1 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, (\delta_{h,k}^{r_1,r_2})_{0,1} = \frac{1}{h^{r_1} k^{r_2}} \begin{pmatrix} -r_2 & 0 & 0 & \dots \\ 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

$$(\delta_{h,k}^{r_1,r_2})_{2,0} = \frac{1}{h^{r_1} k^{r_2}} \begin{pmatrix} \frac{r_1(r_1-1)}{2} & -r_1 & 1 & \dots \\ 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

$$(\delta_{h,k}^{r_1,r_2})_{0,2} = \frac{1}{h^{r_1} k^{r_2}} \begin{pmatrix} \frac{r_2(r_2-1)}{2} & 0 & 0 & \dots \\ -r_2 & 0 & 0 & \dots \\ 1 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

$$(\delta_{h,k}^{r_1,r_2})_{2,1} = \frac{1}{h^{r_1}k^{r_2}} \begin{pmatrix} \frac{-r_2r_1(r_1-1)}{2} & r_2r_1 & -r_2 & \dots \\ \frac{r_1(r_1-1)}{2} & -r_1 & 1 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

$$(\delta_{h,k}^{r_1,r_2})_{2,2} = \frac{1}{h^{r_1}k^{r_2}} \begin{pmatrix} \binom{r_1}{2}\binom{r_2}{2} & -r_1\binom{r_2}{2} & \binom{r_2}{2} & \dots \\ -r_2\binom{r_1}{2} & r_2r_1 & -r_2 & \dots \\ \binom{r_1}{2} & -r_1 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Theorem 20.3.1 *The sets* $\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$, $c_0^2({}_2\Delta_{h,k}^{(r_1,r_2)})$, *and* $c_b^2({}_2\Delta_{h,k}^{(r_1,r_2)})$ *are complete normed linear spaces with the norm defined by*

$$\|x\|_{{}_2\Delta_{h,k}^{(r_1,r_2)}} = \sum_{i=0}^{r_1}\sum_{j=0}^{r_2}|x_{i,j}| + \sup_{m,n}\left|({}_2\Delta_{h,k}^{(r_1,r_2)}x)_{m,n}\right|. \tag{20.13}$$

Proof It is noted that the expression defined in (20.13) without the first term is not a norm. This is due to the fact that if $\sup_{m,n}\left|({}_2\Delta_{h,k}^{(r_1,r_2)}x)_{m,n}\right| = 0$, then it does not imply that $x = \theta$, where $\theta = (\theta_{m,n})$ with $\theta_{m,n} = 0$ for all $m, n \in \mathbb{N}_0$. In fact, the first term suggests that $\|x\|_{{}_2\Delta_{h,k}^{(r_1,r_2)}} = 0$ if and only if $x = \theta$. □

Theorem 20.3.2 *If* r_1 *and* r_2 *are two positive integers, then*

(i) $\mathcal{M}_u \subseteq \ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$,
(ii) $\mathcal{C}_{p0} \subseteq c_0^2({}_2\Delta_{h,k}^{(r_1,r_2)})$,
(iii) $\mathcal{C}_{bp} \subseteq c_b^2({}_2\Delta_{h,k}^{(r_1,r_2)})$,
(iv) $c_0^2({}_2\Delta_{h,k}^{(r_1,r_2)}) \subseteq \ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)}) \subseteq c_b^2({}_2\Delta_{h,k}^{(r_1,r_2)})$.

Proof We provide the proof of (i) only and those of others may require similar arguments. Let $x \in \mathcal{M}_u$, then there exists a constant M, such that $\sup_{m,n}\left|x_{m,n}\right| \leq M$. Therefore, one may directly deduce that

$$\begin{aligned} &\left|\sum_{i=m-r_1}^{m}\sum_{j=n-r_2}^{n}\frac{(-1)^{m+n-(i+j)}}{h^{r_1}k^{r_2}}\binom{r_1}{m-i}\binom{r_2}{n-j}x_{i,j}\right| \\ &\leq \frac{1}{h^{r_1}k^{r_2}}\sum_{i=m-r_1}^{m}\sum_{j=n-r_2}^{n}\left|\binom{r_1}{m-i}\binom{r_2}{n-j}x_{i,j}\right| \\ &\leq \frac{M}{h^{r_1}k^{r_2}}\sum_{i=m-r_1}^{m}\sum_{j=n-r_2}^{n}\left|\binom{r_1}{m-i}\binom{r_2}{n-j}\right| \\ &\leq \frac{M}{h^{r_1}k^{r_2}}2^{r_1+r_2} < \infty. \end{aligned}$$

Thus, $\sup_{m,n} \left|{}_2\Delta^{(r_1,r_2)} x)_{m,n}\right| < \infty$,

This implies that $\mathcal{M}_u \subset \ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$ and equality holds if $r_1 = r_2 = 0$. □

Theorem 20.3.3 *If* $e^2 = (e^2_{m,n})$*, with* $e_{m,n} = 1$ *for all* $m, n \in \mathbb{N}_0$ *and* $\sup_{m,n} |({}_2\Delta^{(-r_1,-r_2)} e^2)_{m,n}| < \infty$*, then*

(i) $\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)}) \subseteq \mathcal{M}_u$,
(ii) $c_0^2({}_2\Delta_{h,k}^{(r_1,r_2)}) \subseteq \mathcal{C}_{p0}$,
(iii) $c_b^2({}_2\Delta_{h,k}^{(r_1,r_2)}) \subseteq \mathcal{C}_{bp}$.

Proof We prove the theorem for the space $\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$. Let $x \in \ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$, then there exists a constant K, such that

$$\sup_{m,n} \left| \sum_{i=m-r_1}^{m} \sum_{j=n-r_2}^{n} \frac{(-1)^{m+n-(i+j)}}{h^{r_1} k^{r_2}} \binom{r_1}{m-i} \binom{r_2}{n-j} x_{i,j} \right| \leq K.$$

But, it is known that

$$\begin{aligned} |x_{m,n}| &= |{}_2\Delta^{(-r_1,-r_2)}(({}_2\Delta_{h,k}^{(r_1,r_2)} x)_{m,n})| \\ &= \left| {}_2\Delta^{(-r_1,-r_2)} \left(\sum_{i=m-r_1}^{m} \sum_{j=n-r_2}^{n} \frac{(-1)^{m+n-(i+j)}}{h^{r_1} k^{r_2}} \binom{r_1}{m-i} \binom{r_2}{n-j} x_{i,j} \right) \right| \\ &\leq K \sup_{m,n} \left| \left({}_2\Delta^{(-r_1,-r_2)} e^2 \right)_{m,n} \right| < \infty, \end{aligned}$$

Therefore, $\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)}) \subset \mathcal{M}_u$, and equality holds for $r_1 = r_2 = 0$. This concludes the proof. □

Theorem 20.3.4 *The sets* $\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$, $c_0^2({}_2\Delta_{h,k}^{(r_1,r_2)})$*, and* $c_b^2({}_2\Delta_{h,k}^{(r_1,r_2)})$ *are linear isomorphic to the spaces* $\mathcal{M}_u, \mathcal{C}_{p0}$*, and* $\mathcal{C}_{bp}$*, respectively.*

Proof We prove the theorem for the space $\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$. We show that there exists a linear bijection between the spaces $\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$ and $\mathcal{M}_u$. Using the notation (20.10), consider the mapping $T : \ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)}) \to \mathcal{M}_u$ defined by $x \mapsto y = Tx$. Then, clearly T is linear and injective. Let $y \in \mathcal{M}_u$ and define a sequence $x = (x_{m,n})$ via $y_{m,n}$ as

$$x_{m,n} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(r_1)_i (r_2)_j}{i! j! h^{-r_1} k^{-r_2}} y_{m-i,n-j}. \tag{20.14}$$

Then, we have

$$\sup_{m,n} |({}_2\Delta_{h,k}^{(r_1,r_2)} x)_{m,n})| = \sup_{m,n} \left|y_{m,n}\right| < \infty.$$

Thus, we obtain that $x \in \ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$ and therefore T is surjective. This completes the proof. □

Now, we discuss the α-dual, $\beta(p)$-dual with respect to p-convergence, and γ-duals of the proposed difference double sequence spaces.

Theorem 20.3.5 *The alpha dual of the space $\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$ is the set D_1, defined by*

$$D_1 = \left\{(a_{m,n}) \in \Omega : \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\left|\sum_{l=m}^{\infty}\sum_{k=n}^{\infty}\frac{(r_1)_{m-l}(r_2)_{n-k}}{(m-k)!(m-n)!h^{-r_1}k^{-r_2}}a_{m,n}\right| < \infty\right\}.$$

Proof Suppose D is the alpha dual of the space $\ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$. Firstly, we show that $D_1 \subset D$. Let $x = (x_{m,n}) \in \ell_\infty^2({}_2\Delta_{h,k}^{(r_1,r_2)})$ and $y = (y_{m,n}) \in \mathcal{M}_u$, then there exists a positive integer $B > 1$ such that

$$|y_{m,n}| < \max\left(1, \sup_{m,n}|y_{m,n}|\right) = B < \infty.$$

Suppose $a = (a_{m,n}) \in D_1$, then we have

$$\begin{aligned}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}|a_{m,n}x_{m,n}| &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\left|\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(r_1)_i(r_2)_j}{i!j!h^{-r_1}k^{-r_2}}y_{m-i,n-j}a_{m,n}\right| \\ &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\left|\sum_{l=m}^{\infty}\sum_{k=n}^{\infty}\frac{(r_1)_{m-l}(r_2)_{n-k}}{(m-l)!(n-k)!h^{-r_1}k^{-r_2}}y_{l,k}a_{m,n}\right| \\ &\le B\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\left|\sum_{l=m}^{\infty}\sum_{k=n}^{\infty}\frac{(r_1)_{m-l}(r_2)_{n-k}}{(m-l)!(n-k)!h^{-r_1}k^{-r_2}}a_{m,n}\right|.\end{aligned}$$

From the above inequality it is concluded that if $a = (a_{m,n}) \in D_1$ whenever $y \in \mathcal{M}_u$, then $a \in D$ which implies that $D_1 \subset D$.

Conversely, suppose that $a \in D$ and $a \notin D_1$, then it is very easy to show that for $y \in \mathcal{M}_u$,

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}|a_{m,n}x_{m,n}| = \infty.$$

This concludes that $a \notin D$, which is contradiction. □

Theorem 20.3.6 *The beta dual with respect to p-convergent and gamma dual of the space* $\ell^2_\infty({}_2\Delta^{(r_1,r_2)}_{h,k})$ *is the set* D_1.

Proof Proof follows from Theorem 20.3.5. □

Theorem 20.3.7 *The alpha, beta dual with respect to p-convergent and gamma dual of the space* $c^2_b({}_2\Delta^{(r_1,r_2)}_{h,k})$ *are the set* $D_1 \cap \mathcal{C}_{bp}$.

Proof Proof is similar to that of Theorem 20.3.5. □

Theorem 20.3.8 *Let* $A = (a^{ij}_{m,n})$ *be a four-dimensional matrix.*
Then $A \in (c^2_b({}_2\Delta^{(r_1,r_2)}_{h,k}, \mathcal{C}_p)$ *if and only if*

$$\sup_{m,n}\left(\sum_{i,j}\sum_{r,s=0,0}^{i,j}\sum_{p,q=0,0}^{r,s}|{}_2\Delta^1((a_{m,n})_{i,j})\delta(r_1,r_2,h,k,p,q,r,s)|\right) < \infty,$$

$$\lim_{j\to\infty}\left(\sum_{r,s=0,0}^{i,j}\sum_{p,q=0,0}^{r,s}\Delta^1_i((a_{m,n})_{i,j})\delta(r_1,r_2,h,k,p,q,r,s)\right) = 0, \text{ for fixed } i \in \mathbb{N}_0,$$

$$\lim_{i\to\infty}\left(\sum_{r,s=0,0}^{i,j}\sum_{p,q=0,0}^{r,s}\Delta^1_j((a_{m,n})_{i,j})\delta(r_1,r_2,h,k,p,q,r,s)\right) = 0, \text{ for fixed } j \in \mathbb{N}_0,$$

$$p-\lim_{m,n\to\infty,\infty}\left(\sum_{r,s=0,0}^{i,j}\sum_{p,q=0,0}^{r,s}{}_2\Delta^1((a_{m,n})_{i,j})\delta(r_1,r_2,h,k,p,q,r,s)\right) = c_{i,j},$$
$$\text{for all } i,j \in \mathbb{N}_0,$$

$$p-\lim_{m,n\to\infty,\infty}\left(\sum_{i,j}\sum_{r,s=0,0}^{i,j}\sum_{p,q=0,0}^{r,s}|{}_2\Delta^1((a_{m,n})_{i,j})\delta(r_1,r_2,h,k,p,q,r,s)|\right) = c,$$

$$p-\lim_{m,n\to\infty,\infty}\left(\sum_{i}\sum_{r,s=0,0}^{i,j}\sum_{p,q=0,0}^{r,s}|{}_2\Delta^1((a_{m,n})_{i,j})\delta(r_1,r_2,h,k,p,q,r,s)|\right) = \sum_j |c_{i,j}|,$$

where

$$\Delta^1_i((a_{m,n})_{i,l}) = a_{m,n,i,l} - a_{m,n,i+1,l}$$
$$\Delta^1_j((a_{m,n})_{k,j}) = a_{m,n,k,j} - a_{m,n,k,j+1}$$
$$\delta(r_1,r_2,h,k,p,q,r,s) = \frac{(r_1)_{r-p}(r_2)_{s-q}}{(r-p)!(s-q)!h^{-r_1}k^{-r_2}}.$$

Proof Suppose that $x \in c_b^2({}_2\Delta_{h,k}^{(r_1,r_2)})$ and define the sequence $y = (y_{m,n})$ by

$$z_{m,n} = \sum_{i,j=0,0}^{m,n} x_{i,j} \ (m, n \in \mathbb{N}_0).$$

Now, taking (k, l)th partial sums of the series

$$(Ax)_{m,n} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_{m,n,i,j}x_{i,j},$$

for each $m, n \in \mathbb{N}_0$, and applying Abel transformation, it becomes

$$\begin{aligned}
(Ax)_{m,n}^{(k,l)} &= \sum_{i,j=0,0}^{k,l} a_{m,n,i,j}x_{i,j} \\
&= \sum_{i,j=0,0}^{k-1,l-1} {}_2\Delta^1((a_{m,n})_{i,j})z_{i,j} + \sum_{i=0}^{k-1} \Delta_i^1((a_{m,n})_{i,l})z_{i,l} \\
&+ \sum_{j=0}^{l-1} \Delta_j^1((a_{m,n})_{k,j})z_{k,j} + a_{m,n,k,l}z_{k,l} \\
&= \sum_{i,j=0,0}^{k-1,l-1} \sum_{r,s=0,0}^{i,j} \sum_{p,q=0,0}^{r,s} {}_2\Delta^1((a_{m,n})_{i,j})\delta(r_1, r_2, h, k, p, q, r, s)y_{p,q} \\
&+ \sum_{i=0}^{k-1} \sum_{r,s=0,0}^{i,l} \sum_{p,q=0,0}^{r,s} \Delta_i^1((a_{m,n})_{i,l})\delta(r_1, r_2, h, k, p, q, r, s)y_{p,q} \\
&+ \sum_{j=0}^{l-1} \sum_{r,s=0,0}^{k,j} \sum_{p,q=0,0}^{r,s} \Delta_j^1((a_{m,n})_{k,j})\delta(r_1, r_2, h, k, p, q, r, s)y_{p,q} \\
&+ a_{m,n,k,l} \sum_{r,s=0,0}^{k,l} \sum_{p,q=0,0}^{r,s} \Delta_i^1((a_{m,n})_{k,l})\delta(r_1, r_2, h, k, p, q, r, s)y_{p,q},
\end{aligned}$$

where

$$\Delta_i^1((a_{m,n})_{i,l}) = a_{m,n,i,l} - a_{m,n,i+1,l},$$

$$\Delta_j^1((a_{m,n})_{k,j}) = a_{m,n,k,j} - a_{m,n,k,j+1}$$

and

$$\delta(r_1, r_2, h, k, p, q, r, s) = \frac{(r_1)_{r-p}(r_2)_{s-q}}{(r-p)!(s-q)!h^{-r_1}k^{-r_2}}.$$

The (k,l)th partial sums of the series $(Ax)_{m,n}^{(k,l)}$ are being expressed as a four-dimensional matrix transformation of the double sequence $y = (y_{m,n})$, i.e.,

$$(Ax)_{m,n}^{(k,l)} = (Ty)_{m,n,k,l}, \tag{20.15}$$

where $T = (t_{k,l,i,j}^{m,n})$ with

$$t_{k,l,i,j}^{m,n} = \begin{cases} \sum_{r,s=0,0}^{i,j} \sum_{p,q=0,0}^{r,s} {}_2\Delta^1((a_{m,n})_{i,j})\delta(r_1, r_2, h, k, p, q, r, s), & (i \le k-1 \text{ and } j \le l-1), \\ \sum_{r,s=0,0}^{i,j} \sum_{p,q=0,0}^{r,s} \Delta_i^1((a_{m,n})_{i,j})\delta(r_1, r_2, h, k, p, q, r, s), & (i = k \text{ and } j \le l-1), \\ \sum_{r,s=0,0}^{i,j} \sum_{p,q=0,0}^{r,s} \Delta_j^1((a_{m,n})_{i,j})\delta(r_1, r_2, h, k, p, q, r, s), & (i \le k-1 \text{ and } j = l), \\ \sum_{r,s=0,0}^{i,j} \sum_{p,q=0,0}^{r,s} ((a_{m,n})_{k,l})\delta(r_1, r_2, h, k, p, q, r, s), & (i = k \text{ and } j = l), \\ 0, & (\text{otherwise}). \end{cases}$$

Therefore, it is concluded by (20.15) that $m, n \in \mathbb{N}_0$, the (k,l)th partial sums of the series $(Ax)_{m,n}^{(k,l)}$ converge (in the Pringsheim's sense) if and only if Ty is bounded and convergent (in the Pringsheim's sense) which is equivalent to the following conditions:

$$\sup_{m,n} \left(\sum_{i,j} \sum_{r,s=0,0}^{i,j} \sum_{p,q=0,0}^{r,s} |{}_2\Delta^1((a_{m,n})_{i,j})\delta(r_1, r_2, h, k, p, q, r, s)| \right) < \infty, \tag{20.16}$$

$$\lim_{j\to\infty} \left(\sum_{r,s=0,0}^{i,j} \sum_{p,q=0,0}^{r,s} \Delta_i^1((a_{m,n})_{i,j})\delta(r_1, r_2, h, k, p, q, r, s) \right) = 0, \text{ for fixed } i \in \mathbb{N}_0, \tag{20.17}$$

$$\lim_{i\to\infty} \left(\sum_{r,s=0,0}^{i,j} \sum_{p,q=0,0}^{r,s} \Delta_j^1((a_{m,n})_{i,j})\delta(r_1, r_2, h, k, p, q, r, s) \right) = 0, \text{ for fixed } j \in \mathbb{N}_0, \tag{20.18}$$

$$p - \lim_{m,n\to\infty,\infty} \left(\sum_{r,s=0,0}^{i,j} \sum_{p,q=0,0}^{r,s} {}_2\Delta^1((a_{m,n})_{i,j})\delta(r_1, r_2, h, k, p, q, r, s) \right) = c_{i,j}, \forall i, j \in \mathbb{N}_0, \tag{20.19}$$

$$p - \lim_{m,n\to\infty,\infty}\left(\sum_{i,j}\sum_{r,s=0,0}^{i,j}\sum_{p,q=0,0}^{r,s}|{}_2\Delta^1((a_{m,n})_{i,j})\delta(r_1,r_2,h,k,p,q,r,s)|\right) = c, \tag{20.20}$$

$$p - \lim_{m,n\to\infty,\infty}\left(\sum_{i}\sum_{r,s=0,0}^{i,j}\sum_{p,q=0,0}^{r,s}|{}_2\Delta^1((a_{m,n})_{i,j})\delta(r_1,r_2,h,k,p,q,r,s)|\right) = \sum_j |c_{i,j}|. \tag{20.21}$$

Combining the above equations together, we complete the proof. □

20.4 Applications

In this section, we provide some applications of the sequence $({}_2\Delta_{h,k}^{(r_1,r_2)}x)$ in approximating the ordinary differential operator and partial differential operators.

Let $f = f(x, y)$ be a differentiable function. Let r_1, r_2 be any positive integers and h, k be two positive constants tending to zero. Then we define the following sequence of function f via double difference operator ${}_2\Delta_{h,k}^{(r_1,r_2)}$ as

$${}_2\Delta_{h,k}^{(r_1,r_2)} f(x, y) = \sum_{i=0}^{r_1}\sum_{j=0}^{r_2}\frac{(-r_1)_i(-r_2)_j}{i!j!h^{r_1}k^{r_2}} f(x - ih, y - jk). \tag{20.22}$$

In particular, for different suitable values of r_1, r_2 and different functions $f(x, y)$, it has been observed that Eq. (20.22) reduces to the following cases of various derivatives:

1. If $r_2 = 0$ and $f(x, y)$ is independent of y, then we have

$${}_2\Delta_{h,k}^{(r_1,0)} f(x) = \sum_{i=0}^{r_1}\frac{(-r_1)_i}{i!h^{r_1}} f(x - ih) = \frac{d^{r_1} f}{dx^{r_1}}.$$

2. If $r_1 = 0$ and $f(x, y)$ is independent of x, then we have

$${}_2\Delta_{h,k}^{(0,r_2)} f(y) = \sum_{j=0}^{r_2}\frac{(-r_2)_j}{j!k^{r_2}} f(y - jh) = \frac{d^{r_2} f}{dy^{r_2}}.$$

3. If $r_2 = 0$ and f is a function with two independent variables x and y, then we have

$$_2\Delta_{h,k}^{(r_1,0)} f(x, y) = \sum_{i=0}^{r_1} \frac{(-r_1)_i}{i!h^{r_i}} f(x - ih, y) = \frac{\partial^{r_1} f}{\partial x^{r_1}}.$$

4. If $r_1 = 0$ and f is a function with two independent variables x and y, then we have

$$_2\Delta_{h,k}^{(0,r_2)} f(x, y) = \sum_{j=0}^{r_2} \frac{(-r_2)_j}{j!k^{r_2}} f(x, y - jh) = \frac{\partial^{r_2} f}{\partial y^{r_2}}.$$

5. If r_1, r_2 are any two positive integers and f is a function with two independent variables x and y, then we have

$$_2\Delta_{h,k}^{(r_1,r_2)} f(x, y) = \sum_{i=0}^{r_1} \sum_{j=0}^{r_2} \frac{(-r_1)_i(-r_2)_j}{i!j!h^{r_1}k^{r_2}} f(x - ih, y - jk) = \frac{\partial^{r_1+r_2} f}{\partial x^{r_1}\partial y^{r_2}}.$$

Now, using formula defined in Eq. (20.22), we illustrate some examples for finding partial derivatives of certain functions.

Example Consider a real valued function $f = f(x, y)$, defined by $f(x, y) = \frac{1}{\sqrt{x^2+y^2}}$. Then for $h, k \to 0$ and $r_1 = 2$, and $r_2 = 0$, Eq. (20.22) reduces to

$$\begin{aligned}
&_2\Delta_{h,k}^{(2,0)} f(x, y)\\
&= \sum_{i=0}^{2} \frac{(-2)_i}{i!h^2} f(x - ih, y - jk)\\
&= \frac{1}{h^2}\left[(x^2 + y^2)^{-1/2} - 2((x - h)^2 + y^2)^{-1/2}) + ((x - 2h)^2 + y^2)^{-1/2}\right]\\
&= \frac{1}{h}\left[\frac{(x^2 + y^2)^{-1/2} - ((x - h)^2 + y^2)^{-1/2}) - ((x - h)^2 + y^2)^{-1/2}) + ((x - 2h)^2 + y^2)^{-1/2}}{h}\right]\\
&= \frac{1}{h}\left[-\frac{x}{(x^2 + y^2)^{3/2}} + \frac{x - h}{((x - h)^2 + y^2)^{3/2}}\right],\\
&= -\frac{1}{((x - h)^2 + y^2)^{3/2}} + \frac{-\frac{x}{(x^2+y^2)^{3/2}} + \frac{x-h}{((x-h)^2+y^2)^{3/2}}}{h},\\
&= -\frac{1}{(x^2 + y^2)^{3/2}} + \frac{3x^2}{(x^2 + y^2)^{5/2}}, \text{(as } h \to 0).
\end{aligned}$$

Example Let us consider a real valued function $f(x, y) = \sin xy + x + y$. Then taking $h, k \to 0$ in Eq. (20.22), it can be easily seen that

$$
\begin{aligned}
&{}_2\Delta_{h,k}^{(2,1)} f(x, y) \\
&= \sum_{i=0}^{2}\sum_{j=0}^{1} \frac{(-2)_i(-1)_j}{i!j!h^2k} f(x - ih, y - jk) \\
&= \frac{1}{h^2k}[\sin xy - 2\sin(x-h)y + \sin(x-2h)y - \sin x(y-k) \\
&\quad - 2\sin(x-h)(y-k) + \sin(x-2h)(y-k) + x + y - 2(x-h) \\
&\quad - 2y + x - 2h + y - x - y + k + 2(x-h) + 2(y-k) - (x-2h) - (y-k)] \\
&= \frac{1}{h^2k}[\sin xy - 2\sin(x-h)y + \sin(x-2h)y - \sin x(y-k) \\
&\quad - 2\sin(x-h)(y-k) + \sin(x-2h)(y-k)] \\
&= \frac{1}{h}\left[\frac{x\cos xy + 2(h-x)\cos(x-h)y + (x-2h)\cos(2x-h)y}{h}\right], \text{(as } k \to 0) \\
&= \frac{x}{h}\left[\frac{\cos xy - 2\cos(x-h)y + \cos(2x-h)y}{h}\right] \\
&= \frac{x}{h}\left[\frac{2\sin\frac{(2x-h)y}{2}\sin\left(-\frac{hy}{2}\right) + 2\sin\frac{(2x-3h)y}{2}\sin\frac{hy}{2}}{h}\right] \\
&= xy\left[\frac{\sin\frac{(2x-3h)y}{2} - \sin\frac{(2x-h)y}{2}}{h}\right], \text{(as } h \to 0) \\
&= xy\left[\frac{2\cos(x-h)y\sin\left(-\frac{hy}{2}\right)}{h}\right] \\
&= -xy^2\cos xy, \text{(as } h \to 0).
\end{aligned}
$$

In a similar way, one can easily find other partial derivatives of $\sin xy + x + y$. Using Matlab, 3D plots of some partial derivative of the above function are mentioned in Figs. 20.1, 20.2, 20.3, 20.4, 20.5, 20.6, and 20.7.

