Understanding Complex Systems

Springer : COMPLEXITY

Victor A. Sadovnichiy
Michael Z. Zgurovsky *Editors*

Modern Mathematics and Mechanics

Fundamentals, Problems and Challenges

Springer

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems—cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The three major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations, and the "SpringerBriefs in Complexity" which are concise and topical working reports, case-studies, surveys, essays and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Understanding Complex Systems

Founding Editor: S. Kelso

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition – typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels – and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals: First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuro- and cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

More information about this series at http://www.springer.com/series/5394

Victor A. Sadovnichiy • Michael Z. Zgurovsky
Editors

Modern Mathematics and Mechanics

Fundamentals, Problems and Challenges

Springer

Editors
Victor A. Sadovnichiy
Lomonosov Moscow State University
Moscow, Russia

Michael Z. Zgurovsky
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
Kyiv, Ukraine

ISSN 1860-0832 ISSN 1860-0840 (electronic)
Understanding Complex Systems
ISBN 978-3-319-96754-7 ISBN 978-3-319-96755-4 (eBook)
https://doi.org/10.1007/978-3-319-96755-4

Library of Congress Control Number: 2018958922

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The given collection of papers has been organized as a result of regular open joint academic panels of research workers from the Faculty of Mechanics and Mathematics of Lomonosov Moscow State University and Institute for Applied Systems Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute." This volume is devoted to the fundamentals of modern mathematics and mechanics. It attracted attention of researchers from leading scientific schools of Brazil, France, Germany, Poland, Russian Federation, Spain, Mexico, Ukraine, the USA, and other countries.

Modern technological applications require development and synthesis of fundamental and applied scientific areas, with a view to reducing the gap that may still exist between theoretical basis used for solving complicated technical problems and implementation of obtained innovations. To solve these problems, mathematicians, mechanics, and engineers from wide research and scientific centers have been working together. Results of their joint efforts, including differential geometry, dynamics of differential and difference equations and applications, solid mechanics, and modern methods of optimization and control, are partially presented here. In fact, serial publication of such collected papers to similar seminars is planned.

This is the sequel of earlier volumes:

- Zgurovsky, Michael Z.; Sadovnichiy, Victor A. (Eds.) Continuous and Distributed Systems: Theory and Applications Series: Solid Mechanics and Its Applications, Vol. 211, 2014, XIX, 333 p. 33 illus., 14 illus. in color.
- Victor A. Sadovnichiy and Michael Z. Zgurovsky (Eds.), Continuous and Distributed Systems: Theory and Applications, Volume II, Studies in Systems, Decision and Control, Volume 30, 2015, Springer, Heidelberg xxiv+375pp
- Victor A. Sadovnichiy and Michael Z. Zgurovsky (Eds.), Advances in Dynamical Systems and Control, Studies in Systems, Decision and Control, Volume 69, 2016, Springer, Heidelberg xxii+471pp

In this volume, we are planning to focus on the fundamentals of modern mathematics and mechanics :

(1) We provide the solutions to modern fundamental problems including the complexity of computing of critical points for set-valued mappings, the behavior of solutions (stability, existence, and long-time behavior of solutions, attractors and repellers, numerical approximations, chaos, entropy, and many other features characterizing the dynamics of solutions) of ordinary differential equations, partial differential equations, and difference equations, the development of abstract theory of global attractors for multi-valued impulsive dynamical systems, etc.;

(2) The abstract mathematical approaches, such as differential geometry, differential equations, and difference equations, are applied to the practical applications in solid mechanics, hydro-, aerodynamics, optimization, decision-making theory, and control theory. In particular, in mechanics: classes of Hamiltonian systems can be studied in terms of Fomenko-Zieschang invariants; in solid mechanics: an algorithm for splitting an equilibrium displacement equation system with bulk forces for a transversely isotropic linearly elastic medium that leads to three uncoupled equations with certain canonical fourth-order differential operators in the three components of the displacement vector is described; in hydrodynamics: a simplified model of the trapped vortex is applied to determine the optimal parameters of the control device and dynamical system analysis is used to explore the performance of this control strategy; in aerodynamics: the effects of airfoil thickness and angle-of-attack on nonlinear wake and wing dynamic characteristics are examined; in optimization: an optimal boundary control problem for the system of nonlinear integro-differential evolution equation (cp. Burgers-Sivansky equation) describing the behavior of the flame front interface under some physical assumptions is solved; in control: the methods of automation of impulse processes control in cognitive maps with multirate sampling of measured vertices coordinates are developed.

(3) We hope that these compilations will be of interest to mathematicians and engineers working at the interface of these fields.

The book is addressed to a wide circle of mathematical, mechanical, and engineering readers.

Moscow, Russian Federation — Victor A. Sadovnichiy
Kyiv, Ukraine — Michael Z. Zgurovsky
May 2018

International Editorial Board of This Volume

Contents

Contributors

Alexander V. Aksenov Lomonosov Moscow State University, Moscow, Russian Federation

M. R. Arias Department of Mathematics, University of Extremadura, Badajoz, Spain

I. Baeza Sampere Department of Applied Economy, University of Valencia, Valencia, Spain

Oleg Barabash State University of Telecommunications, Kyiv, Ukraine

Lesia V. Baranovska Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Volodymyr G. Basovsky Institute of Hydromechanics, National Academy of Sciences of Ukraine, Kyiv, Ukraine

R. Benítez Department of Business Mathematics, University of Valencia, Valencia, Spain

V. J. Bolós Department of Business Mathematics, University of Valencia, Valencia, Spain

J. Calatayud Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Valencia, Spain

Jose S. Cánovas Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, Cartagena, Spain

J.-C. Cortés Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Valencia, Spain

Paulo Henrique da Costa Departamento de Matemática, Universidade de Brasília, Brasília, Brazil

Sergey Dashkovskiy University of Würzburg, Würzburg, Germany

Luu Hoang Duc Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

Institute of Mathematics, Vietnam Academy of Science and Technology, Hanoi, Vietnam

Petro Feketa University of Kaiserslautern, Kaiserslautern, Germany

Anatoly T. Fomenko Lomonosov Moscow State University, Moscow, Russian Federation

Vladimir V. Galatenko Lomonosov Moscow State University, Moscow, Russian Federation

María J. Garrido-Atienza Dpto. Ecuaciones Diferenciales y Análisis Numérico, Facultad de Matemáticas, Universidad de Sevilla, Sevilla, Spain

Dimitri V. Georgievskii Lomonosov Moscow State University, Moscow, Russian Federation

Iryna M. Gorban Institute of Hydromechanics, National Academy of Sciences of Ukraine, Kyiv, Ukraine

Nataliia V. Gorban Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Xiaoying Han Department of Mathematics and Statistics, Auburn University, Auburn, AL, USA

Michael A. Högele Departamento de Matemáticas, Universidad de los Andes, Bogotá, Colombia

Phan Thanh Hong Thang Long University, Hanoi, Vietnam

M. Jornet Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Valencia, Spain

Daryna O. Kapustian Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Olena A. Kapustian Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Oleksiy V. Kapustyan Taras Shevchenko National University of Kyiv, Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Volodymyr O. Kapustyan Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

Pavlo O. Kasyanov Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Olha V. Khomenko Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Oleg I. Klesov National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Department of Mathematical Analysis and Probability Theory, Kyiv, Ukraine

Peter E. Kloeden School of Mathematics and Statistics, Huazhong University of Science & Technology, Wuhan, China

Felix-Klein-Zentrum für Mathematik, TU Kaiserslautern, Kaiserslautern, Germany

Peter I. Kogut Differential Equations Department, Oles Honchar National Dnipro University, Dnipro, Ukraine

Oleg Kopiika Institute of Telecommunications and Global Information Space, Kyiv, Ukraine

Olha P. Kupenko System Analysis and Control Department, Dnipro National Technical University "Dnipro Polytechnics", Dnipro, Ukraine

Institute for Applied and System Analysis of National Technical University of Ukraine "Kiev Polytechnic Institute", Kiev, Ukraine

Oleksiy G. Lebid Institute of Telecommunications and Global Information Space, National Academy of Sciences of Ukraine Kyiv, Ukraine,

Taras P. Lukashenko Lomonosov Moscow State University, Moscow, Russian Federation

Sergej O. Mashchenko Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Yuriy L. Milyavsky Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Viktoria V. Mogilova National Technical University of Ukraine "Igor Sikorsky Kiev Polytechnic Institute", Kiev, Ukraine

I. Molchanov University of Bern, Institute of Mathematical Statistics and Actuarial Science, Bern, Switzerland

F. G. Morillas Jurado Department of Applied Economy, University of Valencia, Valencia, Spain

Andrey Musienko Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

A. Navarro-Quiles Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Valencia, Spain

Liliia S. Paliichuk Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Volodymyr V. Pichkur Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Ivan O. Pyshnograiev World Data Center for Geoinformatics and Sustainable Development, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Victor D. Romanenko Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Iryna V. Romaniuk Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Paulo Regis Ruffino IMECC, Universidade Estadual de Campinas, Campinas, Brazil

Victor A. Sadovnichiy Lomonosov Moscow State University, Moscow, Russian Federation

Björn Schmalfuß Institut für Mathematik, Institut für Stochastik, Jena, Germany

Jacson Simsen Departamento de Matemática e Computacao, Universidade Federal de Itajubá, Itajubá, Minas Gerais, Brazil

Valentyn Sobchuk East-European National University of Lesya Ukrainka, Lutsk, Ukraine

Kirill I. Solodskih Lomonosov Moscow State University, Moscow, Russian Federation

Oleksandr M. Stanzhytskyi Taras Shevchenko National University of Kiev, Kiev, Ukraine

Roman M. Taranets Institute of Applied Mathematics and Mechanics of the National Academy of Sciences of Ukraine, Sloviansk, Ukraine

Alisa O. Tsukanova Taras Shevchenko National University of Kiev, Kiev, Ukraine

Olena A. Tymoshenko National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Department of Mathematical Analysis and Probability Theory, Kyiv, Ukraine

Josè Valero Universidad Miguel Hernandez de Elche, Centro de Investigación Operativa, Elche, Spain

Iryna Zamrii State University of Telecommunications, Kyiv, Ukraine

Michael Z. Zgurovsky National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Part I
Differential Geometry

Chapter 1
Convergence Almost Everywhere of Orthorecursive Expansions in Systems of Translates and Dilates

Vladimir V. Galatenko, Taras P. Lukashenko, and Victor A. Sadovnichiy

Abstract Systems of translates and dilates have been widely studied in the last decades. In particular, V.I. Filippov and P. Oswald obtained conditions on a generating function which guarantee that dyadic translates and dilates of this function form a representation system in $L^p[0, 1]$. A.Yu. Kudryavtsev and A.V. Politov showed that under a slightly harder condition on a generating function each element $f \in L^2[0, 1]$ is represented by its orthorecursive expansion in this system. Here we continue studying orthorecursive expansions in systems of dyadic translates and dilates and present results on convergence almost everywhere of these expansions.

1.1 Introduction

Systems of translates and dilates have been widely studied in the last decades, in particular, in theoretical and applied research related to discrete wavelets [1, 2]. For a case of [0, 1] segment and dyadic translates and dilates these systems can be defined as follows. Let φ be a function defined on [0, 1]. We extend this function to $\mathbb{R}$ by setting its values to zero outside [0, 1], and for all $k \in \mathbb{Z}^+$ and $j \in \{0, 1, \ldots, 2^k - 1\}$ define

$$\varphi_{k,j}(x) = \varphi_{2^k+j}(x) = C_k \varphi\left(2^k x - j\right).$$

Here C_k are positive norming constants which can be either set to 1, or selected in such a way that norms of all functions φ_n ($n \in \{1, 2, 3, \ldots\}$) in the resulting system of translates and dilates of φ equal one.

Wavelet theory generally considers systems of translates and dilates with certain special properties, e.g., orthogonal systems. But it has been shown by V.I. Filippov

V. V. Galatenko (✉) · T. P. Lukashenko · V. A. Sadovnichiy
Lomonosov Moscow State University, Moscow, Russian Federation
e-mail: vgalat@imscs.msu.ru; lukashenko@mail.ru; info@rector.msu.ru

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_1

and P. Oswald [3] that for an arbitrary function $\varphi \in L^p[0,1]$ $(1 \le p < \infty)$ with $\int_0^1 \varphi \, d\mu \neq 0$ the system of its translates and dilates is a representation system in L^p, i.e., for every function $f \in L^p[0,1]$ there exists a sequence of coefficients $\{c_n(f)\}_{n=1}^{\infty}$ such that the series $\sum_{n=1}^{\infty} c_n(f)\varphi_n$ converges to f in L^p-norm.

A.Yu. Kudryavtsev and A.V. Politov showed that at least for $p = 2$ under an additional soft condition on φ, namely, the convergence of $\sum_{k=1}^{\infty} \omega_2^2\left(\varphi, 2^{-k}\right)$, where ω_2 is the integral modulus of continuity in $L^2[0,1]$, coefficients for such representation can be found by orthorecursive expansion [4]. This expansion generalizes classical orthogonal expansions preserving such properties as Bessel's identity, Bessel's inequality, and the equivalence of convergence to an expanded element and Parseval's identity [5]. The scheme of orthorecursive expansion is also utilized in greedy expansions in Hilbert spaces [6, 7], known in signal processing and statistics as Matching Pursuit [8] and Projection Pursuit [9], respectively. However, greedy expansions include the selection of an expanding element from a given dictionary at each step, while in the considered settings the set and the order of elements in the orthorecursive expansion is fixed.

Let us recall the definition of an orthorecursive expansion of an element f from a Hilbert space $\left(H, (\cdot\,,\cdot)\right)$ in a system of non-zero elements $\{\varphi_n\}_{n=1}^{\infty} \subset H$. We set $r_0 = f$, and then inductively define coefficients as remainders as follows:

$$\widehat{f_n} = \frac{(r_{n-1}, \varphi_n)}{(\varphi_n, \varphi_n)}, \quad r_n = r_{n-1} - \widehat{f_n}\varphi_n \quad (n = 1, 2, 3, \dots).$$

The series $\sum_{n=1}^{\infty} \widehat{f_n}\varphi_n$ is called an orthorecursive expansion of f in the system $\{\varphi_n\}_{n=1}^{\infty}$.

As noted in [10], the majority of known results for orthorecursive expansions deal with the convergence with respect to the norm induced by a scalar product (L^2-norm for the case of functional systems), and results related to pointwise convergence are scarce. In this paper we show that under certain conditions orthorecursive expansions in systems of translates and dilates converge to an expanded function not only in L^2-norm, but also almost everywhere.

In order to focus on main ideas, but not on technical details, here we present the result in a simple form, with excessive conditions on a function φ which generates a system of translates and dilates and on an expanded function f. A strengthening of the presented result with essentially weaker conditions imposed on φ and f will be presented in subsequent publications.

The result presented in this publication has been previously announced at a conference level, but its proof has not been published.

1.2 Expansion in a System of Functions with Dyadic Supports

We do not limited the analysis to systems of dyadic translates and dilates, but consider systems $\{\varphi_n\}_{n=1}^{\infty} \subset L^2[0,1]$ of real-valued functions with $\operatorname{supp}\varphi_n \subset \Delta_n$, where Δ_n is the n-th dyadic segment:

$$\Delta_{k,j} = \Delta_{2^k+j} = \left[\frac{j}{2^k}, \frac{j+1}{2^k}\right] \quad (k \in \mathbb{Z}^+,\ j \in \{0,1,\ldots,2^k-1\}).$$

In order to simplify formulas we assume that $\|\varphi_n\|_2 = 1$ for all n (here $\|\cdot\|_2$ is the L^2-norm). For every dyadic-irrational point $x \in [0,1]$ and every $k \in \mathbb{Z}^+$ we define $n_k(x)$ as an index lying in $[2^k, 2^{k+1}-1]$ such that $x \in \Delta_{n_k(x)}$. In other words, $n_k(x)$ is an index of a dyadic segment of scale k (or, equivalently, with length 2^{-k}) that covers x.

For the subsequent analysis of convergence almost everywhere, the following lemma turns out to be useful.

Lemma 1.1 *Let $\{c_n\}_{n=1}^{\infty}$ be a sequence of real numbers with $\sum\limits_{n=1}^{\infty} c_n^2 < \infty$. Then $\sum\limits_{k=0}^{\infty} 2^k c_{n_k(x)}^2$ converges almost everywhere on* $[0,1]$.

As coefficients $\{\widehat{f}_n\}_{n=1}^{\infty}$ of an orthorecursive expansion of an element f from a Hilbert space in a normed system belong to ℓ^2 due to Bessel's inequality, this lemma is applicable to orthorecursive expansions. Note that for the considered class of systems for a dyadic-irrational point x all terms of the expansion $\sum\limits_{n=1}^{\infty} \widehat{f}_n\varphi_n(x)$ with $n \notin \{n_k(x)\}_{k=0}^{\infty}$ equal zero in x. Thus each dyadic-irrational point x is naturally associated with a subset of coefficients $\left\{\widehat{f}_{n_k(x)}\right\}_{k=0}^{\infty}$, and the series in dyadic-irrational points is reduced to $\sum\limits_{k=0}^{\infty} \widehat{f}_{n_k(x)}\varphi_{n_k(x)}(x)$.

The proof of the lemma is quite simple. Let $\chi_n(x)$ denote the indicator function of Δ_n. Let us consider L^1-norm of series terms:

$$\left\|2^k c_{n_k(x)}^2\right\|_1 = \left\|\sum_{j=0}^{2^k-1} 2^k c_{2^k+j}^2 \chi_{2^k+j}(x)\right\|_1 = \sum_{j=0}^{2^k-1} c_{2^k+j}^2 = \sum_{n=2^k}^{2^{k+1}-1} c_n^2.$$

Thus,

$$\sum_{k=0}^{\infty} \int_0^1 2^k c_{n_k(x)}^2 \, d\mu(x) < \infty,$$

and it remains to use Levi's theorem [11, Ch.8, Sect. 30] to complete the proof.

Having this lemma, we can proceed to the following theorem.

Theorem 1.1 *Let all functions of a system* $\{\varphi_n(x)\}_{n=1}^{\infty}$ *satisfy the following conditions:* $\varphi_n \geq 0$*;* $\operatorname{supp}\varphi \subset \Delta_n$ *and* φ_n *is continuous on* Δ_n*;* $\|\varphi_n\|_2 = 1$*. Additionally, let the condition*

$$\sup_{K\in\mathbb{Z}^+} \sum_{k=0}^{K} 2^{-k}\omega_{n_k(x)}^2\left(2^{-K}\right) < \infty$$

(where $\omega_{n_k(x)}(\cdot)$ *is the modulus of continuity of* $\varphi_{n_k(x)}$ *on* $\Delta_{n_k(x)}$*) hold for almost all dyadic-irrational points* $x \in [0, 1]$*. Then for every function* $f(x) \in L^2[0, 1]$ *its orthorecursive expansion in* $\{\varphi_n(x)\}_{n=1}^{\infty}$ *converges to* $f(x)$ *in almost all continuity points of* f*.*

In order to prove the theorem, we first note that for all $n \in \{1, 2, 3, \ldots\}$ and all $x \in \Delta_n^o$ the following estimate holds:

$$|r_n(x)| \leq \operatorname{osc}\left(r_n, \Delta_n^o\right).$$

Here r_n is the n-th remainder of the orthorecursive expansion of f in $\{\varphi_n(x)\}_{n=1}^{\infty}$, $\Delta_n^o = \operatorname{int}\Delta_n$ (i.e., $\Delta_{2^k+j}^o = \left(\frac{j}{2^k}, \frac{j+1}{2^k}\right)$, $k \in \mathbb{Z}^+$, $j \in \{0, 1, \ldots, 2^k - 1\}$), and osc denotes oscillation of a function:

$$\operatorname{osc}(g, A) = \sup_{x_1, x_2 \in A} (g(x_1) - g(x_2)).$$

Indeed, due to the definition of orthorecursive expansion r_n is orthogonal to φ_n, and $\operatorname{supp}\varphi_n \subset \Delta_n$, so

$$\int_{\Delta_n^o} r_n(x)\varphi_n(x)\, d\mu(x) = 0.$$

Furthermore, φ_n is non-negative on Δ_n^o and is strictly positive on a positive measure subset of Δ_n^o, thus, r_n can not be strictly positive everywhere on Δ_n^o as well as strictly negative everywhere on Δ_n^o, i.e., values of r_n on Δ_n^o include both non-negative and non-positive numbers. The estimate directly follows from this fact.

In particular, for every dyadic-irrational $x \in [0, 1]$ and every $k \in \mathbb{Z}^+$ the estimate implies the inequality

$$\left|r_{n_k(x)}(x)\right| \leq \operatorname{osc}\left(r_{n_k(x)}, \Delta_{n_k(x)}^o\right).$$

As for a dyadic-irrational x

$$f(x) - \sum_{n=1}^{N} \widehat{f}_n \varphi_n(x) = r_N(x) = r_{n_K(x)}(x),$$

where $K = \max\{k : n_k(x) \le N\}$, the convergence of the orthorecursive expansion in x to $f(x)$ is equivalent to the convergence of $\left\{r_{n_K(x)}(x)\right\}_{K=0}^{\infty}$ to zero. Due to the above inequality, in order to prove this convergence it is sufficient to show that

$$\operatorname{osc}\left(r_{n_K(x)}, \Delta^o_{n_K(x)}\right) \to 0 \quad (K \to \infty).$$

The oscillations can be trivially estimated as follows:

$$\operatorname{osc}\left(r_{n_K(x)}, \Delta^o_{n_K(x)}\right) = \operatorname{osc}\left(f - \sum_{n=1}^{n_K(x)} \widehat{f}_n \varphi_n(x), \Delta^o_{n_K(x)}\right)$$

$$= \operatorname{osc}\left(f - \sum_{k=0}^{K} \widehat{f}_{n_k(x)} \varphi_{n_k(x)}(x), \Delta^o_{n_K(x)}\right)$$

$$\le \operatorname{osc}\left(f, \Delta_{n_K(x)}\right) + \sum_{k=0}^{K} \left|\widehat{f}_{n_k(x)}\right| \operatorname{osc}\left(\varphi_{n_k(x)}, \Delta_{n_K(x)}\right)$$

$$\le \operatorname{osc}\left(f, \Delta_{n_K(x)}\right) + \sum_{k=0}^{K} \left|\widehat{f}_{n_k(x)}\right| \omega_{n_k(x)}\left(2^{-K}\right).$$

Let us consider an arbitrary continuity point x of f which satisfies the following conditions:

(i) x is dyadic-irrational;

(ii) $\sum_{k=0}^{\infty} 2^k \widehat{f}^2_{n_k(x)} < \infty$;

(iii) $C(x) := \sup_{K \in \mathbb{Z}^+} \sum_{k=0}^{K} 2^{-k} \omega^2_{n_k(x)}\left(2^{-K}\right) < \infty.$

Due to Lemma 1.1 and the conditions of the theorem conditions (i)–(iii) hold for almost all continuity points of f. The term $\operatorname{osc}\left(f, \Delta_{n_K(x)}\right)$ goes to zero as $K \to \infty$ due to continuity. Hence, to complete the proof of the theorem it is sufficient to show

that

$$\sum_{k=0}^{K} \left|\widehat{f}_{n_k(x)}\right| \omega_{n_k(x)}\left(2^{-K}\right)$$

also goes to zero.

To show it, let us take an arbitrary positive ε and find $K_0 \in \mathbb{N}$ such that

$$\sum_{k=K_0+1}^{\infty} 2^k \widehat{f}^2_{n_k(x)} < \frac{\varepsilon^2}{4C(x)}.$$

Next, as for a fixed k due to continuity $\omega_{n_k(x)}\left(2^{-K}\right) \to 0$ $(K \to \infty)$, we find $K_1 > K_0$ such that the inequality

$$\sum_{k=0}^{K_0} \left|\widehat{f}_{n_k(x)}\right| \omega_{n_k(x)}\left(2^{-K}\right) < \frac{\varepsilon}{2}$$

holds for all $K > K_1$. In order to estimate the remaining part of the sum we apply Cauchy's inequality and conditions (iii):

$$\sum_{k=K_0+1}^{K} \left|\widehat{f}_{n_k(x)}\right| \omega_{n_k(x)}\left(2^{-K}\right) \le \left(\sum_{k=K_0+1}^{\infty} 2^k \widehat{f}^2_{n_k(x)}\right)^{\frac{1}{2}} \left(\sum_{k=0}^{K} 2^{-k} \omega^2_{n_k(x)}\left(2^{-K}\right)\right)^{\frac{1}{2}}$$

$$< \frac{\varepsilon}{2\sqrt{C(x)}} \cdot \sqrt{C(x)} = \frac{\varepsilon}{2}.$$

So overall for all $K > K_1$ the inequality

$$\sum_{k=0}^{K} \left|\widehat{f}_{n_k(x)}\right| \omega_{n_k(x)}\left(2^{-K}\right) < \varepsilon$$

holds, and the proof of the theorem is complete.

The condition imposed on moduli of continuity in Theorem 1.1 looks technical, but it can be replaced by weaker conditions which have more natural form. E.g., the condition holds if for all $n \in \{1, 2, 3, \ldots\}$ function φ_n is Lipschitz with constant A_n, and $A_n = O\left(n^{3/2}\right)$ (or, equivalently, $A_n = O\left(|\Delta_n|^{-3/2}\right) = O\left(2^{3k(n)/2}\right)$, where $k(n) = \lfloor \log_2 n \rfloor$ is the scale of Δ_n). Indeed, in this case there exists a constant A such that $A_n \le A \cdot 2^{3k(n)/2}$ for all n, and for all $K \in \mathbb{Z}^+$ and all dyadic-irrational

points x we have

$$\sum_{k=0}^{K} 2^{-k}\omega^2_{n_k(x)}\left(2^{-K}\right) \le \sum_{k=0}^{K} 2^{-k} \cdot A^2_{n_k(x)} 2^{-2K} \le A^2 \sum_{k=0}^{K} 2^{-k-2K} \cdot 2^{3k}$$

$$= A^2 \sum_{k=0}^{K} 4^{k-K} < \frac{4A^2}{3}.$$

Lipshitz condition can be extended to Hölder condition with the same exponent α $(0 < \alpha \le 1)$ for all functions φ_n and constants $A_n = O\left(n^{\alpha+1/2}\right)$.

Clearly Theorem 1.1 can be generalized from a system of dyadic segments to a much wider setting, such as a system of segments $\{\Delta_n\}_{n=1}^{\infty}$ which forms a Vitali covering of $[0, 1]$ (i.e., for every $\delta > 0$ and every $x \in [0, 1]$ there exists a segment Δ_n such that $|\Delta_n| < \delta$, $\Delta_n \ni x$) and satisfies the following additional condition: if $n_1 < n_2$ and segments Δ_{n_1}, Δ_{n_2} overlap, then $\Delta_{n_1} \supset \Delta_{n_2}$ (see [5, Theorem 3] and [12, Cond. $(\Xi 1), (\Xi 2)$]). In this case points excluded from the analysis (similarly to dyadic-rational points) are ends of the segments Δ_n, $\{n_k(x)\}_{k=0}^{\infty}$ is an increasing sequence of all indexes n for which $\Delta_n \ni x$, and the condition on continuity takes the form

$$\sup_{K\in\mathbb{Z}^+} \sum_{k=0}^{K} \left|\Delta_{n_k(x)}\right| \operatorname{osc}^2 \left(\varphi_{n_k(x)}, \Delta_{n_K(x)}\right) < \infty \text{ for almost all dyadic-irrational } x.$$

1.3 Result for Systems of Translates and Dilates

The result on convergence almost everywhere of orthorecursive expansions in systems of dyadic translates and dilates can be obtained as a corollary of Theorem 1.1. It can be formulated as follows.

Theorem 1.2 *Let φ be a continuous non-negative non-zero function on* $[0, 1]$ *which satisfies the following condition:*

$$\sum_{k=0}^{\infty} \omega^2_{\varphi}(2^{-k}) < \infty$$

(here $\omega_{\varphi}(\cdot)$ is the modulus of continuity of φ on $[0, 1]$*). Then for every function $f(x) \in L^2[0, 1]$ its orthorecursive expansion in the system of dyadic translates and dilates of φ converges to $f(x)$ in almost all continuity points of f.*

Corollary 1.1 *Let φ satisfy the conditions of Theorem 1.2, and $f \in L^2[0, 1]$ be continuous almost everywhere on* $[0, 1]$*. Then the orthorecursive expansion of f in*

the system of dyadic translates and dilates of φ converges to $f(x)$ almost everywhere on $[0, 1]$.

Let us derive Theorem 1.2 from Theorem 1.1. Without loss of generality we consider the case $\|\varphi\|_2 = 1$ and scaling constants $C_k = 2^{k/2}$. In this case all functions φ_n in the system of dyadic translates and dilates of φ also have a unit L^2-norm. Non-negativeness of φ_n, continuity of φ_n on Δ_n and inclusion $\operatorname{supp}\varphi_n \subset \Delta_n$ immediately follow from the conditions imposed on φ and the definition of a system of dyadic translates and dilates. Thus, it remains to check the condition of Theorem 1.1 on moduli of continuity. Note that for all $n \in \{1, 2, 3, \ldots\}$

$$\omega_n(\delta) = 2^{k(n)/2}\omega_\varphi\left(2^{k(n)}\delta\right),$$

where $\omega_n(\cdot)$ is the modulus of continuity of φ_n on Δ_n, $k(n)$ is the scale of Δ_n (i.e., $k(n) = \lfloor \log_2(n) \rfloor$), and $0 < \delta \le |\Delta_n| = 2^{-k(n)}$. Hence,

$$\sum_{k=0}^{K} 2^{-k}\omega^2_{n_k(x)}\left(2^{-K}\right) = \sum_{k=0}^{K} 2^{-k} \cdot 2^k\omega^2_\varphi\left(2^{k-K}\right) \le \sum_{k=0}^{\infty}\omega^2_\varphi\left(2^{-k}\right) < \infty.$$

The proof of Theorem 1.2 is complete.