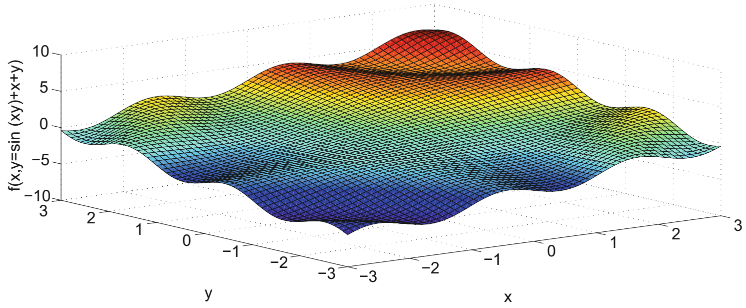

Fig. 20.1 3D graph of the function $f(x, y) = \sin xy + x + y$ with $-3 \leq x \leq 3$ and $-3 \leq y \leq 3$

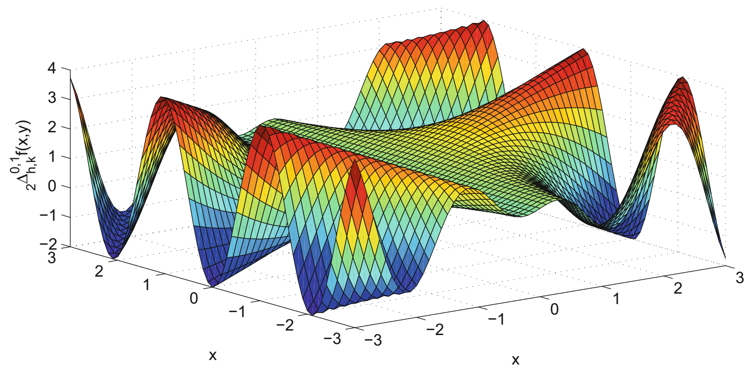

Fig. 20.2 3D graph of the partial derivative $_2\Delta^{0,1}_{h,k}(\sin xy + x + y)$ with $-3 \leq x \leq 3$ and $-3 \leq y \leq 3$

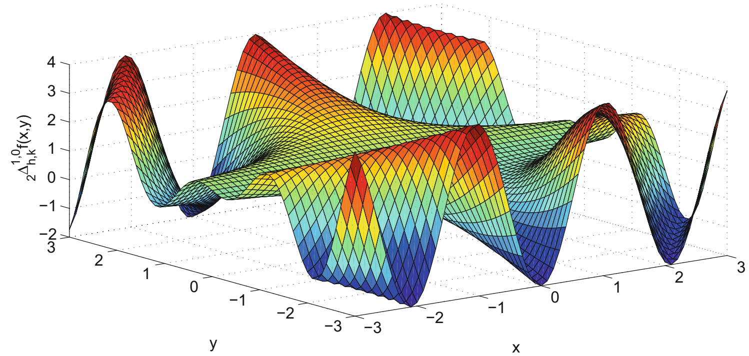

Fig. 20.3 3D graph of the partial derivative $_2\Delta^{1,0}_{h,k}(\sin xy + x + y)$ with $-3 \leq x \leq 3$ and $-3 \leq y \leq 3$

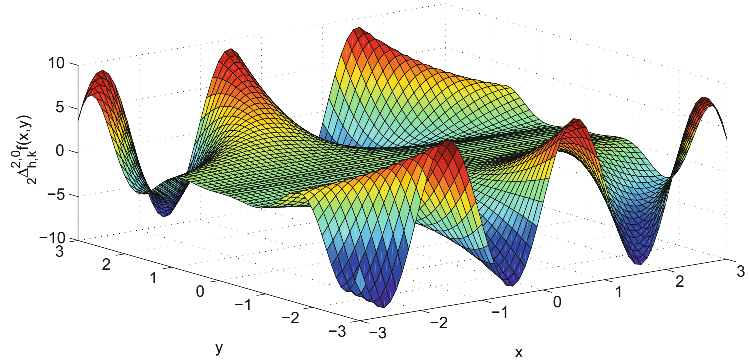

Fig. 20.4 3D graph of the partial derivative ${}_2\Delta^{2,0}_{h,k}(\sin xy + x + y)$ with $-3 \leq x \leq 3$ and $-3 \leq y \leq 3$

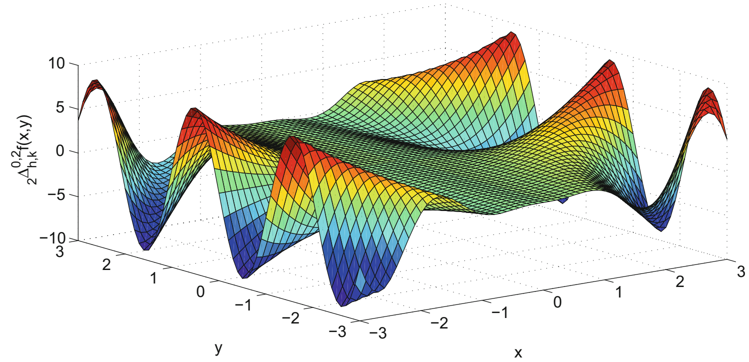

Fig. 20.5 3D graph of the partial derivative ${}_2\Delta^{0,2}_{h,k}(\sin xy + x + y)$ with $-3 \leq x \leq 3$ and $-3 \leq y \leq 3$

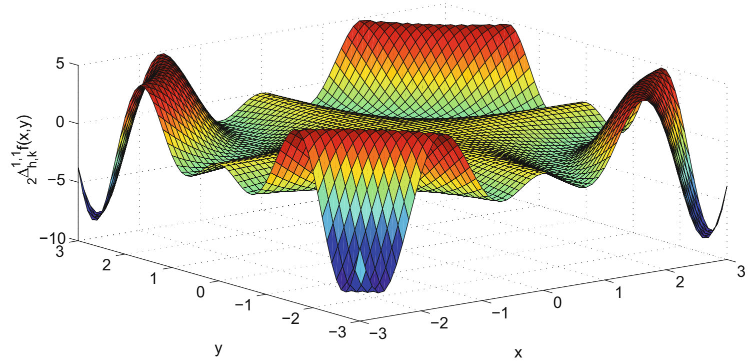

Fig. 20.6 3D graph of the partial derivative ${}_2\Delta^{1,1}_{h,k}(\sin xy + x + y)$ with $-3 \leq x \leq 3$ and $-3 \leq y \leq 3$

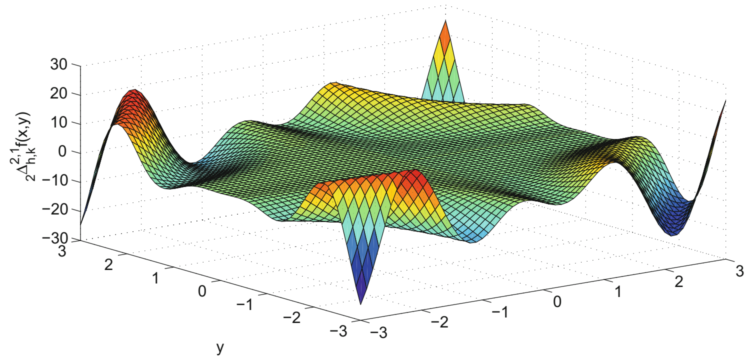

Fig. 20.7 3D graph of the partial derivative ${}_2\Delta^{2,1}_{h,k}(\sin xy + x + y)$ with $-3 \leq x \leq 3$ and $-3 \leq y \leq 3$

Acknowledgements Many thanks to our TEX-pert for developing this class file.

References

1. H. Kızmaz, On certain sequence spaces. Can. Math. Bull. **24**(2), 169–176 (1981)
2. M. Et, R. Colak, On some generalized difference sequence spaces. Soochow J. Math. **21**(4), 377–386 (1995)
3. P. Baliarsingh, Some new difference sequence spaces of fractional order and their dual spaces. Appl. Math. Comput. **219**(18), 9737–9742 (2013)
4. P. Baliarsingh, S. Dutta, A unifying approach to the difference operators and their applications. Bol. Soc. Paran. Mat. **33**(1), 49–57 (2015)
5. P. Baliarsingh, S. Dutta, On the classes of fractional order difference sequence spaces and their matrix transformations. Appl. Math. Comput. **250**, 665–674 (2015)
6. P. Baliarsingh, L. Nayak, A note on fractional difference operators. Alex. Eng. J. (2017) https://doi.org/10.1016/j.aej.2017.02.022
7. P. Baliarsingh, On a fractional difference operator. Alex. Eng. J. **55**(2), 1811–1816 (2016)
8. A. Gökhan, R. Çolak, The double sequence spaces $c_2^P(p)$ and $c_2^{PB}(p)$. Appl. Math. Comput. **157**, 491–501 (2004)
9. M. Mursaleen, Almost strongly regular matrices and a core theorem for double sequences. J. Math. Anal. Appl. **293**(2), 523–531 (2004)
10. M. Mursaleen, O.H.H. Edely, Almost convergence and a core theorem for double sequences. J. Math. Anal. Appl. **293**(2), 532–540 (2004)
11. B. Altay, F. Başar, Some new spaces of double sequences. J. Math. Anal. Appl. **309**(1), 70–90 (2005)
12. A. Pringsheim, Zur theorie der zweifach unendlichen zahlenfolgen. Math. Ann. **53**, 289–321 (1900)
13. G.M. Robinson, Divergent double sequences and series. Trans. Amer. Math. Soc. **28**, 50–73 (1926)
14. F. Başar, *Summability Theory and Its Applications* (Bentham Science Publishers, e-books, Mono- graphs, Istanbul, 2012)
15. M. Zeltser, M. Mursaleen, S.A. Mohiuddine, On almost conservative matrix methods for double sequence spaces. Publ. Math. Debr. **75**, 387–399 (2009)

16. F. Móricz, Extension of the spaces c and c_0 from single to double sequences. Acta Math. Hungar. **57**, 129–136 (1991)
17. R.F. Patterson, Double sequence core theorem, Int. J. Math. Math. Sci. **22**, 785–793 (1999)
18. R.F. Patterson, H. Çakalli, Quasi Cauchy double sequences, Tbilisi Math. J. **8**(2), 211–219 (2015)
19. H. Çakalli, R.F. Patterson, Functions preserving slowly oscillating double sequences. An. Ştiinţ. Univ. Al. I. Cuza Iaşi. Mat. (N.S.) 62 **2**2, 531–536 (2016)
20. H. Çakalli, E. Savaş, Statistical convergence of double sequences in topological groups. J. Comput. Anal. Appl. **12**(2), 421–426 (2010)
21. D. Djurčić, L.D.R. Kočinac, M.R. v Zižzović, Double sequences and selections. Abstr. Appl. Anal. 6 (2012). Article ID: 497594. https://doi.org/10.1155/2012/497594,2012
22. P. Baliarsingh, L. Nayak, P. Beuria, On fractional powers of double band matrices. Proyecciones (Antofagasta) **36**(4), 701–709 (2017)
23. P. Beuria, P. Baliarsingh, L. Nayak, On arbitrary powers of double band matrices. Appl. Math. E-Notes **18**(1), 82–91 (2018)
24. P. Baliarsingh, L. Nayak, V. Kumar, On matrix inversions through difference operators. Iran J. Sci. Technol. Trans. Sci. https://doi.org/10.1007/s40995-017-0161-9
25. L. Nayak, M. Et, P. Baliarsingh, On certain generalized weighted mean fractional difference sequence spaces. Proc. Natl. Acad. Sci., India Sect. A Phys. Sci. (2018). https://doi.org/10.1007/s40010-017-0403-4
26. H. Dutta, L.D.R. Kočinac, On difference sequence spaces defined by Orlicz functions without convexity. Bull. Iran. Math. Soc. **41**(2), 477–489 (2015)
27. A. Aasma, H. Dutta, P.N. Natarajan, *An Introductory Course in Summability Theory*, 1st edn. (John Wiley & Sons, Inc. Hoboken, 2017)
28. H. Dutta, B.E. Rhoades (Eds.), *Current Topics in Summability Theory and Applications*, 1st edn. (Springer, Singapore, 2016)
29. P. Baliarsingh, On difference double sequence spaces of fractional order. Indian J. Math. **58**(3), 287–310 (2016)
30. B. Altay, F. Başar, The fine spectrum and the matrix domain of the difference operator Δ on the sequence space ℓ_p, $(0 < p < 1)$. Commun. Math. Anal. **2**(2), 1–11 (2007)
31. P. Baliarsingh, On double difference operators via four dimensional matrices. J. Indian Math. Soc. **83**(3–4), 209–219 (2016)
32. E. Malkowsky, M. Mursaleen, S. Suantai, The dual spaces of sets of difference sequences of order m and matrix transformations. Acta Math. Sin. (Engl. Ser.) **23**(3), 521–532 (2007)

Chapter 21
Pointwise Convergence Analysis for Nonlinear Double m-Singular Integral Operators

Gümrah Uysal and Hemen Dutta

Abstract In this chapter, m-singularity notion is discussed for double singular integral operators. In this direction, several results concerning pointwise convergence of nonlinear double m-singular integral operators are presented. This chapter is divided into six sections. In the first section, the reasons giving birth to m-singularity notion are explained and related theoretical background is mentioned. Also, the motivations giving inspiration to this note are presented. In the second section, the well-definiteness of the operators which are under the spotlights is shown on their domain. In the third section, an auxiliary result, pointwise convergence theorem, is proved. In the fourth section, main theorem, Fatou type convergence theorem, is proved. In the fifth section, corresponding rates of convergences are evaluated. In the last section, some concluding remarks are given.

Keywords Pointwise convergence · Fatou-type convergence · Nonlinear bivariate integral operators · m-Singularity

21.1 Introduction

Progress in approximation theory techniques concerning approximation by integral operators in the last several decades has turned the researchers work with more general types of operators. In this context, there are at least two important types of integral operators which come to the fore: singular integral operators with approximate identity type kernels and Calderón-Zygmund type singular integral operators. The working areas of both types of operators have almost same power on the theory of real and complex functions too. Also, for both cases, the word

G. Uysal (✉)
Department of Computer Technologies, Division of Technology of Information Security, Karabuk University, Karabuk, Turkey

H. Dutta
Department of Mathematics, Gauhati University, Guwahati, India

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_21

"singular" indicates different concepts. A well-known example for the first type is Gauss-Weierstrass type singular integral operators and Hilbert transform for the second type. For both cases, the word "singular" indicates different concepts. For further information, we refer the reader to [7, 15, 30] and [8], for the first and second types of singular integral operators mentioned above, respectively. Singular integral operators or shortly singular integrals have a wide range of applications in science, engineering and technology. For some applications, we refer the reader to the monograph by Bracewel [6] and work by Papadopoulos [22].

Denoting the points on an open, half-open or closed line segment by $\langle -\pi, \pi \rangle$, in [33], Taberski handled the following two-parameter singular integral operators:

$$L_{\zeta}(f; v) = \int_{-\pi}^{\pi} f(u) K_{\zeta}(u - v) du, \quad v \in \langle -\pi, \pi \rangle, \quad \zeta \in I, \tag{21.1.1}$$

where $K_{\zeta} : \mathbb{R} \to \mathbb{R}_0^+$ stands for a 2π-periodic kernel enriched with some properties for each $\zeta \in I$, and I is a given non-empty set of non-negative numbers with accumulation point ζ_0. In the same work, after giving generalization of Natanson's well-known lemma (see [21]), some pointwise convergence theorems were proved. Then, Gadjiev [11] generalized Taberski's indicated version of Natanson's lemma by the aid of the function $\mu(t)$, which was taken previously as $\mu(t) = t$ in the generalization given by Taberski [33]. This work also contains some approximation theorems for the operators of type (21.1.1) with respect to specific definition of the function $\mu(t)$. Also, the definition μ-generalized Lebesgue point was obtained due to this invention. Particularly, if $\mu(t) = t$, then conventional definition of Lebesgue point is obtained. After this work, Rydzewska [23] studied on the operators of type (21.1.1) and presented some results on the rate of convergence at μ-generalized Lebesgue point of 2π-periodic Lebesgue integrable function $f \in L \langle -\pi, \pi \rangle$. In regard to pointwise approximation by assorted types of linear integral operators, we refer the reader to [3, 7, 17, 31]. Also, for the q-analogues of some linear integral operators, we refer the reader to the monograph by Aral et al. [2].

Musielak [18] focused on the convergence of nonlinear integral operators designated by:

$$T_w(f; y) = \int_{\mathbf{G}} K_w(x - y; f(x)) dx, \quad y \in \mathbf{G}, \quad w \in \Lambda, \tag{21.1.2}$$

where $\mathbf{G}$ denotes locally compact Abelian group along with Haar measure and Λ is a non-empty index set with certain topology. In this study, the innovative ideas were usage of Lipschitz condition for K_w with respect to second variable and a new singularity notion. Therefore, the previously developed solutions and proving methods were applied to nonlinear approximation problems and theorems, respectively. For some studies on nonlinear integral operators in many different settings, we refer the reader to [4, 10, 19, 32].

Mamedov [16] constructed the following m-singular integral operators

$$L_{\lambda}^{[m]}(f;x) = (-1)^{m+1} \int_{\mathbb{R}} \left[\sum_{k=1}^{m} (-1)^{m-k} \binom{m}{k} f(x+kt) \right] K_{\lambda}(t)dt, \qquad (21.1.3)$$

where $x \in \mathbb{R}$, $m \geq 1$ is a finite certain natural number and $\lambda \in \Lambda$ which is a non-empty set of non-negative indices, by employing m-th finite differences. Here, the main aim is approximating the m-th derivatives of the integral of the functions almost everywhere by using these m-singular integral operators (compare with Lebesgue point notion [7]). Then, Karsli [14] studied the Fatou type convergence of nonlinear counterparts of the operators of type (21.1.3) in the following form:

$$T_{\lambda}^{[m]}(f;x) = \int_{\mathbb{R}} K_{\lambda} \left(t, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x+kt) \right) dt, \qquad (21.1.4)$$

where $x \in \mathbb{R}$, $m \geq 1$ is a finite natural number and $\lambda \in \Lambda$ which is a non-empty set of non-negative indices, at $m - p - \mu$-Lebesgue point of the functions $f \in L_p(\mathbb{R})$ $(1 \leq p < \infty)$, where $L_p(\mathbb{R})$ is the space of all measurable functions for which $|f|^p$ has finite integral value on $\mathbb{R}$. For the studies concerning approximation by m-singular integral operators in various settings, we refer the reader to [5, 12, 13, 25, 35].

In the year 1964, Taberski [34] handled the problem of pointwise approximation of functions $f \in L_1(\mathbf{R})$ by convolution type operators of two variables in the following form:

$$L_{\lambda}(f;x,y) = \iint_{\mathbf{R}} f(t,s) K_{\lambda}(t-x, s-y)\, ds dt, \quad (x,y) \in \mathbf{R}, \qquad (21.1.5)$$

where $\mathbf{R}$ denotes an arbitrary bounded rectangle and $K_{\lambda}(t,s)$ is a kernel satisfying some conditions with $\lambda \in \Lambda$, where Λ is a non-empty set of non-negative numbers with accumulation point λ_0. The studies [27, 28] and [25] related to Taberski's study [34] presented some results on the study of pointwise convergence of the operators of type (21.1.5) on some special sets consisting of characteristic points (x_0, y_0) of different types.

Let us consider the nonlinear double singular integral operators of the form:

$$T_{\lambda}(f;x,y) = \iint_{\mathbf{D}} K_{\lambda}(t-x, s-y, f(t,s)) ds dt, \quad (x,y) \in \mathbf{D}, \quad \lambda \in \Lambda, \qquad (21.1.6)$$

where **D** denotes an arbitrary bounded rectangle and Λ is a set of non-negative numbers with accumulation point λ_0. The operators of type (21.1.6) are the nonlinear generalizations of the operators of type (21.1.5). The properties belonging to indicated operators, such as boundedness, well-definiteness, pointwise convergence and modularity, were studied by many approximation theory researchers throughout the years. For some of the related works, we refer the reader to [20, 37].

In [1], certain multidimensional nonlinear integrals in the following form:

$$L_\lambda(u,x) = \frac{\lambda^n}{w_{n-1}} \int\limits_{\mathbb{R}^n} K\left(\lambda\left|t-x\right|, u(t)\right) dt, \quad x \in \mathbb{R}^n, \tag{21.1.7}$$

where $K\left(\lambda\left|t\right|, u(t)\right)$ is a kernel satisfying some properties including differentiability with respect to second variable and λ is a positive parameter, were considered. In order to remove nonlinearity problem in the proofs, two main technics are used: The first technic is obtaining Taylor expansion of kernel with respect to second variable and the second one is using majorization. From this point of view, this work contains very important results.

Incorporating the operators of type (21.1.3) and (21.1.6), the following m-singular integral operators are obtained:

$$T_\zeta^{[m]}(f;x,y) = \iint\limits_{\mathbb{R}^2} K_\zeta\left(t, s, \sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k} f(x+kt, y+ks)\right) ds dt, \tag{21.1.8}$$

where $(x,y) \in \mathbb{R}^2$, $m \geq 1$ is a finite natural number and $\zeta \in \Lambda \subseteq \overline{\mathbb{R}}_+ \cup \{0\}$, which is a non-empty set of indices, at $[m;\mu]$-Lebesgue point of the function $f \in L_1\left(\mathbb{R}^2\right)$, where $L_1\left(\mathbb{R}^2\right)$ is the space of all measurable functions for which $|f|$ has finite integral value on $\mathbb{R}^2$. In operators of type (21.1.8), we consider a direct two-dimensional generalization of m-th finite differences used in [16] and, as a result, these are the two-dimensional counterparts of the operators of type (21.1.4). On the other hand, this study is a continuation of very recent work [36] in which mixed differences were harnessed in order to construct the operators.

The chapter is organized as follows: In Sect. 21.2, we introduce fundamental notions bringing the well-definiteness of the operators. In Sect. 21.3, we prove pointwise convergence of the operators of type (21.1.8). In Sect. 21.4, we present, as a main result, Fatou type convergence theorem for the indicated operators. In Sect. 21.5, we establish the rates of both pointwise and Fatou type convergences by using the results obtained in the previous sections. In Sect. 21.6, we give some concluding remarks.

21.2 Preliminaries

In this section, basic concepts used in this chapter are introduced.

Definition 21.2.1 A point $(x_0, y_0) \in \mathbb{R}^2$ at which the following relations hold:

$$\lim_{(h,k)\to(0,0)} \frac{1}{\mu_1(h)\mu_2(k)} \int_0^{\pm h}\int_0^{\pm k} \left|\Delta^m_{(t,s)} g(x_0, y_0)\right| dsdt = 0, \tag{21.2.1}$$

where

$$\Delta^m_{(t,s)} g(x_0, y_0) = \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} g(x_0 + kt, y_0 + ks)$$

is called $[m; \mu]$-Lebesgue point of $g \in L_1(\mathbb{R}^2)$. *Here,* $\mu_1 : \mathbb{R} \to \mathbb{R}$ is an increasing and absolutely continuous function on $0 \le h \le \delta_1$, where δ_1 is a fixed positive real number with $\mu_1(0) = 0$. Similarly, $\mu_2 : \mathbb{R} \to \mathbb{R}$ is an increasing and absolutely continuous function on $0 \le k \le \delta_2$, where δ_2 is a fixed positive real number with $\mu_2(0) = 0$.

Remark 21.2.2 Definition 21.2.1 is obtained by incorporating the characterization of function μ given by Gadjiev [11] and m-Lebesgue point definition used by Mamedov [16]. On the other hand, for some other μ-generalized Lebesgue point characterizations, we refer the reader to [14, 23, 24].

Definition 21.2.3 (*Class* $\mathcal{A}$) Let ζ_0 be an accumulation point of Λ. A family $(K_\zeta)_{\zeta\in\Lambda}$ which consists of the functions $K_\zeta : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ is called *class* $\mathcal{A}$, if the following conditions hold:

(a) $K_\zeta(t, s, 0) = 0$ for every $(t, s) \in \mathbb{R}^2$ and for each $\zeta \in \Lambda$, and $K_\zeta(., ., u) \in L_1(\mathbb{R}^2)$ for every $u \in \mathbb{R}$ and for each $\zeta \in \Lambda$.

(b) There exists a family $(L_\zeta)_{\zeta\in\Lambda}$ consisting of the globally Lebesgue integrable functions for each ζ $L_\zeta : \mathbb{R}^2 \to \mathbb{R}$ such that the inequality

$$\left|K_\zeta(t, s, u) - K_\zeta(t, s, v)\right| \le L_\zeta(t, s)\,|u - v|$$

holds for every $(t, s) \in \mathbb{R}^2$, $u, v \in \mathbb{R}$, and for each fixed $\zeta \in \Lambda$.

(c) For every $\xi > 0$, $\lim_{\zeta\to\zeta_0}\left[\sup_{\xi\le\sqrt{t^2+s^2}} L_\zeta(t, s)\right] = 0$.

(d) For every $\xi > 0$, $\lim_{\zeta\to\zeta_0}\left[\iint_{\xi\le\sqrt{t^2+s^2}} L_\zeta(t, s)\, dsdt\right] = 0$.

(e) $\|L_\zeta\|_{L_1(\mathbb{R}^2)} \le M < \infty$ for every $\zeta \in \Lambda$.

(f) $L_\zeta(t, s)$ is non-increasing on $[0, \infty)$ and non-decreasing on $(-\infty, 0]$ for each fixed $\zeta \in \Lambda$ as a function of t for all values of s. Similarly, $L_\zeta(t, s)$ is non-increasing on $[0, \infty)$ and non-decreasing on $(-\infty, 0]$ for each fixed $\zeta \in \Lambda$

as a function of s for all values of t. $L_\zeta(t,s)$ is bimonotonically increasing with respect to (t,s) on $[0,\infty)\times[0,\infty)$ and on $(-\infty,0]\times(-\infty,0]$ and bimonotonically decreasing with respect to (t,s) on $[0,\infty)\times(-\infty,0]$ and $(-\infty,0]\times[0,\infty)$ for each fixed $\zeta\in\Lambda$.

(g) $\lim\limits_{\zeta\to\zeta_0}\left|\iint\limits_{\mathbb{R}^2} K_\zeta\left(t,s,\sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k}u\right)dsdt-u\right|=0$ for every $u\in\mathbb{R}$.

Throughout this chapter the kernel function K_ζ belongs to $Class\ \mathcal{A}$.

Remark 21.2.4 The conditions (a)–(g) in Definition 21.2.1, which are compulsory for the proof of main results from the theoretical point of view, were also used in [37] with some modifications (for analogues conditions, see also [36]). The studies [4, 5] shall be seen as main references for $Class\ \mathcal{A}$. For the Lipschitz condition (b), we refer the reader to [4, 18, 19].

Now, we will give a lemma concerning well-definiteness of the operators of type (1.8).

Lemma 21.2.5 *If $f\in L_1\left(\mathbb{R}^2\right)$, then the operators $T_\zeta^{[m]}(f)\in L_1\left(\mathbb{R}^2\right)$ and the inequality*

$$\left\|T_\zeta^{[m]}(f)\right\|_{L_1(\mathbb{R}^2)}\le(2^m-1)\left\|L_\zeta\right\|_{L_1(\mathbb{R}^2)}\|f\|_{L_1(\mathbb{R}^2)}$$

hold for every $\zeta\in\Lambda$.

Proof By condition (a), we can write

$$\begin{aligned}&\left\|T_\zeta^{[m]}(f)\right\|_{L_1(\mathbb{R}^2)}\\&=\iint\limits_{\mathbb{R}^2}\left|\iint\limits_{\mathbb{R}^2}K_\zeta\left(t,s,\sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k}f(x+kt,y+ks)\right)dsdt\right|dydx\\&\le\iint\limits_{\mathbb{R}^2}\left(\iint\limits_{\mathbb{R}^2}\sum_{k=1}^{m}\binom{m}{k}|f(x+kt,y+ks)|\,L_\zeta(t,s)\,dsdt\right)dydx.\end{aligned}$$

Now, applying Fubini's theorem (see, e.g., [26]) and change of variables, we have

$$\begin{aligned}\left\|T_\zeta^{[m]}(f)\right\|_{L_1(\mathbb{R}^2)}&\le\sum_{k=1}^{m}\binom{m}{k}\iint\limits_{\mathbb{R}^2}L_\zeta(t,s)\,dsdt\iint\limits_{\mathbb{R}^2}|f(u,v)|\,dvdu\\&=\left(2^m-1\right)\left\|L_\zeta\right\|_{L_1(\mathbb{R}^2)}\|f\|_{L_1(\mathbb{R}^2)},\end{aligned}$$

where $\sum_{k=1}^{m}\binom{m}{k}=2^m-1$. Therefore, the desired result follows from condition (e). Thus the proof is completed. □

21.3 Pointwise Convergence

Theorem 21.3.1 *If* $(x_0, y_0) \in \mathbb{R}^2$ *is* $[m;\mu]$*-Lebesgue point of the function* $f \in L_1\left(\mathbb{R}^2\right)$, *then*

$$\lim_{\zeta\to\zeta_0}\left|T_\zeta^{[m]}(f;x_0,y_0)-f(x_0,y_0)\right|=0,$$

on any set Z on which the function

$$\int_{-\delta}^{\delta}\int_{-\delta}^{\delta}L_\zeta(t,s)\left|\{\mu_1(|t|)\}_t'\right|\left|\{\mu_2(|s|)\}_s'\right|dsdt,$$

where $0<\delta<\min\{\delta_1,\delta_2\}$, *is bounded as* $\zeta\to\zeta_0$.

Proof Let $(x_0, y_0) \in \mathbb{R}^2$ be a $[m;\mu]$-Lebesgue point of the function $f \in L_1\left(\mathbb{R}^2\right)$. Therefore, for every $\varepsilon>0$, there exists $\delta>0$ such that for every h and for every k satisfying $0<h,k<\delta<\min\{\delta_1,\delta_2\}$, we have the following inequalities by (21.2.1):

$$\int_0^h\int_{-k}^0\left|\Delta_{(t,s)}^m f(x_0,y_0)\right|dsdt\le\varepsilon\mu_1(h)\mu_2(k), \tag{21.3.1}$$

$$\int_{-h}^0\int_{-k}^0\left|\Delta_{(t,s)}^m f(x_0,y_0)\right|dsdt\le\varepsilon\mu_1(h)\mu_2(k), \tag{21.3.2}$$

$$\int_{-h}^0\int_0^k\left|\Delta_{(t,s)}^m f(x_0,y_0)\right|dsdt\le\varepsilon\mu_1(h)\mu_2(k), \tag{21.3.3}$$

$$\int_0^h\int_0^k\left|\Delta_{(t,s)}^m f(x_0,y_0)\right|dsdt\le\varepsilon\mu_1(h)\mu_2(k), \tag{21.3.4}$$

where

$$\Delta^m_{(t,s)} f(x_0, y_0) = \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} f(x_0 + kt, y_0 + ks).$$

Now, set $I(\zeta) = \left| T_\zeta^{[m]}(f; x_0, y_0) - f(x_0, y_0) \right|$. Using condition (g), we obtain

$$\begin{aligned}
I(\zeta) &= \left| \iint\limits_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0 + kt, y_0 + ks) \right) dsdt - f(x_0, y_0) \right| \\
&= \left| \iint\limits_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0 + kt, y_0 + ks) \right) dsdt \right. \\
&\quad - \iint\limits_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0, y_0) \right) dsdt \\
&\quad \left. + \iint\limits_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0, y_0) \right) dsdt - f(x_0, y_0) \right| .
\end{aligned}$$

Furthermore, we obtain the following inequality:

$$\begin{aligned}
I(\zeta) \le &\left| \iint\limits_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0 + kt, y_0 + ks) \right) dsdt \right. \\
&\quad \left. - \iint\limits_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0, y_0) \right) dsdt \right| \\
&\quad + \left| \iint\limits_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0, y_0) \right) dsdt - f(x_0, y_0) \right| \\
&= I_1(\zeta) + I_2(\zeta) .
\end{aligned}$$

By condition (g), $I_2(\zeta) \to 0$ as $\zeta \to \zeta_0$. Now, we focus on the integral $I_1(\zeta)$. By condition (b), the following inequality holds for $I_1(\zeta)$:

$$I_1(\zeta) \leq \iint\limits_{\mathbb{R}^2} \left| \left(\sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0 + kt, y_0 + ks) \right) - \left(\sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0, y_0) \right) \right| \times L_\zeta(t, s)\, ds dt.$$

Let

$$B_\delta := \left\{ (t, s) \in \mathbb{R}^2 : t^2 + s^2 < \delta^2 \right\}.$$

By condition (b), the integral $I_1(\zeta)$ satisfies that

$$\begin{aligned} I_1(\zeta) \leq & \iint\limits_{B_\delta} \left| \left(\sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0 + kt, y_0 + ks) \right) - \left(\sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0, y_0) \right) \right| L_\zeta(t, s)\, ds dt \\ & + \left| \iint\limits_{\mathbb{R}^2 \setminus B_\delta} \sum_{k=1}^{m} \binom{m}{k} \{ |f(x_0 + kt, y_0 + ks)| + |f(x_0, y_0)| \} L_\zeta(t, s)\, ds dt \right| \\ = & I_{11}(\zeta) + I_{12}(\zeta). \end{aligned}$$

Now, we deal with the integral $I_{12}(\zeta)$. The operations for the integral $I_{12}(\zeta)$ are as follows:

$$\begin{aligned} I_{12}(\zeta) \leq \sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta(t, s) & \left| \iint\limits_{\mathbb{R}^2 \setminus B_\delta} \sum_{k=1}^{m} \binom{m}{k} |f(x_0 + kt, y_0 + ks)|\, ds dt \right| \\ & + \sum_{k=1}^{m} \binom{m}{k} |f(x_0, y_0)| \left| \iint\limits_{\mathbb{R}^2 \setminus B_\delta} L_\zeta(t, s)\, ds dt \right|. \end{aligned}$$

Furthermore, we proceed as follows:

$$\begin{aligned} I_{12}(\zeta) &\leq \sup_{\delta\leq\sqrt{t^2+s^2}} L_\zeta(t,s)\left|\iint_{\mathbb{R}^2}\sum_{k=1}^{m}\binom{m}{k}|f(x_0+kt,y_0+ks)|\,dsdt\right| \\ &\quad+\left(2^m-1\right)|f(x_0,y_0)|\iint_{\delta\leq\sqrt{t^2+s^2}} L_\zeta(t,s)\,dsdt \\ &\leq 2^m\left\{\sup_{\delta\leq\sqrt{t^2+s^2}} L_\zeta(t,s)\,\|f\|_{L_1(\mathbb{R}^2)}+|f(x_0,y_0)|\iint_{\delta\leq\sqrt{t^2+s^2}} L_\zeta(t,s)\,dsdt\right\} \\ &= 2^m\left(I_{121}(\zeta)+I_{122}(\zeta)\right), \end{aligned}$$

where

$$\sum_{k=1}^{m}\binom{m}{k}=\left(2^m-1\right).$$

It follows that $I_{121}(\zeta)\to 0$ as $\zeta\to\zeta_0$ and $I_{122}(\zeta)\to 0$ as $\zeta\to\zeta_0$ by conditions (c) and (d), respectively.

Recalling the integral $I_{11}(\zeta)$,we have

$$\begin{aligned} I_{11}(\zeta) &= \iint_{B_\delta}\left|\sum_{k=0}^{m}(-1)^{k-1}\binom{m}{k}f(x_0+kt,y_0+ks)\right|L_\zeta(t,s)\,dsdt \\ &= \iint_{B_\delta}\left|\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}f(x_0+kt,y_0+ks)\right|L_\zeta(t,s)\,dsdt. \end{aligned}$$

It follows that

$$\begin{aligned} I_{11}(\zeta) &\leq \int_{-\delta}^{\delta}\int_{-\delta}^{\delta}\left|\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}f(x_0+kt,y_0+ks)\right|L_\zeta(t,s)\,dsdt \\ &= \left\{\int_{0}^{\delta}\int_{-\delta}^{0}+\int_{-\delta}^{0}\int_{-\delta}^{0}\right\}\left|\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}f(x_0+kt,y_0+ks)\right|L_\zeta(t,s)\,dsdt \end{aligned}$$

$$+\left\{\int_{-\delta}^{0}\int_{0}^{\delta}+\int_{0}^{\delta}\int_{0}^{\delta}\right\}\left|\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}f(x_0+kt,y_0+ks)\right|L_\zeta(t,s)\,dsdt$$

$$=I_{111}+I_{112}+I_{113}+I_{114}.$$

Let us consider the integral I_{111}. Let us define the function $F(t,s)$ by

$$F(t,s)=\int_{0}^{t}\int_{s}^{0}\left|\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}f(x_0+ku,y_0+kv)\right|dvdu.$$

In view of inequality (21.3.1), the following expression

$$|F(t,s)|\leq\varepsilon\mu_1(t)\,\mu_2(-s)\,, \tag{21.3.5}$$

where $0<t<\delta$ and $-\delta<s<0$, holds. In view of Theorem 2.6 in [34], we can write

$$I_{111}=(\mathbf{L})\int_{0}^{\delta}\int_{-\delta}^{0}\left|\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}f(x_0+kt,y_0+ks)\right|L_\zeta(t,s)\,dsdt$$

$$=(\mathbf{LS})\int_{0}^{\delta}\int_{-\delta}^{0}L_\zeta(t,s)\,d\left[-F(t,s)\right],$$

where **LS** denotes Lebesgue-Stieltjes integral. Applying bivariate integration by parts (see Theorem 2.2, p.100 in [34]) to the Lebesgue-Stieltjes integral, we have

$$|I_{111}|=\left|\int_{0}^{\delta}\int_{-\delta}^{0}L_\zeta(t,s)\,d\left[F(t,s)\right]\right|$$

$$\leq\int_{0}^{\delta}\int_{-\delta}^{0}|F(t,s)|\left|dL_\zeta(t,s)\right|+|F(\delta,-\delta)|\,L_\zeta(\delta,-\delta)$$

$$+\int_{0}^{\delta}|F(t,-\delta)|\left|dL_\zeta(t,-\delta)\right|+\int_{-\delta}^{0}|F(\delta,s)|\left|dL_\zeta(\delta,s)\right|.$$

If we apply inequality (21.3.5) to the last inequality, then we have

$$\left|I_{111}\right| \leq \varepsilon \int_{0}^{\delta} \int_{-\delta}^{0} \mu_1(t)\, \mu_2(-s) \left|dL_{\zeta}(t,s)\right| + \varepsilon\mu_1(\delta)\mu_2(\delta) L_{\zeta}(\delta,-\delta)$$
$$+ \varepsilon\mu_2(\delta) \int_{0}^{\delta} \mu_1(t) \left|dL_{\zeta}(t,-\delta)\right| + \varepsilon\mu_1(\delta) \int_{-\delta}^{0} \mu_2(s) \left|dL_{\zeta}(\delta,s)\right|.$$

One more application of integration by parts gives (for the analogues situation, see [24, 34]):

$$\left|I_{111}\right| \leq \varepsilon \int_{0}^{\delta} \int_{-\delta}^{0} \left|\{\mu_1(t)\}_t'\right| \left|\{\mu_2(-s)\}_s'\right| L_{\zeta}(t,s)\, ds dt.$$

Let us consider the integral I_{112}. Let us define the function $E(t,s)$ by

$$E(t,s) = \int_{t}^{0} \int_{s}^{0} \left| \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} f(x_0 + ku, y_0 + kv) \right| dv du.$$

In view of inequality (21.3.2), the following expression

$$|E(t,s)| \leq \varepsilon\mu_1(-t)\, \mu_2(-s), \tag{21.3.6}$$

where $-\delta < t < 0$ and $-\delta < s < 0$, holds. In view of Theorem 2.6 in [34], we can write

$$I_{112} = (\mathbf{L}) \int_{-\delta}^{0} \int_{-\delta}^{0} \left| \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} f(x_0 + kt, y_0 + ks) \right| L_{\zeta}(t,s)\, ds dt$$
$$= (\mathbf{LS}) \int_{-\delta}^{0} \int_{-\delta}^{0} L_{\zeta}(t,s)\, d\left[E(t,s)\right],$$

where **LS** denotes Lebesgue-Stieltjes integral. Applying bivariate integration by parts (see Theorem 2.2, p.100 in [34]) to the Lebesgue-Stieltjes integral, we have

$$\begin{aligned}|I_{112}| &= \left| \int_{-\delta}^{0}\int_{-\delta}^{0} L_{\zeta}(t,s)\, d\left[E(t,s)\right] \right| \\ &\leq \int_{-\delta}^{0}\int_{-\delta}^{0} |E(t,s)| \left| dL_{\zeta}(t,s) \right| + |E(-\delta,-\delta)|\, L_{\zeta}(-\delta,-\delta) \\ &\quad + \int_{-\delta}^{0} |E(t,-\delta)| \left| dL_{\zeta}(t,-\delta) \right| + \int_{-\delta}^{0} |E(-\delta,s)| \left| dL_{\zeta}(-\delta,s) \right|.\end{aligned}$$

If we apply inequality (21.3.6) to the last inequality, then we have

$$\begin{aligned}|I_{112}| &\leq \varepsilon \int_{-\delta}^{0}\int_{-\delta}^{0} \mu_1(-t)\, \mu_2(-s) \left| dL_{\zeta}(t,s) \right| + \varepsilon\mu_1(\delta)\mu_2(\delta) L_{\zeta}(-\delta,-\delta) \\ &\quad + \varepsilon\mu_2(\delta) \int_{-\delta}^{0} \mu_1(t) \left| dL_{\zeta}(t,-\delta) \right| + \varepsilon\mu_1(\delta) \int_{-\delta}^{0} \mu_2(s) \left| dL_{\zeta}(-\delta,s) \right|.\end{aligned}$$

One more application of integration by parts gives:

$$|I_{112}| \leq \varepsilon \int_{-\delta}^{0}\int_{-\delta}^{0} L_{\zeta}(t,s) \left| \{\mu_1(-t)\}_t' \right| \left| \{\mu_2(-s)\}_s' \right| ds dt.$$

Let us consider the integral I_{113}. Let us define the function $G(t,s)$ by

$$G(t,s) = \int_{t}^{0}\int_{0}^{s} \left| \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} f(x_0 + ku, y_0 + kv) \right| dv du.$$

In view of inequality (21.3.3), the following expression

$$|G(t,s)| \leq \varepsilon\mu_1(-t)\, \mu_2(s), \tag{21.3.7}$$

where $-\delta < t < 0$ and $0 < s < \delta$, holds. In view of Theorem 2.6 in [34], we can write

$$I_{113} = (\mathbf{L})\int_{-\delta}^{0}\int_{0}^{\delta}\left|\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k} f(x_0+kt, y_0+ks)\right| L_\zeta(t,s)\,dsdt$$

$$= (\mathbf{LS})\int_{-\delta}^{0}\int_{0}^{\delta} L_\zeta(t,s)\,d\left[-G(t,s)\right],$$

where **LS** denotes Lebesgue-Stieltjes integral. Applying bivariate integration by parts (see Theorem 2.2, p.100 in [34]) to the Lebesgue-Stieltjes integral, we have

$$|I_{113}| = \left|\int_{-\delta}^{0}\int_{0}^{\delta} L_\zeta(t,s)\,d\left[G(t,s)\right]\right|$$

$$\leq \int_{-\delta}^{0}\int_{0}^{\delta} |G(t,s)|\left|dL_\zeta(t,s)\right| + |G(-\delta,\delta)|\,L_\zeta(-\delta,\delta)$$

$$+ \int_{-\delta}^{0}|G(t,\delta)|\left|dL_\zeta(t,\delta)\right| + \int_{0}^{\delta}|G(-\delta,s)|\left|dL_\zeta(-\delta,s)\right|.$$

If we apply inequality (21.3.7) to the last inequality, then we have

$$|I_{113}| \leq \varepsilon\int_{-\delta}^{0}\int_{0}^{\delta}\mu_1(-t)\,\mu_2(+s)\left|dL_\zeta(t,s)\right| + \varepsilon\mu_1(\delta)\mu_2(\delta)L_\zeta(-\delta,\delta)$$

$$+ \varepsilon\mu_2(\delta)\int_{-\delta}^{0}\mu_1(t)\left|dL_\zeta(t,\delta)\right| + \varepsilon\mu_1(\delta)\int_{0}^{\delta}\mu_2(s)\left|dL_\zeta(-\delta,s)\right|.$$

One more application of integration by parts gives:

$$|I_{113}| \leq \varepsilon\int_{-\delta}^{0}\int_{0}^{\delta} L_\zeta(t,s)\left|\{\mu_1(-t)\}_t'\right|\left|\{\mu_2(s)\}_s'\right| dsdt.$$

Let us consider the integral I_{114}. Let us define the function $H(t,s)$ by

$$H(t,s)=\int_0^t\int_0^s\left|\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}f(x_0+ku,y_0+kv)\right|dvdu.$$

In view of inequality (21.3.4), the following expression

$$|H(t,s)|\le\varepsilon\mu_1(t)\,\mu_2(s)\,, \tag{21.3.8}$$

where $0<t<\delta$ and $0<s<\delta$, holds. In view of Theorem 2.6 in [34], we can write

$$\begin{aligned}I_{114}&=(\mathbf{L})\int_0^\delta\int_0^\delta\left|\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}f(x_0+kt,y_0+ks)\right|L_\zeta(t,s)\,dsdt\\&=(\mathbf{LS})\int_0^\delta\int_0^\delta L_\zeta(t,s)\,d\left[H(t,s)\right],\end{aligned}$$

where **LS** denotes Lebesgue-Stieltjes integral. Applying bivariate integration by parts (see Theorem 2.2, p.100 in [34]) to the Lebesgue-Stieltjes integral, we have

$$\begin{aligned}|I_{114}|&=\left|\int_0^\delta\int_0^\delta L_\zeta(t,s)\,d\left[H(t,s)\right]\right|\\&\le\int_0^\delta\int_0^\delta|H(t,s)|\left|dL_\zeta(t,s)\right|+|H(\delta,\delta)|\,L_\zeta(\delta,\delta)\\&\quad+\int_0^\delta|H(t,\delta)|\left|dL_\zeta(t,\delta)\right|+\int_0^\delta|H(\delta,s)|\left|dL_\zeta(\delta,s)\right|.\end{aligned}$$

If we apply inequality (21.3.8) to the last inequality, then we have

$$\begin{aligned}|I_{114}|&\le\varepsilon\int_0^\delta\int_0^\delta\mu_1(t)\,\mu_2(s)\left|dL_\zeta(t,s)\right|+\varepsilon\mu_1(\delta)\mu_2(\delta)L_\zeta(\delta,\delta)\\&\quad+\varepsilon\mu_2(\delta)\int_0^\delta\mu_1(t)\left|dL_\zeta(t,\delta)\right|+\varepsilon\mu_1(\delta)\int_0^\delta\mu_2(s)\left|dL_\zeta(\delta,s)\right|.\end{aligned}$$

One more application of integration by parts gives:

$$|I_{114}| \leq \varepsilon \int_0^{\delta} \int_0^{\delta} L_\zeta (t,s) \left| \{\mu_1 (t)\}_t' \right| \left| \{\mu_2 (s)\}_s' \right| dsdt.$$

Hence, the following inequality is obtained for I_{11} :

$$|I_{11}| \leq \varepsilon \int_{-\delta}^{\delta} \int_{-\delta}^{\delta} L_\zeta (t,s) \left| \{\mu_1 (|t|)\}_t' \right| \left| \{\mu_2 (|s|)\}_s' \right| dsdt.$$

Summarizing all evaluations, we obtain the following general inequality:

$$\begin{aligned} I(\zeta) \leq\; & 2^m \sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta (t,s) \|f\|_{L_1(\mathbb{R}^2)} \\ & + 2^m |f(x_0,y_0)| \iint_{\delta \leq \sqrt{t^2+s^2}} L_\zeta (t,s)\, dsdt \\ & + \left| \iint_{\mathbb{R}^2} K_\zeta \left(t,s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0,y_0) \right) dsdt - f(x_0,y_0) \right| \\ & + \varepsilon \int_{-\delta}^{\delta} \int_{-\delta}^{\delta} L_\zeta (t,s) \left| \{\mu_1 (|t|)\}_t' \right| \left| \{\mu_2 (|s|)\}_s' \right| dsdt. \end{aligned}$$

Since the last term above is bounded by the hypothesis, the assertion follows, that is $I(\zeta) \to 0$ as $\zeta \to \zeta_0$. Thus, the proof is completed. □

21.4 Fatou Type Convergence

In this section, we will prove the Fatou type convergence of the operators of type (21.1.8) (see [9]). Now, we suppose that for a sufficiently small $\delta > 0$ such that the function Ψ_δ given as

$$\begin{aligned} \Psi_\delta(x,y,\zeta) = \sum_{k=1}^{m} \binom{m}{k} \int_{-\delta}^{\delta} \int_{-\delta}^{\delta} & |f(x+kt, y+ks) \\ & - f(x_0+kt, y_0+ks)| \, L_\zeta (t,s)\, dsdt, \end{aligned}$$

where $0 < \delta < \min\{\delta_1, \delta_2\}$, is bounded on the set defined as

$$Z_{C,\delta,m} = \left\{(x, y, \zeta) \in \mathbb{R}^2 \times \Lambda : \Psi_{\delta,m}(x, y, \zeta) < C\right\},$$

where C is positive constant, as (x, y, ζ) tends to (x_0, y_0, ζ_0). For some results giving inspiration to the following theorem, we refer the reader to [14, 29, 34].

Theorem 21.4.1 *Suppose that the hypotheses of Theorem 21.3.1 hold. If* $(x_0, y_0) \in \mathbb{R}^2$ *is a* $[m; \mu]$*-Lebesgue point of the function* $f \in L_1(\mathbb{R}^2)$, *then*

$$\lim_{(x,y,\zeta)\to(x_0,y_0,\zeta_0)} \left|T_\zeta^{[m]}(f; x, y) - f(x_0, y_0)\right| = 0$$

provided that $(x, y, \zeta) \in Z_{C,\delta,m}$.

Proof Let $0 < |x_0 - x| < \frac{\delta}{2}$ and $0 < |y_0 - y| < \frac{\delta}{2}$ for a given $0 < \delta < \min\{\delta_1, \delta_2\}$. Now, we denote $\left|T_\zeta^{[m]}(f; x, y) - f(x_0, y_0)\right|$ by $I_\zeta(x, y)$.

Write

$$I_\zeta(x, y) = \left|\iint_{\mathbb{R}^2} K_\zeta\left(t, s, \sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k} f(x + kt, y + ks)\right) dsdt - f(x_0, y_0)\right|$$

$$= \left|\iint_{\mathbb{R}^2} K_\zeta\left(t, s, \sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k} f(x + kt, y + ks)\right) dsdt\right.$$

$$- \iint_{\mathbb{R}^2} K_\zeta\left(t, s, \sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k} f(x_0 + kt, y_0 + ks)\right) dsdt$$

$$\left.+ \iint_{\mathbb{R}^2} K_\zeta\left(t, s, \sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k} f(x_0 + kt, y_0 + ks)\right) dsdt - f(x_0, y_0)\right|.$$

From above equality, we deduce that

$$\left|I_\zeta(x, y)\right|$$
$$\leq \iint_{\mathbb{R}^2} \sum_{k=1}^{m}\binom{m}{k} |f(x + kt, y + ks) - f(x_0 + kt, y_0 + ks)| L_\zeta(t, s)\, dsdt$$

$$+\left|\iint_{\mathbb{R}^2} K_\zeta\left(t, s, \sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k} f(x_0+kt, y_0+ks)\right) dsdt - f(x_0, y_0)\right|$$
$$= I_1 + I_2.$$

The following inequality is valid for I_1 :

$$|I_1| \leq \iint_{B_\delta} \sum_{k=1}^{m}\binom{m}{k} |f(x+kt, y+ks) - f(x_0+kt, y_0+ks)|$$
$$\times L_\zeta(t,s)\, dsdt$$
$$+ \iint_{\mathbb{R}^2\setminus B_\delta} \sum_{k=1}^{m}\binom{m}{k} |f(x+kt, y+ks) - f(x_0+kt, y_0+ks)|$$
$$\times L_\zeta(t,s)\, dsdt,$$

where

$$B_\delta := \left\{(t,s) \in \mathbb{R}^2 : t^2+s^2 < \delta^2\right\}.$$

Equivalently, we may write

$$I_1 \leq \sum_{k=1}^{m}\binom{m}{k} \int_{-\delta}^{\delta}\int_{-\delta}^{\delta} |f(x+kt, y+ks) - f(x_0+kt, y_0+ks)|\, L_\zeta(t,s)\, dsdt$$
$$+ \iint_{\delta \leq \sqrt{t^2+s^2}} \sum_{k=1}^{m}\binom{m}{k} |f(x+kt, y+ks) - f(x_0+kt, y_0+ks)|\, L_\zeta(t,s)\, dsdt$$
$$= I_{11} + I_{12}.$$

It is easy to see that

$$I_{11} \leq \iint_{\delta \leq \sqrt{t^2+s^2}} \sum_{k=1}^{m}\binom{m}{k} |f(x+kt, y+ks)|\, L_\zeta(t,s)\, dsdt$$
$$+ \iint_{\delta \leq \sqrt{t^2+s^2}} \sum_{k=1}^{m}\binom{m}{k} |f(x_0+kt, y_0+ks)|\, L_\zeta(t,s)\, dsdt$$
$$= I_{111} + I_{112}.$$

Now, we deal with the integral I_{111}. The operations for the integral I_{111} are as follows:

$$I_{111} \leq \sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta(t,s) \left| \iint\limits_{\mathbb{R}^2 \setminus B_\delta} \sum_{k=1}^{m} \binom{m}{k} |f(x+kt, y+ks)| \, ds dt \right|.$$

Furthermore, we proceed as follows:

$$\begin{aligned} I_{111} &\leq \sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta(t,s) \left| \iint\limits_{\mathbb{R}^2} \sum_{k=1}^{m} \binom{m}{k} |f(x+kt, y+ks)| \, ds dt \right| \\ &\leq 2^m \sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta(t,s) \, \|f\|_{L_1(\mathbb{R}^2)} \\ &= 2^m I_{1111}, \end{aligned}$$

where

$$\sum_{k=1}^{m} \binom{m}{k} = \left(2^m - 1\right).$$

It follows that $I_{1111} \to 0$ as $\zeta \to \zeta_0$ by condition (c).

Now, we deal with the integral I_{112}. The operations for the integral I_{112} are as follows:

$$I_{112} \leq \sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta(t,s) \left| \iint\limits_{\mathbb{R}^2 \setminus B_\delta} \sum_{k=1}^{m} \binom{m}{k} |f(x_0+kt, y_0+ks)| \, ds dt \right|.$$

Now, we proceed as follows:

$$\begin{aligned} I_{112} &\leq \sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta(t,s) \left| \iint\limits_{\mathbb{R}^2} \sum_{k=1}^{m} \binom{m}{k} |f(x_0+kt, y_0+ks)| \, ds dt \right| \\ &\leq 2^m \sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta(t,s) \, \|f\|_{L_1(\mathbb{R}^2)} \\ &= 2^m I_{1121}, \end{aligned}$$

where

$$\sum_{k=1}^{m} \binom{m}{k} = \left(2^m - 1\right).$$

It follows that $I_{1121} \to 0$ as $\zeta \to \zeta_0$ by condition (c).

Clearly, by Theorem 21.3.1, $I_2 \to 0$ as ζ tends to ζ_0.

Summarizing all operations, we obtain the following general inequality:

$$I_\zeta(x, y) \le 2^{m+1} \sup_{\delta \le \sqrt{t^2+s^2}} L_\zeta(t, s) \|f\|_{L_1(\mathbb{R}^2)}$$

$$+ \left| \iint_{\mathbb{R}^2} K_\zeta\left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0 + kt, y_0 + ks)\right) ds dt - f(x_0, y_0) \right|$$

$$+ \sum_{k=1}^{m} \binom{m}{k} \int_{-\delta}^{\delta}\int_{-\delta}^{\delta} |f(x + kt, y + ks) - f(x_0 + kt, y_0 + ks)| L_\zeta(t, s) ds dt.$$

Since the last term above is bounded on $Z_{C,\delta,m}$ by the hypothesis, the assertion follows, that is $I_\zeta(x, y) \to 0$ as $(x, y, \zeta) \to (x_0, y_0, \zeta_0)$. Thus, the proof is completed. □

21.5 Rate of Convergence

Theorem 21.5.1 *Suppose that the hypotheses of Theorem 21.3.1 are satisfied. Let*

$$\Delta(\zeta, \delta) = \int_{-\delta}^{\delta}\int_{-\delta}^{\delta} L_\zeta(t, s) \left|\{\mu_1(|t|)\}_t'\right| \left|\{\mu_2(|s|)\}_s'\right| ds dt,$$

where $0 < \delta < \min\{\delta_1, \delta_2\}$, *and the following conditions are satisfied:*

(i) $\Delta(\zeta, \delta) \to 0$ *as* $\zeta \to \zeta_0$ *for some* $\delta > 0$.

(ii) *For every* $\delta > 0$, *we have*

$$\sup_{\delta \le \sqrt{t^2+s^2}} L_\zeta(t, s) = \mathbf{o}(\Delta(\zeta, \delta))$$

as $\zeta \to \zeta_0$.

(iii) *For every* $\delta > 0$, *we have*

$$\iint\limits_{\delta \leq \sqrt{t^2+s^2}} L_\zeta (t,s)\, dsdt = \mathbf{o}(\Delta(\zeta,\delta))$$

as $\zeta \to \zeta_0$.

(iv) *Letting* $\zeta \to \zeta_0$, *we have*

$$\left| \iint\limits_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0,y_0) \right) dsdt - f(x_0,y_0) \right| = \mathbf{o}(\Delta(\zeta,\delta)).$$

Then, at each $[m;\mu]$*-Lebesgue point of* $f \in L_1(\mathbb{R}^2)$, *we have*

$$\left| T_\zeta^{[m]}(f;x_0,y_0) - f(x_0,y_0) \right| = \mathbf{o}(\Delta(\zeta,\delta))$$

as $\zeta \to \zeta_0$.

Proof By the hypotheses of Theorem 21.3.1, we have

$$\left| T_\zeta^{[m]}(f;x_0,y_0) - f(x_0,y_0) \right| \leq 2^m \sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta (t,s) \|f\|_{L_1(\mathbb{R}^2)}$$

$$+ 2^m |f(x_0,y_0)| \iint\limits_{\delta \leq \sqrt{t^2+s^2}} L_\zeta (t,s)\, dsdt$$

$$+ \left| \iint\limits_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0,y_0) \right) dsdt - f(x_0,y_0) \right|$$

$$+ \varepsilon \int_{-\delta}^{\delta}\int_{-\delta}^{\delta} L_\zeta (t,s) \left| \{\mu_1(|t|)\}_t' \right| \left| \{\mu_2(|s|)\}_s' \right| dsdt.$$

The proof is obvious by (i)–(iv). □

Theorem 21.5.2 *Suppose that the hypotheses of Theorem 21.4.1 hold. Let*

$$\Psi_\delta(x,y,\zeta) = \sum_{k=1}^{m} \binom{m}{k} \int_{-\delta}^{\delta}\int_{-\delta}^{\delta} |f(x+kt,y+ks) - f(x_0+kt,y_0+ks)|$$

$$\times L_\zeta (t,s)\, dsdt,$$

where $0 < \delta < \min\{\delta_1, \delta_2\}$, *and the following conditions are satisfied:*

(i) $\Psi_\delta(x, y, \zeta) \to 0$ *as* $(x, y, \zeta) \to (x_0, y_0, \zeta_0)$ *for some* $\delta > 0$.
(ii) *For every* $\delta > 0$, *we have*

$$\sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta\,(t, s) = \mathbf{o}\,(\Psi_\delta(x, y, \zeta))$$

as $(x, y, \zeta) \to (x_0, y_0, \zeta_0)$.
(iii) *Letting* $(x, y, \zeta) \to (x_0, y_0, \zeta_0)$, *we have*

$$\left| \iint_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0 + kt, y_0 + ks) \right) ds dt - f(x_0, y_0) \right|$$
$$= \mathbf{o}\,(\Psi_\delta(x, y, \zeta))\,.$$

Then, at each $[m; \mu]$*-Lebesgue point of* $f \in L_1(\mathbb{R}^2)$, *we have*

$$\left| T_\zeta^{[m]}(f; x, y) - f\,(x_0, y_0) \right| = \mathbf{o}\,(\Psi_\delta(x, y, \zeta))$$

as $(x, y, \zeta) \to (x_0, y_0, \zeta_0)$.