Obviously, the condition on the modulus of continuity in Theorem 1.2 holds for Lipshitz functions and, more generally, Hölder functions with positive exponent. Moreover, the condition holds for functions φ with $\omega_\varphi(\delta) = O\left(|\ln\delta|^{-(1/2+\varepsilon)}\right)$ $(\varepsilon > 0)$.

As one can see, the form of the condition on the modulus of continuity in Theorem 1.2 is similar to the form of the condition imposed by A.Yu. Kudryavtsev and A.V. Politov in their result on convergence in L^2-norm [4].

Clearly, Theorem 1.2 can be generalized to systems of translates and dilates generated simultaneously by a set of functions, e.g., systems in which subsystems of different scales are generated by different generating functions. In this case the condition on the modulus of continuity is formulated is terms of the majorant of continuity moduli of generating functions.

1.4 Concluding Remarks

In this publication we showed that a result on pointwise convergence of orthorecursive expansion in a system of translates and dilates can be obtained using a simple technique based on estimation of local oscillations. However, the simplicity of technique led to excessively hard conditions both on a function φ that generates a system of translates and dilates and on a function f being expended. As noted above, in the subsequent publications we will present results with much softer conditions.

Namely, we are going to relax the condition on non-negativeness of φ and continuity almost everywhere of f.

Acknowledgements The authors thank Dr. Alexey Galatenko for valuable comments and discussions. The research was supported by the Russian Federation Government Grant No. 14.W03.31.0031.

References

1. Novikov, I.Ya., Protasov, V.Yu., Skopina, M.A.: Wavelet Theory. American Mathematical Society, Providence (2011)
2. Akansu, A.N., Serdijn, W.A., Selesnick, I.W.: Emerging applications of wavelets: a review. Phys. Commun. **3**(1), 1–18 (2010)
3. Filippov, V.I., Oswald, P.: Representation in L_p by series of translates and dilates of one function. J. Approx. Theory **82**(1), 15–29 (1995)
4. Politov, A.V.: Orthorecursive expansions in Hilbert spaces. Mosc. Univ. Math. Bull. **65**(3), 95–99 (2010)
5. Lukashenko, T.P.: Properties of orthorecursive expansions in nonorthogonal systems. Mosc. Univ. Math. Bull. **56**(1), 5–9 (2001)
6. Temlyakov, V.N.: Weak greedy algorithms. Adv. Comput. Math. **12**(2–3), 213–227 (2000)
7. Temlyakov, V.: Greedy Algorithms. Cambridge University Press, New York (2011)
8. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Process. **41**(12), 3397–3415 (1993)
9. Friedman, J.H., Stuetzle, W.: Projection pursuit regression. J. Am. Stat. Assoc. **76**, 817–823 (1981)
10. Galatenko, V.V., Lukashenko, T.P., Sadovnichiy, V.A.: Convergence almost everywhere of orthorecursive expansions in functional systems. In: Sadovnichiy, V., Zgurovsky, M. (eds.) Advances in Dynamical Systems and Control. Studies in Systems, Decision and Control, vol. 69. Springer, Cham (2016)
11. Kolmogorov, A.N., Fomin S.V.: Introductory Real Analysis. Dover Publications, New York (1975)
12. Galatenko, V.V.: On the orthorecursive expansion with respect to a certain function system with computational errors in the coefficients. Mat. Sb. **195**(7), 935–949 (2004)

Chapter 2
Three-Dimensional Manifolds of Constant Energy and Invariants of Integrable Hamiltonian Systems

Anatoly T. Fomenko and Kirill I. Solodskih

Abstract This paper is algebraic and topology study of the manifolds of constant energy of integrable Hamiltonian systems with two degrees of freedom. The Liouville foliation defines the topology of isoenergy manifold, but on any isoenergy manifold there are many non-equivalent Hamiltonian systems. We give some review of recent papers on homotopy invariants and their relation with Fomenko-Zieschang invariants. Also, we discuss relatively new results about Reidemeister torsion and applications in the theory of Hamiltonian systems. The last section is efficiency demonstration of Fomenko-Zieschang invariants in concrete mechanic system. Let us note that many known Hamiltonian systems have been investigated in terms of Fomenko-Zieschang invariants.

2.1 Integrable Hamiltonian Systems with Two Degrees of Freedom

2.1.1 Hamiltonian Vector Fields

Let $(\mathbf{M}^4, \omega)$ be a symplectic manifold with symplectic structure ω. The smooth function $H : \mathbf{M}^4 \to \mathbb{R}$ induces the *Hamiltonian vector field* sgrad$H$ as follows

$$(\text{sgrad } H)^i = \frac{\partial H}{\partial x^j}\omega^{ij},$$

where (x^1, x^2, x^3, x^4)—the local coordinates on $\mathbf{M}^4$, ω^{ij}—elements of the inverse matrix of the form ω. Also, the form ω induces the Poisson structure on $\mathbf{M}^4$. For

A. T. Fomenko (✉) · K. I. Solodskih
Lomonosov Moscow State University, Moscow, Russian Federation
e-mail: atfomenko@mail.ru

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*,
Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_2

any two smooth functions f, g on $\mathbf{M}^4$ we have

$$\{f, g\} = \omega(\text{sgrad } f, \text{sgrad } g).$$

Let the smooth function $F: \mathbf{M}^4 \to \mathbb{R}$ be the functionally independent with H on $\mathbf{M}^4$ and $\{F, H\} = 0$, and the vector fields sgrad H, sgrad F are complete.

Definition 2.1 The decomposition of the manifold $\mathbf{M}^4$ into the union of connected components of common level surfaces of the integrals F, H is called the Liouville foliation corresponding to the Hamiltonian system sgrad H.

Consider a common regular level $T_c = \left\{x \in \mathbf{M}^4 \middle| H(x) = c_1, F(x) = c_2\right\}$, where $c = (c_1, c_2)$. By Liouville theorem if T_c is connected and compact, then T_c is diffeomorphic to the 2-dimensional torus T^2 (this torus is called the *Liouville torus*).

Definition 2.2 A Liouville torus T^2 is called non-resonant (irrational) if and only if the closure of every integral trajectory lying on it coincide with the whole torus.

2.1.2 Liouville Equivalence of Hamiltonian Integrable Systems

Consider the regular *isoenergy surface* $\mathbf{Q}^3_h = \left\{x \in \mathbf{M}^4 \middle| H(x) = h\right\}$, and corresponding Liouville foliation on $\mathbf{Q}^3_h$. We will assume further that $\mathbf{Q}^3_h$ is compact and regular. Let us denote the restriction of F on $\mathbf{Q}^3_h$ by the same letter F.

Definition 2.3 Two integrable Hamiltonian systems v_1 and v_2 on symplectic manifolds $\mathbf{M}^4_1$ and $\mathbf{M}^4_2$ (resp. on isoenergy surfaces $\mathbf{Q}^3_1$ and $\mathbf{Q}^3_2$) are called Liouville equivalent iff there exists a diffeomorphism $\mathbf{M}^4_1 \to \mathbf{M}^4_2$ (resp. $\mathbf{Q}^3_1 \to \mathbf{Q}^3_2$) transforming the Liouville foliation of the first system to that of the second one.

Remark 2.1 In non-degenerate case of general position the Liouville equivalent integrable systems have the same closures of their integral trajectories for almost all trajectories. In this case almost all Liouville tori represent the closures of integral trajectories of the system.

Assume that F is a *Bott function*, that is non-degenerate (see [1]), on $\mathbf{Q}^3_h$. And let all critical submanifolds of F on $\mathbf{Q}^3_h$ are the circles. We will define the equivalence relation $\sim$ on $\mathbf{Q}^3_h$. We say that two points $x_1, x_2 \in \mathbf{Q}^3_h$ are equivalent if and only if x_1 and x_2 lie on the same connected component of level surface of F. As $\mathbf{Q}^3_h$ is compact, so $F \in [F_{min}, F_{max}]$ on $\mathbf{Q}^3_h$. The quotient space $\mathbf{G} = \mathbf{Q}^3_h/\sim$ is a some graph. If we consider the standard projection $p: \mathbf{Q}^3_h \to \mathbf{G}$, then we see that the one-parametric set of Liouville tori is projected to some edge of the graph G. Then the preimage of critical value of the map F is projected to some vertex of G (see Fig. 2.1). Let $c \in [F_{min}, F_{max}]$ be a critical value of F on $\mathbf{Q}^3_h$.

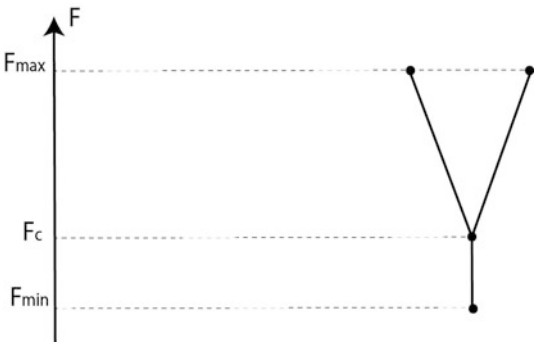


Fig. 2.1 Simple example of quotient space $\mathbf{G} = \mathbf{Q}^3_h / \sim$

Definition 2.4 The class of Liouville equivalence of the preimage $F^{-1}([c - \epsilon, c + \epsilon])$, for sufficiently small ϵ, is called a 3-atom.

The 3-atoms were described by Fomenko in terms of *Seifert manifolds*. To define the Seifert manifold we need to define *standard fibered solid torus* corresponding to a pair of coprime integers (a, b).

Definition 2.5 The standard fibered solid torus of the type (a, b) is the fibration over the circle with the disc-fibers determined by surface bundle of the diffeomorphism of a disk given by rotation by an angle of $\frac{2\pi b}{a}$. Consequently we obtain the foliation of the solid torus over 2-dimensional disc with the fibers, which are homeomorphic to the circle. If $a > 1$ the middle fiber is called exceptional. The pair (a, b) also is called the type of the exceptional fiber.

Definition 2.6 A Seifert manifold is a closed 3-manifold together with a decomposition into a disjoint union of a circles (called fibers) such that each fiber has a tubular neighborhood that forms a standard fibered solid torus.

Theorem 2.1 (A.T. Fomenko) *Any 3-atom V for integrable nondegenerate Hamiltonian system is a Seifert manifold with boundary consisting of tori. Every exceptional fiber of V has type (2, 1).*

Let us mark every vertex of the graph **G** by a symbol of the corresponding 3-atom. This atom describes the corresponding bifurcation of the Liouville tori on this critical level of the integral. This new graph **W** with vertices-atoms is called a *molecule* (rough molecule) of the integrable Hamiltonian system sgrad H on $\mathbf{Q}^3_h$. The graph **W** can be oriented. The orientation of any edge corresponds to integral F increase.

2.1.3 Fomenko-Zieschang Invariants

Consider the edge $(V_1, V_2) \in \mathbf{W}$, where V_1 and V_2 are two atoms. This edge describes the continuous one-parametric family $T^2_t, t \in [t_1, t_2]$ of Liouville tori.

The parameter t on the edge is simply the value of the integral F. The torus $T^2_{t_i}$ is a boundary torus of the atom V_i for $i = 1, 2$. Let us fix the basis $b_i = (\lambda_i, \mu_i)$ in the fundamental group $\pi_1(T^2_{t_i})$, $i = 1, 2$. The basis in $\pi_1(T^2_{t_i})$ determine the basis b'_i in fundamental group $\pi_1(T^2_t)$, where $t \in (t_1, t_2)$. Let us consider the transformation matrix C

$$(\lambda'_2, \mu'_2) = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \lambda'_1 \\ \mu'_1 \end{pmatrix}, \quad C = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

Definition 2.7 The transformation matrix C is called a gluing matrix.

Remark 2.2 The determinant of the matrix C is equal to -1.

To describe the Liouville foliation on $\mathbf{Q}^3_h$ in terms of gluing matrices we have to define the specific set of bases in fundamental groups of boundary tori. Now we will define the *admissible basis* for fundamental group of boundary torus.

2.1.3.1 The Case of Atom A

Atom A is a neighborhood of a stable periodic trajectory S^1 (stable means that F has a local maximum or minimum on S^1). Atom A is homeomorphic to a solid torus. The first basis element λ of $\pi_1(T^2)$ is the homotopy class of the loop, which is presented by a loop contracting to a point on S^1. The second basis element μ can be chosen as arbitrary independent homotopy class of a loop in $\pi_1(T^2)$ (see Fig. 2.2). The orientations of these loops are compatible with orientation of $\mathbf{Q}^3_h$ and with Hamiltonian flow sgrad H.

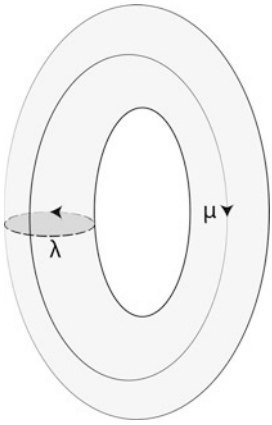


Fig. 2.2 Atom A with admissible basis on boundary torus

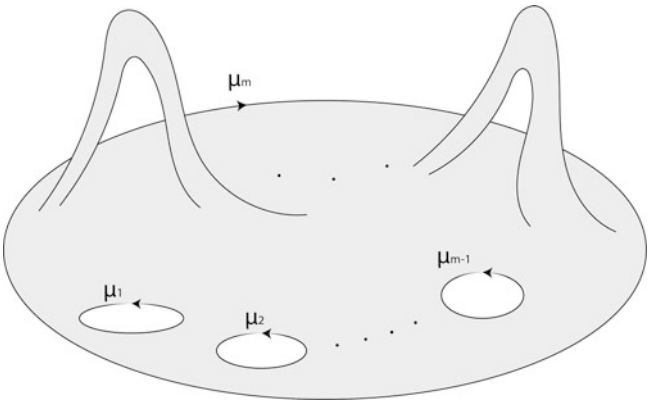


Fig. 2.3 A section of an atom V

2.1.3.2 Case of Saddle Atoms

Definition 2.8 All atoms different from the atom A are called a saddle atoms.

We will consider only saddle atoms without exceptional fibers. In this case saddle atom V of *genius g and complexity m* is homeomorphic to $S^2_{g,m} \times S^1$, where $S^2_{g,m}$ is a 2-sphere with g handles and m holes. The first loop λ_V on the boundary torus T^2_k for $k = 1, \ldots, m$ realizes the homotopy class of $* \times S^1$ in $\pi_1(V)$ under inclusion $T^2 \hookrightarrow V$. This loop is oriented by flow the sgrad H. The second loop μ_k is chosen in a cross-section $S^2_{g,m} \times *$ of the atom V (see Fig. 2.3). Here the star $*$ denotes a point.

2.1.3.3 The Numerical Marks r and ϵ

Let us consider arbitrary edge $(V_1, V_2) \in W$. We fix admissible bases $(\lambda_{V_i}, \mu_{V_i})\, i = 1, 2$ in the fundamental groups of the boundary tori. Then we obtain the gluing matrix C corresponding to the edge (V_1, V_2), namely:

$$(\lambda_{V_2}, \mu_{V_2}) = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \lambda_{V_1} \\ \mu_{V_1} \end{pmatrix}.$$

The numerical marks r and ϵ for the edge (V_1, V_2) are defined as follows

$$r = \frac{\alpha}{\beta} \mod 1, \quad \epsilon = \begin{cases} \operatorname{sign}(\beta), & \text{if } \beta \neq 0; \\ \operatorname{sign}(\alpha), & \text{if } \beta = 0. \end{cases}$$

Lemma 2.1 (See [1]) *The numerical marks r and ϵ are invariant under admissible bases transformations.*

2.1.3.4 Family Mark *n*

Definition 2.9 The edge $(V_1, V_2) \in W$ is called infinite if r-mark on the edge is equal to ∞. Otherwise the edge is called finite.

By deleting all finite edges from the molecule W we obtain some set of subgraphs.

Definition 2.10 Those subgraphs which do not contain vertices A are called the families.

Assume that molecule W has a family W_f. The edges of the molecule W which belong to family W_f or are incident to some vertices of W_f are divided into three classes. We denote this set of edges by $N(W_f)$. Let us consider arbitrary edge $e \in N(W_f)$.

Definition 2.11 The edge e is called inner if $e \in W_f$. By definition both vertices of the edge e belong to the family W_f. The edge e from $N(W_f)$ is called incoming if and only if it's terminal vertex belongs to W_f. If the only initial vertex of e belongs to W_f, then e is called outgoing.

For any edge $e \in N(W_f)$ we define an integer number Θ_e

$$\Theta_e = \begin{cases} \left[\frac{\alpha}{\beta}\right], & \text{if } e \text{ is outgoing;} \\ \left[-\frac{\delta}{\beta}\right], & \text{if } e \text{ is incoming;} \\ \left[-\frac{\gamma}{\alpha}\right], & \text{if } e \text{ is inner.} \end{cases}$$

Let us define n-mark for the family W_f as follows

$$n = \sum_{e \in N(W_f)} \Theta_e.$$

Lemma 2.2 (See [1]) *The mark n is invariant under admissible bases transformations.*

2.1.3.5 Fomenko-Zieschang Theorem and Realization Problem

We have defined the marks r and ϵ for any edge of the molecule W. Also we have defined the marks n for any family of W.

Definition 2.12 The molecule W with the marks r, ϵ and n is called a marked molecule or a Fomenko-Zieschang invariant of integrable system sgrad H on $\mathbf{Q}^3_h$. We denote it as follows

$$W^* = (W, r, \epsilon, n).$$

Let us consider two different integrable Hamiltonian systems sgrad H_1 and sgrad H_2 on $\mathbf{Q}^3_{h_1}$ and $\mathbf{Q}^3_{h_2}$ respectively. Let W_1^* and W_2^* be the marked molecules of these systems.

Theorem 2.2 (A.T. Fomenko, H. Zieschang) *The systems* sgrad H_1 *and* sgrad H_2 *are Liouville equivalent if and only if the marked molecules W_1^* and W_2^* are coincide.*

Let us formulate the important question. Let G be an oriented graph. Let us mark all vertices of G by a symbols corresponding to some 3-atoms. Then we mark all edges by various r-marks and by $\epsilon \in \{1, -1\}$. At last we mark all families by various integer numbers n. Denote the graph G with this marks by G^*. The realization problem: *Is there an integrable Hamiltonian system such that its Fomenko-Zieschang invariant is G^*?*

It turns out, the answer is affirmative.

Theorem 2.3 (A.V. Bolsinov, A.T. Fomenko) *For any marked graph G^* always exists the smooth integrable Hamiltonian system (with Bott integral) with marked molecule G^*.*

2.1.4 Simple Examples of Molecules

Let us consider the molecule A–A. The manifold $\mathbf{Q}^3$ corresponding to this molecule consists of two solid tori which are glued together along the boundary. The following proposition determines a topological type of $\mathbf{Q}^3$ depending on r-mark.

Proposition 2.1

1) *If $r = 0$, then $\mathbf{Q}^3$ is homeomorphic to sphere the S^3.*
2) *If $r = \infty$, then $\mathbf{Q}^3$ is homeomorphic to the direct product $S^1 \times S^2$.*
3) *If $r = \frac{q}{p}$, then $\mathbf{Q}^3$ is homeomorphic to the lens space $L_{p,q}$.*

This proposition demonstrates that Fomenko-Zieschang invariants determine the topological type of $\mathbf{Q}^3$. In case of the concrete mechanical systems the topological type of $\mathbf{Q}^3$ can be often calculated. That is why the study of topology of $\mathbf{Q}^3$ is very important and perspective. Let us denote the class of isoenergy manifolds of integrable Hamiltonian systems by (H). For the details of the theory of Liuoville classification of integrable Hamiltonian systems and its different applications to concrete problems of mechanics, physics and symplectic geometry see the following publications [2–19].

2.2 Homotopy Invariants of Q^3

2.2.1 Fundamental Group $\pi_1(Q^3)$

As Liouville foliation defines the topological type of $\mathbf{Q}^3$, any homotopy and topological invariants can be computed in terms of the marked molecules. The fundamental group $\pi_1(\mathbf{Q}^3)$ has been computed in [20] by A.T. Fomenko and H. Zieschang. In this section we describe the main results of [20]. The idea of the computation is recursive application of Seifert-van Kampen theorem. Let us remind this theorem, its proof can be found in [21].

Theorem 2.4 (Seifert-van Kampen Theorem) *Let X be the path-connected topological space and path-connected open subsets $X_1, X_2 \subset X$ such that $X = X_1 \cup X_2$, $Y = X_1 \cap X_2$ is also path-connected. Then $\pi_1(X, x_0)$ is a free product with amalgamation of the groups $\pi_1(X_1, x_0)$ and $\pi_1(X_2, x_0)$, with respect to induced inclusion homomorphisms i_{1*} and i_{2*}, that is*

$$\pi_1(X, x_0) = \pi_1(X_1, x_0) *_{\pi_1(Y, x_0)} \pi_1(X_2, x_0), \quad i_k \colon Y \hookrightarrow X_k,\ k = 1, 2,$$

where $x_0 \in Y$.

Also, we give some important property of the class (H).

Definition 2.13 The manifold $\mathbf{Q}^3 \in (H)$ is called sufficiently large iff there is a torus $T^2 \in \mathbf{Q}^3$ such that the homomorphism $i_* : \pi_1(T^2) \to \pi_1(\mathbf{Q}^3)$ induced by a standard inclusion is a monomorphism.

Theorem 2.5 (A.T. Fomenko, H. Zieschang) *Any two sufficiently large manifolds $\mathbf{Q}^3_1,\ \mathbf{Q}^3_2 \in (H)$ are homeomorphic iff $\pi_1(\mathbf{Q}^3_1)$ is isomorphic to $\pi_1(\mathbf{Q}^3_2)$.*

Let us consider $\mathbf{Q}^3$ with some Liouville foliation and corresponding molecule W^*. To compute $\pi_1(\mathbf{Q}^3)$ we need to calculate the fundamental groups of all atoms from the molecule W^*. Then we apply the Theorem 2.4. If necessary we move the base point along suitable path.

2.2.1.1 Fundamental Groups of the Atoms

As atom A has homotopy type of the circle S^1, then $\pi_1(A) \cong \mathbb{Z}$. The generator of $\pi_1(A)$ is the image of μ (see Sect. 2.1.3.1) under the inclusion map $i \colon \partial A \hookrightarrow A$. We do not specify base point in A because the $\pi_1(A)$ is commutative.

Let us consider arbitrary saddle atom V (let us note that now we consider only saddle atoms without exceptional fibers). The atom V is homeomorphic to $S^2_{g,m} \times S^1$, then

$$\pi_1(V) \cong \pi_1(S^2_{g,m}) \times \pi_1(S^1).$$

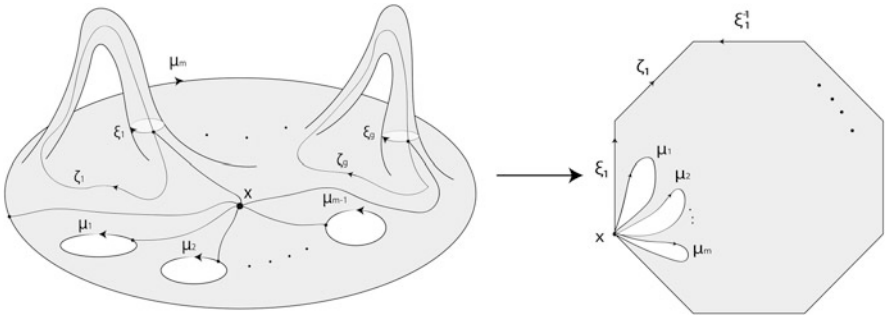


Fig. 2.4 Generators of $\pi_1(S^2_{g,m}, x)$

To calculate the presentation of the group $\pi_1(S^2_{g,m}, x)$ let us fix the following curves:

1. $\xi_1, \zeta_1, \ldots, \xi_g, \zeta_g$—the standard loops on the 2-handles;
2. $\mu_1, \ldots, \mu_m$—the loops corresponding to admissible basis (Sect. 2.1.3.2)
3. $\omega_{\xi_1\zeta_1}, \ldots, \omega_{\xi_g\zeta_g}, \omega_{\mu_1}, \ldots, \omega_{\mu_m}$—the paths from the fix point x to the fixed points on corresponding curves, shown on the Fig. 2.4.

The paths from item (3) determine the tree T on $S^2_{g,m}$. Let us contract the surface $S^2_{g,m}$ (with boundary) along the tree T by contracting all edges to the base point x (see Fig. 2.4). Let us denote its image by Y. Then we construct CW decomposition of Y with single zero-dimensional cell x. We can see that

$$\pi_1(S^2_{g,m}, x) = \left\langle \xi_1, \zeta_1, \ldots, \xi_g, \zeta_g, \mu_1, \ldots, \mu_m \middle| \prod_{k=1}^{g} [\xi_k, \zeta_k]\mu_1 \ldots \mu_m \right\rangle.$$

We do not rename the generators of fundamental group after isomorphism along some path. Finally, we conclude that

$$\begin{aligned} \pi_1(V, x) &= \pi_1(S^2_{g,m} \times S^1, x) \\ &= \left\langle \lambda_V, \xi_1, \zeta_1, \ldots, \xi_g, \zeta_g, \mu_1, \ldots, \mu_m \middle| [\lambda_V, \xi],\ [\lambda_V, \zeta],\right. \\ &\qquad \left. [\lambda_V, \mu], \prod_{k=1}^{g} [\xi_k, \zeta_k]\mu_1 \ldots \mu_m \right\rangle, \quad (2.1) \end{aligned}$$

where $[\lambda_V, \xi]$ are the commutators for all generators $\xi_1, \ldots, \xi_g$ (similarly for all ζ and all μ). The generators of $\pi_1(\mathbf{Q}^3)$ are all generators of fundamental groups of saddle atoms plus some special generators $\{\omega\}$ corresponding to the edges of the molecule. Now we can compute an additional relations which follow from the gluings of two atoms.

2.2.1.2 Gluing of Atoms

At first we glue the atom A with saddle atom V. In this case we obtain only one additional relation, namely $\lambda_V^{\alpha}\mu^{\beta}$, where (α, β) is the first row of the gluing matrix.

In the case when we glue together two saddle atoms V_1 and V_2, we obtain two additional relations:

$$\omega_{(V_1,V_2)}^{-1}\mu_{V_2}\omega_{(V_1,V_2)} = \lambda_{V_1}^{\alpha}\mu_{V_1}^{\beta},$$
$$\omega_{(V_1,V_2)}^{-1}\lambda_{V_2}\omega_{(V_1,V_2)} = \lambda_{V_1}^{\gamma}\mu_{V_1}^{\delta},$$

and the one additional generator $\omega_{(V_1,V_2)}$. Let us consider in molecule W^* arbitrary maximal tree T. If $(V_1, V_2) \in T$, then we add the new relation, namely $\omega_{(V_1,V_2)} = 1$. We see that if W^* is not a tree, then any cycle in W^* corresponds to a non-trivial homotopy class in $\pi_1(\mathbf{Q}^3)$.

2.2.1.3 Example: Seifert Manifolds

In terms of a marked molecules the Seifert manifold is determined by a molecule with single saddle atom V which is homeomorphic to $S^2_{0,m} \times S^1$. We will the molecules of this class as a *simple molecules* (see Fig. 2.5).

The fundamental group $\pi_1(\mathbf{Q}^3)$ of manifold the $\mathbf{Q}^3$ corresponding to a simple molecule, can be easily calculated by the Sects. 2.2.1.1 and 2.2.1.2, described above. We obtain:

$$\pi_1(\mathbf{Q}^3) = \left\langle\, \lambda_V, \mu_1, \ldots \mu_m \;\middle|\; [\lambda_V, \mu_i], \lambda_V^{\alpha_i}\mu_i^{\beta_i}, \mu_1 \ldots \mu_m, \quad i = 1, \ldots, m \,\right\rangle. \tag{2.2}$$

Generally, this group is not commutative and has non-trivial center which is generated by the regular fiber λ_V. We will return to Seifert manifolds later to compute Reidemeister-Franz torsion.

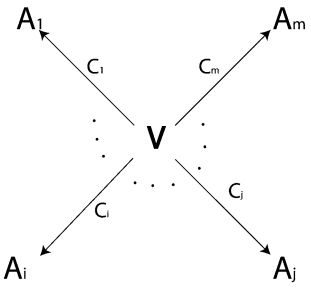


Fig. 2.5 Simple molecule. Here A_i—atom A, C_i—gluing matrix

2.2.2 *Homology Group* $\mathbf{H_1(Q^3, \mathbb{Z})}$

This subsection is devoted to the computation of the first homology group of $\mathbf{Q}^3$ using Mayer-Vietoris technique which has been described by P. Topalov in [22]. It is easy to see that $\mathrm{H}_1(\mathbf{Q}^3, \mathbb{Z})$ depends on the topology of the graph W^* (i.e. molecule corresponding to $\mathbf{Q}^3$) using Mayer-Vietoris exact sequence. Similarly to Sect. 2.2.1, let us present the main ideas of [22]. Let us remind the Mayer-Vietoris theorem. For simplest we formulate this theorem for the case of simplicial homology groups.

Theorem 2.6 (Mayer-Vietoris) *Let X be a simplicial complex and $X_1, X_2 \in X$ are subcomplexes such that $X = X_1 \cup X_2$. Denote $X_1 \cap X_2$ by Y. Then the following sequence is exact*

$$\cdots \to \mathrm{H}_{n+1}(X) \xrightarrow{\partial_*} \mathrm{H}_n(Y) \xrightarrow{(i_{1*},\, i_{2*})} \mathrm{H}_n(X_1) \oplus \mathrm{H}_n(X_2) \xrightarrow{j_{1*} - j_{2*}} \mathrm{H}_n(X) \xrightarrow{\partial_*} \dots,$$

$$i_k \colon Y \hookrightarrow X_k, \;\; j_k \colon X_k \hookrightarrow Y, \quad k = 1, 2,$$

where i_{k*}, j_{k*} are the homomorphisms, induced by corresponding inclusion maps i_k, j_k, and ∂_* is the homomorphism induced by the boundary homomorphism ∂.