Proof Under the hypotheses of Theorem 21.4.1, we may write

$$\left| T_\zeta^{[m]}(f; x, y) - f\,(x_0, y_0) \right| \leq 2^{m+1} \sup_{\delta \leq \sqrt{t^2+s^2}} L_\zeta\,(t, s)\, \|f\|_{L_1(\mathbb{R}^2)}$$
$$+ \left| \iint_{\mathbb{R}^2} K_\zeta \left(t, s, \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} f(x_0 + kt, y_0 + ks) \right) ds dt - f(x_0, y_0) \right|$$
$$+ \sum_{k=1}^{m} \binom{m}{k} \int_{-\delta}^{\delta} \int_{-\delta}^{\delta} |f(x + kt, y + ks) - f(x_0 + kt, y_0 + ks)|\, L_\zeta\,(t, s)\, ds dt.$$

From conditions (i)–(iii), the proof is completed. □

21.6 Concluding Remarks

The concept of singular integral operators arising from Fourier analysis is widely used in many areas of science, including medicine and engineering. Magnetic resonance imaging (MRI) is the well-known application area. Also, the singular integral operator type of operators are used for approximating the solutions of ordinary and partial differential equations in the case of analytical solution cannot be obtained by known technics. Similarly, nonlinear singular integral operators are also employed for approximating the solutions of nonlinear ordinary and partial differential equations. For some applications, we refer the reader to see [6, 22]. In order to approximate derivatives of the indefinite integrals of the integrable functions in the sense of Lebesgue, one may prefer to use m-singular integral operators.

In this work, we start by giving pointwise convergence result for double m-singular integral operators. Using this auxiliary result, the main result, Fatou type convergence theorem, is presented. Finally, corresponding rate of convergences are computed.

References

1. S.E. Almali, A.D. Gadjiev, On approximation properties of certain multidimensional nonlinear integrals. J. Nonlinear Sci. Appl. **9**(5), 3090–3097 (2016)
2. A. Aral, V. Gupta, R.P. Agarwal, *Applications of q-Calculus in Operator Theory* (Springer, New York, 2013)
3. C. Bardaro, On approximation properties for some classes of linear operators of convolution type. Atti Sem. Mat. Fis. Univ. Modena **33**(2), 329–356 (1984)
4. C. Bardaro, J. Musielak, G. Vinti, *Nonlinear Integral Operators and Applications*. De Gruyter Series in Nonlinear Analysis and Applications, vol. 9 (Walter de Gruyter & Co., Berlin, 2003)
5. C. Bardaro, H. Karsli, G. Vinti, On pointwise convergence of Mellin type nonlinear m-singular integral operators. Comm. Appl. Nonlinear Anal. **20**(2), 25–39 (2013)
6. R. Bracewell, *The Fourier Transform and its Applications*, 3rd edn. (McGraw-Hill Science, New York, 1999)
7. P.L. Butzer, R.J. Nessel, *Fourier Analysis and Approximation*, vol. I (Academic, New York, 1971)
8. A.P. Calderón, A. Zygmund, Singular integral operators and differential equations. Am. J. Math. **79**, 901–921 (1957)
9. P. Fatou, Series trigonometriques et series de Taylor. Acta Math. **30**(1), 335–400 (1906)
10. A.D. Gadjiev, On nearness to zero of a family of nonlinear integral operators of Hammerstein. Izv. Akad. Nauk Azerbaĭdžan, SSR Ser. Fiz.-Tehn. Mat. Nauk **2**, 32–34 (1966)
11. A.D. Gadjiev, The order of convergence of singular integrals which depend on two parameters, in *Special Problems of Functional Analysis and Their Applications to the Theory of Differential Equations and the Theory of Functions*. Izdat. Akad. Nauk Azerbaidžan. SSR., Baku 40–44 (1968)
12. E.G. Ibrahimov, On approximation order of functions by m-singular Gegenbauer integrals. Applied problems of mathematics and mechanics (Baku 2002). Proc. Inst. Math. Mech. Natl. Acad. Sci. Azerb. **17**(205), 71–77 (2002)
13. E.J. Ibrahimov, S.A. Jafarova, On convergence and convergence order of Gegenbauer's m-singular integrals. Proc. A. Razmadze Math. Inst. **159**, 21–42 (2012)
14. H. Karsli, Fatou type convergence of nonlinear m-singular integral operators. Appl. Math. Comput. **246**, 221–228 (2014)
15. H. Lebesgue, Sur les intégrales singulières. Annales de la faculté des sciences de Toulouse Sér. **3**(1), 25–117 (1909)

16. R.G. Mamedov, On the order of convergence of m-singular integrals at generalized Lebesgue points and in the space $L_p(-\infty, \infty)$. Izv. Akad. Nauk SSSR Ser. Mat. **27**(2), 287–304 (1963)
17. S.G. Mikhlin, S. Prössdorf, *Singular Integral Operators*. Translated from the German by Albrecht Böttcher and Reinhard Lehmann (Springer, Berlin, 1986)
18. J. Musielak, On some approximation problems in modular spaces, in *Constructive Function Theory, Proceedings of International Conference Varna*, 1–5 June, 1981. Publication House of Bulgarian Academic of Sciences, Sofia 455–461 (1983)
19. J. Musielak, Approximation by nonlinear singular integral operators in generalized Orlicz spaces. Comment. Math. Prace Mat. **31**, 79–88 (1991)
20. J. Musielak, Nonlinear integral operators and summability in $\mathbb{R}^2$. Atti Sem. Mat. Fis. Univ. Modena **48**(1), 249–257 (2000)
21. I.P. Natanson, *Theory of Functions of a Real Variable*, vol. 2. Translated from the Russian by Leo F. Boron. (Frederick Ungar Pub. Co., New York, 1960)
22. M. Papadopoulos, The use of singular integrals in wave propagation problems, with application to the point load source in a semi-infinite elastic medium. Proc. R. Soc. **A276**, 204–237 (1963)
23. B. Rydzewska, Approximation des fonctions par des intégrales singulières ordinaires. Fasc. Math.**7**, 71–81 (1973)
24. B. Rydzewska, Approximation des fonctions de deux variables par des intégrales singulières doubles. Fasc. Math. **8**, 35–45 (1974)
25. B. Rydzewska, Point-approximation des fonctions par des certaines intégrales singuliéres. Fasc. Math. **10**, 13–24 (1978)
26. S. Saks, *Theory of the Integral* (G.E. Stechert & Co., New York, 1937)
27. S. Siudut, On the convergence of double singular integrals. Comment. Math. Prace Mat. **28**(1), 143–146 (1988)
28. S. Siudut, A theorem of Romanovski type for double singular integrals. Comment. Math. Prace Mat. **29**, 277–289 (1989)
29. S. Siudut, On the Fatou type convergence of abstract singular integrals. Comment. Math. Prace Mat. **30**(1) 171–176 (1990)
30. E.M. Stein, *Singular Integrals and Differentiability Properties of Functions* (Princeton Univ. Press, New Jersey, 1970)
31. E.M. Stein, G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces* (Princeton Univ. Press, New Jersey, 1971)
32. T. Swiderski, E. Wachnicki, Nonlinear singular integrals depending on two parameters. Comment. Math. **40**, 181–189 (2000)
33. R. Taberski, Singular integrals depending on two parameters. Prace Mat. **7**, 173–179 (1962)
34. R. Taberski, On double integrals and Fourier series. Ann. Polon. Math. **15**, 97–115 (1964)
35. G. Uysal, Nonlinear m-singular integral operators in the framework of Fatou type weighted convergence. Commun. Fac. Sci. Univ. Ank. Sér. A1 Math. Stat. **67**(1), 262–276 (2018)
36. G. Uysal, On nonlinear bivariate $[m_1, m_2]-$singular integral operators, Math. Methods Appl. Sci. 1–13 (2018). https://doi.org/10.1002/mma.5425
37. G. Uysal, M.M. Yilmaz, E. Ibikli, On pointwise convergence of bivariate nonlinear singular integral operators. Kuwait J. Sci. **44**(2), 46–57 (2017)

Chapter 22
A Survey on p-Adic Integrals

Ugur Duran and Hemen Dutta

Abstract The p-adic numbers are a counterintuitive arithmetic system and were firstly introduced circa end of the nineteenth century. In conjunction with the introduction of these numbers, many mathematicians and physicists started to develop new scientific tools using their available, useful, and applicable properties. Several effects of these researches have emerged in mathematics and physics such as p-adic analysis, string theory, p-adic quantum mechanics, quantum field theory, representation theory, algebraic geometry, complex systems, dynamical systems, and genetic codes. One of the important tools of the mentioned advancements is the p-adic integrals. Intense research activities in such an area like p-adic integrals are principally motivated by their significance in p-adic analysis. Recently, p-adic integrals and its diverse extensions have been studied and investigated in detail by many mathematicians. This chapter considers and investigates multifarious extensions of the p-adic integrals elaborately. q-Analogues with diverse extensions of p-adic integrals are also considered such as the weighted p-adic q-integral on $\mathbb{Z}_p$. The two types of the weighted q-Boole polynomials and numbers are introduced and investigated in detail. As several special polynomials and numbers can be derived from the p-adic integrals, some generalized and classical q-polynomials and numbers are obtained from the aforesaid extensions of p-adic integrals. Finally, the importance of these extensions is analyzed.

U. Duran (✉)
Department of the Basic Concepts of Engineering, Faculty of Engineering and Natural Sciences, İskenderun Technical University, Hatay, Turkey
e-mail: ugur.duran@iste.edu.tr

H. Dutta
Department of Mathematics, Gauhati University, Guwahati, India

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_22

22.1 Introduction

Arnt Volkenborn (cf. [29, 30]) invented p-adic integral with respect to the following Haar measure about the end of the third quarter of the twentieth century. Despite there have been so many scientific studies for these topics since more than four decades (cf. [4, 9, 11, 12, 17, 18, 20, 22, 24–30] and see also the references cited in each of these earlier works), the Volkenborn integral is today a hot topic and still keeps its mystery. Moreover, it is penetrated multifarious mathematical research areas such as special functions, the functional equations of zeta functions, number theory, Stirling numbers, Mittag-Leffler function, and Mahler theory of integration with respect to the ring $\mathbb{Z}_p$ in conjunction with Iwasawa's p-adic L functions.

The fermionic p-adic invariant integral is firstly considered by Taekyun Kim (cf. [14]), a Korean mathematician, in order to investigate several special numbers and polynomials which can be represented by the fermionic p-adic integrals (cf. [1–3, 5–8, 10, 14–16, 18, 19, 21, 23, 25] and see the references cited therein). Then, this integral has been more common, and it is used in many mathematical fields.

In this study, we firstly focus on the Volkenborn integral and fermionic p-adic integral with their properties and reflections on the special polynomials and numbers. We then consider the q-extensions of the aforementioned integrals (p-adic q-integral and fermionic p-adic q-integral) and use these integrals in order to define q-analogues of the classical special polynomials and numbers.

22.2 p-Adic Integrals on $\mathbb{Z}_p$

We study the two-type p-adic integrals: the first is based on the Haar measure and the second is the fermionic p-adic invariant integral.

Imagine that p be a fixed prime number. Throughout this part, $\mathbb{Z}_p$, $\mathbb{Q}$, $\mathbb{Q}_p$, and $\mathbb{C}_p$ will denote the ring of p-adic rational integers, the field of rational numbers, the field of p-adic rational numbers, and the completion of algebraic closure of $\mathbb{Q}_p$, respectively. Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ and $\mathbb{N}^* = \mathbb{N} \cup \{0\}$. The normalized absolute value according to the theory of p-adic analysis is given by $|p|_p = p^{-1}$ (cf. [1–30]).

22.2.1 Volkenborn Integral and Its Some Properties

The Volkenborn integral and its several generalizations have been used to introduce and research some special polynomials and numbers such as Bernoulli, Daehee polynomials and numbers, cf. [4, 9, 11, 12, 17, 18, 20, 22, 24–30] and see also the references cited therein.

Definition 22.1 (Koblitz [22]) Let $f \in C^1(\mathbb{Z}_p \to \mathbb{C}_p) = \{f | f : X \to \mathbb{C}_p$ is continuous$\}$. The uncertain sum of f is denoted Sf and is defined as

$$Sf(n) = \sum_{j=0}^{n} f(j) \qquad (n \in \mathbb{N}).$$

If $f \in C^1(\mathbb{Z}_p \to \mathbb{C}_p)$, then we get $Sf \in C^1(\mathbb{Z}_p \to \mathbb{C}_p)$.

Here we present the definition of the Volkenborn integral and its some properties.

Definition 22.2 (Koblitz [22]) Let $f \in C^1(\mathbb{Z}_p \to \mathbb{C}_p)$ be a function from the p-adic integers taking values in the p-adic numbers. The Volkenborn integral is defined by the limit, if it exists:

$$\int_{\mathbb{Z}_p} f(x)\, d\mu(x) = \lim_{n\to\infty} \frac{1}{p^n} \sum_{x=0}^{p^n-1} f(x). \tag{22.2.1}$$

Some properties of Volkenborn integral are given in theorem below.

Theorem 22.1 (Koblitz [22]) *Let* $d\mu(x) = \mu_{Haar}(x + p^n\mathbb{Z}_p) = \frac{1}{p^n}$. *The following expressions hold:*

$$\int_{\mathbb{Z}_p} f(x)\, d\mu(x) = \lim_{n\to\infty} \frac{Sf(p^n) - Sf(0)}{p^n} = (Sf)'(0).$$

$$\int_{\mathbb{Z}_p} f(x)\, d\mu(x) = \lim_{n\to\infty} \frac{f(0) + f(1) + \cdots + f(p^n - 1)}{p^n}$$

$$\int_{\mathbb{Z}_p} (f(x+1) - f(x))\, d\mu(x) = f'(0) \tag{22.2.2}$$

$$\int_{\mathbb{Z}_p} (f(x+n) - f(x))\, d\mu(x) = \sum_{l=0}^{n-1} f'(l),$$

where

$$(Sf)'(x) = \frac{d(Sf(x))}{dx}.$$

By means of the useful property (22.2.2) of the Volkenborn integral, generating function of a lot of special number type can be found. For example, to get well-known Bernoulli numbers if we take $f(x) = e^{xt}$ in formula (22.2.2), we then have

$$\int_{\mathbb{Z}_p} \left(e^{(x+1)t} - e^{xt}\right) d\mu(x) = t.$$

In conjunction with some mathematical computations, we get

$$\int_{\mathbb{Z}_p} e^{xt} d\mu(x) = \sum_{n=0}^{\infty} B_n \frac{t^n}{n!}.$$

which means, by comparing the coefficient $\frac{t^n}{n!}$ of both sides,

$$\int_{\mathbb{Z}_p} x^n d\mu(x) = B_n.$$

Also choosing $f(x) = (1+t)^{x+y}$ in Volkenborn integral in relation (22.2.2) gives the generating function of the Daehee polynomials given by (cf. [18])

$$\sum_{n=0}^{\infty} D_n(x) \frac{t^n}{n!} = \int_{\mathbb{Z}_p} (1+t)^{x+y} d\mu(y).$$

The main aim of this Volkenborn integral representations of special numbers and polynomials is to get more formulas and identities for the related special numbers and polynomials by making use of the good and useful properties of the Volkenborn integral.

22.2.2 Fermionic p-Adic Integral and Its Some Properties

The fermionic p-adic integral and its diverse extensions have been used to define and explore many special polynomials and numbers such as Euler, Genocchi, Frobenius-Euler, Changhee, Boole polynomials and numbers, cf. [1–3, 5–8, 10, 14–16, 18, 19, 21–23, 25] and see also the references cited therein.

Supposing that p be a fixed odd prime number. Here we give fermionic p-adic integral defined by Kim, South Korean, [14] and its some properties. Then, the Euler numbers and Genocchi numbers are presented by means of the fermionic p-adic integral.

Definition 22.3 (Kim [13]) Let p be an odd prime number and $f \in C^1(\mathbb{Z}_p \to \mathbb{C}_p)$ be a function from the p-adic integers taking values in the p-adic numbers. The fermionic integral of f is defined by the limit, if it exists

$$\int_{\mathbb{Z}_p} f(x) d\mu_{-1}(x) = \lim_{n\to\infty} \sum_{x=0}^{p^n-1} (-1)^x f(x).$$

Theorem 22.2 (cf. [13] and [15]) *Let* $d\mu_{-1}(x) = \mu_{-1}\left(x + p^n\mathbb{Z}_p\right) = (-1)^x$. *The following identities hold true*

$$\int_{\mathbb{Z}_p} (f(x+1) + f(x))\, d\mu_{-1}(x) = 2f(0) \tag{22.2.3}$$

$$\int_{\mathbb{Z}_p} \left(f(x+n) + (-1)^{n-1} f(x)\right) d\mu_{-1}(x) = 2\sum_{l=0}^{n-1} (-1)^{n-1-l} f(l). \tag{22.2.4}$$

By using fermionic p-adic integral generating functions of not only Euler type numbers and polynomials but also Genocchi type numbers and polynomials can be derived. For instance, to obtain the generating function of Euler numbers if we take $f(x) = e^{xt}$ in Eq. (22.2.3), we then have

$$\int_{\mathbb{Z}_p} \left(e^{(x+1)t} + e^{xt}\right) d\mu_{-1}(x) = 2,$$

from which, we deduce that

$$\int_{\mathbb{Z}_p} e^{xt} d\mu_{-1}(x) = \frac{2}{e^t + 1},$$

whose right side gives the generating function of the Euler numbers. By motivating the applications above, we observe that

$$\sum_{n=0}^{\infty} \left(\int_{\mathbb{Z}_p} x^n d\mu_{-1}(x)\right) \frac{t^n}{n!} = \sum_{n=0}^{\infty} E_n \frac{t^n}{n!},$$

which yields to

$$\int_{\mathbb{Z}_p} x^n d\mu_{-1}(x) = E_n \tag{22.2.5}$$

that means the Euler numbers can be shown by means of the fermionic p-adic integral (cf. [14]).

Note that

$$E_n(x) = \frac{G_{n+1}(x)}{n+1},$$

which also implies

$$E_n = \frac{G_{n+1}}{n+1}.$$

Thus, we can write

$$\frac{G_{n+1}}{n+1} = \int_{\mathbb{Z}_p} x^n d\mu_{-1}(x).$$

We refer the reader to look at Refs. [1–30] in order to see some generalizations of the Volkenborn integral and fermionic integral in conjunction with some of their applications.

It is well known that the usual Frobenius-Euler polynomials $H_n(x)$ for $\lambda \in \mathbb{C}$ with $|\lambda| > 1$ were defined by the power series expansion at $t = 0$:

$$\sum_{n=0}^{\infty} H_n(x)\frac{t^n}{n!} = \frac{1-\lambda}{e^t-\lambda}e^{xt}. \tag{22.2.6}$$

Taking $x = 0$ in Eq. (22.2.6), we have $H_n(0) := H_n$ that is widely known as n-th Frobenius-Euler number cf. [3].

In [2], the Frobenius-Euler polynomials are defined by the following p-adic fermionic integral on $\mathbb{Z}_p$, with respect to μ_{-1}:

$$H_n(x \mid -\lambda^{-1}) = \frac{\lambda+1}{2}\int_{\mathbb{Z}_p} \lambda^y (x+y)^n d\mu_{-1}(y). \tag{22.2.7}$$

Upon setting $x = 0$ into Eq. (22.2.7) gives $H_n(0) := H_n$ which are called n-th Frobenius-Euler number.

Moreover, the Changhee and the Boole polynomials can be shown by the fermionic p-adic invariant integral as follows (cf. [2, 5, 7, 8, 19, 21, 23, 28]):

$$\sum_{n=0}^{\infty} Ch_n(x)\frac{t^n}{n!} = \int_{\mathbb{Z}_p} (1+t)^{x+y} d\mu_{-1}(y)$$

and

$$\sum_{n=0}^{\infty} Bl_n(x \mid \omega)\frac{t^n}{n!} = \frac{1}{2}\int_{\mathbb{Z}_p} (1+t)^{x+\omega y} d\mu_{-1}(y) \quad \left(\omega \in \mathbb{Z}_p\right). \tag{22.2.8}$$

22.3 p-Adic q-Integrals on $\mathbb{Z}_p$

We study the two q-extensions of the p-adic integrals: the first is bosonic p-adic q-integral on $\mathbb{Z}_p$ (or called q-Volkenborn integral on $\mathbb{Z}_p$) with respect to the q-Haar measure and the other is the fermionic p-adic q-integral on $\mathbb{Z}_p$.

Let $\mathbb{N} := \{1, 2, 3, \ldots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Along this section, $\mathbb{Z}$ denotes the set of integers, $\mathbb{R}$ denotes the set of real numbers, and $\mathbb{C}$ denotes the set of complex numbers. Let p be chosen as a fixed prime number. The symbols $\mathbb{Z}_p$, $\mathbb{Q}_p$, and $\mathbb{C}_p$ indicate the ring of p-adic rational integers, the field of p-adic rational numbers, and the completion of an algebraic closure of $\mathbb{Q}_p$, respectively. The normalized absolute value according to the theory of p-adic analysis is given by $|p|_p = p^{-1}$. The parameter q can be considered as an indeterminate, a complex number $q \in \mathbb{C}$ with $|q| < 1$, or a p-adic number $q \in \mathbb{C}_p$ with $|q-1|_p < p^{-\frac{1}{p-1}}$ and $q^x = \exp(x \log q)$ for $|x|_p \leq 1$. The q-analogue of x is defined by $[x]_q = (1-q^x)/(1-q)$ (see [1–7, 9–12, 15–18, 20, 21, 23, 24, 26] for more details about q-numbers). It is easy to show that $\lim_{q\to 1}[x]_q = x$ for any x with $|x|_p \leq 1$ in the p-adic case (for details, cf. [1–30]; see also the related references cited therein).

22.3.1 *q*-Volkenborn Integral and Its Some Properties

We say that f is uniformly differentiable function at a point $a \in \mathbb{Z}_p$, which is denoted by $f \in UD(\mathbb{Z}_p)$. From here, Kim defined the q-Volkenborn integral or bosonic p-adic q-integral on $\mathbb{Z}_p$ of a function $f \in UD(\mathbb{Z}_p)$ in [12] as follows:

$$I_q(f) = \int_{\mathbb{Z}_p} f(x)\, d\mu_q(x) = \lim_{N\to\infty} \frac{1}{[p^N]_q} \sum_{x=0}^{p^N-1} f(x) q^x, \tag{22.3.1}$$

where $\mu_q(x)$ implies the q-Haar measure given by

$$\mu_q\left(a + p^N \mathbb{Z}_p\right) = \frac{q^a}{[p^N]_q}.$$

The most famous two properties of the q-Volkenborn integral are (cf. [12])

$$qI_q(f_1) = I_q(f) + (q-1) f(0) + \frac{q-1}{\log q} f'(0)$$

and

$$q^n I_q(f_n) = I_q(f) + (q-1)\sum_{r=0}^{n-1} q^r f(r) + \frac{q-1}{\log q}\sum_{r=0}^{n-1} f'(r)$$

where $f_n(x) = f(x+n)$. For these related issues, see [4, 9, 11, 12, 17, 20, 24, 26] and related references cited therein.

When $q \to 1$, the integral $I_q(f)$ reduces to the usual Volkenborn integral $I_1(f) := I(f)$, see (22.2.1).

The q-Daehee numbers $D_{n,q}$ and q-Daehee polynomials $D_{n,q}(x)$ are defined by means of q-Volkenborn integrals (cf. [4, 20]):

$$D_{n,q} = \int_{\mathbb{Z}_p} (x)_n \, d\mu_q(x) \text{ and } D_{n,q}(x) = \int_{\mathbb{Z}_p} (x+y)_n \, d\mu_q(y) \quad (n \geq 0),$$

where the symbol $(x)_n$ denotes the falling factorial given by (cf. [4, 5, 7, 8, 18–21, 23, 24, 28]):

$$(x)_n = x(x-1)(x-2)\cdots(x-n+1). \tag{22.3.2}$$

The falling factorial $(x)_n$ has the following summation representation:

$$(x)_n = \sum_{k=0}^{n} S_1(n,k)\, x^k, \tag{22.3.3}$$

where $S_1(n,k)$ denotes the Stirling number of the first kind, see [4, 5, 7, 8, 18–21, 23, 24, 28]. The Stirling numbers of second kind are also defined by (cf. [4, 5, 7, 8, 18–21, 23, 24, 28]):

$$\frac{(e^t-1)^k}{k!} = \sum_{n=k}^{\infty} S_2(k,n)\, \frac{t^n}{n!}. \tag{22.3.4}$$

It is obvious that $\lim_{q\to 1} D_{n,q} := D_n$ and $\lim_{q\to 1} D_{n,q}(x) := D_n(x)$, where D_n and $D_n(x)$ denote the classical Daehee numbers and polynomials, respectively.

The q-Bernoulli numbers and polynomials are introduced by the following q-Volkenborn integrals (cf. [4, 20, 24]):

$$B_{n,q} = \int_{\mathbb{Z}_p} x^n d\mu_q(x) \text{ and } B_{n,q}(x) = \int_{\mathbb{Z}_p} (x+y)^n \, d\mu_q(y) \quad (n \geq 0).$$

The q-Daehee and q-Bernoulli polynomials and numbers and their various generalizations have been studied by many mathematicians, cf. [4, 9, 11, 17, 20, 24, 26]; see also the related references cited therein.

22.3.2 Fermionic p-Adic q-Integral and Its Some Properties

In this part, let p be chosen as an odd fixed prime number. The symbols $\mathbb{Z}_p$, $\mathbb{Q}_p$, and $\mathbb{C}_p$ indicate the ring of p-adic rational integers, the field of p-adic rational numbers, and the completion of an algebraic closure of $\mathbb{Q}_p$, respectively. The normalized absolute value according to the theory of p-adic analysis is given by $|p|_p = p^{-1}$. The parameter q can be considered as an indeterminate, a complex number $q \in \mathbb{C}$ with $|q| < 1$, or a p-adic number $q \in \mathbb{C}_p$ with $|q-1|_p < p^{-\frac{1}{p-1}}$ and $q^x = \exp(x \log q)$ for $|x|_p \leq 1$. The q-analogue of x is defined by $[x]_q = (1-q^x)/(1-q)$. It is easy to show that $\lim_{q\to 1}[x]_q = x$ for any x with $|x|_p \leq 1$ in the p-adic case (for details, cf. [1–30]; see also the related references cited therein).

Let f be uniformly differentiable function at a point $a \in \mathbb{Z}_p$, denoted by $f \in UD(\mathbb{Z}_p)$. Kim [13] originally introduced the fermionic q-Volkenborn integral (or fermionic p-adic q-integral on $\mathbb{Z}_p$) of a function $f \in UD(\mathbb{Z}_p)$, as follows:

$$I_{-q}(f) = \int_{\mathbb{Z}_p} f(x)\, d\mu_{-q}(x) = \lim_{N\to\infty} \frac{1}{[p^N]_{-q}} \sum_{k=0}^{p^N-1} (-1)^k f(k)\, q^k. \tag{22.3.5}$$

Let $f_1(x) = f(x+1)$. By (22.3.5), the following interesting integral equation holds true:

$$qI_{-q}(f_1) + I_{-q}(f) = [2]_q\, f(0),$$

which intensely preserves usability in introducing diverse generalizations of several special polynomials such as Euler polynomials, q-Euler polynomials with their assorted extensions, and the families of Changhee polynomials. As a general case of (22.2.4), Kim [13] gave the following integral equation for $f_n(x) = f(x+n)$:

$$q^n I_{-q}(f_n) + (-1)^{n-1} I_{-q}(f) = [2]_q \sum_{r=0}^{n-1} (-1)^{n-r-1} q^r f(r).$$

In [2], Araci et al. defined the q-analogue of Changhee polynomials in terms of the fermionic p-adic q-integral:

$$Ch_{n,q}(x) = \int_{\mathbb{Z}_p} q^{-y} (x+y)_n\, d\mu_{-q}(y) \quad \left(n \geqq 0\right).$$

When $x = 0$, then it yields $Ch_{n,q}(0) := Ch_{n,q}$ being called n-th q-Changhee number. It is obvious that $\lim_{q\to 1} Ch_{n,q}(x) := Ch_n(x)$.

For $n \geqq 0$, the q-Changhee polynomials of the second kind are defined as follows:

$$\widehat{Ch}_{n,q}(x) = \int_{\mathbb{Z}_p} q^{-y} (-x-y)_n\, d\mu_{-q}(y). \tag{22.3.6}$$

Upon setting $x = 0$ in Eq. (22.3.6) yields $\widehat{Ch}_{n,q}(0) := \widehat{Ch}_{n,q}$ being called q-Changhee numbers of the second kind (cf. [2]; see also the related references cited therein).

In [21], Kim et al. considered the q-extension of Boole polynomials the first and second kinds by means of the fermionic q-Volkenborn integrals:

$$Bl_{n,q}\left(x\,|\omega\right) = [2]_q^{-1} \int_{\mathbb{Z}_p} (x + \omega y)_n \, d\mu_{-q}\left(y\right) \quad \left(\omega \in \mathbb{Z}_p \text{ and } n \geqq 0\right), \tag{22.3.7}$$

and

$$\widehat{Bl}_{n,q}\left(x\,|\omega\right) = [2]_q^{-1} \int_{\mathbb{Z}_p} (-x - \omega y)_n \, d\mu_{-q}\left(y\right) \quad \left(\omega \in \mathbb{Z}_p \text{ and } n \geqq 0\right). \tag{22.3.8}$$

Substituting $x = 0$ in Eqs. (22.3.7) and (22.3.8), the polynomials above reduce to the corresponding numbers, namely $Bl_{n,q}\left(0\,|\omega\right) := Bl_{n,q}\left(\omega\right)$ being called n-th q-Boole number of the first kind and $\widehat{Bl}_{n,q}\left(0\,|\omega\right) := \widehat{Bl}_{n,q}\left(\omega\right)$ being called n-th q-Boole number of the second kind. It is readily seen that $\lim_{q\to 1} Bl_{n,q}\left(x\,|\omega\right) := Bl_n\left(x\,|\omega\right)$. In recent years, the Changhee and Boole polynomials and its many generalizations with applications in p-adic analysis and q-analysis have been studied by diverse mathematicians as well, cf. [2, 5, 7, 8, 21, 23].

22.4 Weighted p-Adic Integrals on $\mathbb{Z}_p$

In this section, we present the weighted p-adic q-integral and the weighted fermionic p-adic q-integral and analyze some of their fundamental properties. We then provide the types of the q-Daehee polynomials with weight (α, β) arising from the weighted p-adic q-integral on $\mathbb{Z}_p$ and the two types of the q-Changhee polynomials derived from the weighted fermionic p-adic q-integral on $\mathbb{Z}_p$ with their diverse identities, relations, and formulas. We finally introduce two generalizations of q-Boole numbers and polynomials called q-Boole polynomials and numbers with weight (α, β) and q-Boole polynomials and numbers of second kind with weight (α, β) by means of the weighted fermionic p-adic q-integral on $\mathbb{Z}_p$. Moreover, we acquire multifarious novel and interesting formulas and relations including recurrence relation, symmetric relations and many correlations related to the weighted q-Euler polynomials, the Apostol type weighted q-Euler polynomials, familiar Stirling numbers of first and second kinds, and λ-Stirling numbers of the second kind.

22.4.1 (α, β)-Volkenborn Integral and Its Some Properties

In [4], Araci et al. considered a generalization of Kim's p-adic q-integral on $\mathbb{Z}_p$ including the parameters α and β, as follows:

$$I_q^{(\alpha,\beta)}(f) = \int_{\mathbb{Z}_p} q^{-\beta x} f(x)\, d\mu_{q^\alpha}(x) = \lim_{N\to\infty} \frac{1}{\left[p^N\right]_{q^\alpha}} \sum_{k=0}^{p^N-1} f(k)\, q^{(\alpha-\beta)k}. \tag{22.4.1}$$

Remark 22.1 In the case when $\alpha = 1$ and $\beta = 0$ in Eq. (22.4.1), we have $I_q^{(1,0)}(f) := I_q(f)$ in (22.3.1).

Remark 22.2 As q goes to 1 in Eq. (22.4.1), we have Volkenborn integral in (22.2.1).

By (22.4.1), for, the following relation holds true for $f_n(x) = f(x+n)$ (cf. [4]):

$$q^{(\alpha-\beta)n} I_q^{(\alpha,\beta)}(f_n) - I_q^{(\alpha,\beta)}(f) = \frac{[\alpha]_q}{\alpha}\left((q-1)\sum_{r=0}^{n-1} q^{(\alpha-\beta)r} f(r) + \frac{q-1}{\log q}\sum_{r=0}^{n-1} q^{(\alpha-\beta)r} f'(r)\right). \tag{22.4.2}$$

22.4.1.1 The q-Daehee Polynomials with Weight (α, β)

Taking $f(x) = (x+y)_n$ in (22.4.1), we then have

$$I_q^{(\alpha,\beta)}((x+y)_n) = \int_{\mathbb{Z}_p} q^{-\beta y}(x+y)_n\, d\mu_{q^\alpha}(y). \tag{22.4.3}$$

It follows from Eq. (22.4.3) that

$$\sum_{n=0}^{\infty} \int_{\mathbb{Z}_p} q^{-\beta y}(x+y)_n\, d\mu_{q^\alpha}(y) \frac{t^n}{n!} = \int_{\mathbb{Z}_p} q^{-\beta y}\left(\sum_{n=0}^{\infty}\binom{x+y}{n} t^n\right) d\mu_{q^\alpha}(y) = \int_{\mathbb{Z}_p} q^{-\beta y}(1+t)^{x+y}\, d\mu_{q^\alpha}(y). \tag{22.4.4}$$

Since

$$\int_{\mathbb{Z}_p} q^{-\beta y} (1+t)^{x+y} d\mu_{q^\alpha}(y) = \frac{(q-1)\frac{[\alpha]_q}{\alpha}(\alpha-\beta)+\frac{[\alpha]_q}{\alpha}\frac{q-1}{\log q}\log(1+t)}{tq^{\alpha-2\beta}+q^{\alpha-2\beta}-1}(1+t)^x, \qquad (22.4.5)$$

we are able to state the following definition.

The q-Daehee polynomials with weight (α, β) are defined by the following generating function to be:

$$\sum_{n=0}^{\infty} D_{n,q}^{(\alpha,\beta)}(x)\frac{t^n}{n!} = \frac{(q-1)\frac{[\alpha]_q}{\alpha}(\alpha-\beta)+\frac{[\alpha]_q}{\alpha}\frac{q-1}{\log q}\log(1+t)}{tq^{\alpha-2\beta}+q^{\alpha-2\beta}-1}(1+t)^x. \qquad (22.4.6)$$

From Eqs. (22.4.5) and (22.4.6), we get

$$\sum_{n=0}^{\infty} D_{n,q}^{(\alpha,\beta)}(x)\frac{t^n}{n!} = \int_{\mathbb{Z}_p} q^{-\beta y}(1+t)^{x+y} d\mu_{q^\alpha}(y). \qquad (22.4.7)$$

Upon setting $x = 0$ in (22.4.7), we have $D_{n,q}^{(\alpha,\beta)}(0) := D_{n,q}^{(\alpha,\beta)}$ which are called q-Daehee numbers with weight (α, β) shown by

$$\sum_{n=0}^{\infty} D_{n,q}^{(\alpha,\beta)}\frac{t^n}{n!} = \int_{\mathbb{Z}_p} q^{-\beta y}(1+t)^{y} d\mu_{q^\alpha}(y).$$

From Eqs. (22.3.3) and (22.4.7), it can be easily shown that (cf. [4])

$$D_{n,q}^{(\alpha,\beta)}(x) = \sum_{k=0}^{n} S_1(n,k)\, B_{k,q}^{(\alpha,\beta)}(x),$$

where $B_{k,q}^{(\alpha,\beta)}(x)$ are a new generalization of q-Bernoulli polynomials given by

$$B_{k,q}^{(\alpha,\beta)}(x) = \int_{\mathbb{Z}_p} q^{-\beta y}(x+y)^k d\mu_{q^\alpha}(y).$$

Furthermore, in the case $x = 0$, we obtain

$$D_{n,q}^{(\alpha,\beta)} = \sum_{k=0}^{n} S_1(n,k)\, B_{k,q}^{(\alpha,\beta)}$$

where $B_{k,q}^{(\alpha,\beta)}(0) := B_{k,q}^{(\alpha,\beta)}$ are a new generalization of q-Bernoulli numbers.

There is a relationship among $D_{n,q}^{(\alpha,\beta)}(x)$, $B_{n,q}^{(\alpha,\beta)}(x)$, and $S_2(n,l)$ (cf. [4]):

$$B_{n,q}^{(\alpha,\beta)}(x) = \sum_{k=0}^{n} D_{k,q}^{(\alpha,\beta)}(x) S_2(n,k).$$

The q-Daehee numbers of second kind with weight (α,β) are defined by the following the weighted p-adic q-integral on $\mathbb{Z}_p$:

$$\widehat{D}_{n,q}^{(\alpha,\beta)} = \int_{\mathbb{Z}_p} q^{-\beta y} (-y)_n \, d\mu_{q^\alpha}(y) \quad (n \in \mathbb{N}_0). \tag{22.4.8}$$

By (22.3.3) and (22.4.8), one can get the following result (cf. [4]):

$$\widehat{D}_{n,q}^{(\alpha,\beta)} = \sum_{k=0}^{n} S_1(n,k)(-1)^k B_{k,q}^{(\alpha,\beta)}.$$

The generating function of the weighted q-analogue of Daehee numbers of the second kind is defined in [4] as follows:

$$\begin{aligned}\sum_{n=0}^{\infty} \widehat{D}_{n,q}^{(\alpha,\beta)} \frac{t^n}{n!} &= \sum_{n=0}^{\infty} \left(\int_{\mathbb{Z}_p} q^{-\beta y} (-y)_n \, d\mu_{q^\alpha}(y) \right) \frac{t^n}{n!} \\ &= \int_{\mathbb{Z}_p} q^{-\beta y} \left(\sum_{n=0}^{\infty} \binom{-y}{n} t^n \right) d\mu_{q^\alpha}(y) \\ &= \int_{\mathbb{Z}_p} q^{-\beta y} (1+t)^{-y} \, d\mu_{q^\alpha}(y).\end{aligned}$$

Also, from (22.4.8), we have

$$\int_{\mathbb{Z}_p} q^{-\beta y} (1+t)^{-y} \, d\mu_{q^\alpha}(y) = (q-1) \frac{[\alpha]_q}{\alpha} \frac{\alpha - \beta - \frac{\log(1+t)}{\log q}}{q^{\alpha-2\beta} - t - 1} (1+t).$$

Thus, it can be readily seen that

$$\sum_{n=0}^{\infty} \widehat{D}_{n,q}^{(\alpha,\beta)} \frac{t^n}{n!} = (q-1) \frac{[\alpha]_q}{\alpha} \frac{\alpha - \beta - \frac{\log(1+t)}{\log q}}{q^{\alpha-2\beta} - t - 1} (1+t).$$

Now also the q-Daehee polynomials with weight (α, β) of second kind are defined by means of the following generating function (cf. [4]):

$$(q-1)\,\frac{[\alpha]_q}{\alpha}\,\frac{\alpha-\beta-\frac{\log(1+t)}{\log q}}{q^{\alpha-2\beta}-t-1}\,(1+t)^{1-x}=\sum_{n=0}^{\infty}\widehat{D}_{n,q}^{(\alpha,\beta)}(x)\,\frac{t^n}{n!} \tag{22.4.9}$$

By utilizing (22.4.9), we readily have

$$\int_{\mathbb{Z}_p} q^{-\beta y}\,(1+t)^{-(x+y)}\,d\mu_{q^{\alpha}}(y)=\sum_{n=0}^{\infty}\widehat{D}_{n,q}^{(\alpha,\beta)}(x)\,\frac{t^n}{n!}.$$

From here, it is obvious that

$$\widehat{D}_{n,q}^{(\alpha,\beta)}(x)=\int_{\mathbb{Z}_p} q^{-\beta y}\,(-x-y)_n\,d\mu_{q^{\alpha}}(y)\,. \tag{22.4.10}$$

The following relationships hold true (cf. [4]):

$$\widehat{D}_{n,q}^{(\alpha,\beta)}(x)=\sum_{l=0}^{n}(-1)^l\,S_1(n,l)\,B_{l,q}^{(\alpha,\beta)}(x)$$

and

$$B_{n,q}^{(\alpha,\beta)}(x)=(-1)^n\sum_{k=0}^{n}\widehat{D}_{k,q}^{(\alpha,\beta)}(x)\,S_2(n,k)\,.$$

The following relationships hold for two types of the q-Daehee polynomials with weight (α, β) (cf. [4]):

$$\frac{\widehat{D}_{n,q}^{(\alpha,\beta)}(x)}{n!}(-1)^n=\sum_{k=1}^{n}\binom{n-1}{k-1}\frac{D_{k,q}^{(\alpha,\beta)}(x)}{k!}$$

and

$$\frac{D_{n,q}^{(\alpha,\beta)}(x)}{n!}(-1)^n=\sum_{k=1}^{n}\binom{n-1}{k-1}\frac{\widehat{D}_{k,q}^{(\alpha,\beta)}(x)}{k!}.$$

22.4.2 (α, β)-Fermionic Integral and Its Some Properties

Here assume that $t, q \in \mathbb{C}_p$ with $|q|_p < p^{-\frac{1}{1-p}}$ and $|t|_p < p^{-\frac{1}{1-p}}$. The weighted fermionic p-adic q-integral on $\mathbb{Z}_p$ including the real numbers α and β is given as follows (cf. [7]):

$$I_{-q}^{(\alpha,\beta)}(f) = \int_{\mathbb{Z}_p} q^{-\beta x} f(x)\, d\mu_{-q^\alpha}(x)$$
$$= \lim_{N\to\infty} \frac{1}{\left[p^N\right]_{-q^\alpha}} \sum_{k=0}^{p^N-1} (-1)^k f(k)\, q^{(\alpha-\beta)k}. \qquad (22.4.11)$$

In the case when $\alpha = 1$ and $\beta = 0$ in (22.4.11), it reduces to the $I_{-q}^{(1,0)}(f) := I_{-q}(f)$ given in (22.3.5).

Let $f_n(x) = f(x+n)$. By (22.4.11), we see that (cf. [7])

$$q^{(\alpha-\beta)} I_{-q}^{(\alpha,\beta)}(f_1) = q^{(\alpha-\beta)} \int_{\mathbb{Z}_p} q^{-\beta x} f(x)\, d\mu_{-q^\alpha}(x)$$
$$= -\lim_{N\to\infty} \frac{1}{\left[p^N\right]_{-q^\alpha}} \sum_{k=0}^{p^N-1} (-1)^{k+1} f(k+1)\, q^{(\alpha-\beta)(k+1)}$$
$$= -\lim_{N\to\infty} \frac{1}{\left[p^N\right]_{-q^\alpha}} \sum_{k=0}^{p^N-1} (-1)^k f(k)\, q^{(\alpha-\beta)k}$$
$$+ \left(1+q^\alpha\right) \lim_{N\to\infty} \frac{f\left(p^N\right) q^{(\alpha-\beta)p^N} + f(0)}{1+q^{\alpha p^N}}$$
$$= -I_{-q}^{(\alpha,\beta)}(f) + \left(1+q^\alpha\right) f(0)$$

and

$$q^{2(\alpha-\beta)} I_{-q}^{(\alpha,\beta)}(f_2) = (-1)^2 \lim_{N\to\infty} \frac{1}{\left[p^N\right]_{-q^\alpha}} \sum_{k=0}^{p^N-1} (-1)^{k+2} f(k+2)\, q^{(\alpha-\beta)(k+2)}$$
$$= (-1)^2 I_{-q}^{(\alpha,\beta)}(f) + \left(1+q^\alpha\right)$$
$$\cdot \lim_{N\to\infty} \frac{-f(0) + q^{\alpha-\beta} f(1) - f\left(p^N\right) q^{(\alpha-\beta)p^N} + f\left(p^N+1\right) q^{(\alpha-\beta)\left(p^N+1\right)}}{1+q^{\alpha p^N}}$$
$$= (-1)^2 I_{q}^{(\alpha,\beta)}(f) + \left(1+q^\alpha\right) \sum_{r=0}^{1} (-1)^{r+1} q^{(\alpha-\beta)r} f(r).$$

By continuing this process, we acquire the following result for $f_n(x) = f(x+n)$:

$$q^{(\alpha-\beta)n} I_{-q}^{(\alpha,\beta)}(f_n) - (-1)^{n+1} I_{-q}^{(\alpha,\beta)}(f) = \left(1+q^{\alpha}\right) \sum_{r=0}^{n-1} (-1)^{n-r+1} q^{(\alpha-\beta)r} f(r).$$

This weighted integral is used to generalize and unify the well-known formulas and identities for the some special q-polynomials.

22.4.2.1 The q-Changhee Polynomials with Weight (α, β)

Upon setting $f(x) = (x+y)_n$ in (22.4.11), we then have

$$I_{-q}^{(\alpha,\beta)}((x+y)_n) = \int_{\mathbb{Z}_p} q^{-\beta y} (x+y)_n \, d\mu_{-q^{\alpha}}(y). \tag{22.4.12}$$

It follows from (22.4.12) that

$$\begin{aligned} \sum_{n=0}^{\infty} \int_{\mathbb{Z}_p} q^{-\beta y} (x+y)_n \, d\mu_{-q^{\alpha}}(y) \frac{t^n}{n!} &= \int_{\mathbb{Z}_p} q^{-\beta y} \left(\sum_{n=0}^{\infty} \binom{x+y}{n} t^n \right) d\mu_{-q^{\alpha}}(y) \\ &= \int_{\mathbb{Z}_p} q^{-\beta y} (1+t)^{x+y} \, d\mu_{-q^{\alpha}}(y), \end{aligned} \tag{22.4.13}$$

which yields the following identity:

$$\int_{\mathbb{Z}_p} q^{-\beta y} (1+t)^{x+y} \, d\mu_{-q^{\alpha}}(y) = \frac{1+q^{\alpha}}{1+(1+t)\, q^{\alpha-\beta}} (1+t)^x. \tag{22.4.14}$$

The q-Changhee polynomials with weight (α, β) by the following generating function to be (cf. [7]):

$$\sum_{n=0}^{\infty} Ch_{n,q}^{(\alpha,\beta)}(x) \frac{t^n}{n!} = \frac{1+q^{\alpha}}{1+(1+t)\, q^{\alpha-\beta}} (1+t)^x. \tag{22.4.15}$$

With the help of (22.4.14) and (22.4.15), we get

$$\sum_{n=0}^{\infty} Ch_{n,q}^{(\alpha,\beta)}(x) \frac{t^n}{n!} = \int_{\mathbb{Z}_p} q^{-\beta y} (1+t)^{x+y} \, d\mu_{-q^{\alpha}}(y). \tag{22.4.16}$$

Upon setting $x = 0$ in (22.4.15) and (22.4.16), we have $Ch_{n,q}^{(\alpha,\beta)}(0) := Ch_{n,q}^{(\alpha,\beta)}$ which are called q-Changhee numbers with weight (α, β) generated by (cf. [7])

$$\sum_{n=0}^{\infty} Ch_{n,q}^{(\alpha,\beta)} \frac{t^n}{n!} = \frac{1+q^{\alpha}}{1+(1+t)\,q^{\alpha-\beta}} = \int_{\mathbb{Z}_p} q^{-\beta y} (1+t)^y \, d\mu_{-q^{\alpha}}(y) . \qquad (22.4.17)$$

The weighted q-Euler polynomials for $n \geqq 0$ are introduced in [7] as follows:

$$\mathfrak{E}_{n,q}^{(\alpha,\beta)}(x) = \int_{\mathbb{Z}_p} q^{-\beta y} (x+y)^n \, d\mu_{-q^{\alpha}}(y) . \qquad (22.4.18)$$

Substituting $x = 0$ in (22.4.18) gives $\mathfrak{E}_{n,q}^{(\alpha,\beta)}(0) := \mathfrak{E}_{n,q}^{(\alpha,\beta)}$ dubbed as the weighted q-Euler numbers. Also the weighted q-Euler polynomials and numbers satisfy the following relation:

$$\mathfrak{E}_{n,q}^{(\alpha,\beta)}(x) = \sum_{k=0}^{n} \binom{n}{k} \mathfrak{E}_{k,q}^{(\alpha,\beta)} x^{n-k}.$$

In terms of (22.4.16) and (22.4.18), there are two relationships among $Ch_{n,q}^{(\alpha,\beta)}(x)$, $\mathfrak{E}_{n,q}^{(\alpha,\beta)}(x)$, and $S_2(n,k)$ (cf. [7]):

$$Ch_{n,q}^{(\alpha,\beta)}(x) = \sum_{k=0}^{n} S_1(n,k)\, \mathfrak{E}_{k,q}^{(\alpha,\beta)}(x)$$

and

$$\mathfrak{E}_{n,q}^{(\alpha,\beta)}(x) = \sum_{k=0}^{n} Ch_{k,q}^{(\alpha,\beta)}(x) S_2(n,k) .$$

The q-Changhee numbers of second kind with weight (α, β) are considered in [7] by the following weighted fermionic p-adic q-integral on $\mathbb{Z}_p$:

$$\widehat{Ch}_{n,q}^{(\alpha,\beta)} = \int_{\mathbb{Z}_p} q^{-\beta y} (-y)_n \, d\mu_{-q^{\alpha}}(y) \quad (n \in \mathbb{N}_0) . \qquad (22.4.19)$$

The following result just follows from (22.3.3) and (22.4.19) (cf. [7]):

$$\widehat{Ch}_{n,q}^{(\alpha,\beta)} = \sum_{k=0}^{n} S_1(n,k)\,(-1)^k\, \mathfrak{E}_{k,q}^{(\alpha,\beta)}.$$

Let us consider the exponential generating function of the weighted q-analogue of Changhee numbers of the second kind as follows:

$$\begin{aligned}\sum_{n=0}^{\infty}\widehat{Ch}_{n,q}^{(\alpha,\beta)}\frac{t^n}{n!}&=\sum_{n=0}^{\infty}\left(\int_{\mathbb{Z}_p}q^{-\beta y}(-y)_n\,d\mu_{-q^{\alpha}}(y)\right)\frac{t^n}{n!}\\&=\int_{\mathbb{Z}_p}q^{-\beta y}\left(\sum_{n=0}^{\infty}\binom{-y}{n}t^n\right)d\mu_{-q^{\alpha}}(y)\\&=\int_{\mathbb{Z}_p}q^{-\beta y}(1+t)^{-y}\,d\mu_{-q^{\alpha}}(y).\end{aligned}$$

Also,

$$\int_{\mathbb{Z}_p}q^{-\beta y}(1+t)^{-y}\,d\mu_{-q^{\alpha}}(y)=\frac{(1+q^{\alpha})(1+t)}{1+t+q^{\alpha-\beta}}.$$

so, it can be rewritten as follows:

$$\sum_{n=0}^{\infty}\widehat{Ch}_{n,q}^{(\alpha,\beta)}\frac{t^n}{n!}=\frac{1+q^{\alpha}}{1+t+q^{\alpha-\beta}}(1+t).$$

The q-Changhee polynomials with weight (α,β) of second kind $\widehat{Ch}_{n,q}^{(\alpha,\beta)}(x)$ are defined (cf. [7]) by means of the following exponential generating function:

$$\sum_{n=0}^{\infty}\widehat{Ch}_{n,q}^{(\alpha,\beta)}(x)\frac{t^n}{n!}=\frac{(1+q^{\alpha})}{1+t+q^{\alpha-\beta}}(1+t)^{1-x}\tag{22.4.20}$$

As a result of (22.4.20), it follows

$$\sum_{n=0}^{\infty}\widehat{Ch}_{n,q}^{(\alpha,\beta)}(x)\frac{t^n}{n!}=\int_{\mathbb{Z}_p}q^{-\beta y}(1+t)^{-(x+y)}\,d\mu_{-q^{\alpha}}(y).$$

and then

$$\widehat{Ch}_{n,q}^{(\alpha,\beta)}(x)=\int_{\mathbb{Z}_p}q^{-\beta y}(-x-y)_n\,d\mu_{-q^{\alpha}}(y).\tag{22.4.21}$$

A symmetric relation for $\widehat{Ch}_{n,q}^{(\alpha,\beta)}(x)$ and $Ch_{n,q}^{(\alpha,\beta)}(x)$ is directly derived from (22.4.15) and (22.4.20) with some basic computations (cf. [7]):

$$\widehat{Ch}_{n,q}^{(\alpha,\beta)}(x)=q^{\beta}Ch_{n,q^{-1}}^{(\alpha,\beta)}(1-x).\tag{22.4.22}$$

The rising factorial $x^{\langle n\rangle}$ is defined by (cf. [4, 5, 7, 8, 18–21, 23, 24, 28])

$$x^{\langle n\rangle} = x(x+1)(x+2)\cdots(x+n-1) \qquad \left(n \geqq 0\right)$$

and provides the following expression:

$$x^{\langle n\rangle} = (-1)^n (-x)_n = \sum_{k=0}^{n} S_1(n,k)(-1)^{n-k} x^k. \tag{22.4.23}$$

By (22.4.21), (22.4.23), and (22.4.18), for $n \geqq 0$, the following relations hold true:

$$\widehat{Ch}_{n,q}^{(\alpha,\beta)}(x) = \sum_{k=0}^{n} S_1(n,k)(-1)^k \mathfrak{E}_{k,q}^{(\alpha,\beta)}(x)$$

and

$$\widehat{Ch}_{n,q}^{(\alpha,\beta)}(x) = \sum_{l=0}^{n} (-1)^l S_1(n,l)\, \mathfrak{E}_{l,q}^{(\alpha,\beta)}(x).$$

For $\lambda \in \mathbb{C}$, the λ-Stirling numbers of second kind $S_2(k,n;\lambda)$ are given by the following series expansion (cf. [7] and [28]):

$$\frac{\left(\lambda e^t - 1\right)^k}{k!} = \sum_{n=k}^{\infty} S_2(k,n;\lambda)\frac{t^n}{n!} \qquad \left(n \geqq k\right). \tag{22.4.24}$$

Upon setting $\lambda = 1$, λ-Stirling numbers of second kind $S_2(k,n;\lambda)$ reduce to the Stirling numbers of second kind $S_2(k,n)$ in (22.3.4), cf. [4, 5, 7, 8, 18–21, 23, 24, 28].

A correlation including the λ-Stirling numbers of second kind $S_2(k,n;\lambda)$, q-Changhee polynomials with weight (α,β) of second kind, and the weighted q-Euler polynomials is valid for $n \geqq 0$ (cf. [7]):

$$\mathfrak{E}_{n,q}^{(\alpha,\beta)}(x) = q^{(\alpha-\beta)(1-2x)} \sum_{k=0}^{n} \widehat{Ch}_{k,q}^{(\alpha,\beta)}(1-x)\, S_2\left(k,n;q^{2(\alpha-\beta)}\right).$$

By using the relation (22.4.22), an immediate result of the formula above is as follows:

$$\mathfrak{E}_{n,q}^{(\alpha,\beta)}(x) = q^{(\alpha-\beta)(1-2x)+\beta} \sum_{k=0}^{n} Ch_{k,q^{-1}}^{(\alpha,\beta)}(x)\, S_2\left(k,n;q^{2(\alpha-\beta)}\right).$$

The following formulas provide correlations for the q-Changhee polynomials with weight (α, β) of the both sides (cf. [7]):

$$\widehat{Ch}_{n,q}^{(\alpha,\beta)}(x) = (-1)^n \sum_{k=1}^{n} (n-k)! \binom{n}{k} \binom{n-1}{k-1} Ch_{k,q}^{(\alpha,\beta)}(x)$$

and

$$Ch_{n,q}^{(\alpha,\beta)}(x) = (-1)^n \sum_{k=1}^{n} (n-k)! \binom{n}{k} \binom{n-1}{k-1} \widehat{Ch}_{k,q}^{(\alpha,\beta)}(x).$$

22.4.2.2 The q-Boole Polynomials with Weight (α, β)

In this part, we perform to investigate q-extension of the Boole polynomials with weight (α, β) via the weighted fermionic p-adic q-integral on $\mathbb{Z}_p$.

Throughout this part, assuming that $t, q \in \mathbb{C}_p$ with $|q|_p < p^{-\frac{1}{1-p}}$ and $|t|_p < p^{-\frac{1}{1-p}}$.

We firstly considered choosing a function $f(x) = [2]_q^{-1}(x + \omega y)_n$ in the weighted fermionic p-adic q-integral (22.4.11), we then have

$$I_{-q}^{(\alpha,\beta)}\left([2]_q^{-1}(x+\omega y)_n\right) = [2]_q^{-1} \int_{\mathbb{Z}_p} q^{-\beta y}(x+\omega y)_n \, d\mu_{-q^\alpha}(y). \tag{22.4.25}$$

It follows from (22.4.25) that

$$\begin{aligned}\sum_{n=0}^{\infty} \int_{\mathbb{Z}_p} q^{-\beta y}(x+\omega y)_n \, d\mu_{-q^\alpha}(y) \frac{t^n}{n!} &= \int_{\mathbb{Z}_p} q^{-\beta y} \left(\sum_{n=0}^{\infty} \binom{x+\omega y}{n} t^n \right) d\mu_{-q^\alpha}(y) \\ &= \int_{\mathbb{Z}_p} q^{-\beta y}(1+t)^{x+\omega y} \, d\mu_{-q^\alpha}(y),\end{aligned}$$

which yields the following identity:

$$\int_{\mathbb{Z}_p} q^{-\beta y}(1+t)^{x+\omega y} \, d\mu_{-q^\alpha}(y) = \frac{(1+q^\alpha)}{(1+t)^\omega q^{\alpha-\beta}+1}(1+t)^x. \tag{22.4.26}$$

We are able to consider the following Definition 22.4.

Definition 22.4 We introduce q-Boole polynomials $Bl_{n,q}^{(\alpha,\beta)}(x \mid \omega)$ with weight (α, β) via the following exponential generating function to be:

$$\sum_{n=0}^{\infty} Bl_{n,q}^{(\alpha,\beta)}(x \mid \omega) \frac{t^n}{n!} = \frac{1}{1+q} \frac{1+q^\alpha}{(1+t)^\omega q^{\alpha-\beta}+1}(1+t)^x. \tag{22.4.27}$$

Remark 22.3 In the case when $\alpha = 1$ and $\beta = 0$ in Eq. (22.4.27), we then have $Bl_{n,q}^{(1,0)}(x\,|\omega) := Bl_{n,q}(x\,|\omega)$ in (22.3.7).

Remark 22.4 As q goes to 1 in Eq. (22.4.27), we get the familiar Boole polynomials in (22.2.8).