2.2.2.1 Computation of $\mathbf{H_1(Q^3, \mathbb{Z})}$

In the case of atom A, the basis in the $\mathrm{H}_1(A, \mathbb{Z})$ is the homology class of the loop μ. In the case of the saddle atom V, the basis in $\mathrm{H}_1(V, \mathbb{Z})$ is the homology classes of the loops $\lambda_V, \mu_1, \dots, \mu_{m-1}, \xi_1, \zeta_1, \dots, \xi_g, \zeta_g$ (see Fig. 2.4). Let us denote the i-homology group of X, namely, $\mathrm{H}_i(X, \mathbb{Z})$ with group the coefficients $\mathbb{Z}$ simply by $\mathrm{H}_i(X)$.

Definition 2.14 The edge e of the marked molecule W^* is called external if e is incident to some atom A. Otherwise, e is called inner edge.

Let us add to each inner edge the new vertex K. This vertex K corresponds to the small tubular neighborhood $T^2 \times I$ of some torus T^2 on edge e. Here I is an interval. Now, we separate $\mathbf{Q}^3$ into the union of two subsets X_1 and X_2, namely:

$$X_1 = \text{The union of all saddle atoms of the molecule } W^*,$$

$X_2 =$ The union of all atoms A of the molecule W^* plus all new vertices K (see above).

Let us denote

$$Y = X_1 \cap X_2.$$

Then we have:

$$\mathrm{H}_0(X_1) \cong \mathbb{Z}^n, \;\; \mathrm{H}_0(X_2) \cong \mathbb{Z}^m, \;\; \mathrm{H}_0(Y) \cong \mathbb{Z}^p,$$

where n is the number of vertices of the graph W^* (taking into account the vertices K), which have degree more than 1. Then m is the number of edges of the graph W^*, taking into account the new edges which have appeared under splitting by the vertices K. Then p is the number of edges W^* plus the number of inner edges of W^*, taking into account the new edges. Using Theorem 2.6 we have:

$$\mathbb{Z} \cong \mathrm{H}_0(\mathbf{Q}^3) \cong (\mathrm{H}_0(X_1) \oplus \mathrm{H}_0(X_2))/\mathrm{Im}(i_*), \quad i_* = (i_{1*}, i_{2*}), \tag{2.3}$$

$$\mathrm{H}_1(\mathbf{Q}^3) \cong \left[(\mathrm{H}_1(X_1) \oplus \mathrm{H}_1(X_2))/\mathrm{Im}(i_*)\right] \oplus \mathrm{Im}(\partial_*). \tag{2.4}$$

Comparing the ranks of the groups listed in (2.3), and using the relation (2.4) we conclude that:

$$\mathrm{Im}(\partial_*) \cong \mathbb{Z}^{b_1(W^*)}, \ \mathrm{H}_1(\mathbf{Q}^3) \cong \left[(\mathrm{H}_1(X_1) \oplus \mathrm{H}_1(X_2))/\mathrm{Im}(i_*)\right] \oplus \mathbb{Z}^{b_1(W^*)},$$

where $b_1(W^*)$ is the first Betti number of the graph W^*. We see that any cycle of $\mathbf{W}^*$ corresponds to some non-trivial element of $\mathrm{H}_1(\mathbf{Q}^3)$. The final part of the computation is the calculation of the homomorphism i_*. It is possible to present the matrix of the homomorphism i_* in terms of the gluing matrix (see [22]).

2.2.2.2 Example: Poincare Sphere

Poincare sphere can be realized as marked molecule which is shown on Fig. 2.6 (see [23]). The homology group $\mathrm{H}_1(\mathbf{Q}^3)$ of the manifold $\mathbf{Q}^3$ (Poincare sphere) corresponding to the molecule on Fig. 2.6 is isomorphic to the cokernel of the following homomorphism

$$h\colon \mathbb{Z}^4 \to \mathbb{Z}^4, \quad M_h = \begin{pmatrix} 2 & 0 & 0 & -1 \\ 0 & 3 & 0 & 1 \\ 0 & 0 & 5 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix},$$

where M_h is the matrix of homomorphism h. It is easy to see that $\det M_h$ is equal to 1, i.e. $\mathrm{H}_1(\mathbf{Q}^3)$ is trivial. But the fundamental group $\pi_1(\mathbf{Q}^3)$ is non-trivial. Let us

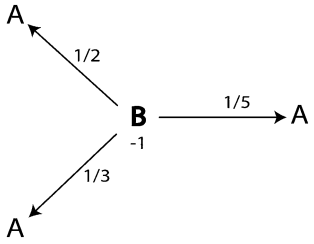


Fig. 2.6 Simple molecule W^* corresponding to Poincare sphere

note that the group $\pi_1(\mathbf{Q}^3)$ has the presentation in the form given by (2.2). Using Tietze transformations we can obtain the following presentation of the same group

$$\pi_1(\mathbf{Q}^3) = \langle\, \mu_2, \mu_3 \mid \mu_3^{-1}\mu_2^{-1}\mu_3^{-1}\mu_2^2, \mu_2^{-1}\mu_3^{-1}\mu_2^{-1}\mu_3^4 \,\rangle\,. \tag{2.5}$$

It is easy to see that group (2.5) is non-trivial. Indeed, assume that group (2.5) is trivial. Then it is obvious that we can add any relations to its presentation. If we add the relation $\mu_3^{-1}\mu_2^{-1} = 1$, then we obtain the group isomorphic to $\mathbb{Z}_7$. We see the contradiction to assumption.

2.3 Reidemeister-Franz Torsion

In this section we will consider an arbitrary field $\mathbb{F}$ and torsion $\tau_h(\mathbf{Q}^3)$ which corresponds to some ring homomorphism $h\colon \mathbb{Z}[\pi_1(\mathbf{Q}^3)] \to \mathbb{F}$. Thereafter, the torsion $\tau_h(\mathbf{Q}^3)$ is the element of $\mathbb{F}^*/\pm h(\pi_1(\mathbf{Q}^3))$. In case when the torsion $\tau_h(\mathbf{Q}^3)$ for the homomorphism h is not defined correctly, then we assume that $\tau_h(\mathbf{Q}^3)$ is equal to 0.

2.3.1 The Torsion of a Simple Molecule

We have considered a simple molecules in Sect. 2.2.1.3. These molecules have only one saddle atom of genus 0. We can compute Reidemeister torsion of a simple molecule for some special class of the ring homomorphisms using torsion of the atoms.

Theorem 2.7 (Solodskikh [24]) *Let ring homomorphism h*

$$h\colon \mathbb{Z}[\pi_1(\mathbf{Q}^3)] \to \mathbb{F}$$

be such that $h(\lambda_V)^{\gamma_k} h(\mu_k)^{\delta_k} \neq 1, k = 1, \ldots, m$. Then the torsion of the manifold $\mathbf{Q}^3$ is not equal to 0 iff $h(\lambda_V) \neq 1$. In case when $h(\lambda_V) \neq 1$, we have:

$$\tau_h(\mathbf{Q}^3) = (h(\lambda_V) - 1)^{m-2} \prod_{k=1}^{m} (h(\lambda_V^{\gamma_k}\mu_k^{\delta_k}) - 1)^{-1} \in \mathbb{F}^*/\pm h(\pi_1(\mathbf{Q}^3)).$$

2.3.1.1 The Torsion of Atoms

The 3-atom A has the same simple homotopy type as well as the circle S^1 (see for example [25]).

Proposition 2.2 *For any ring homomorphism h*

$$h\colon \mathbb{Z}[\pi_1(S^1)] \to \mathbb{F},$$

the torsion $\tau_h(S^1) \neq 0$ *iff* $h(\lambda) \neq 1$, *where* λ *is the generator of* $\pi_1(S^1)$. *If* $h(\lambda) \neq 1$, *then*

$$\tau_h(S^1) = (h(\lambda) - 1)^{-1} \in \mathbb{F}^*/\pm h(\pi_1(S^1)).$$

In the case of saddle atom V of genus 0 and complexity m it is easy to see that atom V has the same simple homotopy type as direct product of the wedge sum of $m-1$ circles and the circle S^1.

Lemma 2.3 ([24]) *Let V be the saddle atom of genus 0 and complexity m. Then*

$$\pi_1(V) = \left\langle \lambda_V, \mu_1, \ldots, \mu_m \;\middle|\; [\lambda_V, \mu_k], \mu_1 \ldots \mu_m, \quad k = 1, \ldots, m \right\rangle.$$

For any ring homomorphism h

$$h\colon \mathbb{Z}[\pi_1(V)] \to \mathbb{F},$$

the torsion $\tau_h(V) \neq 0$ *iff* $h(\lambda_V) \neq 1$. *If* $h(\lambda_V) \neq 1$, *then*

$$\tau_h(V) = (h(\lambda_V) - 1)^{m-2} \in \mathbb{F}^*/\pm h(\pi_1(V)).$$

2.3.2 Corollaries

Using Theorem 2.7 we can establish homeomorphisms between the manifolds corresponding to simple molecules in some special cases. Let us demonstrate some examples.

2.3.2.1 The Case of Zero *r*-Marks

Assume that all r-marks of the simple molecule are equal to 0, and n-mark is not equal to 0, 1, -1. Without loss of generality, the gluing matrices are as follows:

$$C_m = \begin{pmatrix} n & \varepsilon_m \\ \varepsilon_m & 0 \end{pmatrix}, \quad C_i = \begin{pmatrix} 0 & \varepsilon_i \\ \varepsilon_i & 0 \end{pmatrix}, \quad i = 1, \ldots, m-1.$$

As the fundamental group of the manifold $\mathbf{Q}^3$ is cyclic of the order n:

$$\pi_1(\mathbf{Q}^3) = \left\langle \lambda_V \;\middle|\; \lambda_V^n \right\rangle,$$

then the manifold $\mathbf{Q}^3$ is homeomorphic to some lens space $L(n, q)$ (see [26]).

Corollary 2.1 *The manifold* $\mathbf{Q}^3$ *described above is homeomorphic to* $L(n, 1)$.

2.3.2.2 General Case of Lens Spaces

Let us consider a simple molecule such that only two r-marks are not equal to 0. Without loss of generality, the gluing matrices are as follows

$$C_1 = \begin{pmatrix} \alpha_1 & \beta_1 \\ \gamma_1 & \delta_1 \end{pmatrix}, \quad C_2 = \begin{pmatrix} \alpha_2 & \beta_2 \\ \gamma_2 & \delta_2 \end{pmatrix}, \quad C_i = \begin{pmatrix} 0 & \varepsilon_i \\ \varepsilon_i & 0 \end{pmatrix}, \quad i = 3, \ldots, m.$$

The fundamental group of the manifold $\mathbf{Q}^3$ which corresponds to this molecule has the following presentation

$$\pi_1(\mathbf{Q}^3) = \left\langle \lambda_V, \mu_2 \mid [\lambda_V, \mu_2], \lambda_V^{\alpha_1} \mu_2^{-\beta_1}, \lambda_V^{\alpha_2} \mu_2^{\beta_2} \right\rangle.$$

Corollary 2.2 *The manifold* $\mathbf{Q}^3$ *described above is homeomorphic to lens space* $L(p, q)$*, where*

$$p = \alpha_1\beta_2 + \alpha_2\beta_1, \quad q = \alpha_1\gamma_2 + \beta_1\delta_2.$$

2.4 Integrable Geodesic Flows in a Potential Field on the Torus of Revolution

2.4.1 Introduction

In this section we give a short review of the results by D.S. Timonina work (see [27]). This results develop the interesting recent works by Kantonistova (see [28, 29]).

Definition 2.15 A 2-manifold M with a metric g is called a manifold of revolution, if it is invariant under the effective and smooth action of a circle S^1 on M by isometries.

Let us consider the 2-manifold M diffeomorphic to a torus T^2 with the following invariant metric g:

$$ds^2 = d\theta^2 + f^2(\theta)d\phi^2,$$

where θ, ϕ are the standard angular coordinates on T^2, and $f(\theta)$ is a smooth positive function. The differential form $\omega = dp \wedge dq$ defines a symplectic structure on the

cotangent bundle T^*M, where $q = (\theta, \phi)$, $p = (p_\theta, p_\phi)$ are the local coordinates on $T^*_q M$. A geodesic flow with potential $V(q)$ on the torus M is generated by the following Hamiltonian H

$$H = \frac{1}{2} g^{ij}(q) p_i p_j + V(q), \tag{2.6}$$

where g^{ij} is the inverse matrix of the matrix of the metric g.

Let $V(q) = V(\theta)$ then we have the following proposition.

Proposition 2.3 (D.S. Timonina) *The Hamiltonian system with the Hamiltonian in the form (2.6) on the torus T^2 is Liouville integrable for all pairs of functions $(f(\theta), V(\theta))$.*

The smooth function $K = p_\phi$ is an additional independent integral of this system.

2.4.2 Main Results

The main results of the work [27] are formulated in terms of *effective potential*

$$U_h(\theta) = 2f^2(\theta)(h - V(\theta)),$$

on the fixed isoenergy surface $\mathbf{Q}^3_h$.

Lemma 2.4 (D.S. Timonina) *The function K is a Bott function on $\mathbf{Q}^3_h$ iff $U_h(\theta)$ is a Morse function on the circle.*

Lemma 2.5 (D.S. Timonina) *The molecule of Liouville foliation on $\mathbf{Q}^3_h$ consists of the atoms of the following three types only. An atom A, then an atom P_m (see Fig. 2.8) which has two incoming (outgoing) edges and $m > 0$ outgoing (incoming) edges, then atom V_s (see Fig. 2.7), which has one incoming (outgoing) edge and $s > 1$ outgoing (incoming) edges.*

The main result by Timonina is the topological classification of the geodesic flows in a invariant potential field on the torus of revolution in terms of Fomenko-Zieschang invariants (Fig. 2.8).

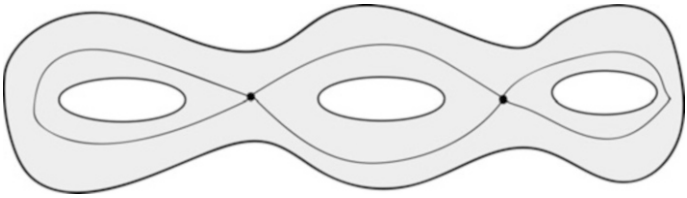


Fig. 2.7 Example: atom V_3

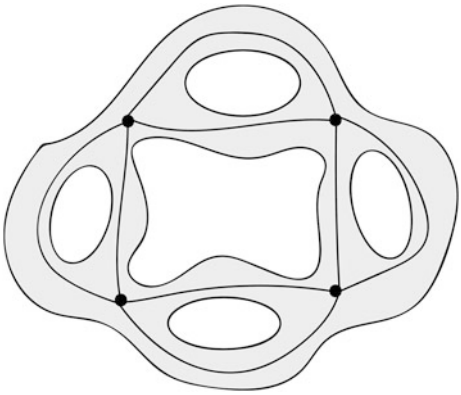


Fig. 2.8 Example: atom P_4

Theorem 2.8 (D.S. Timonina) *If the function $U_h(\theta)$ is positive for all θ then the corresponding molecule has the form W–P_m–P_m–W. The number m is equal to the number of global minima of the function $U_h(\theta)$.*

If the function $U_h(\theta)$ takes negative values then the corresponding molecule of the system on $\mathbf{Q}^3$ takes the form W–W.

Here each graph W is either an atom A or a tree or a forest (the disjoint union of several trees). All the inner vertices of this tree (forest) are the atoms V_s (see Fig. 2.7).

Theorem 2.9 (D.S. Timonina) *The marks on the edges of the type A–$(saddle)$ of the molecule are as follows $r = 0$ and $\epsilon = 1$. The marks on edges of the types: $(saddle$–$saddle)$ or A–A are as follows $r = \infty$ and $\epsilon = \pm 1$. Here $\epsilon = 1$ in case P_m–P_m or in case V_s–V_s, else -1. If the molecule contains some family, then n-mark of this family is equal to 0.*

Recently D.S. Timonina obtain the Liouville classification of the geodesic flows with linear integral in invariant potential field on the two dimensional surfaces of revolution, diffeomorphic to Klein bottle and projective plane.

References

1. Bolsinov, A.V., Fomenko, A.T.: Integrable Hamiltonian Systems. Geometry, Topology, Classification. Chapman & Hall/CRC, Boca Raton (2004). Translated from the 1999 Russian original
2. Fomenko, A.T.: The topology of surfaces of constant energy in integrable Hamiltonian systems, and obstructions to integrability. Math. USSR Izv. **29**, 629–658 (1987)
3. Fomenko, A.T.: Topological invariants of Liouville integrable Hamiltonian systems. Funct. Anal. Appl. **22**(4), 286–296 (1988)
4. Fomenko, A.T., Zieschang, H.: A topological invariant and a criterion for equivalence of integrable Hamiltonian systems with two degrees of freedom. Math. USSR Izv. **36**(3), 567–596 (1991)
5. Bolsinov, A.V., Fomenko, A.T.: Integrable Geodesic Flows on Two-Dimensional Surfaces. Consultants Bureau/Kluwer Academic/Plenum Publishers, New York (2000)

6. Bolsinov, A.V., Fomenko, A.T.: Orbital isomorphism between two classical integrable systems. In: Lie Groups and Lie Algebras. Their Representations, Generalisations and Applications, pp. 359–382. Kluwer Academic Publishers, Dordrecht (1998)
7. Bolsinov, A.V., Matveev, V.S., Fomenko, A.T.: Two-dimensional Riemannian metrics with integrable geodesic flows. Local and global geometry. Sb. Math. **189**(10), 1441–1466 (1998)
8. Bolsinov, A.V., Fomenko, A.T.: On dimension of the space of integrable Hamiltonian systems with two degrees of freedom. In: Dinamicheskie sistemy i smezhnye voprosy. Sbornik statej. K 60-letiyu so dnya rozhdeniya akademika D. V. Anosova, pp. 45–69. Nauka. MAIK Nauka, Moskva (1997)
9. Fomenko, A.T., Kantonistova, E.O.: Topological classification of geodesic flows on revolution 2-surfaces with potential. In: Continuous and Distributed Systems II. Theory and Applications, pp. 11–27. Springer, Cham (2015)
10. Fokicheva, V.V., Fomenko, A.T.: Integrable billiards model important integrable cases of rigid body dynamics. Dokl. Math. **92**(3), 682–684 (2015)
11. Fokicheva, V.V., Fomenko, A.T.: Billiard systems as the models for the rigid body dynamics. In: Advances in Dynamical Systems and Control, pp. 13–33. Springer, Cham (2016)
12. Kudryavtseva, E.A., Nikonov, I.M., Fomenko, A.T.: Maximally symmetric cell decompositions of surfaces and their coverings. Sb. Math. **199**(9), 1263–1353 (2008)
13. Brailov, Y.A., Fomenko, A.T.: Lie groups and integrable Hamiltonian systems. In: Recent Advances in Lie Theory. Selected Contributions to the 1st colloquium on Lie Theory and Applications, Vigo, July 2000, pp. 45–76. Heldermann Verlag, Lemgo (2002)
14. Brailov, A.V., Fomenko, A.T.: The topology of integral submanifolds of completely integrable Hamiltonian systems. Math. USSR Sb. **62**(2), 373–383 (1989)
15. Fomenko, A.T., Konyaev, A.: Algebra and geometry through Hamiltonian systems. In: Continuous and Distributed Systems, pp. 3–21. Springer International Publishing, Cham (2014)
16. Radnović, M., Dragović, V.: Topological invariants for elliptical billiards and geodesics on ellipsoids in the Minkowski space. Fundam. Prikl. Mat. **20**(2), 51–64 (2015)
17. Haghighatdoost, G., Oshemkov, A.A.: The topology of Liouville foliation for the Sokolov integrable case on the Lie algebra so(4). Sb. Math. **200**(6), 899–921 (2009)
18. Fokicheva, V.V.: Classification of billiard motions in domains bounded by confocal parabolas. Sb. Math. **205**(8), 1201–1221 (2014)
19. Konyaev, A.Y.: Classification of Lie algebras with generic orbits of dimension 2 in the coadjoint representation. Sb. Math. **205**(1), 45–62 (2014)
20. Fomenko, A.T., Zieschang, H.: On typical topological properties of integrable Hamiltonian systems. Math. USSR Izv. **32**(2), 385–412 (1989)
21. Hatcher, A.: Algebraic Topology. Cambridge University Press, Cambridge (2002)
22. Topalov, P.: Computation of the fine Fomenko-Zieschang invariant for the main integrable cases of rigid body motion. Sb. Math. **187**(3), 451–468 (1996)
23. Kirby, R.C., Scharlemann, M.G.: Eight faces of the Poincaré homology 3-sphere. Usp. Mat. Nauk **37**(5(227)), 139–159 (1982)
24. Solodskih, K.I.: Graph-manifolds and integrable Hamiltonian systems. Sb. Math. **209**(5), 739–758 (2018)
25. Turaev, V.G.: Reidemeister torsion in knot theory. Russ. Math. Surv. **41**(1), 119–182 (1986)
26. Jankins, M., Neumann, W.D.: Lectures on Seifert Manifolds. Brandeis Lecture Notes. Brandeis University, Waltham (1983)
27. Timonina, D.S.: Topological classification of integrable geodesic flows in a potential field on the torus of revolution. Lobachevskii J. Math. **38**, 1108–1120 (2017, in print)
28. Kantonistova, E.O.: Topological classification of integrable Hamiltonian systems in a potential field on surfaces of revolution. Sb. Math. **207**(3), 358–399 (2016)
29. Kantonistova, E.O.: Liouville classification of integrable Hamiltonian systems on surfaces of revolution. Mosc. Univ. Math. Bull. **70**(5), 220–222 (2015)

Chapter 3
Applying Circulant Matrices Properties to Synchronization Problems

Jose S. Cánovas

Abstract In this chapter, we use circulant matrices to study discrete dynamical systems of higher dimension than one. We show how these matrices are a common framework which is useful to investigate some dynamical properties of some models provided by natural and social sciences. In particular, discrete models from Biology, Economy and Chemistry are considered and analyzed with tools coming from the properties of circulant matrices. More precisely, the special shape of eigenvalues and eigenvectors of circulant matrices is very useful to check whether the dynamics of systems on phase spaces with dimension greater than two can be reduced to that of one dimensional systems.

3.1 Introduction

Discrete dynamics of one dimensional maps has been studied intensively in the past decades and, for smooth enough maps with finite number of extrema, it is well understood. In fact, there are efficient tools to understand the behavior of families of maps, depending on parameters, that commonly appears in models provided by natural and social sciences. We must highlight the existence of finite points that characterize the dynamical behavior of almost every point up to a set of full one dimensional Lebesgue measure.

However, natural and social sciences are providing models depending on several variables, that is, discrete dynamical systems of dimension higher than one, whose dynamics is quite far to be understood. Often, these models have been studied checking the stability of fixed points and simulating some orbits that apparently exhibits a complicated dynamical behavior.

It is also common that the dynamics is analyzed on a one-dimensional invariant set Δ in which the dynamics is known. However, this set Δ has zero two-

J. S. Cánovas (✉)
Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, Cartagena, Spain
e-mail: jose.canovas@upct.es

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_3

dimensional Lebesgue measure and therefore the dynamics is far to be understood on a set of positive Lebesgue measure. The natural question on whether the orbits starting outside Δ converges to an orbit in Δ is quite natural as a first step to understand these systems.

The aim of this paper is to show how circulant matrices can help in solving the above question. Namely, we will show that circulant matrices appears as the Jacobian matrix in several models provided by economy, chemistry and biology, and how the properties of circulant matrices can be used to compute efficiently normal Lyapunov exponents which measures the average distance between orbits outside Δ and this set.

This paper is organized as follows. The next section will be devoted to recall some well-known notions on circulant matrices. Then, we will introduce some basic results on discrete dynamical systems on one and several dimensions. Then, we will introduce the models and analyze them by using both Lyapunov exponents and topological entropy.

3.2 Circulant Matrices: Definitions and Basic Results

Let $\mathbf{A}$ be an $n \times n$ matrix with real coefficients. We denote its rows by $\mathbf{A}_j = (a_{j0}, a_{j1}, \ldots, a_{jn-1})$, $0 \leq j \leq n-1$. This matrix is said to be circulant if $\mathbf{A}_j = \sigma^j(\mathbf{A}_0)$ where σ is the cyclic permutation of length n that carries $(0, 1, .., n-1)$ into $(1, 2, \ldots, n-1, 0)$. In other words,

$$\mathbf{A} = \begin{pmatrix} a_0 & a_1 & a_2 & \ldots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \ldots & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \ldots & a_{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ a_1 & a_2 & a_3 & \ldots & a_0 \end{pmatrix}$$

We denote a circulant matrix by

$$\mathbf{A} = \mathrm{circ}(a_0, a_1, a_2, \ldots, a_{n-1}).$$

Given λ, an eigenvalue of $\mathbf{A}$, we denote its associated eigenspace by V_λ. Circulant matrices have a diagonal form. In particular, it is possible to find all the eigenvalues and eigenvectors. Namely, for any $0 \leq j \leq n-1$, let $\phi_j = e^{i\frac{2\pi j}{n}}$ be a n-th root of unity and put $\mathbf{v}_j = (1, \phi_j, \phi_j^2, \ldots, \phi_j^{n-1})$. Then, applying [22, Th. 3.2.2, p. 72], we obtain easily that $\mathbf{v}_j$ is an eigenvector of $\mathbf{A}$ and

$$\lambda_j = \sum_{k=1}^{n} a_k \phi_j^{k-1}, \;\; j = 1, 2, \ldots, n.$$

is the eigenvalue associated to $\mathbf{v}_j$ for any $0 \leq j \leq n-1$.

The first eigenvalue

$$\lambda_0 = \sum_{k=1}^{n} a_k$$

has an associate eigenvalue $\mathbf{v}_1 = (1, 1, \ldots, 1)$, which generates the eigenspace

$$\Delta = V_{\lambda_0} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1 = x_2 = \ldots = x_n\}.$$

In addition, the orthogonal subspace is

$$\Delta^{\perp} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1 + x_2 + \ldots + x_n = 0\}.$$

Note that the set of vectors

$$\mathscr{B} = \{(1, -1, 0, \ldots, 0), (0, 1, -1, 0, \ldots, 0), \ldots, (0, \ldots 0, 1, -1)\}$$

is a basis of $\Delta^{\perp}$. It is worth to point out that the eigenvalues depend of the matrix coefficients, but the eigenvectors are completely independent of the chosen circulant matrix. This will be important along the chapter.

3.3 Basic Notions on Discrete Dynamical Systems

The models considered in this chapter are given by *difference equations*, which are expressions with the form

$$\begin{cases} x(t+1) = f(x(t)), \\ x(0) = x_0, \end{cases}$$

where $f : X \to X$, is a continuous map on a metric space (X, d) into itself and $x_0 \in X$. The solution of the above difference equation is called *orbit* or *trajectory* of x_0 under f. The pair (X, f) is a *discrete dynamical system*. Then, the orbit of x_0 under f, denoted by $\text{Orb}(x_0, f)$, is given by the sequence $f^t(x_0)$, $t \geq 0$, where $f^t = f \circ f^{t-1}$, $t > 1$, $f^1 = f$, and f^0 is the identity on X.

Although one can study topological properties of dynamical systems, in this chapter we are interested in the case $X = \mathbb{R}^n_{\geq}$, where $\mathbb{R}_{\geq}$ represents the set of non negative real numbers. There is a huge literature on discrete dynamical systems either for the one dimensional case, when $n = 1$ (see e.g. [4, 13] or [23]) or for higher dimensions and even general topological (metric) spaces (see e.g. [5, 24]). Here, we introduce some basic results and notation on dynamical systems on general metric spaces which can be easily carried to real maps.

3.3.1 Periodic Orbits and Topological Dynamics

To understand the dynamics of f, we have to introduce some definitions which have topological roots (see e.g. [13] or [50]). A point $x \in X$ is *periodic* when $f^t(x) = x$ for some $t \geq 1$. The smallest positive integer satisfying this condition is called the *period* of x. Periodic points of period 1 are called *fixed points*. Denote by $\mathrm{F}(f)$, $\mathrm{P}(f)$ and $\mathrm{Per}(f)$ the sets of fixed and periodic points and periods of f, respectively.

Periodic orbits are the simplest orbits that a discrete dynamical system can generate. However, there can exist many other different orbits that may produce a richer dynamics. For $x \in X$, define its *ω-limit set*, $\omega(x, f)$, as the set of limit points of its orbit $\mathrm{Orb}(x, f)$. If $\omega(x, f)$ is finite, then it is a periodic orbit, but often, the dynamical behavior of a single orbit can be very complicated or unpredictable, and usually the word chaos is used to refer to dynamical systems which are able to produce such complicated orbits as we discuss below.

Previously, note that to understand the dynamics it is enough to do it on small subsets of X called *attractors*, which are non-empty compact sets A that attract all the trajectories starting in some neighborhood $\mathscr{U}$ of A, that is, for all $x \in \mathscr{U}$ we have that

$$\lim_{t\to\infty} dist(f^t(x), A) = 0,$$

where $dist(x, A) = \min\{d(x, y) : y \in A\}$. When $\mathscr{U}$ is the whole space X the set A is a *global attractor*. The existence and approximate location of attractors are usually given by the *absorbing sets*, namely, a subset $B \subset X$ is an *absorbing set* if for any bounded set D of X there is $t_0 = t_0(D)$ such that $f^t(D) \subset B$ for all $t \geq t_0$.

There are many definitions of chaos, but we will focus our interest in the following well-known ones. The map f is *chaotic in the sense of Li and Yorke* (*LY-chaotic*) [40] if there is an uncountable set $S \subset X$ (called *scrambled set* of f) such that for any $x, y \in S$, $x \neq y$, we have that

$$\liminf_{t\to\infty} d(f^t(x), f^t(y)) = 0,$$

$$\limsup_{t\to\infty} d(f^t(x), f^t(y)) > 0.$$

Li and Yorke's definition of chaos became famous because of the result *period three implies chaos* which linked periodic orbits and unpredictable dynamical behavior for continuous interval maps. Note that the definition requires the comparison between two orbits or limit points of orbits. Another well-known definition of chaos, inspired by the notion of sensitivity with respect to the initial conditions [30], was given by Devaney [24] as follows. The map f is said to be *chaotic in the sense of Devaney* (*D-chaotic*) if it fulfills the following properties:

- The map f is *transitive*, which in absence of isolated points means that there is a $x \in X$ such that $\omega(x, f) = X$.