In terms of (22.4.26) and (22.4.27), we can write

$$\sum_{n=0}^{\infty} Bl_{n,q}^{(\alpha,\beta)}(x\,|\omega)\frac{t^n}{n!} = \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}(1+t)^{x+\omega y}\,d\mu_{-q^{\alpha}}(y). \tag{22.4.28}$$

If we put $x = 0$ in (22.4.27) and (22.4.28), we then have $Bl_{n,q}^{(\alpha,\beta)}(0\,|\omega) := Bl_{n,q}^{(\alpha,\beta)}(\omega)$ that are called q-Boole numbers with weight (α,β) having the following generating function:

$$\begin{aligned}\sum_{n=0}^{\infty} Bl_{n,q}^{(\alpha,\beta)}(\omega)\frac{t^n}{n!} &= \frac{1}{1+q}\frac{1+q^{\alpha}}{(1+t)^{\omega}q^{\alpha-\beta}+1}\\ &= \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}(1+t)^{\omega y}\,d\mu_{-q^{\alpha}}(y).\end{aligned}$$

By (22.4.25) and (22.4.28), it is obvious that

$$Bl_{n,q}^{(\alpha,\beta)}(x\,|\omega) = [2]_q^{-1}\int_{\mathbb{Z}_p} q^{-\beta y}(x+\omega y)_n\,d\mu_{-q^{\alpha}}(y) \tag{22.4.29}$$

and

$$Bl_{n,q}^{(\alpha,\beta)}(\omega) = [2]_q^{-1}\int_{\mathbb{Z}_p} q^{-\beta y}(\omega y)_n\,d\mu_{-q^{\alpha}}(y). \tag{22.4.30}$$

In terms of (22.3.3), (22.4.18), and (22.4.29), we observe that

$$\begin{aligned}Bl_{n,q}^{(\alpha,\beta)}(x\,|\omega) &= [2]_q^{-1}\int_{\mathbb{Z}_p} q^{-\beta y}(x+\omega y)_n\,d\mu_{-q^{\alpha}}(y)\\ &= [2]_q^{-1}\int_{\mathbb{Z}_p} q^{-\beta y}\sum_{k=0}^{n} S_1(n,k)(x+\omega y)^k\,d\mu_{-q^{\alpha}}(y)\\ &= \sum_{k=0}^{n} S_1(n,k)\,\omega^k\,[2]_q^{-1}\int_{\mathbb{Z}_p} q^{-\beta y}\left(\frac{x}{\omega}+y\right)^k d\mu_{-q^{\alpha}}(y)\\ &= \sum_{k=0}^{n} S_1(n,k)\,\omega^k\,[2]_q^{-1}\,\mathfrak{E}_{k,q}^{(\alpha,\beta)}\left(\frac{x}{\omega}\right)\end{aligned}$$

Thus we state the following theorem.

Theorem 22.3 *Let n be a nonnegative integer. We have*

$$Bl_{n,q}^{(\alpha,\beta)}(x\,|\omega) = \sum_{k=0}^{n} S_1(n,k)\,\omega^k\,[2]_q^{-1}\,\mathfrak{E}_{k,q}^{(\alpha,\beta)}\left(\frac{x}{\omega}\right). \tag{22.4.31}$$

In the special case $x = 0$ in Eq. (22.4.31), we get

$$Bl_{n,q}^{(\alpha,\beta)}(\omega) = \sum_{k=0}^{n} S_1(n,k)\,\omega^k\,[2]_q^{-1}\,\mathfrak{E}_{k,q}^{(\alpha,\beta)}.$$

Now, we provide a formula including $Bl_{n,q}^{(\alpha,\beta)}(x\,|\omega)$, $\mathfrak{E}_{n,q}^{(\alpha,\beta)}(x)$, and $S_2(n,k)$ which we discuss below.

Theorem 22.4 *For $n \in \mathbb{N}_0$, we have*

$$\mathfrak{E}_{n,q}^{(\alpha,\beta)}\left(\frac{x}{\omega}\right) = \frac{[2]_q}{\omega^n}\sum_{k=0}^{n} Bl_{k,q}^{(\alpha,\beta)}(x\,|\omega)S_2(n,k). \tag{22.4.32}$$

Proof Substituting t by $e^t - 1$ in (22.4.27), we then get

$$\begin{aligned}
[2]_q^{-1}\frac{(1+q^\alpha)}{q^{\alpha-\beta}e^{t\omega}+1}e^{tx} &= \sum_{n=0}^{\infty} Bl_{n,q}^{(\alpha,\beta)}(x\,|\omega)\frac{\left(e^t-1\right)^n}{n!} \\
&= \sum_{n=0}^{\infty} Bl_{n,q}^{(\alpha,\beta)}(x\,|\omega)\sum_{k=n}^{\infty} S_2(n,k)\frac{t^k}{k!} \\
&= \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} Bl_{k,q}^{(\alpha,\beta)}(x\,|\omega)S_2(n,k)\right)\frac{t^n}{n!}
\end{aligned}$$

and via (22.4.18), we also have

$$\sum_{n=0}^{\infty}\mathfrak{E}_{n,q}^{(\alpha,\beta)}\left(\frac{x}{\omega}\right)\omega^n\frac{t^n}{n!} = \frac{1+q^\alpha}{q^{\alpha-\beta}e^{t\omega}+1}e^{tx},$$

which completes the proof of this theorem. □

The immediate result of Eq. (22.4.32) is as follows:

$$\mathfrak{E}_{n,q}^{(\alpha,\beta)} = \frac{[2]_q}{\omega^n}\sum_{k=0}^{n} Bl_{k,q}^{(\alpha,\beta)}(\omega)S_2(n,k).$$

Let us consider the other type of the Boole numbers arising from the weighted fermionic p-adic q-integral on $\mathbb{Z}_p$.

Definition 22.5 The q-Boole numbers of second kind with weight (α, β) are introduced by the following weighted fermionic p-adic q-integral on $\mathbb{Z}_p$:

$$\widehat{Bl}_{n,q}^{(\alpha,\beta)}(\omega) = \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}(-\omega y)_n \, d\mu_{-q^{\alpha}}(y) \quad \left(\omega \in \mathbb{Z}_p \text{ and } n \in \mathbb{N}_0\right). \tag{22.4.33}$$

In view of (22.3.3) and (22.4.33), we acquire the following relation stated in Theorem 22.5.

Theorem 22.5 *We have, for $\omega \in \mathbb{Z}_p$ and $n \geqq 0$,*

$$\widehat{Bl}_{n,q}^{(\alpha,\beta)}(\omega) = \frac{1}{1+q}\sum_{k=0}^{n} S_1(n,k)(-\omega)^k \, \mathfrak{E}_{k,q}^{(\alpha,\beta)}.$$

Proof Using (22.3.3), we obtain

$$\begin{aligned}
\widehat{Bl}_{n,q}^{(\alpha,\beta)}(\omega) &= \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}(-\omega y)_n \, d\mu_{-q^{\alpha}}(y) \\
&= \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}\left(\sum_{k=0}^{n} S_1(n,k)(-\omega y)^k\right) d\mu_{-q^{\alpha}}(y) \\
&= \frac{1}{1+q}\sum_{k=0}^{n} S_1(n,k)(-\omega)^k \int_{\mathbb{Z}_p} q^{-\beta y} y^k d\mu_{-q^{\alpha}}(y) \\
&= \frac{1}{1+q}\sum_{k=0}^{n} S_1(n,k)(-\omega)^k \, \mathfrak{E}_{k,q}^{(\alpha,\beta)},
\end{aligned}$$

hence, we attain the asserted result. □

Here we perform to investigate the exponential generating function of the weighted q-Boole numbers of the second kind as follows:

$$\begin{aligned}
\sum_{n=0}^{\infty} \widehat{Bl}_{n,q}^{(\alpha,\beta)}(\omega)\frac{t^n}{n!} &= \sum_{n=0}^{\infty}\left(\frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}(-\omega y)_n \, d\mu_{-q^{\alpha}}(y)\right)\frac{t^n}{n!} \\
&= \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}\left(\sum_{n=0}^{\infty}\binom{-\omega y}{n} t^n\right) d\mu_{-q^{\alpha}}(y) \\
&= \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}(1+t)^{-\omega y} \, d\mu_{-q^{\alpha}}(y).
\end{aligned} \tag{22.4.34}$$

Also, by means of (22.4.11), we have

$$\int_{\mathbb{Z}_p} q^{-\beta y} (1+t)^{-\omega y} \, d\mu_{-q^{\alpha}}(y) = \frac{(1+q^{\alpha})}{(1+t)^{\omega} + q^{\alpha-\beta}} (1+t)^{\omega}.$$

Thus let us rewrite Definition 22.5 as

$$\sum_{n=0}^{\infty} \widehat{Bl}_{n,q}^{(\alpha,\beta)}(\omega) \frac{t^n}{n!} = \frac{1+q^{\alpha}}{1+q} \frac{(1+t)^{\omega}}{(1+t)^{\omega} + q^{\alpha-\beta}}.$$

Now, we can consider the q-Boole polynomials with weight (α, β) of second kind by the following definition.

Definition 22.6 The q-Boole polynomials $\widehat{Bl}_{n,q}^{(\alpha,\beta)}(x \mid \omega)$ with weight (α, β) of second kind $\widehat{Ch}_{n,q}^{(\alpha,\beta)}(x)$ are introduced via the following series expansion:

$$\sum_{n=0}^{\infty} \widehat{Bl}_{n,q}^{(\alpha,\beta)}(x \mid \omega) \frac{t^n}{n!} = \frac{1+q^{\alpha}}{1+q} \frac{(1+t)^{\omega+x}}{(1+t)^{\omega} + q^{\alpha-\beta}} \tag{22.4.35}$$

Remark 22.5 Upon setting $\alpha = 1$ and $\beta = 0$ in Eq. (22.4.35), we then have $\widehat{Bl}_{n,q}^{(1,0)}(x \mid \omega) := \widehat{Bl}_{n,q}(x \mid \omega)$ in (22.3.8).

As a result of (22.4.34), we readily have

$$\sum_{n=0}^{\infty} \widehat{Bl}_{n,q}^{(\alpha,\beta)}(x \mid \omega) \frac{t^n}{n!} = \frac{1}{1+q} \int_{\mathbb{Z}_p} q^{-\beta y} (1+t)^{x-\omega y} \, d\mu_{-q^{\alpha}}(y), \tag{22.4.36}$$

then,

$$\begin{aligned} \sum_{n=0}^{\infty} \widehat{Bl}_{n,q}^{(\alpha,\beta)}(x \mid \omega) \frac{t^n}{n!} &= \frac{1}{1+q} \int_{\mathbb{Z}_p} q^{-\beta y} \left(\sum_{n=0}^{\infty} \binom{x-\omega y}{n} t^n \right) d\mu_{-q^{\alpha}}(y) \\ &= \sum_{l=0}^{\infty} \left(\frac{1}{1+q} \int_{\mathbb{Z}_p} q^{-\beta y} (x-\omega y)_n \, d\mu_{-q^{\alpha}}(y) \right) \frac{t^n}{n!}. \end{aligned}$$

Hence, the following expression:

$$\widehat{Bl}_{n,q}^{(\alpha,\beta)}(x \mid \omega) = \frac{1}{1+q} \int_{\mathbb{Z}_p} q^{-\beta y} (x-\omega y)_n \, d\mu_{-q^{\alpha}}(y) \tag{22.4.37}$$

holds true for $\omega \in \mathbb{Z}_p$ and $n \in \mathbb{N}_0$.

A formula for $\widehat{Bl}_{n,q}^{(\alpha,\beta)}(x \mid \omega)$ and $Bl_{n,q}^{(\alpha,\beta)}(x \mid \omega)$ is a direct result of (22.4.27) and (22.4.35) with some basic computations.

Theorem 22.6 *The following symmetric relation holds true for $n \geqq 0$:*

$$\widehat{Bl}_{n,q}^{(\alpha,\beta)}(x\,|-\omega) = Bl_{n,q}^{(\alpha,\beta)}(x\,|\omega).$$

From (22.4.37) and (22.4.23), it is easily seen that

$$\begin{aligned}\widehat{Bl}_{n,q}^{(\alpha,\beta)}(-x\,|-\omega) &= \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}(-x-\omega y)_n\,d\mu_{-q^\alpha}(y)\\ &= \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}(-1)^n(x+\omega y)^{\langle n\rangle}\,d\mu_{-q^\alpha}(y)\\ &= \frac{1}{1+q}\sum_{k=0}^{n} S_1(n,k)(-1)^k\int_{\mathbb{Z}_p} q^{-\beta y}(x+\omega y)^k\,d\mu_{-q^\alpha}(y),\end{aligned}$$

which yields the following theorem with (22.4.18).

Theorem 22.7 *The following relation*

$$\widehat{Bl}_{n,q}^{(\alpha,\beta)}(-x\,|-\omega) = \frac{1}{1+q}\sum_{k=0}^{n} S_1(n,k)(-\omega)^k\,\mathfrak{E}_{k,q}^{(\alpha,\beta)}\left(\frac{x}{\omega}\right)$$

holds true for $n \geqq 0$ and $\omega \in \mathbb{Z}_p$.

An immediate result of the theorem above is stated below:

$$\widehat{Bl}_{n,q}^{(\alpha,\beta)}(-\omega) = \frac{1}{1+q}\sum_{k=0}^{n} S_1(n,k)(-\omega)^k\,\mathfrak{E}_{k,q}^{(\alpha,\beta)}.$$

We here give the following correlation.

Theorem 22.8 *The following relationship holds true:*

$$\widehat{Bl}_{n,q}^{(\alpha,\beta)}(x\,|-\omega) = \frac{1}{1+q}\sum_{k=0}^{n} S_1(n,k)(-\omega 1)^k\,\mathfrak{E}_{k,q}^{(\alpha,\beta)}\left(-\frac{x}{\omega}\right).$$

Proof Indeed,

$$\begin{aligned}\widehat{Bl}_{n,q}^{(\alpha,\beta)}(x\,|-\omega) &= \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}(x-\omega y)_n\,d\mu_{q^\alpha}(y)\\ &= \frac{1}{1+q}\int_{\mathbb{Z}_p} q^{-\beta y}\left(\sum_{k=0}^{n} S_1(n,k)(x-\omega y)^k\right)d\mu_{-q^\alpha}(y)\end{aligned}$$

$$= \frac{1}{1+q} \sum_{k=0}^{n} S_1(n,k)(-\omega)^k \int_{\mathbb{Z}_p} q^{-\beta y} \left(-\frac{x}{\omega}+y\right)^k d\mu_{-q^\alpha}(y)$$

$$= \frac{1}{1+q} \sum_{k=0}^{n} S_1(n,k)(-\omega 1)^k \mathfrak{E}_{k,q}^{(\alpha,\beta)}\left(-\frac{x}{\omega}\right).$$

□

We now introduce the Apostol type weighted q-Euler polynomials $\mathfrak{E}_{n,q}^{(\alpha,\beta)}(x\,|\gamma)$ and numbers $\mathfrak{E}_{n,q}^{(\alpha,\beta)}(\gamma)$ by the following exponential generating functions:

$$\sum_{n=0}^{\infty} \mathfrak{E}_{n,q}^{(\alpha,\beta)}(x\,|\gamma) \frac{t^n}{n!} = \frac{1+q^\alpha}{q^{\alpha-\beta}\gamma e^{t\omega}+1} e^{t\omega x} \tag{22.4.38}$$

and

$$\sum_{n=0}^{\infty} \mathfrak{E}_{n,q}^{(\alpha,\beta)}(\gamma) \frac{t^n}{n!} = \frac{1+q^\alpha}{q^{\alpha-\beta}\gamma e^{t\omega}+1} e^{t\omega x}.$$

When $\gamma = 1$, the Apostol type weighted q-Euler polynomials $\mathfrak{E}_{n,q}^{(\alpha,\beta)}(x\,|\gamma)$ and numbers $\mathfrak{E}_{n,q}^{(\alpha,\beta)}(\gamma)$ reduce to the classical corresponding polynomials and numbers mentioned in the previous part, see Eq. (22.4.18).

A relationship covering the λ-Stirling numbers of second kind $S_2(k,n;\lambda)$ in (22.4.24), the q-Boole polynomials with weight (α,β) of second kind in (22.4.36), and the weighted q-Euler polynomials in (22.4.18) is stated in the following theorem.

Theorem 22.9 *The following relationship is valid for $n \geqq 0$:*

$$\mathfrak{E}_{n,q}^{(\alpha,\beta)}(x) = q^{(\alpha-\beta)(1-2x)} \sum_{k=0}^{n} \widehat{Ch}_{k,q}^{(\alpha,\beta)}(1-x)\, S_2\left(k,n;q^{2(\alpha-\beta)}\right).$$

Proof By replacing t by $\left(q^{(\alpha-\beta)}e^t-1\right)$ in (22.4.35), we get

$$\frac{1+q^\alpha}{1+q} \frac{q^{(\alpha-\beta)(\omega+x)}e^{t(\omega+x)}}{q^{(\alpha-\beta)\omega}e^{t\omega}+q^{\alpha-\beta}}$$

$$= \sum_{n=0}^{\infty} \widehat{Bl}_{n,q}^{(\alpha,\beta)}(x\,|\omega) \frac{\left(q^{(\alpha-\beta)}e^t-1\right)^n}{n!}$$

$$\frac{q^{(\alpha-\beta)(\omega+x-1)}}{1+q} \frac{1+q^\alpha}{q^{(\alpha-\beta)(\omega-2)}q^{(\alpha-\beta)}e^{t\omega}+1} e^{t\omega(1+x/\omega)}$$

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} \widehat{Bl}_{n,q}^{(\alpha,\beta)}(x\,|\omega) \frac{\left(q^{(\alpha-\beta)}e^{t}-1\right)^{n}}{n!} \\
&\quad \frac{q^{(\alpha-\beta)(\omega+x-1)}}{1+q} \sum_{n=0}^{\infty} \mathfrak{E}_{n,q}^{(\alpha,\beta)}\left(1+\frac{x}{\omega}\,\Big|q^{(\alpha-\beta)(\omega-2)}\right)\frac{t^{n}}{n!} \\
&= \sum_{n=0}^{\infty} \widehat{Bl}_{n,q}^{(\alpha,\beta)}(x\,|\omega) \sum_{k=n}^{\infty} S_{2}\left(k,n;q^{\alpha-\beta}\right)\frac{t^{k}}{k!} \\
&\quad \frac{q^{(\alpha-\beta)(\omega+x-1)}}{1+q} \sum_{n=0}^{\infty} \mathfrak{E}_{n,q}^{(\alpha,\beta)}\left(1+\frac{x}{\omega}\,\Big|q^{(\alpha-\beta)(\omega-2)}\right)\frac{t^{n}}{n!} \\
&= \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} \widehat{Bl}_{k,q}^{(\alpha,\beta)}(x\,|\omega)\, S_{2}\left(k,n;q^{\alpha-\beta}\right)\right)\frac{t^{n}}{n!}
\end{aligned}
$$

and from (22.4.38), we arrive

$$
\mathfrak{E}_{n,q}^{(\alpha,\beta)}\left(1+\frac{x}{\omega}\,\Big|q^{(\alpha-\beta)(\omega-2)}\right) = \frac{[2]_{q}}{q^{(\alpha-\beta)(\omega+x-1)}} \sum_{k=0}^{n} \widehat{Bl}_{k,q}^{(\alpha,\beta)}(x\,|\omega)\, S_{2}\left(k,n;q^{\alpha-\beta}\right).
$$

□

We finally give the following relationships for two types of the q-Boole polynomials with weight (α,β).

Theorem 22.10 *The following expressions are valid:*

$$
Bl_{n,q}^{(\alpha,\beta)}(x\,|\omega) = (-1)^{n}\sum_{k=1}^{n}(n-k)!\binom{n}{k}\binom{n-1}{k-1}\widehat{Bl}_{k,q}^{(\alpha,\beta)}(-x\,|\omega)
$$

and

$$
\widehat{Bl}_{n,q}^{(\alpha,\beta)}(x\,|\omega) = (-1)^{n}\sum_{k=1}^{n}(n-k)!\binom{n}{k}\binom{n-1}{k-1}Bl_{k,q}^{(\alpha,\beta)}(-x\,|\omega).
$$

Proof With the help of the known binomial identities

$$
\binom{m}{n} = (-1)^{n}\binom{-m+n-1}{n} \text{ and } \binom{-m+n-1}{n} = \sum_{k=1}^{n}\binom{n-1}{n-k}\binom{-m}{k},
$$

and via (22.4.29) and (22.4.37), we acquire

$$
\begin{aligned}
\frac{Bl_{n,q}^{(\alpha,\beta)}(x\,|\omega)}{n!}(-1)^n &= (-1)^n \int_{\mathbb{Z}_p} q^{-\beta y}\binom{x+\omega y}{n} d\mu_{-q^\alpha}(y) \\
&= \int_{\mathbb{Z}_p} q^{-\beta y}\binom{-x-\omega y+n-1}{n} d\mu_{-q^\alpha}(y) \\
&= \sum_{k=1}^{n}\binom{n-1}{n-k}\int_{\mathbb{Z}_p} q^{-\beta y}\binom{-x-\omega y}{k} d\mu_{-q^\alpha}(y) \\
&= \sum_{k=1}^{n}\binom{n-1}{k-1}\frac{\widehat{Bl}_{k,q}^{(\alpha,\beta)}(-x\,|\omega)}{k!}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\widehat{Bl}_{n,q}^{(\alpha,\beta)}(x\,|\omega)}{n!}(-1)^n &= (-1)^n \int_{\mathbb{Z}_p} q^{-\beta y}\binom{x-\omega y}{n} d\mu_{-q^\alpha}(y) \\
&= \int_{\mathbb{Z}_p} q^{-\beta y}\binom{-x+\omega y+n-1}{n} d\mu_{-q^\alpha}(y) \\
&= \sum_{k=1}^{n}\binom{n-1}{n-k}\int_{\mathbb{Z}_p} q^{-\beta y}\binom{-x+\omega y}{k} d\mu_{-q^\alpha}(y) \\
&= \sum_{k=1}^{n}\binom{n-1}{k-1}\frac{Bl_{k,q}^{(\alpha,\beta)}(-x\,|\omega)}{k!}.
\end{aligned}
$$

□

22.5 Conclusion

The Volkenborn integral was firstly introduced by, a German mathematician, Arnt Volkenborn, see [29] and [30]. With the introduction of the mentioned integral, this integral has been utilized to define and investigate several special polynomials and numbers such as Daehee and Bernoulli numbers and polynomials along with their diverse generalizations. The fermionic p-adic integral is firstly introduced by Korean mathematician Taekyun Kim in 2005. In conjunction with the introduction of this integral, the foregoing integral and its many extensions have been used to consider and analyze several special numbers and polynomials such as Euler, Genocchi, Frobenius-Euler, Eulerian, Changhee, Boole and their many extensions polynomials and numbers. Nowadays, the Volkenborn integral and the fermionic

p-adic integral have been more common, and they are used in multifarious mathematical research areas and physical research areas.

The p-adic q-integral and the fermionic p-adic q-integral were introduced more than a decade ago. Since the introductions of these integrals, the problems have been considered and investigated by using the useful properties of the aforementioned integrals which intensely preserve usability in introducing several extensions of diverse special polynomials such as q-Euler, q-Genocchi, q-Frobenius-Euler, q-Eulerian, q-Changhee, q-Boole polynomials and numbers with their assorted extensions.

In the last section, the weighted p-adic q-integral and the weighted fermionic p-adic q-integral are given and examined with some of their basic properties. The types of the q-Daehee polynomials with weight (α, β) arising from the weighted p-adic q-integral on $\mathbb{Z}_p$ and the two types of the q-Changhee polynomials derived from the weighted fermionic p-adic q-integral on $\mathbb{Z}_p$ are presented in conjunction with their several properties. Then, two generalizations of q-Boole numbers and polynomials called q-Boole polynomials and numbers with weight (α, β) and q-Boole polynomials and numbers of second kind with weight (α, β) are introduced by means of the weighted fermionic p-adic q-integral on $\mathbb{Z}_p$. Multifarious new and interesting formulas and relations including recurrence relation, symmetric relation, and many correlations related to the weighted q-Euler polynomials, the Apostol type weighted q-Euler polynomials, familiar Stirling numbers of first and second kinds, and λ-Stirling numbers of the second kind are derived properly.

References

1. S. Araci, D. Erdal, J.J. Seo, A study on the fermionic p-adic q-integral representation on associated with weighted q-Bernstein and q-Genocchi polynomials. Abst. Appl. Anal. **2011** (2011). Article ID 649248. https://doi.org/10.1155/2011/649248
2. S. Araci, E. Ağyüz, M. Acikgoz, On a q-analog of some numbers and polynomials. J. Inequal. Appl. 19 (2015). https://doi.org/10.1186/s13660-014-0542-y
3. S. Araci, U. Duran, M. Acikgoz, Symmetric identities involving q-Frobenius-Euler polynomials under Sym (5). Turkish J. Anal. Number Theory **3**(3), 90–93 (2015)
4. S. Araci, M. Acikgoz, U. Duran, On weighted q-Daehee numbers and polynomials, Indag. Math. **30**(2), 365–374 (2019)
5. D.V. Dolgy, T. Kim, S.-H. Rim, J.-J. Seo, A note on Changhee polynomials and numbers with q-parameters. Int. J. Math. Anal. **8**(26), 1255–1264 (2014)
6. D.V. Dolgy, T. Kim, S.-H. Rim, S.-H. Lee, Some symmetric identities for h-extension of q-Euler polynomials under third dihedral group D_3. Int. J. Math. Anal. **8**(48), 2369–2374.3 (2014)
7. U. Duran, M. Acikgoz, On applications for Mahler expansion associated with p-adic q-integrals. Int. J. Number Theory **15**(1), 67–84 (2019)
8. F.G. Gurkan, M. Acikgoz, E. Agyuz, A study on the new mixed-type polynomials related to Boole polynomials. Afr. Mat. **28**(1–2), 279–290 (2017)
9. Ö.Ç. Havara, H. Menken, On the Volkenborn integral of the q-extension of the p-adic gamma function. J. Math. Anal. **8**(2), 64–72 (2017)

10. L.-C. Jang, T. Kim, q-Genocchi numbers and polynomials associated with fermionic p-adic invariant integrals on $\mathbb{Z}_p$. Abstr. Appl. Anal. **2008**, 8 pp. (2008). Article ID 232187
11. Y.S. Jang, T. Kim, S.-H. Rim, J.-J. Seo, Symmetry identities for the generalized higher-order q-Bernoulli polynomials under S_3. Int. J. Math. Anal. **8**(38), 1873–1879 (2014)
12. T. Kim, q-Volkenborn integration. Russ. J. Math. Phys. **9**, 288–299 (2002)
13. T. Kim, A note on q-Volkenborn integration. Proc. Jangjeon Math. Soc. **8**(1), 13–17 (2005)
14. D.S. Kim, A note on p-adic invariant integral in the rings of p-adic integers. Adv. Stud. Contemp. Math. **13**, 95–99 (2006)
15. T. Kim, q-Euler numbers and polynomials associated with p-adic q-integrals. J. Nonlinear Math. Phys. **14**(1), 15–27 (2007). https://doi.org/10.2991/jnmp.2007.14.1.3
16. T. Kim, Some identities on the q-Euler polynomials of higher order and q-Stirling numbers by the fermionic p-adic integral on $\mathbb{Z}_p$. Russ. J. Math. Phys. **16**, 484–491 (2009)
17. T. Kim, On the weighted q-Bernoulli numbers and polynomials. Adv. Stud. Contemp. Math. **21**, 207–215 (2011)
18. D.S. Kim, T. Kim, Daehee numbers and polynomials. Appl. Math. Sci. **7**(120), 5969–5976 (2013)
19. D.S. Kim, T. Kim, A note on Boole polynomials. Integral Transforms Spec. Funct. **25**(8), 627–633 (2014)
20. T. Kim, S.-H. Lee, T. Mansour, J.-J. Seo, A note on q-Daehee polynomials and numbers. Adv. Stud. Contemp. Math. **24**(2), 155–160 (2014)
21. D.S. Kim, T. Kim, J.J. Seo, A note on q-analogue of Boole polynomials. Appl. Math. Inf. Sci. **9**(6), 3153–3158 (2015)
22. N. Koblitz, *p-Adic Numbers, p-Adic Analysis, and Zeta Functions* (Springer, New York, 1977)
23. J. Kwon, J.-W. Park, A note on (h, q)-Boole polynomials. Adv. Differ. Equ. **2015**, 198 (2015). https://doi.org/10.1186/s13662-015-0536-1
24. H. Ozden, I.N. Cangul, Y. Sımsek, Remarks on q-Bernoulli numbers associated with Daehee numbers. Adv. Stud. Contemp. Math. **18**(1), 41–48 (2009)
25. A.M. Robert, *A Course in p-Adic Analysis* (Springer, New York, 2000)
26. C.S. Ryoo, Symmetric identities for Carlitz's twisted q-Bernoulli polynomials associated with p-adic q-integral on $\mathbb{Z}_p$. Appl. Math. Sci. **9**(72), 3569–3575 (2015)
27. W.H. Schikhof, *Ultrametric Calculus: An Introduction to p-Adic Analysis*. Cambridge Studies in Advanced Mathematics, vol. 4 (Cambridge University Press, Cambridge, 1984)
28. Y. Simsek, A. Yardimci, Applications on the Apostol-Daehee numbers and polynomials associated with special numbers, polynomials, and p-adic integrals. Adv. Differ. Equ. 308 (2016). https://doi.org/10.1186/s13662-016-1041-x
29. A. Volkenborn, Ein p-adisches integral und seine Anwendungen I. Manuscripta Math. **7**(4), 341–373 (1972)
30. A. Volkenborn, Ein p-adisches integral und seine Anwendungen II. Manuscripta Math. **12**(1), 17–46 (1974)

Chapter 23
On Statistical Deferred Cesàro Summability

Hemen Dutta, S. K. Paikray, and B. B. Jena

Abstract This chapter consists of five sections. The first section is introductory, where from the concept of infiniteness to the development of summability methods are presented. In the second section, ordinary and statistical versions of Cesàro and deferred Cesàro summability methods have been introduced and accordingly some basic terminologies are considered. In the third section, we have applied our proposed deferred Cesàro mean to prove a Korovkin-type approximation theorem for the set of functions $1, e^{-x}$, and e^{-2x} defined on a Banach space and demonstrated that our theorem is a non-trivial extension of some well-known Korovkin-type approximation theorems. In the fourth section, we have established a result for the rate of our statistical deferred Cesàro summability mean with the help of the modulus of continuity. Finally, in the last section, we have given some concluding remarks and presented some interesting examples in support of our definitions and results.

Keywords Infinite series · Natural density · Statistical convergence · Statistical deferred Cesàro convergence · Statistical deferred Cesàro summability · Korovkin-type approximation theorem · Modulus of continuity · Rate of statistical deferred Cesàro summability

23.1 Introduction

The concept of counting seems to excite the human thought right from the time man put his feet on the earth as Homo-sapiens. Later with the introduction of arithmetic operations in the field of number system there emerged the concept of infinite

---

H. Dutta
Department of Mathematics, Gauhati University, Guwahati, Assam, India

S. K. Paikray (✉) · B. B. Jena
Department of Mathematics, Veer Surendra Sai University of Technology, Burla, Odisha, India
e-mail: skpaikray_math@vssut.ac.in

H. Dutta et al. (eds.), *Current Trends in Mathematical Analysis and Its Interdisciplinary Applications*,
https://doi.org/10.1007/978-3-030-15242-0_23

series. Its scope has been widened in the present days. The study of such series is very interesting with regard to the notion of their convergence and divergence. The study of summability has always been of great interest for modern mathematicians, as it occupies a very prominent position in analysis. It is worth mentioning here that the concept of summability is nothing but a generalization of the concept of convergence.

Carl Friedrich Gauss, the German Mathematician, was the pioneer in the introduction of concept of infinite process into Mathematical Analysis. However, Augustine-Louis Cauchy (1789–1859), a French Mathematician, was the first to look into the clear concept of sum of an infinite series strictly in terms of limit. He introduced ideas regarding convergence and divergence in his famous book *Analysis Algebrique* (Published in 1821, Paris), which was the first book on analysis written in modern spirit. The sum of a series by Cauchy is known as *natural sum* or *Cauchy's sum* of a series.

Let $\{a_n\}$ be a given real or complex-valued sequence. Then an expression of the form

$$a_1 + a_2 + a_3 + a_4 + \cdots$$

is called an *infinite series* and is in brief generally denoted by

$$\sum_{n=1}^{\infty} a_n \quad \text{or} \quad \sum a_n.$$

If all of the terms of the sequence $\{a_n\}$ after a certain number are zero, then the expression

$$a_1 + a_2 + a_3 + a_4 + \cdots + a_m$$

is called a *finite series* and is written simply as

$$\sum_{n=1}^{m} a_n.$$

An expression of the form

$$\sum_{n=1}^{\infty} a_n = \sum a_n = a_1 + a_2 + a_3 + a_4 + \cdots,$$

which involves the addition of infinitely many terms, has indeed no meaning, as there is no way to sum an infinite number of terms. However, in order to accord some plausible meaning to such an expression, Cauchy uses the concept of limits. For this Cauchy forms a sequence of partial sums of the series and defines the sum

$$a_1 + a_2 + a_3 + a_4 + \cdots$$

as the limiting value of the partial sums as the number of terms tends to infinity.

Let $\sum a_n$ be an infinite series with real or complex terms and let, for $n = 1, 2, 3, \ldots$,

$$s_n = a_1 + a_2 + a_3 + a_4 + \cdots + a_n.$$

Then s_n is called the n-th partial sum of the series and the sequence $\{s_n\}$, thus obtained, is called the sequence of partial sums of the series $\sum a_n$. An infinite series $\sum a_n$ is said to converge, diverge, or oscillate, according as its sequence of partial sums $\{s_n\}$ converges, diverges, or oscillates. According to Cauchy the infinite series $\sum a_n$ has the sum s (known as a Cauchy's sum) if and only if there exists a finite real number s such that, for every $\epsilon > 0$, there exists a natural number n_0 such that

$$|s_n - s| < \epsilon, \quad \text{for every} \quad n \geq n_0.$$

That is to say, $\lim_{n\to\infty} s_n = s$. A series for which Cauchy's sum exists (that is, $\lim_{n\to\infty} s_n = s$, a finite number) is termed as convergent.

The series which were not convergent, that is, the series having no sum in the sense of Cauchy, were termed as divergent. According to Cauchy, divergent series do not belong to the understandable domain of mathematics, and the convergent series were the only valid mathematical entities. Before Cauchy series, convergent and divergent both were in use and no distinction was made between the two. This leads to paradoxes and irreconcilable situations. But Cauchy in one stroke removed all the contradictions and paradoxes by out casting divergent series from the valid domain of mathematics. It brought the much needed relief to the then mathematicians, whose faith in their methodology was badly shaken owing to frequent appearances of paradoxes and contradictions. After this it began to be regarded that the problem of the sum of an infinite series had fully and finally been resolved. Thus, even though the divergent series were mostly used for good purposes earlier by such eminent mathematicians as Leibnitz, Euler, and others, they were thrown out from the valid domain of mathematics without hesitation. The concept of sum of an infinite series, as derived by Cauchy was so natural, so efficacious that mathematicians thought the problem of sum of infinite series had finally been settled once for all. Niels Henrik Abel (Norwegian Mathematician, 1802–1829) was another important contributor for giving the ideas concerning convergence and divergence in the early part of the nineteenth century. He was so excited with the discovery that in a letter to Holmbee he expresses his conviction in such telling that divergent series are in general, something quite calamitous, and it is a shame that any one dares to base a proof of them.

As mathematics is based on principles of reasoning, any slightest deviation from the right track of the flow of mathematical ideas would ultimately end in disharmony. Even after the theory, propounded by Cauchy, having received the stamp of finality of almost all mathematicians of the time, it did face same disharmonies particularly in the field of orthogonal expansion of continuous functions and product series. It was noticed that certain non-convergent series (Fourier series) behaved very much the same way with regard to arithmetical operations on them as convergent ones and the calculation based on certain asymptotic series, not convergent in the sense of Cauchy, used in dynamical astronomy, were quite valid and verifiable otherwise. All these facts, in course of time led mathematicians to conclude that Cauchy method of assigning sum to an infinite series was of a far reaching importance and yet was not the last word on the subject and divergent series were not that devilish as they were earlier made out to be. All these stirred the imagination of several inquisitive mathematicians to delve deep into the character of the sum of an infinite series, over and above that of Cauchy of assigning sum. Persistent efforts made by a number of eminent mathematicians led to the discovery of alternative methods which were closely connected with that of Cauchy, yet associated sum even to divergent series, particularly to those whose partial sums oscillate. By the close of the nineteenth century, several alternative methods of assigning sum to infinite series were invented by mathematicians. These methods of summation were termed as *summability methods*. Some of the most familiar methods of summability are those that are associated with the name of great mathematicians like Abel, Borel, Cesàro, Euler, Hausdroff, Hölder, Lambert, Nörlund, Riesz, Riemann, and Lebesgue. Thus, by the third decade of the last century, a very rich and fruitful theory of summability had been introduced. This theory found applications even in such remote fields as the probability and the theory of numbers. Norbert Wiener applied Lambert's method of summation to prove the prime number theorem. As Cauchy's concept of sum of convergent series well withstood all the rigors of mathematics, the framework of the summability methods was in general so devised as to assign convergent series, the same sum as that assigned by Cauchy.

A *summability method* or *summation method* is a function from the set of sequence of partial sums of series to a value. In other words, in its broadest meaning, summability theory or in short summability is the theory of assignment of limits, which is fundamental in analysis, function theory, topology, and functional analysis. Moreover, a summability method is said to be *regular* if the method sums all convergent series to its Cauchy's sum and it is said to be *consistent* if it assigns same sum to same series. Thus, regular methods of summability may be regarded as the generalization of Cauchy's concept of convergence. Just as the concept of ordinary convergence has been generalized into that of summability, commonly termed "ordinary summability," the concept of absolute convergence too has been extended similarly into the concept called as *absolute summability*. The simplest method of summability is the method of *first arithmetic mean*, that is, $(C, 1)$-mean (Cesàro mean of order one).

23.2 Cesàro and Deferred Cesàro Summability Methods

In 1890, Cesàro made a study about multiplication of series and introduced arithmetic mean method for sequence of partial sum of infinite series to find the sum of divergent series. In fact, he was the first mathematician who introduced $(C, 1)$-summability. Let $\sum u_n$ be an infinite series of real or complex terms and let the n-th partial sum of the series be

$$s_n = \sum_{k=0}^{n} u_k, \quad n = 0, 1, \ldots .$$

Let

$$\sigma_n = \frac{s_0 + s_1 + s_2 + \cdots + s_n}{n+1} = \frac{1}{n+1} \sum_{k=0}^{n} x_k,$$

also this σ_n is known as the first arithmetic mean of the Cesàro transform of order one (in short $(C, 1)$ transform) of the sequence $\{s_n\}$.

If $\lim_{n\to\infty} \sigma_n = s$, then the series $\sum u_n$ is $(C, 1)$ summable to s.

In this case, we write

$$\sum u_n = s(C, 1).$$

Further, if $\{\sigma_n\}$ belongs to B.V., that is to say,

$$\sum |\sigma_n - \sigma_{n-1}| < \infty,$$

then the series $\sum u_n$ or the sequence $\{s_n\}$ is said to be absolutely $(C, 1)$-summable or $|(C, 1)|$-summable to s. Later he extended this method for the positive integral order α. It is symbolically denoted as (C, α) method. Subsequently, Knop (1911) extended the scope of this summability method to positive fractional orders. Further, Chapman (1910) extended it to negative index α, for $\alpha > -1$, which increases the scope of summability even beyond convergence. For Cesàro transform of order α, the sequence-to-sequence transformation $\{\sigma_n\}$ is defined as

$$\sigma_n = \frac{1}{A_n^{\alpha}} \sum_{k=0}^{n} A_{n=k}^{\alpha-1} s_k,$$

where

$$A_n^{\alpha} = \frac{\Gamma(n+\alpha+1)}{\Gamma(\alpha+1)\Gamma(n+1)},$$

and it is also known as the nth Cesàro-mean of order α, or simply the (C, α) mean of the series $\sum u_n$, or of the sequence $\{s_n\}$. The series $\sum u_n$ is said to be summable by Cesàro-method of order α, or simply (C, α)-summable, to the sum s, if $\lim_{n\to\infty} t_n = s$, where s is a finite number.

Further, if $\{\sigma_n\}$ belongs to B.V., that is to say,

$$\sum |\sigma_n - \sigma_{n-1}| < \infty,$$

then the series $\sum u_n$ or the sequence $\{s_n\}$ is said to be absolutely (C, α)-summable or $|(C, \alpha)|$-summable to s.

Example Consider a sequence $(x_n) = (-1)^n$, $(n \in \mathbb{N} \cup \{0\})$ with sequence of partial sum (s_n). We have

$$s_0 = x_0 = 1$$

$$s_1 = x_0 + x_1 = 0$$

$$s_2 = x_0 + x_1 + x_2 = 1$$

$$\cdots \quad \cdots \quad \cdots$$

$$s_{2n-1} = x_0 + x_1 + x_2 + \cdots + x_{2n-1} = 0$$

$$s_{2n} = x_0 + x_1 + x_2 + \cdots + x_{2n-1} + x_{2n} = 1.$$

The Cesàro transform of order one is given by

$$\sigma_n = \frac{s_0 + s_1 + s_2 + \cdots + s_n}{n+1} \quad (n \geqq 0).$$

Clearly,

$$\lim_{n\to\infty} \sigma_{2n} = \lim_{n\to\infty} \frac{s_0 + \cdots + s_{2n}}{2n+1} = \lim_{n\to\infty} \frac{n}{2n+1} = \frac{1}{2},$$

and

$$\lim_{n\to\infty} \sigma_{2n-1} = \lim_{n\to\infty} \frac{s_1 + \cdots + s_{2n-1}}{2n} = \lim_{n\to\infty} \frac{n-1}{2n} = \frac{1}{2}.$$

This implies,

$$\lim_{n\to\infty} \sigma_n = \frac{1}{2}.$$

Hence, (x_n) is Cesàro summable to $\frac{1}{2}$.

In the study of sequence space, classical convergence has got innumerable applications where the convergence of a sequence requires that almost all elements are to satisfy the convergence condition, that is, every element of the sequence needs to be in some neighborhood (arbitrarily small) of the limit. However, such restriction is relaxed in statistical convergence, where the set having a few elements that are not in the neighborhood of the limit is discarded subject to the condition that the natural density of the set is zero, and at the same time the condition of convergence is valid for the rest majority of the elements. The notion of statistical convergence was introduced by Fast [7] and Steinhaus [27]. Recently, statistical convergence has been a dynamic research area due to the fact that it is more general than classical convergence and such theory is discussed in the study of Fourier analysis, number theory, and approximation theory. For more details, see [8, 9, 12, 20, 23, 25] and [26].

Let $\mathbb{N}$ be the set of natural numbers and $K \subseteq \mathbb{N}$. Also, let

$$K_n = \{k : k \leq n \text{ and } k \in K\}$$

with $|K_n|$ as the cardinality of K_n. The *natural density* of K is defined by

$$\delta(K) = \lim_{n\to\infty} \frac{|K_n|}{n} = \lim_{n\to\infty} \frac{1}{n}|\{k : k \leq n \text{ and } k \in K\}|,$$

provided the limit exists.

A given sequence (x_n) is said to be *statistically convergent* to ℓ, if for each $\epsilon > 0$, the set

$$K_\epsilon = \{k : k \in \mathbb{N} \text{ and } |x_k - \ell| \geq \epsilon\}$$

has zero natural density (see [7, 27]). That is, for each $\epsilon > 0$,

$$\delta(K_\epsilon) = \lim_{n\to\infty} \frac{|K_\epsilon|}{n} = \lim_{n\to\infty} \frac{1}{n}|\{k : k \leq n \text{ and } |x_k - \ell| \geq \epsilon\}| = 0.$$

In this case, we write

$$\text{stat} \lim_{n\to\infty} x_n = \ell.$$

Now we present an example to show that every convergent sequence is statistically convergent but the converse is not true in general.

Example Consider a sequence $x = (x_n)$ by

$$x_n = \begin{cases} n & \text{when } n = m^2, \text{ for all } m \in \mathbb{N} \\ \frac{1}{n} & \text{otherwise.} \end{cases}$$

It is easy to see that the sequence (x_n) is divergent in the ordinary sense, while 0 is the statistical limit of (x_n) since $\delta(K) = 0$, where $K = \{m^2, \text{ for all } m = 1, 2, 3, \ldots\}$.

In 2002, Móricz [15] introduced the fundamental idea of statistical $(C, 1)$ summability and recently Mohiuddine et al. [14] have established the statistical $(C, 1)$ summability as follows.

A sequence (x_n) is said to be *statistical* $(C, 1)$ *summable* to ℓ, if for each $\epsilon > 0$, the set

$$\{k : k \in \mathbb{N} \text{ and } |\sigma_k - \ell| \geq \epsilon\}$$

has zero Cesàro density. That is, for each $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n} |\{k : k \leq n \text{ and } |\sigma_k - \ell| \geq \epsilon\}| = 0.$$

In this case, we write

$$\text{stat} - \lim_{n\to\infty} \sigma_n = \ell \text{ or } C_1(\text{stat}) - \lim_{n\to\infty} x_n = \ell.$$

In the following example, we illustrate that a sequence is statistically $(C, 1)$ summable even if it is not statistically convergent.

Example Define a sequence x_k by

$$x_k = \begin{cases} 1 & \text{if } k = m^2 - m, m^2 - m + 1, \ldots, m^2 - 1, \\ -m & \text{if } k = m^2 \ \ (m = 2, 3, 4, \ldots,) \\ 0 & \text{otherwise.} \end{cases}$$

The $(C, 1)$ transform

$$\sigma_n = \frac{1}{n+1} \sum_{k=0}^{n} x_k,$$

yields

$$\sigma_n = \begin{cases} \frac{s+1}{n+1} & \text{if } n = m^2 - m + s; s = 0, 1, 2, \ldots m - 1; m = 2, 3, 4, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that $\lim_{n\to\infty} \sigma_n = 0$, which implies that stat $- \lim_{n\to\infty} \sigma_n = 0$. Clearly, the sequence x_n is statistically $(C, 1)$ summable to 0; however, it is not statistically convergent to 0.

In the year 2008, Özarslan et al. [18] established certain results on statistical approximation for Kantorovich-type operators involving some special polynomials, and then Braha et al. [5] investigated a Korovkin-type approximation theorem for periodic functions via the statistical summability of the generalized de la Vallée Poussin mean. Very recently, Kadak et al. [10] have established some approximation theorems by statistical weighted $\mathcal{B}$-summability, and then Srivastava and Et [21] established a result on lacunary statistical convergence and strongly lacunary summable functions of order α. Furthermore, Srivastava et al. [24] have proved some interesting results on approximation theorems involving the q-Szàsz–Mirakjan–Kantorovich type operators via Dunkl's generalization.

Motivated essentially by the above-mentioned works, in view of establishing certain new approximation results, we now recall the deferred Cesàro $D(a_n, b_n)$ summability mean as follows.

Let (a_n) and (b_n) be sequences of non-negative integers such that (i) $a_n < b_n$ and (ii) $\lim\limits_{n\to\infty} b_n = \infty$, then the deferred Cesàro $D(x_n)$ mean is defined by (Agnew [1, p. 414]),

$$D(x_n) = \frac{x_{a_n+1} + x_{a_n+2} + \cdots + x_{b_n}}{b_n - a_n} = \frac{1}{b_n - a_n} \sum_{k=a_n+1}^{b_n} x_k \tag{23.1}$$

It is well known that $D(x_n)$ is *regular* under conditions (i) and (ii) (see Agnew [1]).

Also, very recently Srivastava et al. [23] have introduced deferred weighted mean, $D_a^b(\overline{N}, p, q)$ as,

$$t_n = \frac{1}{R_{a_n+1}^{b_n}} \sum_{m=a_n+1}^{b_n} p_m q_m x_m.$$

It will be interesting to see that, for $p_m = q_m = 1$, t_n is same as $D(x_n)$. Thus, *deferred Cesàro mean* is very fundamental in the study of such type of means. Here, we have considered the statistical summability via deferred Cesàro mean in order to establish certain approximation theorems.

Let us now introduce the following definitions in support of our proposed work.

Definition 23.2.1 A sequence (x_n) is said to be *deferred Cesàro convergent* to ℓ if, for every $\epsilon > 0$, the set

$$\{k : a_n < k \leq b_n \text{ and } |x_n - \ell| \geq \epsilon\}$$

has deferred Cesàro density zero [12, 29] that is,

$$\lim_{n\to\infty} \frac{1}{b_n - a_n} |\{k : a_n < k \leq b_n \ \text{ and } \ |x_n - \ell| \geq \epsilon\}| = 0.$$

In this case, we write

$$\text{stat}_{DC} \lim_{n\to\infty} x_n = \ell.$$

Definition 23.2.2 A sequence (x_n) is said to be *statistical deferred Cesàro summable* to ℓ if, for every $\epsilon > 0$, the set

$$\{k : a_n < k \leq b_n \ \text{ and } \ |D(x_n) - \ell| \geq \epsilon\}$$

has deferred Cesàro summable density zero, that is,

$$\lim_{n\to\infty} \frac{1}{b_n - a_n} |\{k : a_n < k \leq b_n \ \text{ and } \ |D(x_n) - \ell| \geq \epsilon\}| = 0.$$

In this case, we write

$$\text{stat} - \lim_{n\to\infty} D(x_n) = \ell \ \text{ or } \ DC_1(\text{stat}) - \lim_{n\to\infty} x_n = \ell.$$

Clearly, above definition can be viewed as the generalization of some existing definitions.

We now prove the following theorem which determines the inclusion relation between the deferred Cesàro statistical convergence and the statistical deferred Cesàro summability.

Theorem 23.2.3 *Let a sequence (x_n) is deferred Cesàro statistical convergent to a number ℓ, then it is statistical deferred Cesàro summable to the same number ℓ, but the converse is not true.*

Proof Suppose (x_n) is deferred Cesàro statistically convergent to ℓ. By the hypothesis, we have

$$\lim_{n\to\infty} \frac{1}{b_n - a_n} |\{k : a_n < k \leq b_n \ \text{ and } \ |x_n - \ell| \geq \epsilon\}| = 0.$$

Consider two sets as follows:

$$\mathcal{K}_\epsilon = \lim_{n\to\infty} |\{k : a_n < k \leq b_n \ \text{ and } \ |x_n - \ell| \geq \epsilon\}|$$

and

$$\mathcal{K}_\epsilon^c = \lim_{n\to\infty} |\{k : a_n < k \leq b_n \ \text{ and } \ |x_n - \ell| < \epsilon\}|.$$

Now,

$$\begin{aligned}
|D(x_n) - \ell| &= \left| \frac{1}{b_n - a_n} \sum_{m=a_n+1}^{b_n} (x_k) - \ell \right| \\
&\leqq \left| \frac{1}{b_n - a_n} \sum_{m=a_n+1}^{b_n} (x_k - \ell) \right| + |\ell| \left| \frac{1}{b_n - a_n} \sum_{m=a_n+1}^{b_n} 1 - 1 \right| \\
&\leqq \frac{1}{b_n - a_n} \sum_{\substack{m=a_n+1 \\ (k \in \mathcal{K}_\epsilon)}}^{b_n} |x_k - \ell| + \frac{1}{b_n - a_n} \sum_{\substack{m=a_n+1 \\ (k \in \mathcal{K}_\epsilon^c)}}^{b_n} |x_k - \ell| + \frac{|\ell|}{b_n - a_n} \\
&\leqq \frac{1}{b_n - a_n} |\mathcal{K}_\epsilon| + \frac{1}{b_n - a_n} \mathcal{K}_\epsilon^c + 0 \\
&\to 0 \text{ as } n \to \infty \quad (\because \lim_{n \to \infty} b_n = \infty),
\end{aligned}$$

which implies that $D(x_n) \to \ell$. Hence, the sequence (x_n) is statistically deferred Cesàro summable to the same number ℓ. □

In order to prove that the converse is not true, we consider an example (below).

Example Suppose that

$$a_n = 2n - 1 \text{ and } b_n = 4n - 1,$$

and also consider a sequence (x_n) by

$$x_n = \begin{cases} 0 & (n \text{ is even}) \\ 1 & (n \text{ is odd}). \end{cases} \tag{23.2}$$

It is easy to see that (x_n) is neither convergent nor statistical convergent. Further, we have

$$\frac{1}{b_n - a_n} \sum_{m=a_n+1}^{b_n} x_m = \frac{1}{2n} \sum_{m=2n}^{4n-1} x_m = \frac{1}{2n} \frac{2n}{2} = \frac{1}{2}.$$

That is, (x_n) deferred Cesàro summable to $\frac{1}{2}$, and so also, statistical deferred Cesàro summable to $\frac{1}{2}$; however, it is not deferred Cesàro statistical convergent.

23.3 A Korovkin-Type Theorem

The theory of approximation of functions has been originated from a well-known theorem of Weierstrass, it has become an exciting interdisciplinary field of study for the last 130 years. Later, the theory of approximation was enriched by Korovkin-type approximation results. Korovkin-type theorems furnish simple and useful tools for ascertaining whether a given sequence of positive linear operators acting on some function space is an approximation process or, equivalently, converges strongly to the identity operator. Roughly speaking, these theorems exhibit a variety of test subsets of functions which guarantee that the approximation (or the convergence) property holds in the whole space provided it holds on them. The custom of calling these kinds of results "Korovkin-type theorems" refers to P. P. Korovkin who in 1953 discovered such a property for the functions $1,\ x$, and x^2 in the space $C([0,1])$ of all continuous functions on the real interval as well as for the functions 1, $\cos x$, and $\sin x$ in the space of all continuous 2π-periodic functions on the real line. Several mathematicians have worked on extending or generalizing the Korovkin-type theorems in many ways and to several settings, including function spaces, abstract Banach lattices, Banach algebras, Banach spaces, and so on. This theory is very useful in real analysis, functional analysis, harmonic analysis, measure theory, probability theory, summability theory, and partial differential equations. Recently, Mohiuddine [13] has obtained an application of almost convergence for single sequences in Korovkin-type approximation theorem and proved some related results. For the function of two variables, such type of approximation theorems are proved in [3] by using almost convergence of double sequences. Quite recently, in [16] and [17] the Korovkin-type theorem is proved for statistical λ-convergence and statistical lacunary summability, respectively. For some recent work on this topic, we refer to [4, 6, 19, 20, 23] and [26]. Recently, Mohiuddine et al. [14] have proved the Korovkin theorem on $\mathcal{C}[0,\infty)$ by using the test functions $1,\ e^{-x}$, and e^{-2x}. In this paper, we generalize the result of Mohiuddine, Alotaibi, and Mursaleen via the notion of statistical deferred Cesàro summability for the same test functions $1,\ e^{-x}$, and e^{-2x}. We also present an example to justify that our result is stronger than that of Mohiuddine, Alotaibi, and Mursaleen (see [14]).

Let $\mathcal{C}(X)$ be the space of all real-valued continuous functions defined on $[0,\infty)$ under the norm $\|.\|_\infty$. Also, $\mathcal{C}[0,\infty)$ is a Banach space. We have, for $f \in \mathcal{C}[0,\infty)$, the norm of f denoted by $\|f\|$ is given by,

$$\|f\|_\infty = \sup_{x\in[0,\infty)} \{|f(x)|\}$$

with

$$\omega(\delta, f) = \sup_{0\le|h|\le\delta} \|f(x+h) - f(x)\|_\infty,\ \ f \in \mathcal{C}[0,\infty).$$

The quantities $\omega(\delta, f)$ is called the modulus of continuity of f.

Let $L : \mathcal{C}[0, \infty) \to \mathcal{C}[0, \infty)$ be a linear operator. Then, as usual, we say that L is a positive linear operator provided that $f \geq 0$ implies $L(f) \geq 0$. Also, we denote the value of $L(f)$ at a point $x \in [0, \infty)$ by $L(f(u); x)$ or, briefly, $L(f; x)$.

The classical Korovkin theorem states as follows [11].

Let $L_n : \mathcal{C}[a, b] \to \mathcal{C}[a, b]$ be a sequence of positive linear operators and let $f \in \mathcal{C}[0, \infty)$. Then

$$\lim_{n\to\infty} \|L_n(f; x) - f(x)\|_\infty = 0 \Longleftrightarrow \lim_{n\to\infty} \|L_n(f_i; x) - f_i(x)\|_\infty = 0,$$

where

$$f_0(x) = 1, \quad f_1(x) = x \text{ and } f_2(x) = x^2 \ (i = 0, 1, 2).$$

The statistical Cesàro summability version for the theorem established by Mohiuddine et al. [14] states as follows.

Let $L_n : \mathcal{C}[0, \infty) \to \mathcal{C}[0, \infty)$ be a sequence of positive linear operators and let $f \in \mathcal{C}[0, \infty)$. Then

$$C_1(\text{stat}) - \lim_{n\to\infty} \|L_n(f; x) - f(x)\|_\infty = 0$$

if and only if

$$C_1(\text{stat}) - \lim_{n\to\infty} \|L_n(f_i; x) - f_i(x)\|_\infty = 0 \ (i = 0, 1, 2),$$

where

$$f_0(x) = 1, \quad f_1(x) = e^{-x} \text{ and } f_2(x) = e^{-2x}.$$

Now we prove the following theorem by using the notion of statistical deferred Cesàro summability.

Theorem 23.3.1 *Let $L_m : \mathcal{C}[0, \infty) \to \mathcal{C}[0, \infty)$ be a sequence of positive linear operators. Then for all $f \in \mathcal{C}[0, \infty)$*

$$DC_1(\text{stat}) - \lim_{m\to\infty} \|L_m(f; x) - f(x)\|_\infty = 0, \tag{23.3}$$

if and only if

$$DC_1(\text{stat}) - \lim_{m\to\infty} \|L_m(1; x) - 1\|_\infty = 0, \tag{23.4}$$

$$DC_1(\text{stat}) - \lim_{m\to\infty} \|L_m(e^{-s}; x) - e^{-x}\|_\infty = 0 \tag{23.5}$$

$$DC_1(\text{stat}) - \lim_{m\to\infty} \|L_m(e^{-2s}; x) - e^{-2x}\|_\infty = 0. \tag{23.6}$$

Proof Since each of $f_i(x) = \{1, e^{-x}, e^{-2x}\} \in \mathcal{C}(X)$ $(i = 0, 1, 2)$ is continuous, the implication (23.3)$\Longrightarrow$(23.4)–(23.6) is obvious. In order to complete the proof of the theorem we first assume that (23.4)–(23.6) hold true. Let $f \in \mathcal{C}[X]$, then there exists a constant $\mathcal{K} > 0$ such that $|f(x)| \leq \mathcal{K}$, $\forall\, x \in X = [0, \infty)$.