- The set of periodic points $\mathrm{P}(f)$ is dense on X.
- It has *sensitive dependence on initial conditions*, that is, there is $\varepsilon > 0$ such that for any $x \in X$ there is an arbitrarily close $y \in X$ and $t \in \mathbb{N}$ such that $d(f^t(x), f^t(y)) > \varepsilon$.[1]

Both, Li–Yorke chaos and sensitivity to initial conditions are in the dynamical systems folklore.

It is interesting to explain what we understand by simple dynamics. In fact, sometimes, the chaotic behavior can be also taken as the opposite of simple (or ordered) behavior. We say that f is *strongly simple* (*ST-simple*) if any ω-limit set is a periodic orbit of f. We say that an orbit $\mathrm{Orb}(x, f)$, $x \in X$, is approximated by periodic orbits if for any $\varepsilon > 0$ there is $y \in \mathrm{P}(f)$ and $t_0 \in \mathbb{N}$ such that $d(f^t(x), f^t(y)) < \varepsilon$ for all $t \geq t_0$. The map f is *LY-simple* [53] if any orbit is approximated by periodic orbits. Finally f is *Lyapunov stable* (*L-simple*) [27] if it has equicontinuous powers.

The above definitions are quite difficult to verify and, specially when we are working with models which may depend on several parameters, we need some practical methods to try to measure the dynamical complexity of the system. One of them is given by *topological entropy*, which was introduced in the setting of continuous maps on compact topological spaces by Adler et al. [1] and by Bowen [16][2] for uniformly continuous maps on metric spaces. It is remarkable that both definitions agree when the set X is metric and compact. It is a conjugacy invariant[3] which is usually taken as a criterion to decide whether the dynamic is complicated or not according to the topological entropy $h(f)$, which will be defined below, is greater than zero or not. Here we introduce the equivalent definitions by Bowen [16] when (X, d) is a compact metric space. Given $\varepsilon > 0$, we say that a set $E \subset X$ is (t, ε, f)-*separated* if for any $x, y \in E$, $x \neq y$, there exists $k \in \{0, 1, \ldots, t-1\}$ such that $d(f^k(x), f^k(y)) > \varepsilon$. Denote by $s(t, \varepsilon, f)$ the biggest cardinality of any maximal (t, ε, f)-separated set in X. Then the topological entropy of f is

$$h(f) = \lim_{\varepsilon \to 0} \limsup_{t \to \infty} \frac{1}{t} \log s(t, \varepsilon, f).$$

The definition does not depend on the metric d, and gives us a nice interpretation of topological entropy (see [4, p. 188]) as follows. Imagine that we have a magnifying glass through which we can distinguish two points if and only if they

[1] It is proved in [9] that the first two conditions in Devaney's definition implies the third one. The definitions are presented in the original form because of the dynamical meaning of sensitive dependence on initial conditions.

[2] Dinaburg [25] gave simultaneously a Bowen like definition for continuous maps on a compact metric space.

[3] Two continuous maps $f : X \to X$ and $g : Y \to Y$ are said to be topologically conjugate if there is an homeomorphism $\varphi : X \to Y$ such that $g \circ \varphi = \varphi \circ f$. In general, conjugate maps share many dynamical properties.

are more than ε-apart. If we know t points of two orbits given by x and y, that is, $(x, f(x), \ldots, f^{t-1}(x))$ and $(y, f(y), \ldots, f^{t-1}(y))$, then we can distinguish between x and y iff $\max_{0\leq i\leq t-1} d(f^i(x), f^i(y)) > \varepsilon$. Hence, $s(t, \varepsilon, f)$ gives us how many points of the space X we can see if we know the pieces of orbits of length t. Then we take the exponential growth rate with t of this quantity, and finally the limit of this as we take better and better magnifying glasses. Then we obtain the topological entropy.

In general, the above chaos definitions are not equivalent and their relations with topological entropy are not homogeneous. For instance, it has been proved that D-chaotic maps are LY-chaotic [32], but the converse is false [53]. On the other hand, positive topological entropy implies LY-chaos [12][4] and the converse is also false [53]. In [7] and [39] it is studied the relationship between topological entropy and D-chaos. ST-simple maps are LY-simple maps but the converse is false [53].

More popular than topological entropy are the so-called Lyapunov exponents (see [45]), which are defined when differentiable structures are considered. Namely, assume that X is a smooth finite dimensional manifold and $f : X \to X$ is a $C^{1+\alpha}$ map. Denote, as usual, by T_xX the tangent space at x and the derivative $d_x f : T_xX \to T_{f(x)}X$. The Lyapunov exponent at $x \in X$ in the direction of $\mathbf{v} \in T_xX \setminus \{\mathbf{0}\}$ is given by

$$\text{lyex}(x, \mathbf{v}) = \lim_{t\to\infty} \frac{1}{t} \log ||d_x f^t(\mathbf{v})||$$

if this limit exists. An invariant measure μ is a probability measure on the Borel sets of X such that $\mu(f^{-1}(A)) = \mu(A)$ for any Borel set $A \subseteq X$. This invariant measure μ is ergodic if the equality $f^{-1}(A) = A$ implies that $\mu(A)$ is either 0 or 1. The multiplicative ergodic Theorem states that the above limit exists for μ-almost all point in X. We use Lyapunov exponents in particular cases where the existence of chaos is linked to the property of having positive Lyapunov exponents.

Next, we study the particular case of real maps, starting by the one dimensional case. We will see how the above results are sharpened in this setting. In addition, we will give some notions on the dynamics of real maps on phase spaces with dimension higher than one.

3.3.2 *Dynamics of Continuous Interval Maps*

In general, for one dimensional maps, the relevant results are given when $X = [a, b] \subset \mathbb{R}$ is a compact interval, which by conjugacy can be taken to be $[0, 1]$. In

[4]See also [54] which almost simultaneously states the same result for C^2 diffeomorphisms on compact manifolds of dimension greater than one.

this setting, Sharkovsky's Theorem is a remarkable result which helps to distinguish between simple and complicated dynamics. Recall Sharkovsky's order of natural numbers

$$3 >_s 5 >_s 7 >_s \ldots >_s 2 \cdot 3 >_s 2 \cdot 5 >_s \ldots >_s 2^2 \cdot 3 >_s 2^2 \cdot 5 >_s \ldots$$

$$\ldots >_s 2^k \cdot 3 >_s 2^k \cdot 5 >_s \ldots >_s 2^3 >_s 2^2 >_s 2 >_s 1.$$

Applying Sharkovsky's Theorem (see [50] or [4]. Also [26] for an "easy" proof) one can see that for any continuous map $f : \mathbb{R} \to \mathbb{R}$ with one periodic point it is held that either $\mathrm{Per}(f) = \mathrm{S}(m) = \{k : m >_s k\} \cup \{m\}$, with $m \in \mathbb{N}$, or $\mathrm{Per}(f) = \mathrm{S}(2^\infty) = \{2^n : n \in \mathbb{N} \cup \{0\}\}$. A map is of *type* $m \in \mathbb{N} \cup \{2^\infty\}$ if $\mathrm{Per}(f) = \mathrm{S}(m)$. A map f is called *S-chaotic* if $\mathrm{Per}(f) = \mathrm{S}(m)$, $m = 2^r q$, $r \geq 0$ and $q > 1$ odd.

On the other hand, for one dimensional dynamics the topological entropy is an useful tool to check the dynamical complexity of a map because it is strongly connected with the notion of *horseshoe* (see [4, p. 205]). We say that the map $f : [0, 1] \to [0, 1]$ has a k-horseshoe, $k \in \mathbb{N}$, $k \geq 2$, if there are k disjoint subintervals J_i, $i = 1, \ldots, k$, such that $J_1 \cup \ldots \cup J_k \subseteq f(J_i)$, $i = 1, \ldots, k$.[5]

The following result shows some equivalences among the above definitions of chaos and order (see [13, 50, 53]). Note that the situation is simpler than in the general case.

Theorem 3.1 *Let $f : [0, 1] \to [0, 1]$ be a continuous map. Then*

(a) The map f has positive topological entropy if and only if the map f is S-chaotic.
(b) If f is D-chaotic, then $h(f) > 0$.
(c) If f is either ST-simple or L-simple, then $h(f) = 0$.
(d) If $h(f) > 0$, then f is LY-chaotic, but the converse is false in general. If f is LY-simple, then $h(f) = 0$. The union of LY-chaotic and LY-simple continuous maps is the set of continuous interval maps.

We remark the topological nature of the above result. If we consider another points of view, we can obtain more information giving rise to apparently strange paradoxes. For instance, there exist maps with positive entropy, and therefore chaotic in some sense, such that the orbit of almost all points in [0, 1] (with respect to the Lebesgue measure) converges to a periodic orbit.

Although we will come back to this point later, let us show how to get such example. Consider f a C^3 *unimodal* map such that $f(0) = f(1) = 0$. Recall that a map f is said to be unimodal if there is $c \in [0, 1]$, called *turning point* such that $f|_{[0,c)}$ is strictly increasing and $f|_{[c,1]}$ is strictly decreasing. The *Schwarzian*

[5] Since Smale's work (see [52]), horseshoes have been in the core of chaotic dynamics, describing what we could call random deterministic systems.

derivative (see [51] or [55]) is then given by

$$S(f)(x) = \frac{f'''(x)}{f'(x)} - \frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2,$$

at those points whose first derivative does not vanish. Assume that $S(f)(x) < 0$ and that there is a locally attracting periodic orbit, that is, a periodic orbit $P = \{x_1, \ldots, x_p\}$ for which there exists a neighborhood V of P such that for any $x \in V$ the distance $d(f^t(x), P) = \min_{1 \le i \le p} d(f^t(x), x_i)$ tends to zero as t tends to infinity. The logistic map $f(x) = 3.83\,x(1-x)$ is a good example of such behavior; almost all trajectory converges to a periodic orbit of period 3, while the topological entropy is positive (see e.g. [15]). This example, and many others in the literature, shows that it is important to study the dynamics from several points of view.

3.3.3 Piecewise Monotone Maps: Entropy and Attractors

Usually, one dimensional difference equations models in science are given by piecewise monotone maps. A continuous interval map is *piecewise monotone* if there is a finite partition of $[0, 1]$, $0 = x_0 < x_1 < \ldots < x_k = 1$, such that $f|_{[x_i, x_{i+1}]}$ is monotone for $0 \le i < k$. Note that a piecewise monotone map may have constant pieces. The extreme points of f, which can be isolated or contained in a subinterval of extreme points, will be called *turning points* (turning intervals if the extreme points form a subinterval). For a piecewise monotone map f, let $c(f)$ denote the number of pieces of monotonicity of f. If g is another piecewise monotone map, it is easy to see that $c(f \circ g) \le c(f)c(g)$. Hence, the sequence $c(f^t)$ gives the number of monotonicity pieces of f^t and the following result due to Misiurewicz and Szlenk (see [44]), shows that for piecewise monotone maps topological entropy can be easily understood.

Theorem 3.2 *Let $f : [0, 1] \to [0, 1]$ be a continuous and piecewise monotone map. Then*

$$h(f) = \lim_{t \to \infty} \frac{1}{t} \log c(f^t).$$

Note that $c(f^t) \le c(f)^t$, and so $h(f) \le \log c(f)$. Hence, a consequence of Misiurewicz–Szlenk Theorem is that homeomorphisms on the interval have zero topological entropy. On the other hand, following Theorem 3.2, we can easily see that the logistic map $f(x) = 4x(1 - x)$ and the tent map $g(x) = 1 - |2x - 1|$ have topological entropy $\log 2$, since $c(f^t) = c(g^t) = 2^t$ for all $t \in \mathbb{N}$. However, computing topological entropy can be a very complicated task, but we will see how to make these computations for a suitable class of maps.

The dynamics of smooth enough piecewise monotone maps are well-known in the following sense. Following [43], a metric attractor is a subset $A \subset [0, 1]$ such that $f(A) \subseteq A$, $O(A) = \{x : \omega(x, f) \subset A\}$ has positive Lebesgue measure, and there is no proper subset $A' \subsetneqq A$ with the same properties. The set $O(A)$ is called the *basin* of the attractor.

By van Strien and Vargas [56], the regularity properties of f imply that there are three possibilities for its metric attractors for a class of piecewise monotone maps, called *multimodal maps*, fulfilling the following assumptions. There are $c_1 < c_2 < \ldots < c_k$, creating a partition on $[0, 1]$, such that f is strictly monotone on each element of the partition. f is C^3 and f is non flat on the turning points $c_1, \ldots, c_k$, that is, for x close to c_i, $i = 1, 2, \ldots, k$,

$$f(x) = \pm|\phi_i(x)|^{\beta_i} + f(c_i),$$

where ϕ_i is C^3, $\phi_i(c_i) = 0$ and $\beta_i > 0$. Then, the metric attractors of such multimodal maps can be of one of the following types:

(A1) A periodic orbit.
(A2) A solenoidal attractor, which is basically a Cantor set in which the dynamic is quasi periodic. More precisely, the dynamic on the attractor is conjugated to a minimal translation, in which each orbit is dense on the attractor. The dynamic of f restricted to the attractor is simple, neither positive topological entropy nor Li–Yorke chaos can be obtained. Its dynamic is often known as quasi-periodic.
(A3) A union of periodic intervals $J_1, \ldots, J_k$, such that $f^k(J_i) = J_i$ and $f^k(J_i) = J_j$, $1 \leq i < j \leq k$, and such that f^k is topologically mixing. Topologically mixing property implies the existence of dense orbits on each periodic interval (under the iteration of f^k).

Moreover, if f has an attractor of type (A2) and (A3), then they must contain the orbit of a turning point, and therefore its number is bounded by the turning points. In addition, if $Sf(x) < 0$, then the total number of attractors is bounded by k. From a practical point of view, in a computer simulation we are able to show the existence of attractors of type (A1) and (A3), and only attractors of type (A3) are able to exhibit unpredictable dynamics. As a conclusion of this, if all the turning points of f are attracted by periodic orbits, then the map f will not exhibit physically observable chaos, although it can be topologically chaotic.

The Lyapunov exponents on the image of the turning points can be computed by

$$\text{lyex}(c_i) = \lim_{t\to\infty} \frac{1}{t} \log |(f^t)'(f(c_i))| = \lim_{t\to\infty} \frac{1}{t} \sum_{j=1}^{t} \log |f'(f^j(c_i))|,$$

for $i = 1, 2, \ldots, k$, and all of them are negative when the map f is free of attractors of type (A3). So, positive Lyapunov exponents imply the existence of observable chaos.

3.3.4 Computing Topological Entropy

The above definition of topological entropy is not useful in practice, and counting monotone pieces of an iterated map f^t is not easy. In addition, an exact computation of topological entropy for continuous interval maps cannot be done in general, but there are several papers devoted to compute it approximately for unimodal maps (see [15]) bimodal maps, that is, with three monotone pieces (see [14]) and four monotone pieces (see [18]). In general, it is possible to make computations for arbitrarily large monotone pieces whenever the number of different kneading sequences of turning points is smaller than 4 (see [18, 19]).

Now, we introduce the unimodal case where the topological entropy can be computed by using kneading sequences as follows. Let f be an unimodal map with maximum (turning point) at c. Let $k(f) = (k_1, k_2, k_3, \ldots)$ be its kneading sequence given by the rule

$$k_i = \begin{cases} R & \text{if } f^i(c) > c, \\ C & \text{if } f^i(c) = c, \\ L & \text{if } f^i(c) < c. \end{cases}$$

We fix that $L < C < R$. For two different unimodal maps f_1 and f_2, we fix their kneading sequences $k(f_1) = (k_n^1)$ and $k(f_2) = (k_n^2)$. We say that $k(f_1) \leq k(f_2)$ provided there is $m \in \mathbb{N}$ such that $k_i^1 = k_i^2$ for $i < m$ and either an even number of $k_i^{1\prime}s$ are equal to R and $k_m^1 < k_m^2$ or an odd number of $k_i^{1\prime}s$ are equal to R and $k_m^2 < k_m^1$. Then it is proved in [15] that if $k(f_1) \leq k(f_2)$, then $h(f_1) \leq h(f_2)$. In addition, if $k_m(f)$ denotes the first m symbols of $k(f)$, then if $k_m(f_1) < k_m(f_2)$, then $h(f_1) \leq h(f_2)$.

The algorithm for computing the topological entropy is based on the fact that the tent family

$$g_k(x) = \begin{cases} kx & if\ x \in [0, 1/2], \\ -kx + k & if\ x \in [1/2, 1], \end{cases}$$

with $k \in [1, 2]$, holds that $h(g_k) = \log k$. The idea of the algorithm is to bound the topological entropy of an unimodal maps between the topological entropies of two tent maps. The algorithm is divided in four steps:

Step 1. Fix $\varepsilon > 0$ (fixed accuracy) and an integer n such that $\delta = 1/n < \varepsilon$.
Step 2. Find the least positive integer m such that $k_m(g_{1+i\delta})$, $0 \leq i \leq n$, are distinct kneading sequences.
Step 3. Compute $k_m(f)$ for a fixed unimodal map f.
Step 4. Find r the largest integer such that $k_m(g_{1+r\delta}) < k_m(f)$. Hence $\log(1 + r\delta) \leq h(f) \leq \log(1 + (r + 2)\delta)$.

The algorithm is easily programmed. We usually use Mathematica, which has the advantage of computing the kneading invariants of tent maps without round off errors, improving in practice the accuracy of the method.

3.3.5 Dynamics in Higher Dimension

Things are more complicated when $n > 1$, and discrete dynamical systems are far to be understood in an analytic way. The common agreement among researchers is that in general one dimensional results cannot be extended to general higher dimension dynamical systems. As a keynote example, one can easily check that Sharkovsky's Theorem does not hold for two dimensional maps: rational rotations on the plane are a good example of that. Although there are some results on limit sets (see [2] or [3]) and good results for some types of two dimensional maps like triangular or skew product ones (see [35, 36] or [37]) and antitriangular ones (see [17] or [8]), the dynamics on two dimensional maps is still quite unexplored and usually papers dealing with models constructed on higher dimensional spaces have to show numerical experiments and simulations.

In this paper we are going to analyze the models trying to reduce the dimension. This can be done if the system has a global attractor with dimension smaller than that of the whole space. Another way is to work with models that have some symmetry properties. The problem can be stated as follows. Assume that $f : X \to X$ is a $C^{1+\alpha}$ map on a manifold X with dimension n and there is a submanifold $Y \subset X$ with dimension m such that $f(Y) \subseteq Y$, due to symmetric properties. Hence, any orbit starting with an initial condition $x \in Y$ will remain in Y along the trajectory. Since $m < n$, it is possible that the dynamics of f on Y can be understood, and from this knowledge, we can derive some properties on the dynamics of f on the whole space X.

For instance, assume that $f|_Y$ is chaotic in the sense of Li and Yorke, that is, there exists an uncountable scrambled subset $S \subset Y \subset X$. It is clear that f itself is also chaotic in the sense of Li and Yorke. The same happens if $f|_Y$ has positive topological entropy, but it is not true in general if $f|_Y$ is Devaney chaotic. Moreover, it may happen that $f|_Y$ is Li–Yorke chaotic, but for any neighborhood N of Y one has that the trajectory of any $x \in N \setminus Y$ converges to a periodic orbit, which will make unobserved the existence of Li–Yorke chaos. So, we are interested in analyzing not only the dynamics of $f|_Y$ but also when the trajectories outside Y may converge or synchronize with trajectories inside Y.

This can be easily done if the attractor inside Y is a periodic orbit, because at least locally, Jacobian matrices along the periodic orbit give you the key: the spectral radius of the product of such Jacobian matrices has modulus smaller than one.[6]

[6] A periodic orbit can be an attractor when spectral radius has modulus one, but in general the converse is not true.

The problem arises when the attractor is chaotic. This paper mainly considers one-dimensional chaotic attractors of piecewise monotone maps.

Following [6, §2.1], for a given $v \in \mathbb{R}^n$, we define the tangential Lyapunov exponent at $x \in Y$ in the direction of v to be

$$lyex(f, x, v) = \lim_{n\to\infty} \frac{1}{n} \log \|\Pi_{T_{f^t(x)}Y} \circ df_x^n \circ \Pi_{T_xY}(v)\|,$$

where Π_V means the projection of a vector of $\mathbb{R}^n$ in the subspace V and df_x^t denotes the differential of f^t at x.

On the other hand, again following [6, §2.1], we define the normal Lyapunov exponent at $x \in Y$ in the direction of v to be

$$tyex(f, x, v) = \lim_{n\to\infty} \frac{1}{n} \log \|\Pi_{T_{f^t(x)}Y^\perp} \circ df_x^n \circ \Pi_{T_xY^\perp}(v)\|.$$

In the spirit of [6], our simulations will show that the system locally *synchronizes* to the set Y, that is, the system is locally attracted by the attractor of $f|_Y$, whenever the normal Lyapunov exponent is negative. If in addition the Lyapunov exponent (tangential) is positive (and hence topological entropy is also positive), then the system produces a chaotic synchronization.

3.4 Application to Oligopoly Dynamics

In oligopoly models, a number of firms compete in a market in such a way that the interaction between them plays a crucial role in the market evolution. The rules are usually given by economic assumptions on e.g. demand functions, cost functions or decisions on future productions. For a wide range of different scenarios the reader can see [11]. The basic idea is that, if we have n firms and Π_i is the profit for the ith-firm, which will be assumed to be smooth enough, the optimization of profits can be the key for describing ways of how firms organize their future productions. At the end of the process, we have a system of difference equations

$$x_i(t+1) = f_i(x_1(t), \ldots, x_n(t)), \ i = 1, \ldots, n,$$

where x_i is the variable that firm i wants to control, basically quantities or price, and f_i are called reaction functions, which give you the evolution of variables x_i with time.

Of course, the reaction functions need not be linear maps and then, the analysis of the dynamics of the systems, that is, the evolution with time of all the possible initial states is quite hard to be analyzed in general. However, there are several ideas that can be used to give some partial results on the dynamics, as it is shown below.

To fix ideas, we will always assume that q_i are the quantities of goods produced by each firm.[7] The reaction functions can have different forms according to the way that firms will organize their future productions. To obtain them, we have to maximize the concave profit function solving the equation

$$\frac{\partial \Pi_i}{\partial q_i}(q_1, \ldots, q_n) = 0,$$

to obtain the reaction function for firm i, which will have the general form

$$f_i(q_1, \ldots, q_n) = g_{\mathbf{C}_i}(Q - q_i),$$

where $Q = \sum_{i=1}^{n} q_i$ is the total market supply, and $g_{\mathbf{C}_i} : \mathbb{R}^+ \to \mathbb{R}^+$, $\mathbb{R}^+ = [0, \infty)$, is a one-dimensional map depending on parameter $\mathbf{C}_i \in \mathbb{R}^m$. In addition, the closure of the support of the map $g_{\mathbf{C}_i}$ is given by

$$\mathrm{Cl}\{q \in \mathbb{R}^+ : g_{\mathbf{C}_i}(q) > 0\} = [0, q_{\mathbf{C}_i}],$$

and therefore the map $g_{\mathbf{C}_i}$ has an absolute maximum value q_{M_i}.

Perhaps the simplest case of planning productions is the naive expectations on future productions in which firms produces at time $t + 1$ the maximum obtained for time t. There are more sophisticated ways of generating the functions f_i, $i = 1, 2, \ldots, n$. For instance, under adaptive expectations we may assume that

$$f_i(q_1, \ldots, q_n) = (1 - \lambda_i)q_i + \lambda_i g_{\mathbf{C}_i}(Q - q_i),$$

where $\lambda_i \in [0, 1]$. Notice that $\lambda_i = 1$ gives us the case of naive expectations.[8]

Here, we assume that we are working with maps $f : \mathbb{R}^+ \to \mathbb{R}^+$ smooth enough and such that the following hypothesis are fulfilled:

h0. There is a map $g : \mathbb{R}^+ \to \mathbb{R}^+$ such that there is $q_M \in (0, q_0)$ such that $g|_{(0,q_M)}$ is strictly increasing, $g|_{(q_M,q_0)}$ is strictly decreasing and $g(q_0) = 0$. The maximum q_M is called the turning point of g.

h1. $g^{-1}(0) = \{0, q_0\}$, $q_0 \in \mathbb{R}^+$.

[7] From now on, we denote the production with the letter "q" instead of "x" because this is the usual notation for that.

[8] Another alternatives which do not imply a optimization process can be see e.g. [10] as for instance

$$f_i(q_1, \ldots, q_n) = q_i + \lambda_i \Pi_i(q_1, \ldots, q_n),$$

or

$$f_i(q_1, \ldots, q_n) = q_i + \lambda_i \frac{\partial \Pi_i}{\partial q_i}(q_1, \ldots, q_n).$$

h2. $g(q) > 0$ for $q \in (0, q_0)$.
h3. $f(q) = \max\{0, g(q)\}$.

If $Q(t)$ is the total output at time t and $Q_i(t) = Q(t) - q_i(t)$ for $i = 1, 2, \ldots, n$, the oligopoly is defined under naive expectations by

$$q_i(t+1) = f(Q_i(t)),\ i = 1, \ldots, n,$$

and under adaptive expectations by

$$q_i(t+1) = \max\{0, (1-\lambda)q_i(t) + \lambda g(Q_i(t))\},\ i = 1, \ldots, n.$$

In [20], Puu's oligopoly [46] and Kopel's oligopoly [38] were analyzed. In this work, we consider the Puu–Norin's oligopoly (see [47]) which is a modification of that from [46]. Before writing the reaction functions, we give a general framework for analyzing these models. If all the firms are homogeneous, the space

$$\Delta = \{(q, q, \ldots, q) \in \mathbb{R}^n : q \geq 0\}$$

is invariant by the model. On Δ the model reads as

$$q(t+1) = f((n-1)q(t)) = \max\{0, g((n-1)q(t))\}$$

under naive expectations and

$$q(t+1) = \max\{0, (1-\lambda)q(t) + \lambda g((n-1)q(t))\}$$

for adaptive expectations. Let $q_0 > 0$ be such that $g(q_0) = 0$. Then, under naive expectations we get positive productions when $q \in (0, \frac{q_0}{n-1})$. Since g is unimodal, let q_M be the turning point of g, and note that $g(q_M)$ is the maximum output given by the system. When $g(q_M) \geq \frac{q_0}{n-1}$, we have that g has a 2-horseshoe, and therefore we prove that the dynamics on Δ is topologically chaotic because its topological entropy is equal to $\log 2$. In addition, if $g(q_M) > \frac{q_0}{n-1}$, then there is an interval J containing the turning point such that $f(J) = \{0\}$ and numerical simulations will show that all the orbits go eventually to zero. One could expect that the set $\cup_{n\geq 0} f^{-n}(J)$ have full one dimensional Lebesgue measure on $(0, \frac{q_0}{n-1})$ and therefore the chaotic dynamics lies in a set of zero one dimensional Lebesgue measure. Moreover, when we take initial conditions outside Δ, numerical simulations show that all the orbits go eventually to zero.

When adaptive expectations are assumed, the linear part $(1-\lambda)q$ goes to infinite as q tends to infinite, while the nonlinear part $\lambda g((n-1)q)$ tends to minus infinite, and so the existence of q_0 for the map $(1-\lambda)q + \lambda g((n-1)q)$ is not guaranteed. If such number q_0 exists, that is, there is $q_0 = q_0(n)$ such that $(1-\lambda)q_0 + \lambda g((n-1)q_0) = 0$, then there is $q_M \in (0, q_0)$ such that $(1-\lambda)q + \lambda g((n-1)q)$ attains its

maximum value at q_M, and the above reasoning made for naive expectations makes sense in the adaptive expectations case.

On Δ, the Jacobian matrix is

$$\mathbf{J} = \text{circ}\left(1-\lambda, \lambda g'((n-1)q), \lambda g'((n-1)q), \ldots, \lambda g'((n-1)q)\right)$$

under adaptive expectations. Naive expectations are obtained by putting $\lambda = 1$. Then $(1 - \lambda + \lambda(n-1)g'((n-1)q))$ is an eigenvalue of $\mathbf{J}$ associated to the eigenvector $(1, 1, \ldots, 1)$. On the other hand, the subspace $\Delta^{\perp}$ is generated by

$$\mathscr{B} = \{(1,-1,0,0,\ldots,0),(0,1,-1,0,\ldots,0),\ldots,(0,0,\ldots 0,1,-1)\}$$

and each vector of $\mathscr{B}$ is a eigenvector of the eigenvalue $(1 - \lambda - \lambda g'((n-1)q))$. Hence, it is easy to see that, on Δ, the tangential Lyapunov exponent is given by

$$lyex(q) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \log|1 - \lambda + \lambda(n-1)g'((n-1)q(i))|$$

while the normal Lyapunov exponent is

$$tyex(q) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \log|1 - \lambda - \lambda g'((n-1)q(i))|$$

where $q(i)$ ranges the orbit of $(q, q, \ldots, q)$ for some $q \geq 0$. Fixing an invariant measure μ on Δ, note that $lyex(q)$ and $tyex(q)$ are well defined for almost all q related to μ. In addition, $lyex(q)$ will provide us information on the dynamics on Δ, while $tyex(q)$ informs us on when initial conditions outside Δ converge to attractors on Δ, that is, when firms locally synchronize (see [6] for more information). The fact that firms do synchronize is very important because it is a commonly accepted fact, although we will show in our example that sometimes they fail to synchronize.

3.4.1 Puu–Norin's Oligopoly

The reaction function of Puu–Norin's oligopoly [47] is given by

$$g_u(q) = \frac{1}{2}\sqrt{4uq + 5q^2} - \frac{3}{2}q$$

which implies that the system evolves under the difference equations

$$q_i(t+1) = (1-\lambda)q_i(t) + \lambda g_u(Q_i(t)),\ i = 1, \ldots, n.$$

Note that for $\lambda = 1$ we have naive expectations model and adaptative expectations otherwise. The parameter $u > 0$ is related to the cost functions. To make more significative the optimization process in the model, we will consider that λ ranges the interval $[0.5, 1]$. In addition, the map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\varphi(q_1, \ldots, q_n) = u(q_1, \ldots, q_n), \ (q_1, \ldots, q_n) \in \mathbb{R}^n$$

is a conjugacy that allows us to normalize $u = 1$, and then we make $g_1 = g$ and the model reads as

$$q_i(t+1) = (1-\lambda)q_i(t) + \lambda g(Q_i(t)), \ i = 1, \ldots, n,$$

and therefore, the model restricted to the invariant set Δ is given by the difference equation

$$q(t+1) = (1-\lambda)q(t) + \lambda g((n-1)q(t)).$$

For naive expectations it is easy to see that $q_0 = 1$ and $q_M = 1/5$. Then, the inequality

$$g(q_M) \geq \frac{q_0}{n-1}$$

gives us the condition $n \geq 6$. The equality is obtained for $n = 6$. So, the system becomes chaotic when the number of firms is greater or equal than 6, but such chaotic behavior cannot be observed when it is strictly greater than 6. For adaptative expectations, the existence of

$$q_0 = \frac{\lambda^2(n-1)}{1 + \lambda(\lambda(n-1)(n+2) - 3n - 5)}$$

is not guaranteed (notice that q_0 can be negative for some values of λ and n, which is not allowed since outputs must be non-negative) and therefore the model can be given by a strictly increasing map, providing a non chaotic model, or by unimodal maps, which are able to produce chaotic phenomena.

We concentrate our attention on the dynamics on Δ and when the dynamics on it attracts the dynamics outside this set. Figures 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, and 3.7 show the computation of topological entropy with accuracy 10^{-6} and estimations of Lyapunov exponents when the number of firms increases. The topological entropy can be positive when the number of firms is greater or equal than 6. For $n \geq 8$, there is a parameter value λ_0 such that if $\lambda > \lambda_0$, then the topological entropy is equal to $\log 2$ but the chaotic behavior cannot be observed in numerical simulations, as in the bifurcation diagrams. For $n = 5$ the fixed point can be destabilized to get a two periodic point attractor.

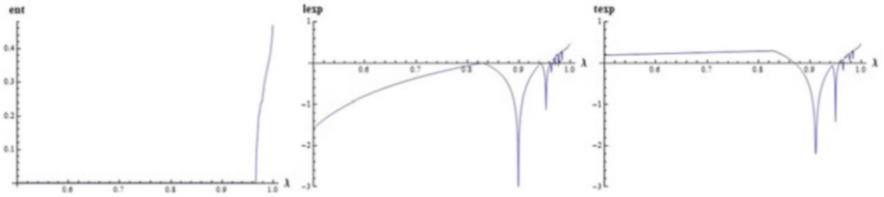


Fig. 3.1 For n=6 we show the computation of topological entropy, Lyapunov exponent and transversal Lyapunov exponent on Δ

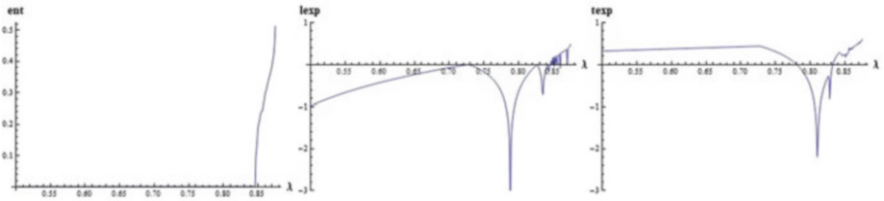


Fig. 3.2 For n=7 we show the computation of topological entropy, Lyapunov exponent and transversal Lyapunov exponent on Δ

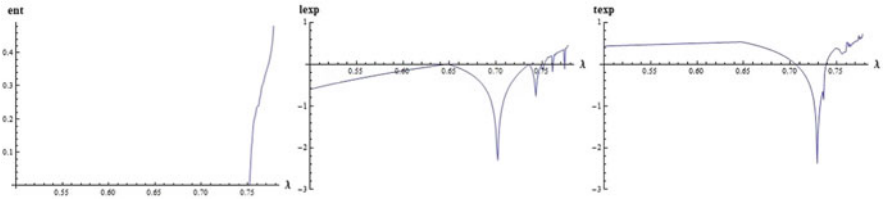


Fig. 3.3 For n=8 we show the computation of topological entropy, Lyapunov exponent and transversal Lyapunov exponent on Δ

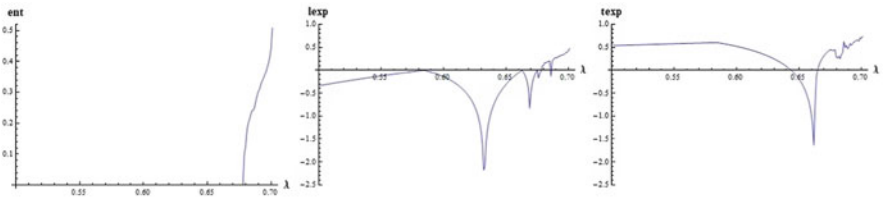


Fig. 3.4 For n=9 we show the computation of topological entropy, Lyapunov exponent and transversal Lyapunov exponent on Δ

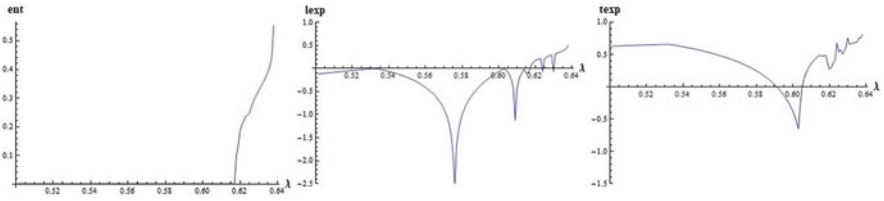


Fig. 3.5 For n=10 we show the computation of topological entropy, Lyapunov exponent and transversal Lyapunov exponent on Δ

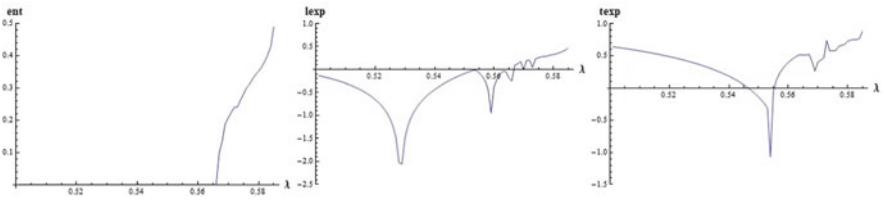


Fig. 3.6 For n=11 we show the computation of topological entropy, Lyapunov exponent and transversal Lyapunov exponent on Δ

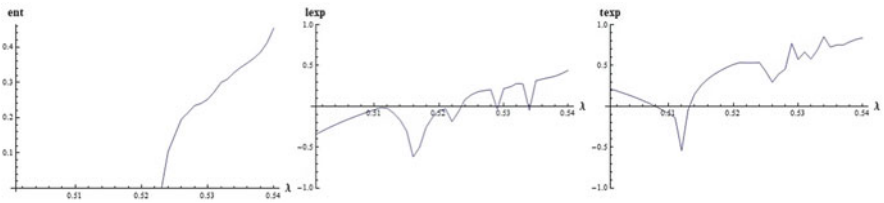


Fig. 3.7 For n=12 we show the computation of topological entropy, Lyapunov exponent and transversal Lyapunov exponent on Δ

Figure 3.8 shows the bifurcation diagrams. We construct them by iterating 100,000 times and drawing the last 250 points. Our study reveals that complicated dynamics appears when $n \geq 6$. When the number of firms increases has the effect that complicated dynamics can be obtained for smaller values of the parameter λ. It is interesting to point out that our computations reveal that the normal Lyapunov exponent cannot be negative if the tangential Lyapunov exponent is positive. Hence, firms synchronization is detected only in the non-chaotic case. This situation is completely different for the case analyzed in [20].

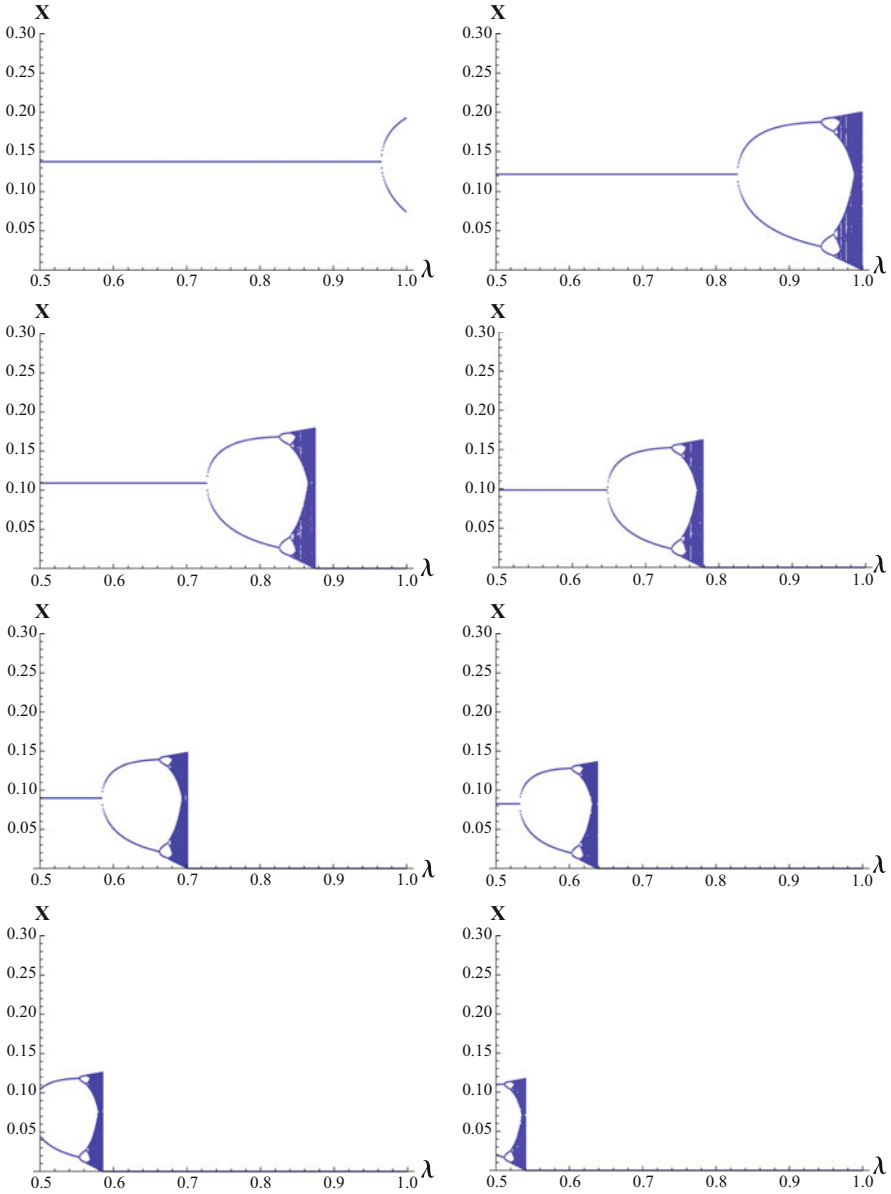


Fig. 3.8 We show the bifurcation diagrams on Δ when the number of firms changes from 5 (top-left) to 12 (bottom-right). We realize that we have more complexity when the number of firms increases, but this complexity is more difficult to be observed in a numerical simulation. Note that for some values of λ, only the fixed point 0 is observed

3.5 Coupled Maps Lattice Models

3.5.1 *Chemical Reactions: Belushov–Zhabotinsky Chemical Reaction*

The main aim of this section is to analyze the system of difference equations given by

$$x_{t+1}^m = (1-\varepsilon) f(x_t^m) + \frac{\varepsilon}{2}\left(f(x_t^{m-1}) + f(x_t^{m+1})\right), \tag{3.1}$$

where $(x_t^1, \ldots, x_t^n) \in \mathbb{R}^n$ for each $t \geq 0$, and it satisfies the boundary condition $x_t^{n+1} = x_t^1$. The parameter $\varepsilon \in [0,1]$ is called the coupling constant and $f : [0,1] \to [0,1]$ is a continuous map, which is usually unimodal.

The model (3.1) is related with the Belushov–Zhabotinsky chemical reaction [57] and the Kaneko's works [33, 34].

Our model (3.1) is similar to these given by Kaneko. When studying its dynamical behavior, we can focus our attention in the notion of chaos. It is worth mentioning that several authors have investigated the existence of chaos in this model (see e.g. [28, 29, 41, 42, 58] or [59]).

To find chaotic dynamics in the system (3.1) is not really complicated for all the coupling parameters ε: just restrict the phase space to the invariant set Δ and consider an unimodal map f such that the associated first order difference equation $x_{t+1} = f(x_t)$ has positive topological entropy.

On Δ, the Jacobian matrix reads as

$$\mathbf{J} = \operatorname{circ}(a, b, 0, 0, \ldots, b) \tag{3.2}$$

where $a := (1-\varepsilon) f'(x)$ and $b := \frac{\varepsilon}{2} f'(x)$. Then $f'(x)$ is an eigenvalue on the direction of Δ. On the other hand, for $0 \leq j \leq n-1$, let

$$\phi_j = e^{i\frac{2\pi j}{n}},\ \mathbf{v}_j = (1, \phi_j, \phi_j^2, \ldots, \phi_j^{M-1}),\ \alpha_j = a + 2b\cos\left(\frac{2\pi j}{n}\right). \tag{3.3}$$

Then, applying [22, Th. 3.2.2, p. 72], we obtain easily that $\mathbf{v}_j$ is an eigenvector of $\mathbf{J}$ and α_j is an eigenvalue associated to $\mathbf{v}_j$ for any $0 \leq j \leq n-1$. Proceeding as in [21], we easily check that, if we denote by F the map defining the whole system, the tangential Lyapunov exponent at Δ is given by

$$lyex(F, x, v) = \lim_{t\to\infty} \frac{1}{t} \log \|\Pi_\Delta \circ dF_x^t \circ \Pi_\Delta(v)\| \tag{3.4}$$

$$= \lim_{n\to\infty} \frac{1}{t} \sum_{i=0}^{t-1} \log |f'(f^i(x))| \tag{3.5}$$

which agrees with the Lyapunov exponent of f at x. On the other hand, the normal Lyapunov exponent is computed as

$$\begin{aligned} tyex(F,x,v) &= \lim_{t\to\infty}\frac{1}{t}\log\|\Pi_{\Delta^\perp}\circ dF^t_{\tilde{x}}\circ\Pi_{\Delta^\perp}(v)\| \\ &= \lim_{t\to\infty}\max_{1\le j\le n-1}\frac{1}{t}\sum_{k=0}^{t-1}\log\left|\left[1-\varepsilon+\varepsilon\cos\left(\frac{2\pi j}{n}\right)\right]f'(f^k(x))\right| \\ &= \max_{1\le j\le t-1}\log\left|1-\varepsilon+\varepsilon\cos\left(\frac{2\pi j}{n}\right)\right|+\lim_{t\to\infty}\frac{1}{t}\sum_{k=0}^{t-1}\log\left|f'(f^k(x))\right| \\ &= \max_{1\le j\le n-1}\log\left|1-\varepsilon\left[1-\cos\left(\frac{2\pi j}{n}\right)\right]\right|+lyex(F,x,v) \end{aligned}$$

Before computing the tangential and normal Lyapunov exponents for a suitable model, we will introduce an application from biology which is strongly connected with this model, and whose mathematical analysis is completely similar.

3.5.2 Application to Biological Systems

The growth of species under dispersal on smaller regions has been modeled by a difference equation with the form

$$x^m_{t+1}=\sum_{j=1}^{n}d_{mj}f(x^j_t),\ m=1,2,\ldots,n, \tag{3.6}$$

where x^m_t is a population of a species in a region with dispersal rates $d_{mj}\ge 0$, $1\le m,j\le n$. We will assume that population does not decrease because of dispersal, that is,

$$\sum_{j=1}^{n}d_{mj}=1,\ m=1,2,\ldots,n.$$

In addition, the evolution of local populations, that is, when $d_{mm}=1$, is given by

$$x^m_{t+1}=f(x^m_t),\ m=1,2,\ldots,n.$$

For more information on this kind of models see [31] or [49].

In this chapter, we choose the well-known Ricker model [48], given by

$$f(x) = xe^{r-x}, \tag{3.7}$$

which depends on a parameter $r > 0$. Additionally, to simplify our computations, we will consider that conditions on coupling parameters d_{ij} such that the system (3.6) can be written as the system (3.1).

3.5.3 Mathematical Analysis of the Models

It is important to point out that both, chemical and biological models can be analyzed with the same mathematical tools. As in the oligopoly case, we concentrate our analysis in studying the dynamics on the invariant set Δ and whether the dynamics outside this set can converge to the dynamics on Δ. Firstly, note that the dynamics on Δ is given by the Ricker model (3.7). Figure 3.9 shows the computation of topological entropy with accuracy 10^{-6}, estimations of Lyapunov exponents and bifurcation diagrams.

On the other hand, Fig. 3.10 shows the evolution when n increases of the parameter region (r, ε) such that the normal Lyapunov exponent is negative, that is, when the dynamics outside of Δ can synchronize, and when such synchronization can be made in a chaotic way. As we can see, the parameters region allowing this synchronization is reduced when n increases.

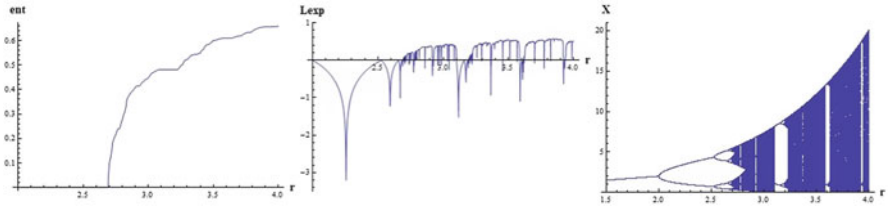


Fig. 3.9 We show the computation of topological entropy, Lyapunov exponent and bifurcation diagram for the Ricker family

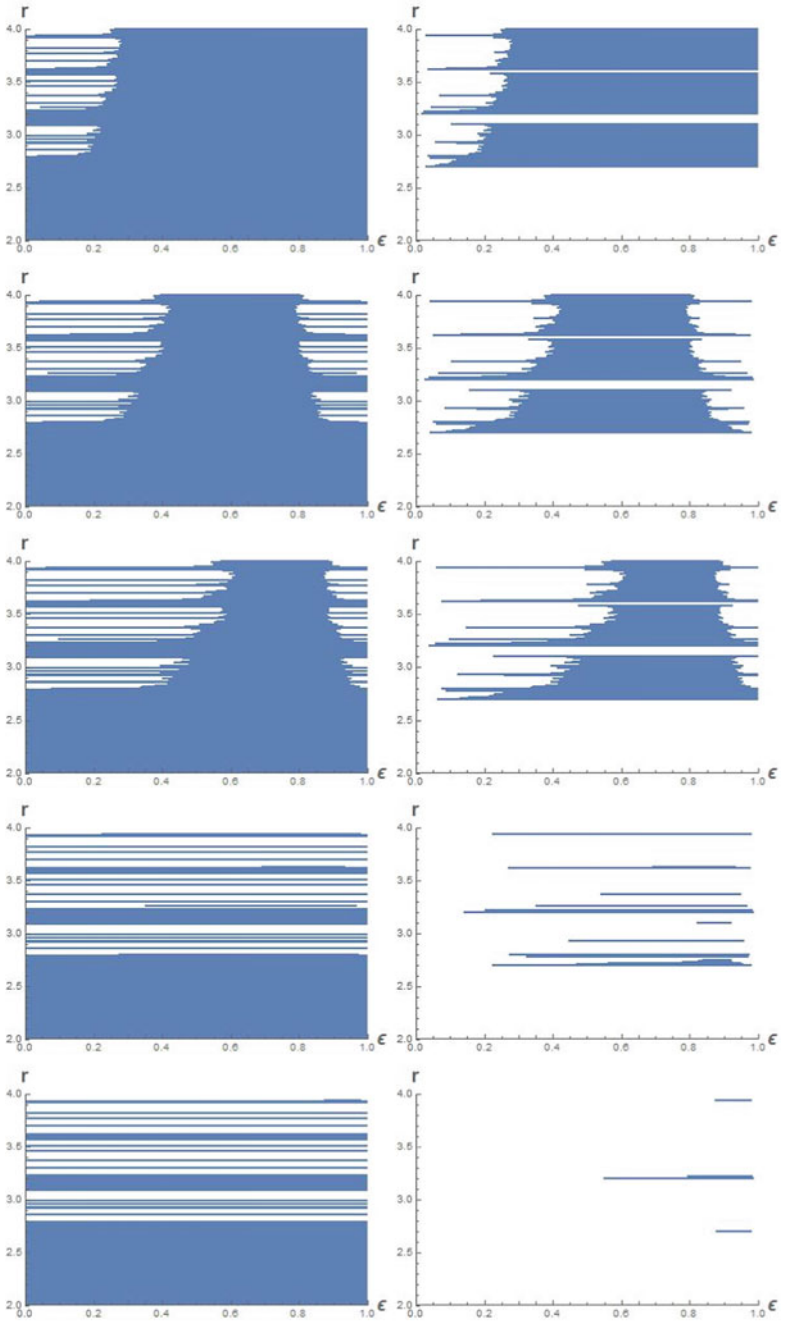


Fig. 3.10 We show the set on the parameter space (ε, r) on which the model synchronizes (left) and chaotically synchronizes (right) for $n = 3, 4, 5, 10, 20$ from top to down. We see that synchronization is more difficult when the number n increases. In particular, for $n = 26$ our simulations show that chaotic synchronization is not possible

Acknowledgements This work has been supported by the grants MTM2014-52920-P and MTM 2017-84079-P from Ministerio de Economía y Competitividad (Spain).

References

1. Adler, R.L., Konheim, A.G., McAndrew, M.H.: Topological entropy. Trans. Am. Math. Soc. **114**, 309–319 (1965)
2. Agronsky, S., Ceder, J.: What sets can be - limit sets in E^n? Real Anal. Exch. **17**, 97–109 (1991–1992)
3. Agronsky, S., Ceder, J.: Each Peano subspace of E^k is an ω–limit set. Real Anal. Exch. **17**, 371–378 (1991–1992)
4. Alsedá, L., Llibre, J., Misiurewicz, M.: Combinatorial Dynamics and Entropy in Dimension One. World Scientific Publishing, Singapore (1993)
5. Aoki, N., Hiraide, K.: Topological Theory of Dynamical Systems: Recent Advances. North-Holland, Amsterdam (1994)
6. Ashwin, P., Buescu, J., Stewart, I.: From attractor to chaotic saddle: a tale of transverse instability. Nonlinearity **9**, 703–737 (1996)
7. Balibrea, F., Snoha, L.: Topological entropy of Devaney chaotic maps. Topol. Appl. **133**, 225–239 (2003)
8. Balibrea, F., Cánovas, J.S., Linero, A.: On ω–limit sets of antitriangular maps. Topology Appl. **137**, 13–19 (2004)
9. Banks, J., Brooks, J., Cairns, G., Davis, G., Stacey, P.: On Devaney's definition of chaos. Am. Math. Mon. **99**, 332–334 (1992)
10. Bischi, G.I., Cerboni Baiardi, L.: Fallacies of composition in nonlinear marketing models. Commun. Nonlinear Sci. Numer. Simul. **20**, 209–228 (2015)
11. Bischi, G.I., Chiarella, C., Kopel, M., Szidarovszky, F.: Nonlinear Oligopolies. Springer, Berlin (2010)
12. Blanchard, F., Glasner, E., Kolyada, S., Maass, A.: On Li-Yorke pairs. J. Reine Angew. Math. **547**, 51–68 (2002)
13. Block, L.S., Coppel, W.A.: Dynamics in One Dimension. Lectures Notes in Mathematics, vol. 1513. Springer, Berlin (1992)
14. Block, L., Keesling, J.: Computing the topological entropy of maps of the interval with three monotone pieces. J. Stat. Phys. **66**, 755–774 (1992)
15. Block, L., Keesling, J., Li, S., Peterson, K.: An improved algorithm for computing topological entropy. J. Stat. Phys. **55**, 929–939 (1989)
16. Bowen, R.: Entropy for group endomorphism and homogeneous spaces. Trans. Am. Math. Soc. **153**, 401–414 (1971)
17. Cánovas, J.S., Linero, A.: Topological dynamic classification of duopoly games. Chaos, Solitons Fractals **12**, 1259–1266 (2001)
18. Cánovas, J.S., Muñoz Guilermo, M.: Computing topological entropy for periodic sequences of unimodal maps. Commun. Nonlinear Sci. Numer. Simul. **19**, 3119–3127 (2014)
19. Cánovas, J.S., Muñoz Guilermo, M.: Computing the topological entropy of continuous maps with at most three different kneading sequences with applications to Parrondo's paradox. Chaos, Solitons Fractals **83**, 1–17 (2016)
20. Cánovas, J.S., Muñoz Guilermo, M.: Dynamics on large sets and its applications to oligopoly dynamics. In: Complex Networks and Dynamics. Springer, Berlin (2016)
21. Cánovas, J.S., Linero, A., Soler López, G.: Chaotic synchronization in a coupled lattice related with Belousov–Zhabotinsky reaction. Commun. Nonlinear Sci. Numer. Simul. **62**, 418–428 (2018)
22. Davis, P.J.: Circulant Matrices. Wiley, New York (1979)
23. de Melo, W., van Strien, S.: One–Dimensional Dynamics. Springer, New York (1993)

24. Devaney, R.L.: An Introduction to Chaotic Dynamical Systems. Addison-Wesley, Redwood City (1989)
25. Dinaburg, E.I.: The relation between topological entropy and metric entropy. Sov. Math. **11**, 13–16 (1970)
26. Du, B.S.: A simple proof of Sharkovsky's theorem. Am. Math. Mon. **111**, 595–599 (2004)
27. Fedorenko, V.V., Sharkovsky, A.N., Smítal, J.: Characterizations of weakly chaotic maps of the interval. Proc. Am. Math. Soc., **110**, 141–148 (1990)
28. García Guirao, J.L., Lampart, M.: Positive entropy of a coupled lattice system related with Belusov–Zhabotinskii reaction. J. Math. Chem. **48**, 66–71 (2010)
29. García Guirao, J.L., Lampart, M.: Chaos of a coupled lattice system related with Belusov–Zhabotinskii reaction. J. Math. Chem. **48**, 159–164 (2010)
30. Guckhenheimer, J.: Sensitive dependence to initial conditions for one-dimensional maps. Commun. Math. Phys. **70**, 133–160 (1979)
31. Hastings, A.: Complex interactions between dispersal and dynamics: lessons from coupled logistic equations. Ecology **74**, 1362–1372 (1993)
32. Huang, W., Ye, X.: Devaney's chaos or 2-scattering implies Li-Yorke's chaos. Topol. Appl. **117**, 259–272 (2002)
33. Kaneko, K.: Period-doubling of kink-antikink patterns, quasiperiodicity and antiferro-like structures and spatial intermittency in coupled logistic lattice. Prog. Theor. Phys. **72**, 480–486 (1984)
34. Kaneko, K.: Globally coupled chaos violates law of large numbers, but not the Central-Limit Theorem. Phys. Rev. Lett. **65**, 1391–1394 (1990). See also Errata. Phys. Rev. Lett. **66**, 243 (1991)
35. Kloeden, P.E.: On Sharkovsky's cycle coexistence ordering. Bull. Aust. Math. Soc. **20**, 171–177 (1979)
36. Kolyada, S.F.: On dynamics of triangular maps of the square. Ergod. Theory Dyn. Syst. **12**, 749–768 (1992)
37. Kolyada, S.F., Snoha, Ľ., On ω–limit sets of triangular maps. Real Anal. Exch. **18**, 115–130 (1992–1993)
38. Kopel, M.: Simple and complex adjustment dynamics in Cournot duopoly models. Chaos, Solitons Fractals **7**, 2031–2048 (1996)
39. Kwietniak, D., Misiurewicz, M.: Exact devaney chaos and entropy. Qual. Theory Dyn. Syst. **6**, 169–179 (2005)
40. Li, T.Y., Yorke, J.A.: Period three implies chaos. Am. Math. Mon. **82**, 985–992 (1975)
41. Li, R., Wang, J., Lu, T., Jiang, R.: Remark on topological entropy and $\mathscr{P}$–chaos of a coupled lattice system with non-zero coupling constant related with Belusov-Zhabotinskii reaction. J. Math. Chem. **54**, 1110–1116 (2016)
42. Liu, J., Lu, T., Li, R.: Topological entropy and $\mathscr{P}$-chaos of a coupled lattice system with non-zero coupling constant related with Belousov–Zhabotinsky reaction. J. Math. Chem. **53**, 1220–1226 (2015)
43. Milnor, J.: On the concept of attractor. Commun. Math. Phys. **99**, 177–195 (1985)
44. Misiurewicz, M., Szlenk, W.: Entropy of piecewise monotone mappings. Stud. Math. **67**, 45–63 (1980)
45. Oseledets, V.I.: A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. Trans. Mosc. Math. Soc. **19**, 197–231 (1968)
46. Puu, T.: Chaos in duopoly pricing. Chaos, Solitons Fractals **1**, 573–581 (1991)
47. Puu, T., Norin, A.: Cournot duopoly when the competitors operate under capacity constraints. Chaos, Solitons Fractals **18**, 577–592 (2003)
48. Ricker, W.E.: Stock and recruitment. J. Fish. Res. Board Can. **11**, 559–623 (1954)
49. Ruíz Herrera, A.: Analysis of dispersal effects in metapopulation models. J. Math. Biol. **72**, 683–698 (2016)
50. Sharkovsky, A.N., Kolyada, S.F., Sivak, A.G., Fedorenko, V.V.: Dynamics of One–Dimensional Maps. Kluwer Academic Publishers, Dordrecht (1997)

51. Singer, D.: Stable orbits and bifurcation of maps of the interval. SIAM J. Appl. Math. **35**, 260–267 (1978)
52. Smale, S.: Differentiable dynamical systems. Bull. Am. Math. Soc. **73**, 747–817 (1967)
53. Smítal, J.: Chaotic functions with zero topological entropy. Trans. Am. Math. Soc. **297**, 269–282 (1986)
54. Sumi, N.: Diffeomorphisms with positive entropy and chaos in the sense of Li–Yorke. Ergod. Theory Dyn. Syst. **23**, 621–635 (2003)
55. Thunberg, H.: Periodicity versus chaos in one–dimensional dynamics. SIAM Rev. **43**, 3–30 (2001)
56. van Strien, S., Vargas, E.: Real bounds, ergodicity and negative Schwarzian for multimodal maps. J. Am. Math. Soc. **17**, 749–782 (2004)
57. Winfree, A.T.: The prehistory of the Belousov–Zhabotinsky oscillator. J. Chem. Educ. **61**, 661–663 (1984)
58. Wu, X., Zhu, P.: Li–Yorke chaos in a coupled lattice system related with Belusov–Zhabotinskii reaction. J. Math. Chem. **50**, 1304–1308 (2012)
59. Wu, X., Zhu, P.: The principal measure and distributional (p, q)-chaos of a coupled lattice system related with Belousov–Zhabotinsky reaction. J. Math. Chem. **50**, 2439–2445 (2012)

Chapter 4
Existence and Invariance of Global Attractors for Impulsive Parabolic System Without Uniqueness

Sergey Dashkovskiy, Petro Feketa, Oleksiy V. Kapustyan, and Iryna V. Romaniuk

Abstract In this paper, we apply the abstract theory of global attractors for multi-valued impulsive dynamical systems to weakly-nonlinear impulsively perturbed parabolic system without uniqueness of a solution to the Cauchy problem. We prove that for a sufficiently wide class of impulsive perturbations (including multi-valued ones) the global attractor of the corresponding multi-valued impulsive dynamical system has an invariant non-impulsive part.