Thus,

$$|f(s) - f(x)| \leq 2\mathcal{K}, \ s, x \in X. \tag{23.7}$$

Clearly, for given $\epsilon > 0$, there exists $\delta > 0$ such that

$$|f(s) - f(x)| < \epsilon \tag{23.8}$$

whenever $|e^{-s} - e^{-x}| < \delta$, for all $s, x \in X$.

Let us choose $\varphi_1 = \varphi_1(s, x) = (e^{-s} - e^{-x})^2$. If $|e^{-s} - x^{-x}| \geq \delta$, then we obtain

$$|f(s) - f(x)| < \frac{2\mathcal{K}}{\delta^2}\varphi_1(s, x). \tag{23.9}$$

From Eqs. (23.8) and (23.9), we get

$$|f(s) - f(x)| < \epsilon + \frac{2\mathcal{K}}{\delta^2}\varphi_1(s, x),$$

This implies that,

$$-\epsilon - \frac{2\mathcal{K}}{\delta^2}\varphi_1(s, x) \leq f(s) - f(x) \leq \epsilon + \frac{2\mathcal{K}}{\delta^2}\varphi_1(s, x). \tag{23.10}$$

Now since $L_m(1; x)$ is monotone and linear, so by applying the operator $L_m(1; x)$ to this inequality, we have

$$\begin{aligned} L_m(1; x)\left(-\epsilon - \frac{2\mathcal{K}}{\delta^2}\varphi_1(s, x)\right) &\leq L_m(1; x)(f(s) - f(x)) \\ &\leq L_m(1; x)\left(\epsilon + \frac{2\mathcal{K}}{\delta^2}\varphi_1(s, x)\right). \end{aligned} \tag{23.11}$$

Note that x is fixed and so $f(x)$ is a constant number. Therefore,

$$\begin{aligned} -\epsilon L_m(1; x) - \frac{2\mathcal{K}}{\delta^2}L_m(\varphi_1; x) &\leq L_m(f; x) - f(x)L_m(1; x) \\ &\leq \epsilon L_m(1; x) + \frac{2\mathcal{K}}{\delta^2}L_m(\varphi_1; x). \end{aligned} \tag{23.12}$$

But

$$L_m(f;x) - f(x) = [L_m(f;x) - f(x)L_m(1;x)] \\ + f(x)[L_m(1;x) - 1]. \tag{23.13}$$

Using (23.12) and (23.13), we have

$$L_m(f;x) - f(x) < \epsilon L_m(1;x) + \frac{2\mathcal{K}}{\delta^2} L_m(\varphi_1;x) \\ + f(x)[L_m(1;x) - 1]. \tag{23.14}$$

Now, estimate $L_m(\varphi_1;x)$ as,

$$\begin{aligned} L_m(\varphi_1;x) &= L_m((e^{-s} - e^{-x})^2;x) = L_m(e^{-2s} - 2e^{-x}e^{-s} + e^{-2x};x) \\ &= L_m(e^{-2s};x) - 2e^{-x}L_m(e^{-s};x) + e^{-2s}L_m(1;x) \\ &= [L_m(e^{-2s};x) - e^{-2x}] - 2e^{-x}[L_m(e^{-s};x) - e^{-x}] \\ &\quad + e^{-2x}[L_m(1;x) - 1]. \end{aligned}$$

Using (23.14), we obtain

$$\begin{aligned} L_m(f;x) - f(x) &< \epsilon L_m(1;x) + \frac{2\mathcal{K}}{\delta^2}\{[L_m(e^{-2s};x) - e^{-2s}] \\ &-2e^{-x}[L_m(e^{-s};x) - e^{-x}] + e^{-2s}[L_m(1;x) - 1]\} \\ &+ f(x)[L_m(1;x) - 1]. \\ &= \epsilon[L_m(1;x) - 1] + \epsilon + \frac{2\mathcal{K}}{\delta^2}\{[L_m(e^{-2s};x) - e^{-2x}] \\ &-2e^{-x}[L_m(e^{-s};x) - e^{-x}] + e^{-2x}[L_m(1;x) - 1]\} \\ &+ f(x)[L_m(1;x) - 1]. \end{aligned}$$

Since ϵ is arbitrary, we can write

$$\begin{aligned} |L_m(f;x) - f(x)| &\le \epsilon + \left(\epsilon + \frac{2\mathcal{K}}{\delta^2} + \mathcal{K}\right)|L_m(1;x) - 1| \\ &\quad + \frac{4\mathcal{K}}{\delta^2}|L_m(e^{-s};x) - e^{-x}| + \frac{2\mathcal{K}}{\delta^2}|L_m(e^{-2s};x) - e^{-2x}| \\ &\le B(|L_m(1;x) - 1| + |L_m(e^{-s};x) - e^{-x}| \\ &\quad + |L_m(e^{-2s};x) - e^{-2x}|) \end{aligned} \tag{23.15}$$

where

$$B = \max\left(\epsilon + \frac{2\mathcal{K}}{\delta^2} + \mathcal{K}, \frac{4\mathcal{K}}{\delta^2}, \frac{2\mathcal{K}}{\delta^2}\right).$$

Now replacing $L_m(f; x)$ by $\mathfrak{L}(f; x)$, where

$$\mathfrak{L}(f; x) = \frac{1}{b_n - a_n} \sum_{m=a_n+1}^{b_n} T_m(f; x)$$

and next, for a given $r > 0$, there exists $\epsilon > 0$ with $\epsilon < r$ and by setting

$$\Psi_m(x; r) = \{m : a_n < m \leq b_n \text{ and } |\mathfrak{L}(f; x) - f(x)| \geq r\}$$

and

$$\Psi_{i,m}(x; r) = \left\{m : a_n < m \leq b_n \text{ and } |\mathfrak{L}(f_i; x) - f_i(x)| \geq \frac{r - \epsilon}{3B}\right\},$$

Eq. (23.15) yields

$$\Psi_m(x; r) \leq \sum_{i=0}^{2} \Psi_{i,m}(x; r).$$

Clearly,

$$\frac{\|\Psi_m(x; r)\|_{\mathcal{C}(X)}}{b_n - a_n} \leq \sum_{i=0}^{2} \frac{\|\Psi_{i,m}(x; r)\|_{\mathcal{C}(X)}}{b_n - a_n}. \tag{23.16}$$

Now, using the above assumption about the implications in (23.4)–(23.6) and by Definition 23.2.2, the right-hand side of (23.16) is seen to tend to zero as $n \to \infty$. Consequently, we get

$$\lim_{n\to\infty} \frac{\|\Psi_m(x; r)\|_{\mathcal{C}(X)}}{b_n - a_n} = 0 \ (r > 0).$$

Therefore, the implication (23.3) holds true. This completes the proof of Theorem 23.3.1. □

Remark 23.3.2 By taking $a_n = 0, b_n = n, \ \forall\, n$ in Theorem 23.3.1, one can obtain the statistical Cesàro summability version of Korovkin-type approximation for the set of functions $1, \ e^{-x}$, and e^{-2x} established by Mohiuddine et al. [14].

Now we present below an illustrative example for the sequence of positive linear operators that does not satisfy the conditions of the Korovkin approximation theorems due to Mohiuddine et al. [14] and Boyanov and Veselinov [4] but satisfies the conditions of our Theorem 23.3.1. Thus, our theorem is stronger than the results established by both Mohiuddine et al. [14] and Boyanov and Veselinov [4].

Here we consider the operator

$$x(1 + xD) \quad \left(D = \frac{d}{dx}\right)$$

which was used by Al-Salam [2] and, more recently, by Viskov and Srivastava [28] (see also the monograph by Srivastava and Manocha [22] for various general families of operators of this kind). Here, we use this operator over the Baskakov operators.

Example Let $L_m : \mathcal{C}[0, \infty) \to \mathcal{C}[0, \infty)$ be defined by

$$L_m(f; x) = (1 + x_m)x(1 + xD)V_m(f; x), \tag{23.17}$$

where

$$V_m(f; x) = \sum_{k=0}^{\infty} f\left(\frac{k}{m}\right)\binom{m-k+1}{k}x^k \cdot (1 + x)^{-n-k}$$

and (x_m) is a sequence as defined in Eq. (23.2).

Recall that (see [4]),

$$V_m(1; x) = 1, \quad V_m(e^{-s}; x) = (1 + x - xe^{-\frac{1}{m}})^{-m}$$

and

$$V_m(f; x) = (1 + x^2 - x^2e^{-\frac{1}{m}})^{-m}.$$

Now, we have

$$\begin{aligned}
L_m(1; x) &= [1 + x_m]x(1 + xD)1 = [1 + x_m]x, \\
L_m(e^{-s}; x) &= [1 + x_m]x(1 + xD)(1 + x - xe^{-\frac{1}{m}})^{-m} \\
&= [1 + x_m]x(1 + x - xe^{-\frac{1}{m}})^{-m} \\
&\quad \cdot \left(1 - mx(1 - e^{-\frac{1}{m}})(1 + x - xe^{-\frac{1}{m}})^{-1}\right), \\
L_m(e^{-2s}; x) &= [1 + x_m]x(1 + xD)(1 + x^2 - x^2e^{-\frac{1}{m}})^{-m} \\
&= [1 + x_m]x(1 + x^2 - x^2e^{-\frac{1}{m}})^{-m} \\
&\quad \cdot \left(1 - 2mx^2(1 - e^{-\frac{1}{m}})(1 + x^2 - x^2e^{-\frac{1}{m}})^{-1}\right).
\end{aligned}$$

Thus, we obtain

$$DC_1(\text{stat}) - \lim_{m\to\infty} \|L_m(1; x) - 1\|_\infty = 0,$$
$$DC_1(\text{stat}) - \lim_{m\to\infty} \|L_m(e^{-s}; x) - e^{-x}\|_\infty = 0$$

and

$$DC_1(\text{stat}) - \lim_{m\to\infty} \|L_m(e^{-2s}; x) - e^{-2x}\|_\infty = 0,$$

that is, the sequence $L_m(f; x)$ satisfies the conditions (23.4)–(23.6). Therefore by Theorem 23.3.1, we have

$$DC_1(\text{stat}) - \lim_{m\to\infty} \|L_m(f; x) - f\|_\infty = 0.$$

Hence, it is statistically deferred Cesàro summable; however, since (x_m) is neither statistically convergent nor statistically Cesàro summable, so we conclude that earlier works under [14] and [4] are not valid for the operators defined by (23.17), while our Theorem 23.3.1 still works.

23.4 Rate of Statistical Deferred Cesàro Summability

In this section, we study the rates of statistical deferred Cesàro summability of a sequence of positive linear operators $L(f; x)$ defined on $\mathcal{C}[0, \infty)$ with the help of modulus of continuity.

We now presenting the following definition.

Definition 23.4.1 Let (u_n) be a positive non-increasing sequence. A given sequence $x = (x_m)$ is statistically deferred Cesàro summable to a number ℓ with rate $o(u_n)$, if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{u_n(b_n - a_n)} |\{m : a_n < m \leq b_n \text{ and } |D(x_m) - \ell| \geq \epsilon\}| = 0.$$

In this case, we may write

$$x_m - \ell = DC_1(\text{stat}) - o(u_n).$$

We now prove the following basic lemma.

Lemma 23.4.2 *Let (u_n) and (v_n) be two positive non-increasing sequences. Let $x = (x_m)$ and $y = (y_m)$ be two sequences such that*

$$x_m - \ell_1 = DC_1(stat) - o(u_n)$$

and

$$y_m - \ell_2 = DC_1(stat) - o(v_n)$$

respectively. Then the following conditions hold true:

(i) $(x_m + y_m) - (\ell_1 + \ell_2) = DC_1(stat) - o(w_n)$*;*
(ii) $(x_m - \ell_1)(y_m - \ell_2) = DC_1(stat) - o(u_n v_n)$*;*
(iii) $\lambda(x_m - \ell_1) = DC_1(stat) - o(u_n)$ *(for any scalar* λ*);*
(iv) $\sqrt{|x_m - \ell_1|} = DC_1(stat) - o(u_n)$*,*

where

$$w_n = \max\{u_n, v_n\}.$$

Proof (i) In order to prove the condition (i), for $\epsilon > 0$ and $x \in [0, \infty)$, we define the following sets:

$$A_n(x;\epsilon) = |\{m : a_n < m \leq b_n \text{ and } |D(x_m) + D(y_m) - (\ell_1 + \ell_2)| \geq \epsilon\}|,$$
$$A_{0,n}(x;\epsilon) = \left|\left\{m : a_n < m \leq b_n \text{ and } |D(x_m) - \ell_1| \geq \frac{\epsilon}{2}\right\}\right|,$$

and

$$A_{1,n}(x;\epsilon) = \left|\left\{m : a_n < m \leq b_n \text{ and } |D(y_m) - \ell_2| \geq \frac{\epsilon}{2}\right\}\right|.$$

Clearly, we have

$$A_n(x;\epsilon) \subseteq A_{0,n}(x;\epsilon) \cup A_{1,n}(x;\epsilon).$$

Moreover, since

$$w_n = \max\{u_n, v_n\},$$

by condition (23.3) of Theorem 23.3.1, we obtain

$$\frac{\|A_m(x;\epsilon)\|_\infty}{w_n(b_n - a_n)} \leq \frac{\|A_{0,n}(x;\epsilon)\|_\infty}{u_n(b_n - a_n)} + \frac{\|A_{1,n}(x;\epsilon)\|_\infty}{v_n(b_n - a_n)}. \tag{23.18}$$

Now, by conditions (23.4)–(23.6) of Theorem 23.3.1, we obtain

$$\frac{\|A_n(x;\epsilon)\|_\infty}{w_n(b_n - a_n)} = 0, \tag{23.19}$$

which establishes (i). □

Proof (ii) In order to prove the condition (ii), for $\epsilon > 0$ and $x \in [0, \infty)$, we define the following sets:

$$G_n(x;\epsilon) = |\{m : a_n < m \leq b_n \text{ and } |D(x_m)D(y_m) - (\ell_1\ell_2)| \geq \epsilon\}|,$$

$$G_{0,n}(x;\epsilon) = \left|\left\{m : a_n < m \leq b_n \text{ and } |D(x_m) - \ell_1| \geq \frac{\epsilon}{2}\right\}\right|,$$

and

$$G_{1,n}(x;\epsilon) = \left|\left\{m : a_n < m \leq b_n \text{ and } |D(y_m) - \ell_2| \geq \frac{\epsilon}{2}\right\}\right|.$$

Clearly, we have

$$G_n(x;\epsilon) \subseteq G_{0,n}(x;\epsilon) \cup G_{1,n}(x;\epsilon).$$

Moreover, since

$$w_n = \max\{u_n, v_n\},$$

by condition (23.3) of Theorem 23.3.1, we obtain

$$\frac{\|G_m(x;\epsilon)\|_\infty}{u_n v_n(b_n - a_n)} \leq \frac{\|G_{0,n}(x;\epsilon)\|_\infty}{u_n(b_n - a_n)} + \frac{\|G_{1,n}(x;\epsilon)\|_\infty}{v_n(b_n - a_n)}.$$

Now, by conditions (23.4)–(23.6) of Theorem 23.3.1, we obtain

$$\frac{\|G_n(x;\epsilon)\|_\infty}{u_n v_n(b_n - a_n)} = 0,$$

which establishes (ii).

Since the proofs of other conditions (iii)–(iv) are similar, we omit them. □

Further, we recall that the modulus of continuity of a function $f \in \mathcal{C}[0, \infty)$ is defined by

$$\omega(f,\delta) = \sup_{|y-x|\leq\delta: x,y\in X} |f(y) - f(x)| \quad (\delta > 0),$$

this implies that

$$|f(y) - f(x)| \leq \omega(f,\delta)\left(\frac{|x-y|}{\delta} + 1\right). \tag{23.20}$$

Now we state and prove a new result in the form of the following theorem.

Theorem 23.4.3 *Let* $[0, \infty) \subset \mathbb{R}$ *and let* $L_m : \mathcal{C}[0, \infty) \to \mathcal{C}[0, \infty)$ *be a sequence of positive linear operators. Assume that the following conditions hold true:*

(i) $\|L_m(1; x) - 1\|_\infty = DC_1(stat) - o(u_n)$,
(ii) $\omega(f, \lambda_m) = DC_1(stat) - o(v_n)$,
where

$$\lambda_m = \sqrt{L_m(\varphi^2; x)} \ \textit{and} \ \varphi_1(y, x) = (e^{-y} - x^{-x})^2.$$

Then, for all $f \in \mathcal{C}[0, \infty)$*, the following statement holds true:*

$$\|L_m(f; x) - f\|_\infty = DC_1(stat) - o(w_n), \tag{23.21}$$

$w_n = \max\{u_n, v_n\}$.

Proof Let $f \in \mathcal{C}[0, \infty)$ and $x \in [0, \infty)$. Using (23.20), we have

$$\begin{aligned}
|L_m(f; x) - f(x)| &\leq L_m(|f(y) - f(x)|; x) + |f(x)||L_m(1; x) - 1| \\
&\leq L_m\left(\frac{|e^{-x} - e^{-y}|}{\lambda_m} + 1; x\right)\omega(f, \lambda_m) + |f(x)||L_m(1; x) - 1| \\
&\leq L_m\left(1 + \frac{1}{\lambda_m^2}(e^{-x} - e^{-y})^2; x\right)\omega(f, \lambda_m) \\
&\quad + |f(x)||L_m(1; x) - 1| \\
&\leq \left(L_m(1; x) + \frac{1}{\lambda_m^2}L_m(\varphi_x; x)\right)\omega(f, \lambda_m) \\
&\quad + |f(x)||L_m(1; x) - 1|.
\end{aligned}$$

Putting $\lambda_m = \sqrt{L_m(\varphi^2; x)}$, we get

$$\begin{aligned}
\|L_m(f; x) - f(x)\|_\infty &\leq 2\omega(f, \lambda_m) + \omega(f, \lambda_m)\|L_m(1; x) - 1\|_\infty \\
&\quad + \|f(x)\|\|L_m(1; x) - 1\|_\infty \\
&\leq \mathcal{M}\{\omega(f, \lambda_m) + \omega(f, \lambda_m)\|L_m(1; x) - 1\|_\infty \\
&\quad + \|L_m(1; x) - 1\|_\infty\},
\end{aligned}$$

where

$$\mathcal{M} = \{\|f\|_\infty, 2\}.$$

Thus,

$$\left\| \frac{1}{b_n - a_n} \sum_{m=a_n+1}^{b_n} L_m(f; x) - f(x) \right\|_\infty \leq \mathcal{M} \left\{ \omega(f, \lambda_m) \frac{1}{b_n - a_n} \right.$$

$$\left. + \omega(f, \lambda_m) \left\| \frac{1}{b_n - a_n} \sum_{m=a_n+1}^{b_n} L_m(f; x) - f(x) \right\|_\infty \right\}$$

$$+ \mathcal{M} \left\{ \left\| \frac{1}{b_n - a_n} \sum_{m=a_n+1}^{b_n} L_m(f; x) - f(x) \right\|_\infty \right\}.$$

Now, by using the conditions (i) and (ii) of Theorem 23.4.3, in conjunction with Lemma 23.4.2, we arrive at the statement (23.21) of Theorem 23.4.3.

This completes the proof of Theorem 23.4.3. □

23.5 Concluding Remarks and Observations

In this concluding section of our investigation, we present several further remarks and observations concerning to various results which we have proved here.

Remark 23.5.1 Let $(x_m)_{m \in \mathbb{N}}$ be a sequence given in Eq. (23.2). Then, since

$$DC_1(\text{stat}) - \lim_{m \to \infty} x_m \to 0 \ \text{ on } \ [0, \infty),$$

we have

$$DC_1(\text{stat}) - \lim_{m \to \infty} \|L_m(f_i; x) - f_i(x)\|_\infty = 0. \tag{23.22}$$

Thus, we can write (by Theorem 23.3.1)

$$DC_1(\text{stat}) - \lim_{m \to \infty} \|L_m(f; x) - f(x)\|_\infty = 0, \tag{23.23}$$

where

$$f_0(x) = 1, \quad f_1(x) = e^{-x} \ \text{ and } \ f_2(x) = e^{-2x}.$$

However, since (x_m) is not ordinarily convergent and so also it does not converge uniformly in the ordinary sense. Thus, the classical Korovkin theorem does not work here for the operators defined by (23.17). Hence, this application clearly indicates that our Theorem 23.3.1 is a non-trivial generalization of the classical Korovkin-type theorem (see [11]).

Remark 23.5.2 Let $(x_m)_{m\in\mathbb{N}}$ be a sequence as given in Eq. (23.2). Then, since

$$DC_1(\text{stat}) - \lim_{m\to\infty} x_m \to 0 \text{ on } [0, \infty),$$

so (23.22) holds true. Now by applying (23.22) and Theorem 23.3.1, condition (23.23) holds true. However, since (x_m) does not statistical Cesàro summable, so Theorem 23.3.1 of Mohiuddine et al. (see [14]) does not work for our operator defined in (23.17). Thus, our Theorem 23.3.1 is also a non-trivial extension of Theorem 23.3.1 of Mohiuddine et al. [14] (see also [4] and [11]). Based upon the above results, it is concluded here that our proposed method has successfully worked for the operators defined in (23.17) and therefore it is stronger than the classical and statistical version of the Korovkin-type approximation (see [4, 14] and [11]) established earlier.

Remark 23.5.3 Let us suppose that we replace the conditions (i) and (ii) in Theorem 23.4.3 by the following condition:

$$|L_m(f_i; x) - f_i| = DC_1(\text{stat}) - o(u_{n_i}) \quad (i = 0, 1, 2). \tag{23.24}$$

Then, since

$$L_m(\varphi^2; x) = e^{-2x}|L_m(1; x) - 1| - 2e^{-x}|L_m(e^{-x}; x) - e^{-x}|$$
$$+|L_m(e^{-2x}; x) - e^{-2x}|,$$

we can write

$$L_m(\varphi^2; x) \leq M \sum_{i=0}^{2} |L_m(f_i; x) - f_i(x)|_\infty, \tag{23.25}$$

where

$$M = \{\|f_2\|_\infty + 2\|f_1\|_\infty + 1\}.$$

Now it follows from (23.24), (23.25) and Lemma 23.4.2 that,

$$\lambda_m = \sqrt{L_m(\varphi^2)} = DC_1(\text{stat}) - o(d_n), \tag{23.26}$$

where

$$o(d_n) = \max\{u_{n_0}, u_{n_1}, u_{n_2}\}.$$

This implies,

$$\omega(f, \delta) = DC_1(\text{stat}) - o(d_n).$$

Now using (23.26) in Theorem 23.4.3, we immediately see that, for $f \in \mathcal{C}[0, \infty)$,

$$L_m(f; x) - f(x) = DC_1(\text{stat}) - o(d_n). \tag{23.27}$$

Therefore, if we use the condition (23.24) in Theorem 23.4.3 instead of (i) and (ii), then we obtain the rates of statistical deferred Cesàro summability of the sequence of positive linear operators in Theorem 23.3.1.

Acknowledgements The authors would like to express their heartfelt thanks to the editors and anonymous referees for their most valuable comments and constructive suggestions which leads to the significant improvement of the earlier version of the manuscript.

References

1. R.P. Agnew, On deferred Cesàro means. Ann. Math. **33**, 413–421 (1932)
2. W.A. Al-Salam, Operational representations for the Laguerre and other polynomials. Duke Math. J. **31**, 127–142 (1964)
3. G.A. Anastassiou, M. Mursaleen, S.A. Mohiuddine, Some approximation theorems for functions of two variables through almost convergence of double sequences. J. Comput. Anal. Appl. **13**, 37–40 (2011)
4. B.D. Boyanov, V.M. Veselinov, A note on the approximation of functions in an infinite interval by linear positive operators. Bull. Math. Soc. Sci. Math. Roum. **14**, 9–13 (1970)
5. N.L. Braha, H.M. Srivastava, S.A. Mohiuddine, A Korovkin's type approximation theorem for periodic functions via the statistical summability of the generalized de la Vallée Poussin mean. Appl. Math. Comput. **228**, 162–169 (2014)
6. A.A. Das, B.B. Jena, S.K. Paikray, R.K. Jati, Statistical deferred weighted summability and associated Korovkin-type approximation theorem. Nonlinear Sci. Lett. A **9**(3), 238–245 (2018)
7. H. Fast, Sur la convergence statistique. Colloq. Math. **2**, 241–244 (1951)
8. B.B. Jena, S.K. Paikray, U.K. Misra, Inclusion theorems on general convergence and statistical convergence of $(L, 1, 1)$-summability using generalized Tauberian conditions. Tamsui Oxf. J. Inf. Math. Sci. **31**, 101–115 (2017)
9. B.B. Jena, S.K. Paikray, U.K. Misra, Statistical deferred Cesaro summability and its applications to approximation theorems. Filomat **32**, 2307–2319 (2018)
10. U. Kadak, N.L. Braha, H.M. Srivastava, Statistical weighted $\mathcal{B}$-summability and its applications to approximation theorems. Appl. Math. Comput. **302**, 80–96 (2017)
11. P.P. Korovkin, Convergence of linear positive operators in the spaces of continuous functions (in Russian). Doklady Akad. Nauk. SSSR (New Ser.) **90**, 961–964 (1953)
12. M. Küçükaslan, M. Yılmaztürk, On deferred statistical convergence of sequences. Kyungpook Math. J. **56**, 357–366 (2016)
13. S.A. Mohiuddine, An application of almost convergence in approximation theorems. Appl. Math. Lett. **24**, 1856–1860 (2011)
14. S.A. Mohiuddine, A. Alotaibi, M. Mursaleen, Statistical summability $(C, 1)$ and a Korovkin type approximation theorem. J. Inequal. Appl. **2012**, 1–8 (2012). Article ID 172

15. F. Móricz, Tauberian conditions under which statistical convergence follows from statistical summability $(C, 1)$. J. Math. Anal. Appl. **275**, 277–287 (2002)
16. M. Mursaleen, A. Alotaibi, Statistical summability and approximation by de la Vallée-Poussin mean. Appl. Math. Lett. **24**, 320–324 (2011)
17. M. Mursaleen, A. Alotaibi, Statistical lacunary summability and a Korovkin type approximation theorem. Ann. Univ. Ferrara **57**, 373–381 (2011)
18. M.A. Özarslan, O. Duman, H.M. Srivastava, Statistical approximation results for Kantorovich-type operators involving some special polynomials. Math. Comput. Model. **48**, 388–401 (2008)
19. S.K. Paikray, B.B. Jena, U.K. Misra, Statistical deferred Cesàro summability mean based on (p, q)-integers with application to approximation theorems, in *Advances in Summability and Approximation Theory* (Springer Nature Singapore Pte Ltd., Singapore, 2018)
20. T. Pradhan, S.K. Paikray, B.B. Jena, H. Dutta, Statistical deferred weighted $\mathcal{B}$-summability and its applications to associated approximation theorems. J. Inequal. Appl. **2018**, 1–21 (2018). Article Id: 65
21. H.M. Srivastava, M. Et, Lacunary statistical convergence and strongly lacunary summable functions of order α. Filomat **31**, 1573–1582 (2017)
22. H.M. Srivastava, H.L. Manocha, *A Treatise on Generating Functions*. Halsted Press (Ellis Horwood Limited/Wiley, Chichester/New York, 1984)
23. H.M. Srivastava, B.B. Jena, S.K. Paikray, U.K. Misra, Generalized equi-statistical convergence of the deferred Nörlund summability and its applications to associated approximation theorems. Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Math. **112**, 1487–1501 (2017)
24. H.M. Srivastava, M. Mursaleen, A.M. Alotaibi, Md. Nasiruzzaman, A.A.H. Al-Abied, Some approximation results involving the q-Szász-Mirakjan-Kantorovich type operators via Dunkl's generalization. Math. Methods Appl. Sci. **40**, 5437–5452 (2017)
25. H.M. Srivastava, B.B. Jena, S.K. Paikray, U.K. Misra, A certain class of weighted statistical convergence and associated Korovkin type approximation theorems for trigonometric functions. Math. Methods Appl. Sci. **41**, 671–683 (2018)
26. H.M. Srivastava, B.B. Jena, S.K. Paikray, U.K. Misra, Deferred weighted $\mathcal{A}$-statistical convergence based upon the (p, q)-Lagrange polynomials and its applications to approximation theorems. J. Appl. Anal. **24**, 1–16 (2018)
27. H. Steinhaus, Sur la convergence ordinaire et la convergence asymptotique. Colloq. Math. **2**, 73–74 (1951)
28. O.V. Viskov, H.M. Srivastava, New approaches to certain identities involving differential operators. J. Math. Anal. Appl. **186**, 1–10 (1994)
29. M. Yılmaztürk, M. Küçükaslan, On strongly deferred Cesàro summability and deferred statistical convergence of the sequences. Bitlis Eren Univ. J. Sci. Technol. **3**, 22–25 (2011)