4.1 Introduction

The paper studies qualitative behavior of solutions to impulsive dynamical systems (DS), i.e. autonomous systems whose trajectories undergo impulsive perturbations at the moments of intersection of the trajectories with a certain surface in the phase space. For finite-dimensional systems we refer the reader to the works [1, 3, 8, 9, 13, 14, 24–26, 29, 30, 33–35] in which stability properties and long-time behavior of solutions have been studied. Stability of infinite dimensional impulsive systems with external perturbations was studied in [10].

For infinite-dimensional dissipative dynamical systems the theory of global attractors [7, 36] proved to be an effective tool to describe qualitative behavior of solutions. For multi-valued dynamical systems in the case of non-uniqueness of a

S. Dashkovskiy
University of Würzburg, Würzburg, Germany
e-mail: sergey.dashkovskiy@mathematik.uni-wuerzburg.de

P. Feketa
University of Kaiserslautern, Kaiserslautern, Germany
e-mail: petro.feketa@mv.uni-kl.de

O. V. Kapustyan (✉) · I. V. Romaniuk
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
e-mail: alexkap@univ.kiev.ua

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_4

solution to the Cauchy problem this theory has been further developed in [2, 15, 17–21, 23, 27, 28, 37].

A lack of continuous dependence on initial data in impulsive dynamical systems requires a new concept of global attractor for such systems. One approach has been proposed in [4–6]. The core idea of those papers is to keep the invariance property in the definition of attractor. However, this approach sets very strong constraints (so-called "tube condition") on the behavior of trajectories of the given nonlinear system in a neighborhood of the impulsive set. This does not allow to apply this approach effectively and extend it on wider classes of infinite-dimensional nonlinear evolution problems. Another approach based on the notion of global attractor for non-autonomous systems [7, 20] has been developed in [11, 12, 22, 32]. In particular, it exploits the notion of global attractor for the systems with impulsive effects at fixed moments of time [16, 31]. This notion requires minimality property of the global attractor instead of the invariance. It allows to obtain results on the existence and study properties of global attractors for impulsive dynamical systems with infinite number of impulsive points under natural assumptions on systems' parameters.

In the present paper, we extend the mentioned results on the following problem:

$$\frac{du}{dt} = F(u),\ u \notin M, \tag{4.1}$$

$$u|_{t=0} = u_0 \in X, \tag{4.2}$$

$$\triangle u|_{u\in M} \in Iu - u, \tag{4.3}$$

where (4.1), (4.2) is an infinite-dimensional evolution system (in fact, two-dimensional parabolic system) in the phase space X for which the uniqueness of solutions is not assumed, $\triangle u(t) = u(t+0) - u(t-0)$ denotes the instantaneous increment of the state variable u, and M is some subset of the phase space X.

The solution $u = u(t)$ to the problem (4.1)–(4.3) is right-continuous function satisfying (4.1) $\forall t \neq \tau$, where τ is defined by the equation $u(\tau - 0) \in M$, and jumps to the state $u(\tau) \in Iu(\tau - 0)$ at the moment of time τ, where $I : M \mapsto X$ is a given (maybe, multi-valued) map. The set M is called *impulsive set*. The map I is called *impulsive map*, points from the set IM are called *impulsive points*.

We will show that the problem (4.1)–(4.3) generates multi-valued dynamical system $G : \mathbb{R}_+ \times X \to P(X)$ under some natural assumptions (see Definition 4.1). In this paper, we study the existence and invariance properties of global attractor of dynamical system G. A lack of continuous dependence on initial data for the problem (4.1)–(4.3) leads to the discontinuity of the map $G(t, \cdot)$. This requires a reformulation of the classical definition of the global attractor [27]. Therefore, we consider the global attractor as a compact minimal uniformly attracting set $\Theta \subset X$ (see Definition 4.2). Under natural assumptions on the impulsive parameters M and I, we prove the existence of Θ and invariance of the set $\Theta \setminus M$ for two-dimensional impulsively perturbed weakly-nonlinear parabolic system without uniqueness of solutions to the Cauchy problem.

4.2 Global Attractors of Abstract Multi-Valued Impulsive Dynamical Systems

In this section, we present basic concepts and results of global attractor theory for abstract multi-valued impulsive DS. This presentation is based on the recently obtained results [11, 12].

Let (X, ρ) be a complete metric space, $P(X)$ $(\beta(X))$ be a set of all non-empty (non-empty bounded) subsets of X, and for any $A, B \in P(X)$ we denote

$$dist(A, B) = \sup_{x \in A} \inf_{y \in B} \rho(x, y).$$

Definition 4.1 A multi-valued map $G : \mathbb{R}_+ \times X \to P(X)$ is called a multi-valued DS (MDS), if

1) $\forall x \in X \quad G(0, x) = x$;
2) $\forall x \in X \quad \forall t, s \geq 0 \quad G(t + s, x) \subseteq G(t, G(s, x))$.

The MDS is called strict if in 2) the equality takes place.

Remark 4.1 If G is a single-valued map, Definition 4.1 coincides with the definition of classical semigroup.

Definition 4.2 A non-empty subset $\Theta \subset X$ is called a global attractor of the MDS G if

1) Θ is compact;
2) Θ is uniformly attracting, i.e.,

$$\forall\, B \in \beta(X) \;\; dist(G(t, B), \Theta) \to 0, \; t \to \infty;$$

3) Θ is minimal among all closed uniformly attracting sets.

Remark 4.2 If global attractor exists then it is unique.

Note that we do not assume any continuity properties for the map $G(t, \cdot)$. Therefore it seems to be natural to change classical definition of the global attractor and require minimality condition 3) instead of the invariance property. On the other hand, if the MDS G has global attractor in classical sense, i.e., if there exists a set $\Theta_1 \subset X$ which satisfies 1), 2) and $\Theta_1 \subset G(t, \Theta_1)$ $\forall t \geq 0$ then, clearly, Θ_1 satisfies 3).

Next, we recall the existence criterion of the global attractor for the dissipative MDS from [11]:

Lemma 4.1 *Assume that the MDS G satisfies the dissipativity condition:*

$$\exists B_0 \in \beta(X) \quad \forall B \in \beta(X) \quad \exists T = T(B) \geq 0 \quad \forall t \geq T \quad G(t, B) \subset B_0. \tag{4.4}$$

Then the following conditions are equivalent:

1) *MDS G has the global attractor Θ;*
2) *MDS G is asymptotically compact, i.e., $\forall t_n \nearrow \infty \quad \forall B \in \beta(X)$*

$$\text{every sequence } \{\xi_n \in G(t_n, B)\} \text{ is precompact in } X. \tag{4.5}$$

Moreover, under condition (4.4) it holds that

$$\Theta = \omega(B_0) := \bigcap_{\tau>0} \overline{\bigcup_{t\geq\tau} G(t, B_0)}. \tag{4.6}$$

Now we introduce a special subclass of MDS called impulsive MDS. To this purpose, we impose additional structural properties on MDS.

Let K be some family of continuous maps $\varphi : [0, +\infty) \to X$ and the following properties hold:

K1) $\forall x \in X \quad \exists \varphi \in K : \varphi(0) = x$;
K2) $\forall \varphi \in K \quad \forall s \geq 0 \quad \varphi(\cdot + s) \in K$.

We denote

$$K_x = \{\varphi \in K \mid \varphi(0) = x\}.$$

In most applications, K is some set of solutions to a particular autonomous problem.

Remark 4.3 If in assumption K1) for every $x \in X$ there exists a unique $\varphi \in K$ such that $\varphi(0) = x$, then K_x consists of a single trajectory φ and the equality $V(t, x) = \varphi(t)$ defines a classical semigroup $V : \mathbb{R}_+ \times X \mapsto X$.

In all further arguments we will understand impulsive MDS as an MDS G consisting of a given family of maps K with properties $K1)$, $K2)$, a given set $M \subset X$, and a given map $I : M \to P(X)$. We denote it by $G = (K, M, I)$. Such impulsive MDS describes the following behavior: a phase point moves along trajectories of K and jumps onto a new position from the set IM when it meets the set M.

For the "well-posedness" of impulsive problem we assume the following conditions:

$$M \subset X \text{ is a closed set, } I : M \mapsto P(X) \text{ is a compact-valued map;} \tag{4.7}$$

$$M \cap IM = \emptyset; \tag{4.8}$$

$$\forall x \in M \quad \forall \varphi \in K_x \quad \exists \tau = \tau(\varphi) > 0 \quad \forall t \in (0, \tau) \quad \varphi(t) \notin M. \tag{4.9}$$

Also, we introduce the following notation:

$$\text{for } x \in M \; x^+ \text{ means some element from } Ix;$$

$$\text{for } \varphi \in K \; M^+(\varphi) = \bigcup_{t>0} \varphi(t) \cap M.$$

The following result is crucial for construction of impulsive DS [3, 24].

Lemma 4.2 ([12]) *If conditions (4.7)–(4.9) hold then for every $\varphi \in K$ either $M^+(\varphi) = \emptyset$ or $\exists s = s(\varphi) > 0$ such that*

$$\varphi(s) \in M \text{ and } \varphi(t) \notin M \; \forall t \in (0, s). \tag{4.10}$$

According to (4.10) we can define function $s : K \to (0, +\infty]$ as

$$s(\varphi) = \begin{cases} s, & \text{if} \quad M^+(\varphi) \neq \emptyset, \\ +\infty, & \text{if} \quad M^+(\varphi) = \emptyset. \end{cases} \tag{4.11}$$

Now, we construct an impulsive trajectory $\tilde{\varphi}$ starting from $x_0 \in X$. Let $\varphi_0 \in K_{x_0}$. If $M^+(\varphi_0) = \emptyset$ then we define $\tilde{\varphi}$ on $[0, +\infty)$ by the equality

$$\tilde{\varphi}(t) = \varphi_0(t) \; \forall t \geq 0.$$

If $M^+(\varphi_0) \neq \emptyset$, then for $s_0 = s(\varphi_0) > 0$, $x_1 = \varphi_0(s_0) \in M$ and for $x_1^+ \in Ix_1$ we define $\tilde{\varphi}$ on $[0, s_0]$ by the equality

$$\tilde{\varphi}(t) = \begin{cases} \varphi_0(t), & t \in [0, s_0), \\ x_1^+, & t = s_0. \end{cases}$$

Let $\varphi_1 \in K_{x_1^+}$. If $M^+(\varphi_1) = \emptyset$ then

$$\tilde{\varphi}(t) = \varphi_1(t - s_0) \; \forall t \geq s_0.$$

If $M^+(\varphi_1) \neq \emptyset$, then for $s_1 = s(\varphi_1) > 0$, $x_2 = \varphi_1(s_1) \in M$ and $x_2^+ \in Ix_2$ we define $\tilde{\varphi}$ on $[s_0, s_0 + s_1]$ by the equality

$$\tilde{\varphi}(t) = \begin{cases} \varphi_1(t - s_0), & t \in [s_0, s_0 + s_1), \\ x_2^+, & t = s_0 + s_1. \end{cases}$$

Continuing this process we obtain impulsive trajectory $\tilde{\varphi}$ with finite or infinite number of impulsive points $\{x_n^+\}_{n\geq 1} \subset IM$, corresponding durations between impulses $\{s_n\}_{n\geq 0} \subset (0, \infty)$ and functions $\{\varphi_n\}_{n\geq 0} \subset K$.

We denote impulsive trajectory $\tilde{\varphi}$ by

$$\tilde{\varphi} = \tilde{\varphi}(\{x_n^+\}, \{s_n\}, \{\varphi_n\}).$$

and define

$$t_0 = 0, \quad t_{n+1} := \sum_{k=0}^{n} s_k.$$

If $\tilde{\varphi}$ has infinite number of jumps, then it is defined by the formula

$$\forall n \geq 0 \ \ \forall t \geq 0 \ \ \tilde{\varphi}(t) = \begin{cases} \varphi_n(t - t_n), & t \in [t_n, t_{n+1}), \\ x_{n+1}^+, & t = t_{n+1}. \end{cases} \tag{4.12}$$

Let $\tilde{K}_x$ denote the set of all impulsive trajectories starting from $x \in X$. We assume that every impulsive trajectory is defined on $[0, +\infty)$, i.e.,

$$\forall x \in X \quad \text{every } \tilde{\varphi} \in \tilde{K}_x \quad \text{is defined on} \quad [0, +\infty). \tag{4.13}$$

Remark 4.4 Assumption (4.13) means that for every impulsive trajectory

$$\tilde{\varphi} = \tilde{\varphi}(\{x_n^+\}, \{s_n\}, \{\varphi_n\})$$

either the number of its impulsive points is finite and the continuous dynamics does not lead to finite escape time or $\sum_{n=0}^{\infty} s_n = \infty$ i.e., there is no accumulation of impulses on a finite time interval.

Remark 4.5 There are some trivial situations when assumption (4.13) is satisfied. For example, if conditions (4.7)–(4.9), K1), and K2) hold, the set IM is compact and the family K satisfies some additional assumption (see (4.25)) then condition (4.13) is satisfied. In the single-valued case it was discussed in [5].

Remark 4.6 From (4.8) and (4.12) we have

$$\forall x \in X \ \ \forall \tilde{\varphi} \in \tilde{K}_x \ \ \forall t > 0 \ \ \tilde{\varphi}(t) \notin M. \tag{4.14}$$

Now we are ready to present a rigorous definition of impulsive MDS.

Definition 4.3 A multi-valued map $G : \mathbb{R}_+ \times X \to P(X)$ defined by the formula

$$G(t, x) = \{\tilde{\varphi}(t) \mid \tilde{\varphi} \in \tilde{K}_x\}, \tag{4.15}$$

where $\tilde{K}_x$ denotes the set of all impulsive trajectories starting from $x \in X$, is called impulsive MDS, which will be denoted by $G = (K, M, I)$.

Remark 4.7 If in assumption K1), for every $x \in X$ there exists a unique $\varphi \in K$ such that $\varphi(0) = x$, and the impulsive map $I : M \mapsto X$ is single-valued, then (4.15) defines single-valued impulsive DS [24] by the formula

$$\forall n \geq 0 \ \ \forall t \geq 0 \ \ G(t,x) = \begin{cases} V(t-t_n, x_n^+), & t \in [t_n, t_{n+1}), \\ x_{n+1}^+, & t = t_{n+1}, \end{cases} \tag{4.16}$$

where $x_{n+1}^+ = IV(t_{n+1} - t_n, x_n^+)$, and $V : \mathbb{R}_+ \times X \mapsto X$ is a semigroup generated by K.

Lemma 4.3 ([11]) *Let conditions K1),K2), (4.7)–(4.9), and (4.13) hold. Then*

$$\forall t, s \geq 0 \, \forall x \in X \ \ G(t+s, x) \subset G(t, G(s, x)),$$

i.e., formula (4.15) defines an MDS.

If, additionally, $\forall \varphi, \psi \in K$, $\forall s > 0$ *such that* $\varphi(0) = \psi(s)$ *it holds that*

$$\theta(p) = \begin{cases} \psi(p), & p \in [0, s), \\ \varphi(p-s), & p \geq s \end{cases} \in K, \tag{4.17}$$

then G is strict.

Remark 4.8 Property (4.17) obviously holds if for every $x \in X$ there exists a unique $\varphi \in K$ such that $\varphi(0) = x$.

Remark 4.9 We are also able to prove the following property:

$$\forall x \in X \, \forall \varphi \in \tilde{K}_x \, \forall s \geq 0 \ \varphi(\cdot + s) \in \tilde{K}_{\varphi(s)}.$$

Definition 4.4 We will say that the problem (4.1), (4.3) *generates an impulsive MDS* (by formula (4.15)) if solutions to (4.1) generate a set K of maps $\varphi : [0, +\infty) \mapsto X$ satisfying $K1)$, $K2)$ and for a given set $M \subset X$ and for a given map $I : M \mapsto P(X)$ conditions (4.7)–(4.9), (4.13) are satisfied.

The following examples show that we cannot expect invariance of global attractor for impulsive DS.

In bounded domain $\Omega \subset \mathbb{R}^n$, $n \geq 1$ we consider the problem

$$\begin{cases} \frac{\partial y}{\partial t} = \Delta y, \ \ (t,x) \in (0, \infty) \times \Omega, \\ y|_{\partial\Omega} = 0, \end{cases} \tag{4.18}$$

which generates DS with phase space $X = L^2(\Omega)$ and semigroup V:

$$\text{for } y_0 = \sum_{i=1}^{\infty} c_i \psi_i \ \ V(t, y_0) = y(t) = \sum_{i=1}^{\infty} c_i e^{-\lambda_i t} \psi_i. \tag{4.19}$$

Here and after, $\{\psi_i\}_{i=1}^{\infty}$ is orthonormal basis in $L^2(\Omega)$ such that

$$-\Delta\psi_i = \lambda_i\psi_i,\ \psi_i \in H_0^1(\Omega), \tag{4.20}$$

$\|\cdot\|$ and $(\cdot,\cdot)$ are norm and scalar product in $L^2(\Omega)$.

Let us consider the following impulsive problem with fixed numbers $a > 0$ and $\mu > 0$

$$M = \{y \in X|\ (y,\psi_1) = a\}, \tag{4.21}$$

$$M' = \{y \in X|\ (y,\psi_1) = a(1+\mu)\},$$

$$I : M \mapsto M' \text{ such that for } y = \sum_{i=1}^{\infty} c_i\psi_i$$

$$Iy = (\mu+1)c_1\psi_1 + \sum_{i=2}^{\infty} c_i\psi_i. \tag{4.22}$$

Lemma 4.4 ([32]) *For every $a > 0$, $\mu > 0$ impulsive problem (4.18), (4.21), (4.22) generates an impulsive DS G, which has global attractor Θ in the phase space $X = L^2(\Omega)$ and*

$$\Theta = \bigcup_{t\in[0,\ln(1+\mu)]} \left\{(1+\mu)ae^{-t}\psi_1\right\} \cup \{0\}. \tag{4.23}$$

It is easy to see that Θ is not invariant with respect to the impulsive flow G, i.e. both statements

$$\forall\, t \geq 0\ \ \Theta \subset G(t,\Theta),$$

$$\forall\, t \geq 0\ \ G(t,\Theta) \subset \Theta$$

are not true. The reason is that every impulsive trajectory has infinite number of impulsive points and, therefore, $\Theta \cap M \neq \emptyset$.

On the other hand, it is easy to verify that

$$\forall\, t \geq 0\ \ \Theta \setminus M = G(t,\Theta \setminus M). \tag{4.24}$$

The aim of the paper is to prove such an equality for general weakly-nonlinear two-dimensional impulsive-perturbed parabolic system which will be considered in the next section. For this purpose we need some additional restrictions on K, which we will formulate with the help of function (4.11).

Let Θ be a global attractor of G.

$$\begin{gathered}\forall x_n \to x \ \ \forall \varphi_n \in K_{x_n} \ \ \exists \varphi \in K_x \text{ such that on some subsequence} \\ \varphi_n \to \varphi \text{ uniformly on every } [a,b] \subset [0,\infty). \\ \text{If for } x_n \to x \in \Theta \setminus M, \ \varphi_n \in K_{x_n} \text{ and } \varphi \in K_x \text{ we have} \\ s(\varphi_n) < \infty \text{ and } \varphi_n(t) \to \varphi(t) \ \forall t \geq 0,\end{gathered} \tag{4.25}$$

then

$$s(\varphi) < \infty \text{ and } s(\varphi_n) \to s(\varphi). \tag{4.26}$$

The following Lemma will play a crucial role in proving of (4.24) for particular impulsive problem, considered in the next section. This result firstly has appeared in [5] for single-valued case and under restrictive "tube conditions" on impulsive semiflow in the neighborhood of impulsive set M.

Lemma 4.5 ([12]) *Assume that impulsive MDS $G = (K, M, I)$ satisfies K1), K2), (4.7)–(4.9), (4.13), (4.25), (4.26) and impulsive map $I : M \to P(X)$ is upper-semicontinuous.*

Let Θ be a global attractor of G. Then if $x_n \to x \in \Theta \setminus M$ and $\tilde{\varphi}_n \in \tilde{K}_{x_n}$ has infinite number of impulsive points, then

$$\forall\, t \geq 0 \ \exists\, \tilde{\varphi} \in \tilde{K}_x \ \exists\, \eta_n \to 0+ \ \text{ such that } \tilde{\varphi}_n(t+\eta_n) \to \tilde{\varphi}(t). \tag{4.27}$$

Moreover, for $\alpha_n \to 0+$

$$\tilde{\varphi}_n(\alpha_n) \to x. \tag{4.28}$$

Proof The proof follows directly from the proofs of Lemmas 5, 7 in [12].

4.3 Application to Impulsive Parabolic System

In this section, we apply the obtained abstract results to a weakly-nonlinear two-dimensional parabolic systems whose trajectories have jumps when they reach a certain subset of the phase space. The main result of this paper is to prove the existence, asymptotically precise formula and invariance of non-impulsive part for global attractor of the system. The existence result for another type of two-dimensional parabolic impulsive system was recently obtained in [22].

Let $\Omega \subset \mathbb{R}^n$, $n \geq 1$ be a bounded domain. For unknown functions $u(t,x), v(t,x)$ in $(0,+\infty) \times \Omega$ we consider the following problem:

$$\begin{cases} \frac{\partial u}{\partial t} = a_1 \Delta u + \varepsilon f_1(u,v), \\ \frac{\partial v}{\partial t} = a_2 \Delta v + \varepsilon f_2(u,v), \\ u|_{\partial\Omega} = v|_{\partial\Omega} = 0, \end{cases} \tag{4.29}$$

where $\varepsilon > 0$ is a small parameter, $a_2 \geq a_1 > 0$. Nonlinear functions $f_i : \mathbb{R}^2 \mapsto \mathbb{R}$, $i = 1,2$ are continuous and satisfy the following condition:

$$\exists C > 0 \quad \forall u, v \in \mathbb{R} \quad |f_1(u,v)| + |f_2(u,v)| \leq C. \tag{4.30}$$

The phase space of the problem (4.29) is the space $H = L^2(\Omega) \times L^2(\Omega)$ with the norm $\|z\| = \sqrt{\|u\|^2 + \|v\|^2}$. It is known [7] that for every $\varepsilon > 0$, $z_0 \in H$ there exists at least one solution $z = \begin{pmatrix} u \\ v \end{pmatrix} \in C([0,+\infty); H)$ to the problem (4.29) with $z(0) = z_0$.

Thus, problem (4.29) generates a family of continuous maps

$$K^\varepsilon = \{z : [0,+\infty) \to H \mid z \text{ is a solution of (4.29)}\},$$

which satisfies K1) and K2). Moreover, the condition (4.25) holds.

As a direct generalization of (4.21) on two-dimensional case we consider the following impulsive problem for fixed numbers $\alpha > 0, \beta > 0, \mu > 0, \gamma \in (0, \frac{1+\mu}{\alpha+\beta})$

$$M = \left\{ z = \begin{pmatrix} u \\ v \end{pmatrix} \in H \mid (u,\psi_1) \geq 0,\ (v,\psi) \geq 0,\ \alpha(u,\psi_1) + \beta(v,\psi_1) = 1 \right\} \tag{4.31}$$

$$M' = \left\{ z = \begin{pmatrix} u \\ v \end{pmatrix} \in H \mid (u,\psi_1) \geq 0,\ (v,\psi) \geq 0,\ \alpha(u,\psi_1) + \beta(v,\psi_1) = 1 + \mu \right\}$$

$I : M \mapsto P(M')$ is closed-valued impulsive map, such that

$$\text{for } z = \sum_{i=1}^{\infty} \begin{pmatrix} c_i \\ d_i \end{pmatrix} \psi_i \in M$$

$$Iz \subseteq I_0 z = \left\{ \begin{pmatrix} c_1' \\ d_1' \end{pmatrix} \psi_1 + \sum_{i=2}^{\infty} \begin{pmatrix} c_i \\ d_i \end{pmatrix} \psi_i \mid\ c_1' \geq \gamma,\ d_1' \geq \gamma,\ \alpha c_1' + \beta d_1' = 1 + \mu \right\}. \tag{4.32}$$

Remark 4.10 For every $z \in M$ the set Iz from (4.32) is compact, i.e. impulsive map I is compact-valued. In particular, conditions (4.7) are satisfied.

As a particular example we can consider the following continuous single-valued map $I : M \mapsto M'$ defined by

$$I\left(\sum_{i=1}^{\infty}\begin{pmatrix}c_i\\ d_i\end{pmatrix}\psi_i\right)=\begin{pmatrix}c_1+\frac{\mu}{2\alpha}\\ d_1+\frac{\mu}{2\beta}\end{pmatrix}\psi_1+\sum_{i=2}^{\infty}\begin{pmatrix}c_i\\ d_i\end{pmatrix}\psi_i.$$

Another example is upper-semicontinuous compact-valued map $I \equiv I_0$.

Theorem 4.1 *Under conditions (4.30) for sufficiently small ε the problem (4.29)–(4.32) generates an impulsive MDS $G_\varepsilon : \mathbb{R}_+ \times H \to P(H)$, which has global attractor Θ_ε and*

$$dist(\Theta_\varepsilon, \Theta) \to 0,\ \varepsilon \to 0, \tag{4.33}$$

where

$$\Theta = \bigcup_{t\in[0,\tau],\ c_1\geq\gamma, d_1\geq\gamma}\left\{\begin{pmatrix}c_1e^{-a_1\lambda_1 t}\psi_1\\ d_1e^{-a_2\lambda_1 t}\psi_1\end{pmatrix}\ |\alpha c_1+\beta d_1 = 1+\mu,\right.$$

$$\left.\alpha c_1e^{-a_1\lambda_1\tau}+\beta d_1e^{-a_2\lambda_1\tau}=1\right\}\cup\begin{pmatrix}0\\ 0\end{pmatrix}$$

is global attractor of the unperturbed ($\varepsilon = 0$) impulsive problem (4.29)–(4.32).

If, additionally, the impulsive map $I : M \mapsto P(M')$ is upper-semicontinuous, then

$$\forall t \geq 0\ \ \Theta_\varepsilon \setminus M = G_\varepsilon(t, \Theta_\varepsilon \setminus M). \tag{4.34}$$

Remark 4.11 In all further arguments the phrase "*for sufficiently small ε*" means that there exists $\varepsilon_1 > 0$ that depends only on the parameters of the problem (4.29)–(4.32) such that some property holds for every $\varepsilon \in [0, \varepsilon_1]$.

Proof Under conditions (4.30) and Poincare inequality for an arbitrary solution $z \in K^\varepsilon$ to the problem (4.29) and for a.a. $t > 0$ the following inequality holds:

$$\frac{1}{2}\frac{d}{dt}\|z(t)\|_H^2 + a_1\lambda_1\|z(t)\|_H^2 \leq \varepsilon\sqrt{2}C\|z(t)\|_H. \tag{4.35}$$

Then, there exists $\delta > 0$ such that for a sufficiently small ε we obtain

$$\forall\, z \in K^\varepsilon\ \ \forall\, t \geq 0\ \ \ \|z(t)\|_H^2 \leq \|z(0)\|_H^2 e^{-\delta t} + 1. \tag{4.36}$$

Moreover, for every $z = \begin{pmatrix} u \\ v \end{pmatrix} \in K^\varepsilon$ and $\forall i \geq 1$ we get the following equalities:

$$(u(t), \psi_i) = (u(0), \psi_i)e^{-a_1\lambda_i t} + \varepsilon \int_0^t e^{-a_1\lambda_i(t-s)}(f_1(u(s), v(s)), \psi_i)ds, \tag{4.37}$$

$$(v(t), \psi_i) = (v(0), \psi_i)e^{-a_2\lambda_i t} + \varepsilon \int_0^t e^{-a_2\lambda_i(t-s)}(f_2(u(s), v(s)), \psi_i)ds. \tag{4.38}$$

Further, in order to simplify the relations we denote

$$\psi := \psi_1, \quad \lambda := \lambda_1,$$

$$F_\varepsilon^1(t) = \int_0^t e^{-a_1\lambda(t-s)}(f_1(u(s), v(s)), \psi)ds,$$

$$F_\varepsilon^2(t) = \int_0^t e^{-a_2\lambda(t-s)}(f_2(u(s), v(s)), \psi)ds,$$

$$F_\varepsilon(t) = \alpha F_\varepsilon^1(t) + \beta F_\varepsilon^2(t),$$

where $F_\varepsilon^1, F_\varepsilon^2, F_\varepsilon \in C^1([0, \infty))$, $F_\varepsilon^1(0) = F_\varepsilon^2(0) = F_\varepsilon(0) = 0$. These functions depend on $z \in K^\varepsilon$, but $\exists\, C_1 > 0 \ \ \forall\, \varepsilon \in (0, 1)$

$$\sup_{t\geq 0}\left(|F_\varepsilon^1(t)| + |(F_\varepsilon^1)'(t)| + |F_\varepsilon^2(t)| + |(F_\varepsilon^2)'(t)| + |F_\varepsilon(t)| + |F_\varepsilon'(t)|\right) \leq C_1. \tag{4.39}$$

For $z \in K^\varepsilon$ we consider the function

$$g_\varepsilon(t) = \alpha(u(t), \psi) + \beta(v(t), \psi) \tag{4.40}$$

$$= \alpha e^{-a_1\lambda t}(u(0), \psi) + \beta e^{-a_2\lambda t}(v(0), \psi) + \varepsilon F_\varepsilon(t).$$

Let us verify condition (4.9). For $z(0) \in M$ we deduce

$$g_\varepsilon'(0) \leq -a_1\lambda + \varepsilon F_\varepsilon'(0). \tag{4.41}$$

From the last inequality and (4.39) we deduce that for a sufficiently small ε there exists $\tau = \tau(z(0), \varepsilon) > 0$ such that $\forall t \in (0, \tau) \ \ g_\varepsilon(t) < 1$, which means that (4.9) is satisfied.

Let us verify condition (4.13). This condition directly follows from the estimate (4.36) if z does not intersect the set M. Let us consider other situation. We take $z \in K^\varepsilon$ with $\|z(0)\| \leq R$ such that $s = s(z) < \infty$, where function $s(\cdot)$ is

defined by (4.11). Then from the equality $g_\varepsilon(s) = 1$ for sufficiently small ε we deduce

$$s \le T = T(R) := \frac{1}{a_1\lambda} ln2(\alpha+\beta)R. \tag{4.42}$$

So, without loss of generality we can consider $z \in K^\varepsilon$ with $z(0) = z_0 \in IM$. Then

$$g_\varepsilon(0) = 1+\mu,\ (u_0,\psi) \ge \gamma,\ (v_0,\psi) \ge \gamma.$$

Therefore, $\exists\, s_\varepsilon > 0$ such that

$$\forall\, t \in (0, s_\varepsilon)\ g_\varepsilon(t) > 1,\ g_\varepsilon(s_\varepsilon) = 1.$$

Then for sufficiently small ε we have the next inequality

$$\left(1+\frac{\mu}{2}\right) e^{a_2\lambda s_\varepsilon} \ge 1+\mu. \tag{4.43}$$

From (4.43) we obtain

$$s_\varepsilon \ge \bar{s} = \frac{1}{a_2\lambda} ln \frac{1+\mu}{1+\frac{\mu}{2}}, \tag{4.44}$$

which implies (4.13).

Let us prove that $z(s_\varepsilon) \in M$, i.e.

$$(u(s_\varepsilon),\psi) \ge 0,\ (v(s_\varepsilon),\psi) \ge 0.$$

Indeed, for $t \ge 0$

$$(u(t),\psi) \ge \gamma e^{-a_1\lambda t} - \varepsilon C_1,\ (v(t),\psi) \ge \gamma e^{-a_2\lambda t} - \varepsilon C_1.$$

Hence, $\forall\, t \in [0, \frac{1}{a_2\lambda} \ln \frac{\gamma}{\varepsilon C_1}]$

$$(u(t),\psi) \ge 0,\ (v(t),\psi) \ge 0. \tag{4.45}$$

On the other hand, from the equality $g_\varepsilon(s_\varepsilon) = 1$ we deduce that

$$\alpha(u_0,\psi)e^{-a_1\lambda s_\varepsilon} + \beta(v_0,\psi)e^{-a_2\lambda s_\varepsilon} \ge \frac{1}{2}.$$

Hence,

$$\frac{1}{2}e^{a_1\lambda s_\varepsilon} \leq 1+\mu,$$

$$s_\varepsilon \leq \hat{s} = \frac{1}{a_1\lambda} ln2(1+\mu). \tag{4.46}$$

So, for a sufficiently small ε we get $z(s_\varepsilon) \in M$. These arguments show that in our impulsive problem (4.29)–(4.32) we have only two possibilities: either impulsive trajectory has no impulsive points or it has infinitely many ones and inequalities (4.44), (4.46) hold.

Finally, combining the estimate (4.36) with the following one

$$\forall z \in H \ \forall z^+ \in I(z) \quad \|z^+\|_H^2 \leq (1+\mu)^2 \left(\frac{1}{\alpha^2}+\frac{1}{\beta^2}\right) + \|z\|_H^2, \tag{4.47}$$

we can apply standard arguments [16, 31] for differential equations with impulses at fixed moments of time and obtain dissipativity property (4.4) with the set B_0, which does not depend on ε.

Let us prove that G_ε is asymptotically compact. The most arguments with slight changes are the same as in [22], so we give only the sketch of the proof. For an arbitrary solution $z = \begin{pmatrix} u \\ v \end{pmatrix}$ of the problem (4.29) we consider every equation in (4.29) as a linear parabolic equation with right-hand side $h_i(t) = \varepsilon f_i(u(t), v(t))$, $i = 1, 2$. Then, from [36] we deduce, that there exists constant $C_2 > 0$ that depends only on the parameters of the problem (4.29) and does not depends on ε such that for almost all $t > 0$

$$\frac{d}{dt}\left(\|u(t)\|_{H_0^1}^2 + \|v(t)\|_{H_0^1}^2\right) + a_1\left(\|\Delta u(t)\|^2 + \|\Delta v(t)\|^2\right) \leq C_2, \tag{4.48}$$

$$\frac{d}{dt}\left(\|u(t)\|^2 + \|v(t)\|^2\right) + a_1\left(\|u(t)\|_{H_0^1}^2 + \|v(t)\|_{H_0^1}^2\right) \leq C_2. \tag{4.49}$$

From (4.48), (4.49) and Uniform Gronwall Lemma [36] we obtain

$$\forall\, t > 0 \quad \left(\|u(t)\|_{H_0^1}^2 + \|v(t)\|_{H_0^1}^2\right) \leq C_2 t + \frac{\|u(0)\|^2 + \|v(0)\|^2}{a_1 t} + \frac{2C_2}{a_1}. \tag{4.50}$$

Now, let $z_0^{(n)} = \sum_{i=1}^{\infty} \begin{pmatrix} c_i^{(n)} \\ d_i^{(n)} \end{pmatrix} \cdot \psi_i$, $\|z_0^{(n)}\|_H \leq R$ be an arbitrary bounded sequence of initial data, $\xi_n \in G_\varepsilon(t_n, z_0^{(n)})$, $t_n \nearrow +\infty$. Then, $\xi_n = z_n(t_n)$, $z_n \in \tilde{K}^\varepsilon_{z_0^{(n)}}$. If z_n does not have impulsive points, then for function $y_n(t) = z_n(t + t_n - 1)$, $t \geq 0$ we

obtain the following:

$$y_n \in \tilde{K}^{\varepsilon}_{z_n(t_n-1)},\ \xi_n = z_n(t_n) = y_n(1).$$

From (4.36) we obtain that

$$\|z_n(t_n-1)\| \le \sqrt{2} \quad \forall n \ge N(R).$$

Therefore, combining with estimate (4.50), the sequence $\{y_n(1) = \xi_n\}$ is bounded in $H_0^1(\Omega) \times H_0^1(\Omega)$, so it is precompact in H.

Otherwise, without loss of generality, we can assume that $z_0^{(n)} \in IM$, $\|z_0^{(n)}\|_H \le R$ and $z_n = \begin{pmatrix} u_n(\cdot) \\ v_n(\cdot) \end{pmatrix} \in \tilde{K}^{\varepsilon}_{z_0^{(n)}}$ has infinite number of jumps. Let $\{T_{i+1}^{(n)} = \sum_{k=0}^{i} s_k^{(n)}\}_{i=0}^{\infty}$ be the moments of impulsive perturbation for $z_n(\cdot)$ and

$$\{\eta_i^{(n)+} = z_n(T_i^{(n)})\}_{i=1}^{\infty} \subset IM$$

be the corresponding impulsive points. Let us prove the precompactness of the sequence $\{\eta_i^{(n)+}\}$. From the dissipativity condition (4.4), the estimate

$$\bar{s} \le s_k^{(n)} \le \hat{s}, \tag{4.51}$$

and the estimate (4.50) we obtain the existence of constant $C(R)$, which does not depend on ε, such that

$$\forall\, i \ge 1\ \forall\, n \ge 1\ \ \|u_n(T_i^{(n)} - 0)\|^2_{H_0^1} + \|v_n(T_i^{(n)} - 0)\|^2_{H_0^1} \le C(R). \tag{4.52}$$

Then, from (4.32) for all $i \ge 1,\ n \ge 1$ we deduce that

$$\|u_n(T_i^{(n)})\|^2_{H_0^1} + \|v_n(T_i^{(n)})\|^2_{H_0^1} \le C(R) + \lambda(1+\mu)^2\left(\frac{1}{\alpha^2} + \frac{1}{\beta^2}\right). \tag{4.53}$$

The estimate (4.53) and compact embedding $H_0^1(\Omega) \subset L^2(\Omega)$ imply precompactness of the sequence $\{\eta_i^{(n)+} \mid i \ge 1,\ n \ge 1\}$ in H. Then, for the sequence $\xi_n \in G_\varepsilon(t_n, z_0^{(n)})$ for every $n \ge 1$ there exists a number $i = i(n)$, $i(n) \to \infty$, $n \to \infty$ such that

$$t_n \in [T_{i(n)}^{(n)}, T_{i(n)+1}^{(n)}).$$

Hence, from the inclusion

$$\xi_n = z_n(t_n) \in G_\varepsilon(t_n - T^{(n)}_{i(n)}, \eta^{(n)+}_{i(n)}) \tag{4.54}$$

we get that $\xi_n = y_n(\tau_n)$, where $\tau_n := t_n - T^{(n)}_{i(n)}$, $y_n \in K^\varepsilon$ is a sequence of solutions to the (non-perturbed) problem (4.29), where $y_n(0) = \eta^{(n)+}_{i(n)}$. From the previous arguments on some subsequence

$$\eta^{(n)+}_{i(n)} \to \eta \text{ in } H, \ \tau_n \to \tau \in [0, \hat{s}]. \tag{4.55}$$

Hence, from regularity property (4.25) of solutions of the problem (4.29) we deduce the following result:

$$y_n(\tau_n) \to y(\tau) \text{ in } H, \text{ where } y \in K^\varepsilon, \ y(0) = \eta. \tag{4.56}$$

Therefore, sequence $\{\xi_n\}$ is precompact in H and from Lemma 4.1 we deduce the existence of global attractor

$$\Theta_\varepsilon = \bigcap_{s>0} \overline{\bigcup_{t \geq s} G_\varepsilon(t, B_0)}, \tag{4.57}$$

where $B_0 = \{z \in H \mid \|z\| \leq R_0\}$ is dissipative set, which does not depend on ε.

Let us prove the convergence (4.33). For this purpose it is sufficient to show that for $\varepsilon_k \to 0$, $\xi^{(k)} \in \Theta_{\varepsilon_k}$ on subsequence

$$\xi^{(k)} \to \xi \in \Theta \text{ in } H, \ k \to \infty. \tag{4.58}$$

From (4.57) there exist sequences $\{t_k \nearrow \infty\}$, $\{z^0_k\} \subset B_0$, $z_k \in \tilde{K}^{\varepsilon_k}_{z^0_k}$ such that

$$\forall\, k \geq 1 \ \ \|\xi^{(k)} - z_k(t_k)\| \leq \frac{1}{k}.$$

If z_k do not have impulsive perturbations, then using (4.35) we obtain the estimate

$$\forall\, t \geq 0 \ \ \|z_k(t)\|^2_H \leq \|z^0_k\|^2_H e^{-\delta t} + \frac{2\varepsilon_k^2 C^2}{\delta^2}, \tag{4.59}$$

from which it follows that $\xi^{(k)} \to 0$ in H. Now, let us consider the case when every z_k has infinite number of impulsive perturbations. Then, from the previous arguments, for $\xi_k = z_k(t_k)$ we obtain

$$\xi_k = y_k(\tau_k), \ y_k \in K^{\varepsilon_k}_{\eta^+_k}, \ \tau_k = t_k - T^{(k)}_{i(k)} \in [0, s^{(k)}_{i(k)}],$$

$$\tau_k \to \tau, \ \eta^+_k := \eta^{(k)+}_{i(k)} \to \eta, \ i(k) \to \infty, \ k \to \infty.$$

Let us denote $s_k := s^{(k)}_{i(k)}$. Then

$$\alpha e^{-a_1\lambda s_k}(u_0^k, \psi) + \beta e^{-a_2\lambda s_k}(v_0^k, \psi) + \varepsilon_k F_{\varepsilon_k}(s_k) = 1, \tag{4.60}$$

where u_0^k, v_0^k are components of the vector $\eta_k^+ \in IM$. So, for $k \to \infty$ we obtain that on subsequence $s_k \to s$, where $\tau \in [0, s]$ and

$$\alpha e^{-a_1\lambda s}(u_0, \psi) + \beta e^{-a_2\lambda s}(v_0, \psi) = 1, \tag{4.61}$$

$$(u_0, \psi_1) \geq \gamma,\ (v_0, \psi_1) \geq \gamma,\ \alpha(u_0, \psi_1) + \beta(v_0, \psi_1) = 1 + \mu. \tag{4.62}$$

Using (4.25) we obtain

$$\xi_k = \begin{pmatrix} u_k(\tau_k) \\ v_k(\tau_k) \end{pmatrix} \to \xi = y(\tau) = \begin{pmatrix} u(\tau) \\ v(\tau) \end{pmatrix} \text{ in } H, \text{ where } y \in K^\varepsilon,\ y(0) = \eta.$$

Due to (4.37), (4.38) as $k \to \infty$ we obtain

$$(u(\tau), \psi_1) = (u_0, \psi_1)e^{-a_1\lambda_1\tau},\ (v(\tau), \psi_1) = (v_0, \psi_1)e^{-a_2\lambda_1\tau}, \tag{4.63}$$

where $\tau \in [0, s]$, s is a unique root of Eq. (4.61) under fixed u_0, v_0 from (4.62).

Using (4.37), (4.38) and taking into account the "non-impulsive" character of coordinates $j \geq 2$ along each impulsive trajectory, we have:

$$\forall j \geq 2\ \ |(u_0^k, \psi_j)| + |(v_0^k, \psi_j)| \leq 2e^{-\lambda_j T^{(k)}_{i(k)}} + \frac{2C_1\varepsilon_k}{1 - e^{-\lambda_j \bar{s}}} \to 0,\quad k \to \infty. \tag{4.64}$$

Then, from (4.63), (4.64) we obtain that $\xi \in \Theta_\varepsilon$ and (4.33) hold.

Before proving (4.34) according to Lemma 4.5 we need to verify property (4.26). We consider $z_n^0 \to z^0 \notin M$, $z_n \in K_{z_n^0}$ with $s_n := s(z_n) < \infty$. From (4.25) we can assume that $z_n(t) \to z(t)$ in H uniformly on compacts from $[0, +\infty)$. From (4.42) we also can assume that $s_n \to s$. The inclusion $z_n(s_n) \in M$ means that for

$$g_\varepsilon^n(t) := \alpha(u_n(t), \psi) + \beta(v_n(t), \psi)$$

we have

$$g_\varepsilon^n(s_n) = \alpha e^{-a_1\lambda s_n}(u_n(0), \psi) + \beta e^{-a_2\lambda s_n}(v_n(0), \psi) + \varepsilon F_\varepsilon^{(n)}(s_n) = 1,$$

$$e^{-a_1\lambda s_n}(u_n(0), \psi) + \varepsilon F_\varepsilon^{(1n)}(s_n) \geq 0, \tag{4.65}$$

$$e^{-a_2\lambda s_n}(v_n(0), \psi) + \varepsilon F_\varepsilon^{(2n)}(s_n) \geq 0.$$

Passing to the limits we obtain the same relationships for function z. It means that $z(s) \in M$. Since $z^0 \notin M$ so $s > 0$. Let us prove that $s = s(z)$, i.e. let us prove that for trajectory z moment s is the first moment of reaching the impulsive set M. If it is not true then there exists $s_1 \in (0, s)$ such that

$$z(s_1) \in M \text{ and } \forall\, t \in (s_1, s)\ z(t) \notin M.$$

On the other hand, for function $g_\varepsilon(t) := \alpha(u(t), \psi) + \beta(v(t), \psi)$ we have

$$g'_\varepsilon(s) = -a_1\lambda(1 - \varepsilon F_\varepsilon(s)) + (a_1 - a_2)\lambda\beta(v(0), \psi)e^{-a_2\lambda s} + \varepsilon F'_\varepsilon(s).$$

Therefore, from the third inequality in (4.65) for sufficiently small ε

$$g'_\varepsilon(s) \le -\frac{a_1\lambda}{2}. \tag{4.66}$$

In the same way

$$g'_\varepsilon(s_1) \le -\frac{a_1\lambda}{2}. \tag{4.67}$$

So there exists $s_2 \in (s_1, s)$ such that $g_\varepsilon(s_2) = 1$, $g'_\varepsilon(s_2) \ge 0$. On the other hand, for sufficiently small ε

$$g'_\varepsilon(s_2) = -a_1\lambda(1 - \varepsilon F_\varepsilon(s_2))$$

$$+(a_1 - a_2)\lambda\beta(v(0), \psi)e^{-a_2\lambda s_1}e^{-a_2\lambda(s_2 - s_1)} + \varepsilon F'_\varepsilon(s_2) \le -\frac{a_1\lambda}{2}$$

and we obtain a contradiction.

Now let us prove (4.34). Let $\xi \in \Theta_\varepsilon \setminus M$, $t > 0$ be fixed. Then $\xi = \lim \xi_n$, where $\xi_n = \varphi_n(t_n)$, $t_n \nearrow \infty$, $\varphi_n \in \tilde{K}_{x_n}$, $\{x_n\} \subset B_0$. We will assume that $t_n \ge T(R_0)$, where number $T(R_0)$ is taken from (4.42). For sufficiently large n we can consider functions

$$\psi_n(p) = \varphi_n(p + t_n - t),\ \ p \ge 0.$$

From Lemma 4.1 G is asymptotically compact. Therefore, up to subsequence

$$y_n := \varphi_n(t_n - t) \to y \in \Theta_\varepsilon.$$

Hence

$$\psi_n \in \tilde{K}_{y_n},\ \ y_n \notin M,\ \ y_n \to y,\ \ \xi_n = \psi_n(t).$$

If φ_n has no impulsive points then $\varphi_n \in K_{x_n}$ and from estimate (4.59) for sufficiently small ε

$$\|y_n\| \le \frac{1}{2(\alpha+\beta)}, \ \|y\| \le \frac{1}{2(\alpha+\beta)}.$$

In particular, $y_n, y \notin M$ and every trajectory starting from y_n and y is non-impulsive. Therefore,

$$\xi_n = \psi_n(t) \to \xi = \psi(t) \in G_\varepsilon(t, y) \subset G_\varepsilon(t, \Theta_\varepsilon \setminus M).$$

In other case we have two cases: $y \notin M$ and $y \in M$.

If $y \notin M$, then from Lemma 4.5 $\exists \psi \in \tilde{K}_y$, $\exists \eta_n \to 0+$ such that

$$\psi_n(t+\eta_n) \to \psi(t).$$

On the other hand, for $\theta_n(p) = \psi_n(p+t)$, $p \ge 0$ we have

$$\theta_n \in \tilde{K}_{\xi_n}, \ \xi_n \to \xi \notin M, \ \eta_n \to 0+.$$

Therefore, from Lemma 4.5

$$\theta_n(\eta_n) = \psi_n(t+\eta_n) \to \xi.$$

So,

$$\xi = \psi(t) \in G_\varepsilon(t, \Theta_\varepsilon \setminus M).$$

Let $y \in M$. For further arguments we need to verify the following condition:

$$\text{If for } x_n \to x \in \Theta_\varepsilon \cap M, \ \varphi_n \in K_{x_n} \text{ we have } s(\varphi_n) < \infty$$

then up to subsequence

$$s(\varphi_n) \to 0. \tag{4.68}$$

So let us consider $z_n^0 \to z^0 \in M$, $z_n \in K_{z_n^0}$ with $s_n := s(z_n) < \infty$. From (4.25) we have that $z_n(t) \to z(t)$ in H uniformly on compacts from $[0, +\infty)$. From (4.42) we also have that $s_n \to s \ge 0$. Let us prove that $s = 0$. Indeed, if $s > 0$ then passing to the limit in (4.65) and take into account that $z^0 \in M$ we obtain

$$g_\varepsilon(s) = \alpha e^{-a_1 \lambda s}(u(0), \psi) + \beta e^{-a_2 \lambda s}(v(0), \psi) + \varepsilon F_\varepsilon(s) = 1,$$

$$g_\varepsilon(0) = \alpha(u(0), \psi) + \beta(v(0), \psi) = 1, \ (u(0), \psi) \ge 0, \ (v(0), \psi) \ge 0.$$

Then for sufficiently small ε

$$s \leq T := \frac{1}{a_1 \lambda} \ln 2.$$

On the other hand, for $t \in [0, s]$

$$g_\varepsilon'(t) = -a_1 \lambda \alpha e^{-a_1 \lambda t}(u(0), \psi) - a_1 \lambda \beta e^{-a_2 \lambda t}(v(0), \psi) + \varepsilon F_\varepsilon'(t)$$

$$\leq -a_1 \lambda \alpha e^{-a_2 \lambda t} + \varepsilon F_\varepsilon'(t).$$

Therefore, for sufficiently small ε for $t \in [0, s]$ $g_\varepsilon'(t) < 0$ and we have a contradiction.

So, for $\psi_0^{(n)}$—the first component of ψ_n, we have that up to subsequence

$$\tau_n := s(\psi_0^{(n)}) \to 0, \quad z_n := \psi_0^{(n)}(\tau_n) \to y.$$

Then semi-continuity of impulsive map I follows

$$\psi_n(\tau_n) = z_n^+ \in I z_n, \ z_n^+ \to z^+ \in I y.$$

But $\psi_n(\tau_n) = \varphi_n(\tau_n + t_n - t)$, therefore $z^+ \in \Theta_\varepsilon \setminus M$.

Now we consider functions $\beta_n(p) = \psi_n(p + \tau_n)$, $p \geq 0$. Then

$$\beta_n \in \tilde{K}_{(z_n)^+}, \ (z_n)^+ \to z^+ \notin M,$$

so from Lemma 4.5 $\exists \beta \in \tilde{K}_{z^+}$ $\exists \eta_n \to 0+$ such that on some subsequence:

$$\beta_n(t + \eta_n) = \psi_n(t + \eta_n + \tau_n) \to \beta(t) \in G(t, \Theta_\varepsilon \setminus M).$$

But $\eta_n + \tau_n \to 0+$, so

$$\beta(t) = \xi \in G_\varepsilon(t, \Theta_\varepsilon \setminus M).$$

Thus we obtain

$$\forall t \geq 0 \ \ \Theta_\varepsilon \setminus M \subseteq G_\varepsilon(t, \Theta_\varepsilon \setminus M).$$

Let us prove equality. From the strictness of G_ε we have that $\forall s \geq 0$

$$G_\varepsilon(s, \Theta_\varepsilon \setminus M) \subset G_\varepsilon(t + s, \Theta_\varepsilon \setminus M) \subset O_\delta(\Theta_\varepsilon)$$

for arbitrary small $\delta > 0$ if t is large enough. So,

$$G_\varepsilon(s, \Theta_\varepsilon \setminus M) \subset \Theta_\varepsilon \setminus M,$$

which implies the required result.

Theorem is proved. □

Acknowledgements This work was partially supported by the German Academic Exchange Service (DAAD). Oleksiy Kapustyan was partially supported by the State Fund For Fundamental Research, Grant of President of Ukraine.

References

1. Akhmet, M.: Principles of Discontinuous Dynamical Systems. Springer, Berlin (2010)
2. Ball, J.M.: Continuity properties and attractors of generalized semiflows and the Navier-Stokes equations. J. Nonlinear Sci. **7**(5), 475–502 (1997)
3. Bonotto, E.M.: Flows of characteristic 0+ in impulsive semidynamical systems. J. Math. Anal. Appl. **332**, 81–96 (2007)
4. Bonotto, E.M., Demuner, D.P.: Attractors of impulsive dissipative semidynamical systems. Bull. Sci. Math. **137**, 617–642 (2013)
5. Bonotto, E.M., Bortolan, M.C., Carvalho, A.N., Czaja, R.: Global attractors for impulsive dynamical systems – a precompact approach. J. Differ. Equ. **259**, 2602–2625 (2015)
6. Bonotto, E.M., Bortolan, M.C., Collegari, R., Czaja, R.: Semicontinuity of attractors for impulsive dynamical systems. J. Differ. Equ. **261**, 4338–4367 (2016)
7. Chepyzhov, V.V., Vishik, M.I.: Attractors of Equations of Mathematical Physics. Colloquium Publications, vol. 49. American Mathematical Society, Providence (2002)
8. Ciesielski, K.: On stability in impulsive dynamical systems. Bull. Pol. Acad. Sci. Math. **52**, 81–91 (2004)
9. Dashkovskiy, S., Feketa, P.: Input-to-state stability of impulsive systems and their interconnections. Nonlinear Anal. Hybrid Syst. **26**, 190–200 (2017)
10. Dashkovskiy, S., Mironchenko, A.: Input-to-state stability of nonlinear impulsive systems. SIAM J. Control Optim. **51**(3), 1962–1987 (2013)
11. Dashkovskiy, S., Kapustyan, O., Romanjuk, I.: Global attractors of impulsive parabolic inclusions. Discrete Contin. Dyn. Syst. Ser. B **22**(5), 1875–1886 (2017)
12. Dashkovskiy, S., Feketa, P., Kapustyan, O., Romaniuk, I.: Invariance and stability of global attractors for multi-valued impulsive dynamical systems. J. Math. Anal. Appl. **458**(1), 193–218 (2018)
13. Feketa, P., Bajcinca, N.: Stability of nonlinear impulsive differential equations with non-fixed moments of jumps. In: Proceedings of 17th European Control Conference, Limassol, Cyprus, 900–905 (2018)
14. Feketa, P., Perestyuk, Yu.: Perturbation theorems for a multifrequency system with pulses. J. Math. Sci. (N.Y.) **217**(4), 515–524 (2016)
15. Gorban, N.V., Kapustyan, O.V., Kasyanov, P.O.: Uniform trajectory attractor for non-autonomous reactiondiffusion equations with Caratheodorys nonlinearity. Nonlinear Anal. **98**, 13–26 (2014)
16. Iovane, G., Kapustyan, O.V., Valero, J.: Asymptotic behaviour of reaction-diffusion equations with non-damped impulsive effects. Nonlinear Anal. **68**, 2516–2530 (2008)
17. Kapustyan, A.V.: Global attractors of non-autonomous reaction-diffusion equation. Diff. Equ. **38**, 1467–1471 (2002)

18. Kapustyan, A.V., Melnik, V.S.: On global attractors of multivalued semidynamical systems and their approximations. Dokl. Akad. Nauk. **366**(4), 445–448 (1999)
19. Kapustyan, O.V., Shkundin, D.V.: Global attractor of one nonlinear parabolic equation. Ukr. Math. J. **55**(4), 446–455 (2003)
20. Kapustyan, O.V., Kasyanov, P.O., Valero, J.: Pullback attractors for some class of extremal solutions of 3D Navier-Stokes system. J. Math. Anal. Appl. **373**, 535–547 (2011)
21. Kapustyan, O.V., Kasyanov, P.O., Valero, J.: Regular solutions and global attractors for reaction-diffusion systems without uniqueness. Commun. Pure Appl. Anal. **13**, 1891–1906 (2014)
22. Kapustyan, O., Perestyuk, M., Romaniuk, I.: Global attractor of weakly nonlinear parabolic system with discontinuous trajectories. Mem. Differ. Equ. Math. Phys. **72**, 59–70 (2017)
23. Kasyanov, P.O.: Multivalued dynamics of solutions of autonomous differential-operator inclusion with pseudomonotone nonlinearity. Cybern. Syst. Anal. **47**(5), 800–811 (2011)
24. Kaul, S.K.: On impulsive semidynamical systems. J. Math. Anal. Appl. **150**(1), 120–128 (1990)
25. Kaul, S.K.: Stability and asymptotic stability in impulsive semidynamical systems. J. Appl. Math. Stoch. Anal. **7**(4), 509–523 (1994)
26. Lakshmikantham, V., Bainov, D.D., Simeonov, P.S.: Theory of Impulsive Differential Equations. World Scientific, Singapore (1989)
27. Melnik, V.S.: Families of multi-valued semiflows and their attractors. Dokl. Math. **55**, 195–196 (1997)
28. Melnik, V.S., Valero, J.: On attractors of multi-valued semi-flows and differential inclusions. Set-Valued Var. Anal. **6**, 83–111 (1998)
29. Perestyuk, M.O., Feketa, P.V.: Invariant manifolds of one class of systems of impulsive differential equations. Nonlinear Oscil. **13**(2), 260–273 (2010)
30. Perestyuk, M., Feketa, P.: Invariant sets of impulsive differential equations with particularities in ω-limit set. Abstr. Appl. Anal. **2011**, ID 970469, 14 pp. (2011)
31. Perestyuk, M.O., Kapustyan, O.V.: Long-time behavior of evolution inclusion with non-damped impulsive effects. Mem. Differ. Equ. Math. Phys. **56**, 89–113 (2012)
32. Perestyuk, M.O., Kapustyan, O.V.: Global attractors of impulsive infinite-dimensional systems. Ukr. Math. J. **68**(4), 517–528 (2016)
33. Pichkur, V.V., Sasonkina, M.S.: Maximum set of initial conditions for the problem of weak practical stability of a discrete inclusion. J. Math. Sci. **194**, 414–425 (2013)
34. Rozko, V.: Stability in terms of Lyapunov of discontinuous dynamic systems. Differ. Uravn. **11**(6), 1005–1012 (1975)
35. Samoilenko, A.M., Perestyuk, N.A.: Impulsive Differential Equations. World Scientific, Singapore (1995)
36. Temam, R.: Infinite-Dimensional Dynamical Systems in Mechanics and Physics. Springer, Berlin (1988)
37. Zgurovsky, M.Z., Kasyanov, P.O., Kapustyan, O.V., Valero, J., Zadoianchuk, N.V.: Evolution Inclusions and Variation Inequalities for Earth Data Processing III. Long-Time Behavior of Evolution Inclusions Solutions in Earth Data Analysis. Springer, Berlin, 330 pp. (2012)

Chapter 5
Fraktal and Differential Properties of the Inversor of Digits of Q_s-Representation of Real Number

Oleg Barabash, Oleg Kopiika, Iryna Zamrii, Valentyn Sobchuk, and Andrey Musienko

Abstract This paper aims at introducing and studying a continuous function $I(x)$ that depends on the $s-1$ parameters, $I(x)$ is called inversor of digits of Q_s-representation of real number. This representation is determined by stochastic vector $(q_0, q_1, \ldots, q_{s-1})$ with positive entries and for an arbitrary $x \in [0; 1]$ there exists a sequence (α_n), $\alpha_n \in \{0, 1, \ldots, s-1\} \equiv A_s$, such that

$$x = \beta_{\alpha_1} + \sum_{k=2}^{\infty} \left[\beta_{\alpha_k} \prod_{j=1}^{k-1} q_{\alpha_j} \right] = \Delta^{Q_s}_{\alpha_1 \alpha_2 \ldots \alpha_n \ldots},$$

where $\beta_0 = 0$, $\beta_k = \sum_{i=0}^{k-1} q_i$, it is generalization of the classical s-adic representation (because it coincides with the last-mentioned if $q_i = \frac{1}{s}$, $i \in A_s$).

The differential and fractal properties of the inversor of digits of Q_s-representation of real number are described.

O. Barabash (✉) · I. Zamrii
State University of Telecommunications, Kyiv, Ukraine
e-mail: bar64@ukr.net

O. Kopiika
Institute of Telecommunications and Global Information Space, Kyiv, Ukraine
e-mail: kopiika@nas.gov.ua

V. Sobchuk
East-European National University of Lesya Ukrainka, Lutsk, Ukraine

A. Musienko
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_5

5.1 Introduction

Metric space $C_{[0;1]}$ of continuous functions with a uniform metric rich in functions with a complex local structure and "massive" sets of various features. These include continuous nonmonotonous, twisting, non-differential functions (in particular, the Banakh-Mazurkevych's (1931) and Kozyriev's (1983) theorems), which determine the topological wealth of their families (which are sets of the second category of Ber), as well as singular functions (a continuous function, which is different from the constant, is called singular if it is equal to zero practically everywhere according to Lebesgue measure). The set of all singular functions is topologically massive, as evidenced by Zamfirescu's (1981) theorem: singular functions in the metric space of all continuous monotone functions with a supremummetric form the sets of the second category of Ber.

A well-known theorem of Lebesgue asserts that every real function of a real variable with limited variation is either a jump function, or an absolutely continuous function, or a singular or nontrivial combination of the preceding three pure types of Lebesgue. In addition, each continuous function is either completely continuous, or singular, or summation. Among the pure types of Lebesgue, the singular functions are the least studied class, although this has been a separate topic of research for more than 100 years. The first researchers who have introduced the singular functions: G. Cantor, E. Hellinger, W. Sierpinski, H. Minkowski and others. The proposed constructions concern monotonic and even strictly increasing functions (with the exception of the Cantor function is the non-decreasing function of the "uniform" probability distribution on the Cantor sets, which grows exclusively at the points of this set). There were other very interesting examples of singularly strictly increasing functions, for example, proposed by the authors of Freilich [4], Gelbaum and Olmsted [5, pp. 96–98], Hewitt and Stromberg [6, pp. 278–282], Riesz and Nagy [21, pp. 48–49], Takacs [23]. In recent years, the desire to study singular functions and to apply them in various fields of science also has not disappeared, in support of this work [1–3, 8–10, 13, 17, 19, 20, 24, 25, 28]. But the impression is that for a long time there was an unpublished competition for constructing the simplest example of a singular function. Once thought that this is an example of Salem [22]. But, in our opinion, a more simple examples are the inversor of digits for Q_2-representation of real number [16], and the inversor of digits for Q_3-representation of real number [27], which we are generalizing in this paper.

Monotone singular functions are closely related to singular probability distributions (these distributions are concentrated on zero sets of Lebesgue). Domination of such distributions is convincingly proved in different classes of random variables, whose images in one or another numerical encoding system are independent.

There are a number of problems associated with singular functions, one of which is the problem of effective methods for their definition and research. Recently, various systems of representation of real numbers are used for this purpose, both with a finite, and with an infinite alphabet, one of which is the Q-representation of numbers first introduced in 1986 by M.V. Pratsiovytyi. It was used to study singular

distribution functions. We use Q-representation of numbers to study monotonic singular functions, too, but we participate in the fractal part of the study.

The fractal geometry from a group point of view is a theory of invariants of a group of transformations of the metric space, which retain fractal dimension, in particular, Hausdorff-Besicovitch. Its useful ideas are the ideas of self-similarity, self-affinity, and so on. Graphs of functions with complex local structure as plural of space R^2 potentially have fractal properties.

5.2 Q_s-Representation of Real Number

The encoding of numbers with [0, 1] means of the alphabet A (finite or infinite) is called surjective transformation

$$\varphi : A \times A \times \ldots \times A \times \ldots \equiv L \to [0, 1].$$

In this case, the symbolic entry of the number $x = \varphi((a_n))$, which is an direct image of the sequence (a_n), in the form $\Delta^{\varphi}_{a_1 a_2 \ldots a_n \ldots}$ is called its φ-representation (or φ-code).

One of the easiest ways to encode numbers $x \in [0, 1]$ by the means of the alphabet A_s, $2 \leq s \in N$ is the s-adic representation of number:

$$x = \frac{\alpha_1}{s} + \frac{\alpha_2}{s^2} + \ldots + \frac{\alpha_n}{s^n} + \ldots \equiv \Delta^{s}_{\alpha_1 \alpha_2 \ldots \alpha_n \ldots}.$$

In the traditional sense, the geometry of numbers involves the solution of theoretical numerical problems using geometric means. In recent years, a new field of research has shown itself that is the geometry of various representation of real numbers, which describes the geometric change of numbers, metric relations, topological-metric properties number sets, determined by conditions on their representation, etc. and applications to the construction of different mathematical objects and complex (non-homogeneous) local structure [12].

Definition 5.1 Let $(c_1, c_2, \ldots, c_m)$ be a fixed ordered set of elements of the alphabet A. The cylinder of rank m with the basis of $c_1 c_2 \ldots c_m$ in encoding φ is the set $\Delta^{\varphi}_{c_1 c_2 \ldots c_m}$ of all numbers $x \in [0, 1]$, which have the following φ-representation

$$x = \Delta^{\varphi}_{c_1 c_2 \ldots c_m \alpha_{m+1} \ldots \alpha_{m+k} \ldots}, \ \alpha_{m+i} \in A.$$

The segment [0, 1] itself is called the cylinder of zero rank and denoted by Δ.

Directly from this definition, the following properties of the cylinders follow:

1. $\Delta^{\varphi}_{c_1 c_2 \ldots c_m} = \bigcup\limits_{i \in A} \Delta^{\varphi}_{c_1 c_2 \ldots c_m i}$;
2. $\Delta = \bigcup\limits_{i_1 \in A} \bigcup\limits_{i_2 \in A} \ldots \bigcup\limits_{i_n \in A} \Delta^{\varphi}_{i_1 i_2 \ldots i_n}$ for every n.

For the s-adic representation of real numbers of cylinders there are segments, namely:

$$\Delta^s_{c_1c_2\ldots c_m} = \left[\sum_{i=1}^{m}\frac{c_i}{s^i}; \frac{1}{s^m} + \sum_{n=m+1}^{\infty}\frac{c_n}{s^n}\right].$$

It is said that the coding system has zero redundancy if all numbers or the overwhelming majority of numbers have a single representation and only a small part of them has two representation.

Examples of encoding numbers by means of an infinite alphabet are L-, E-, and S-representation, which are based on the expanding numbers in the series of positive terms of Luroth [7, 11, 29], Engel [15], Silvester [18], respectively. As with s-adic representation, they all have zero redundancy, since each number with (0, 1] has a single L-, E-, and S-representation.

Coding is called continuous if the cylinder is an open interval (a segment, a half-interval, or a half-open interval), and for any sequence (a_n), $a_n \in A$, intersection $\bigcap\limits_{m=1}^{\infty} \Delta^{\varphi}_{a_1a_2\ldots a_m} \equiv \Delta^{\varphi}_{a_1a_2\ldots a_m\ldots}$ is a number (point), and moreover

$$\Delta^{\varphi}_{a_1a_2\ldots a_m\ldots} = x \to x' = \Delta^{\varphi}_{a'_1a'_2\ldots a'_m\ldots} \equiv m \to \infty,$$

where $a_m \neq a'_m$, but $a_i = a'_i$ for $i < m$.

Continuous φ-representation is called Q-representation if the alphabet A is finite and for every $i \in A \equiv A_s$ performed $\sup \Delta^{Q}_{c_1c_2\ldots c_m i} = \inf \Delta^{Q}_{c_1c_2\ldots c_m[i+1]}$, and metric relationship

$$\frac{|\Delta^{\varphi}_{c_1c_2\ldots c_m i}|}{|\Delta^{\varphi}_{c_1c_2\ldots c_m}|} \equiv q_i = const,$$

which is called the main (most important in the metric theory).

A meaningful introduction of a Q_s-representation gives the following statement [14]: for any $x \in [0, 1]$ there is a sequence (α_n), $\alpha_n \in A_s$, such that

$$x = \beta_{\alpha_1} + \sum_{k=2}^{\infty}\left[\beta_{\alpha_k}\prod_{j=1}^{k-1} q_{\alpha_j}\right] = \Delta^{Q}_{\alpha_1\alpha_2\ldots\alpha_n\ldots}, \tag{5.1}$$

where $\beta_0 = 0$, $\beta_i = q_0 + q_1 + \ldots + q_{i-1}$, $0 < i \leq s$.

For Q_s-representation have the following relationship:

3. $q_0 + q_1 + \ldots + q_{s-1} = 1$;
4. $|\Delta^{Q}_{c_1c_2\ldots c_m}| = \prod\limits_{i=1}^{m} q_{c_i} \to 0$ $(m \to \infty)$.

The classic s-adic representation is a Q_s-representation if $q_0 = q_1 = \ldots = q_{s-1} = \frac{1}{s}$.

The period in Q_s-representation of the number (if it exists) is denoted by the parentheses. There are numbers that have two Q_s-representation. These are numbers with period (0) or $(s-1)$, moreover $\Delta^{Q_s}_{c_1\ldots c_{m-1}c_m(0)} = \Delta^{Q_s}_{c_1\ldots c_{m-1}[c_m-1](s-1)}$. These numbers are called Q_s-rational, set of Q_s-rational numbers is countable. The rest of the numbers are called Q_s-irrational.

5.3 Inversor of Digits of Q_s-Representation for Fractional Part of Real Number

Definition 5.2 Function defined [0, 1] by equality

$$I(\Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\ldots}) = \Delta^{Q_s}_{[s-1-\alpha_1][s-1-\alpha_2]\ldots[s-1-\alpha_n]\ldots} \tag{5.2}$$

is called an inversor I of digits the Q_s-representation of a real number $x = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\ldots}$ (or simply inversor).

The function is correctly defined in each Q_s-rational points, and hence at each point of the segment [0, 1].

Theorem 5.1 *For inversor I of digits the Q_s-representation of a real number $x \in [0, 1]$ there is equality*

$$\Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\ldots} + \Delta^{Q'_s}_{[s-1-\alpha_1][s-1-\alpha_2]\ldots[s-1-\alpha_n]\ldots} = 1,$$

where $Q'_s = \{q'_0 = q_{s-1}; q'_1 = q_{s-2}; \ldots; q'_{s-1} = q_0\}$.

Proof For the numbers $0 = \Delta^{Q_s}_{(0)}$ and $1 = \Delta^{Q_s}_{(s-1)}$, the assertion is obvious.

Let $x = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\ldots}$ be some number belongs to (0; 1) and $1 - x \equiv x'$.

Let $\Delta^{Q_s}_{c_1c_2\ldots c_m} \equiv \Delta^{Q'_s}_{c'_1c'_2\ldots c'_m}$, where $c'_i = s - 1 - c_i$.

Since x belongs to the system of embedded cylinders $\Delta^{Q_s}_{\alpha_1}, \Delta^{Q_s}_{\alpha_1\alpha_2}, \ldots, \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n}, \ldots$, then $x' = 1 - x$ belongs to the system of embedded cylinders $\Delta^{Q'_s}_{\alpha'_1}, \Delta^{Q'_s}_{\alpha'_1\alpha'_2}, \ldots, \Delta^{Q'_s}_{\alpha'_1\alpha'_2\ldots\alpha'_n}, \ldots$.

Then

$$x' = \bigcap_{n=1}^{\infty} \Delta^{Q'_s}_{\alpha'_1\alpha'_2\ldots\alpha'_n} = \bigcap_{n=1}^{\infty} \Delta^{Q'_s}_{[s-1-\alpha_1][s-1-\alpha_2]\ldots[s-1-\alpha_n]} = \Delta^{Q'_s}_{[s-1-\alpha_1][s-1-\alpha_2]\ldots[s-1-\alpha_n]\ldots},$$

which was required to prove.

Corollary 5.1 $I(\Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\ldots}) = 1 - \Delta^{Q'_s}_{\alpha_1\alpha_2\ldots\alpha_n\ldots}$.

Corollary 5.2 $I(x) = 1 - x$ *for all* $x \in [0, 1]$, *when* $q_i = q_{s-1-i}$.

Proof If $q_i = q_{s-1-i}$ then $Q_s = Q'_s$ and $\Delta^{Q'_s}_{\alpha_1\alpha_2\ldots\alpha_n\ldots} = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\ldots}$. So

$$I(\Delta^{Q_s}_{\alpha_1(x)\alpha_2(x)\ldots\alpha_n(x)\ldots}) = 1 - \Delta^{Q'_s}_{\alpha_1(x)\alpha_2(x)\ldots\alpha_n(x)\ldots} = 1 - x.$$

Let $q_i \neq q_{s-1-i}$ and $x = \Delta^{Q_s}_{2(s-1)}$ then

$$\begin{aligned} I(x) = I(\Delta^{Q_s}_{2(s-1)}) &= \beta_{s-3} + \beta_0 q_{s-3} + \beta_0 q_{s-3} q_0 + \ldots \\ &= \beta_{s-3} = q_0 + q_1 + \ldots + q_{s-4}, \\ 1 - x = 1 - \Delta^{Q_s}_{2(s-1)} &= 1 - \beta_2 - \beta_{s-1} q_2 - \beta_{s-1} q_2 q_{s-1} - \ldots \\ &= 1 - \beta_2 - \frac{\beta_{s-1} q_2}{1 - q_{s-1}} \neq I(x). \end{aligned}$$

5.4 Differential Properties of the Inversor

The property Φ of the element x in the set K is called *normal* if the overwhelming majority of the elements of this set (almost all) have it.

Such concepts as potency, measure, Hausdorff-Besicovitch's dimension, the category of Ber and so others [14], allow to interpret unambiguously the words "almost all".

Let $N_i(x, k)$ be the number of digits i in the Q_s-representation x to the k-th place inclusive. Then the lim (if it exists) $\lim\limits_{k\to\infty} \dfrac{N_i(x,k)}{k} \equiv \nu_i(x) \equiv \nu_i^{Q_s}(x)$ is called the *frequency (asymptotic frequency) of the numbers* i in the Q_s-representation of the number x.

The number $x = \Delta^{Q_s}_{\alpha_1\ldots\alpha_k\ldots}$ for which the equations are valid

$$\nu_0^{Q_s}(x) = q_0, \quad \nu_1^{Q_s}(x) = q_1, \quad \ldots, \quad \nu_{s-1}^{Q_s}(x) = q_{s-1}$$

called Q_s*-normal*. It is known that the Lebesgue measure of the set of all Q_s-normal numbers is $[0, 1]$ equal to 1.

Lemma 5.1 *If $I'(x)$ exists then:*

1) if s is an even number, the derivative equals

$$I'(x)=\lim_{n\to\infty}\left(\left(\frac{q_0}{q_{s-1}}\right)^{\nu_{s-1}(x)-\nu_0(x)}\cdot\left(\frac{q_1}{q_{s-2}}\right)^{\nu_{s-2}(x)-\nu_1(x)}\cdot\ldots\cdot\left(\frac{q_{\frac{s}{2}-1}}{q_{\frac{s}{2}}}\right)^{\nu_{\frac{s}{2}}(x)-\nu_{\frac{s}{2}-1}(x)}\right)^n; \tag{5.3}$$

2) if s is an odd number, then the derivative equals

$$I'(x)=\lim_{n\to\infty}\left(\left(\frac{q_0}{q_{s-1}}\right)^{\nu_{s-1}(x)-\nu_0(x)}\cdot\left(\frac{q_1}{q_{s-2}}\right)^{\nu_{s-2}(x)-\nu_1(x)}\cdot\ldots\cdot\left(\frac{q_{\frac{s-3}{2}}}{q_{\frac{s+1}{2}}}\right)^{\nu_{\frac{s+1}{2}}(x)-\nu_{\frac{s-3}{2}}(x)}\right)^n. \tag{5.4}$$

Proof Obviously, the derivative equals $I'(x)=\lim\limits_{n\to\infty}\dfrac{\mu_I\left(\Delta^{Q_s}_{\alpha_1(x)\alpha_2(x)\ldots\alpha_n(x)}\right)}{|\Delta^{Q_s}_{\alpha_1(x)\alpha_2(x)\ldots\alpha_n(x)}|}$.

$$I'(x)=\lim_{n\to\infty}\frac{\mu_I\left(\Delta^{Q_s}_{\alpha_1(x)\alpha_2(x)\ldots\alpha_n(x)}\right)}{|\Delta^{Q_s}_{\alpha_1(x)\alpha_2(x)\ldots\alpha_n(x)}|}=\lim_{n\to\infty}\frac{\prod\limits_{i=1}^{n}q_{[s-1-\alpha_i]}}{\prod\limits_{i=1}^{n}q_{\alpha_i}}$$

$$=\lim_{n\to\infty}\frac{q_0^{N_{s-1}(x,n)}q_1^{N_{s-2}(x,n)}\ldots q_{s-1}^{N_0(x,n)}}{q_0^{N_0(x,n)}q_1^{N_1(x,n)}\ldots q_{s-1}^{N_{s-1}(x,n)}}$$

$$=\lim_{n\to\infty}\left(\left(\frac{q_0}{q_{s-1}}\right)^{\frac{N_{s-1}(x,n)-N_0(x,n)}{n}}\cdot\left(\frac{q_1}{q_{s-2}}\right)^{\frac{N_{s-2}(x,n)-N_1(x,n)}{n}}\cdot\ldots\right)^n.$$

From the definition of the frequency of the numbers i in the Q_s-representation for an even number s, we obtain

$$I'(x)=\lim_{n\to\infty}\left(\left(\frac{q_0}{q_{s-1}}\right)^{\nu_{s-1}(x)-\nu_0(x)}\cdot\left(\frac{q_1}{q_{s-2}}\right)^{\nu_{s-2}(x)-\nu_1(x)}\cdot\ldots\cdot\left(\frac{q_{\frac{s}{2}-1}}{q_{\frac{s}{2}}}\right)^{\nu_{\frac{s}{2}}(x)-\nu_{\frac{s}{2}-1}(x)}\right)^n;$$

for an odd number s, we obtain

$$I'(x)=\lim_{n\to\infty}\left(\left(\frac{q_0}{q_{s-1}}\right)^{\nu_{s-1}(x)-\nu_0(x)}\cdot\left(\frac{q_1}{q_{s-2}}\right)^{\nu_{s-2}(x)-\nu_1(x)}\cdot\ldots\cdot\left(\frac{q_{\frac{s-3}{2}}}{q_{\frac{s+1}{2}}}\right)^{\nu_{\frac{s+1}{2}}(x)-\nu_{\frac{s-3}{2}}(x)}\right)^n.$$

What should have been proved.

Theorem 5.2 *The derivative $I'(x_0)$ does not exist if:*

1) *x_0 is a Q_s-rational points and $q_0 \neq q_{s-1}$;*
2) *$x_0 = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n(j)}$, where $0 \neq j \neq s-1$ and $q_i \neq q_{[s-1-i]}$ to all $i \in A_s$;*
3) *there exists frequencies $\nu_0(x_0), \nu_1(x_0), \ldots, \nu_{s-1}(x_0)$ of digits $0, 1, \ldots, s-1$ in the Q_s-representation of the number x_0 and*

$$\begin{cases} \begin{cases} (\nu_{s-1}(x_0) - \nu_0(x_0))(q_0 - q_{s-1}) < 0, \\ (\nu_{s-2}(x_0) - \nu_1(x_0))(q_1 - q_{s-2}) < 0, \\ \ldots, \\ (\nu_{\frac{s}{2}}(x_0) - \nu_{\frac{s}{2}-1}(x_0))(q_{\frac{s}{2}-1} - q_{\frac{s}{2}}) < 0, \end{cases} & \text{if } s \text{ is an even number;} \\ \begin{cases} (\nu_{s-1}(x_0) - \nu_0(x_0))(q_0 - q_{s-1}) < 0, \\ (\nu_{s-2}(x_0) - \nu_1(x_0))(q_1 - q_{s-2}) < 0, \\ \ldots, \\ (\nu_{\frac{s+1}{2}}(x_0) - \nu_{\frac{s-3}{2}}(x_0))(q_{\frac{s-3}{2}} - q_{\frac{s+1}{2}}) < 0, \end{cases} & \text{if } s \text{ is an odd number.} \end{cases}$$

Proof

1) Let $x_0 \in [0; 1]$ be some Q_s-rational number

$$x_0 = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n[\alpha_{n+1}-1](s-1)} \equiv \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\alpha_{n+1}(0)} = x_0',$$

where is the digit $\alpha_{n+1} \neq 0$.

Choose the sequence (x_m) such that $x_m = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots[\alpha_{n+1}-1]\underbrace{s-1\ldots s-1}_{m}(0)}$.

Obviously, $x_m \to x_0$ for $m \to \infty$. Then

$$\lim_{m\to\infty} \frac{I(x_0) - I(x_m)}{x_0 - x_m}$$

$$= \lim_{m\to\infty} \frac{\Delta^{Q_s}_{[s-1-\alpha_1]\ldots[s-1-\alpha_n][s-1-\alpha_{n+1}+1](0)} - \Delta^{Q_s}_{[s-1-\alpha_1]\ldots[s-1-\alpha_n][s-1-\alpha_{n+1}+1]\underbrace{0\ldots0}_{m}(s-1)}}{q_{[\alpha_{n+1}-1]}q_{s-1}^m \prod_{i=1}^{n} q_{\alpha_i}(\beta_{s-1} + \beta_{s-1}q_{s-1} + \beta_{s-1}q_{s-1}^2 + \ldots - \beta_0 - \beta_0 q_0 - \beta_0 q_0^2 - \ldots)}$$

$$= \lim_{m\to\infty} \frac{q_0^m q_{[s-\alpha_{n+1}]} \prod_{i=1}^{n} q_{[s-1-\alpha_i]}(0 - \beta_{s-1} - \beta_{s-1}q_{s-1} - \beta_{s-1}q_{s-1}^2 - \ldots)}{q_{[\alpha_{n+1}-1]}q_{s-1}^m \prod_{i=1}^{n} q_{\alpha_i} \dfrac{\beta_{s-1}}{1 - q_{s-1}}}$$

$$= - \lim_{m\to\infty} \frac{q_{[s-\alpha_{n+1}]}}{q_{[\alpha_{n+1}-1]}} \left(\frac{q_0}{q_{s-1}}\right)^m \prod_{i=1}^{n} \frac{q_{[s-1-\alpha_i]}}{q_{\alpha_i}} = w_1.$$

Choose the sequence (x'_m) such that $x'_m = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\alpha_{n+1}\underbrace{0\ldots0}_{m}(s-1)}$. Obviously, $x'_m \to x'_0$ for $m \to \infty$ and

$$\lim_{m\to\infty} \frac{I(x'_0) - I(x'_m)}{x'_0 - x'_m}$$

$$= \lim_{m\to\infty} \frac{\Delta^{Q_s}_{[s-1-\alpha_1]\ldots[s-1-\alpha_{n+1}+1](s-1)} - \Delta^{Q_s}_{[s-1-\alpha_1]\ldots[s-1-\alpha_{n+1}+1]\underbrace{s-1\ldots s-1}_{m}(0)}}{q_{[\alpha_{n+1}-1]} q_0^m \prod_{i=1}^{n} q_{\alpha_i} (\beta_0 + \beta_0 q_0 + \beta_0 q_0^2 + \ldots - \beta_{s-1} - \beta_{s-1} q_{s-1} - \beta_{s-1} q_{s-1}^2 - \ldots)}$$

$$= \lim_{m\to\infty} \frac{q_{[s-\alpha_{n+1}]} q_{s-1}^m \left(\frac{\beta_{s-1}}{1-q_{s-1}}\right) \prod_{i=1}^{n} q_{[s-1-\alpha_i]}}{-\frac{\beta_{s-1}}{1-q_{s-1}} q_{[\alpha_{n+1}-1]} q_0^m \prod_{i=1}^{n} q_{\alpha_i}}$$

$$= - \lim_{m\to\infty} \frac{q_{[s-\alpha_{n+1}]}}{q_{[\alpha_{n+1}-1]}} \left(\frac{q_{s-1}}{q_0}\right)^m \prod_{i=1}^{n} \frac{q_{[s-1-\alpha_i]}}{q_{\alpha_i}} = w_2.$$

The derivative $I'(x_0)$ exists at Q_s-rational points, if $w_1 = w_2$. Then

$$w_1 = w_2 \;\Leftrightarrow\; \lim_{m\to\infty} \left(\frac{q_0}{q_{s-1}}\right)^m = \lim_{m\to\infty} \left(\frac{q_{s-1}}{q_0}\right)^m.$$

If $\frac{q_0}{q_{s-1}} < 1$, then $w_1 = 0$ and $w_2 = \infty$. If $\frac{q_{s-1}}{q_0} < 1$, then $w_1 = \infty$ and $w_2 = 0$.

Since $x'_0 = x_0$, then does not exist derivative $I'(x_0)$. So, the function I does not have a derivative in any Q_s-rational point for $q_0 \neq q_{s-1}$.

2) Let $x_0 = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n(j)}$ and $0 \neq j \neq s-1$.

Let $x_1 = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\underbrace{j\ldots j}_{m}(0)}$, $x_2 = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\underbrace{j\ldots j}_{m}(s-1)}$. So

$$\lim_{k\to\infty} \frac{I(x_0) - I(x_1)}{x_0 - x_1} = \lim_{k\to\infty} \frac{q_{[s-1-j]}^m \left(\frac{\beta_{[s-1-j]}}{1-q_{[s-1-j]}} - \frac{\beta_{s-1}}{1-q_{s-1}}\right) \prod_{i=1}^{n} q_{[s-1-\alpha_i]}}{q_j^m \left(\frac{\beta_j}{1-q_j}\right) \prod_{i=1}^{n} q_{\alpha_i}} = w_3.$$

On the other hand

$$\lim_{k\to\infty}\frac{I(x_0)-I(x_2)}{x_0-x_2}=\lim_{k\to\infty}\frac{q^m_{[s-1-j]}\left(\frac{\beta_{[s-1-j]}}{1-q_{[s-1-j]}}\right)\prod\limits_{i=1}^{n}q_{[s-1-\alpha_i]}}{q^m_j\left(\frac{\beta_j}{1-q_j}-\frac{\beta_{s-1}}{1-q_{s-1}}\right)\prod\limits_{i=1}^{n}q_{\alpha_i}}=w_4.$$

It is obvious that $w_3=w_4$ for $\dfrac{\frac{\beta_{[s-1-j]}}{1-q_{[s-1-j]}}-\frac{\beta_{s-1}}{1-q_{s-1}}}{\frac{\beta_j}{1-q_j}}=\dfrac{\frac{\beta_{[s-1-j]}}{1-q_{[s-1-j]}}}{\frac{\beta_j}{1-q_j}-\frac{\beta_{s-1}}{1-q_{s-1}}}$.

Having simplified the last equality, we obtain

$$\left(\frac{\beta_{[s-1-j]}}{1-q_{[s-1-j]}}-1\right)\cdot\left(\frac{\beta_j}{1-q_j}-1\right)=\frac{\beta_{[s-1-j]}}{1-q_{[s-1-j]}}\cdot\frac{\beta_j}{1-q_j},$$

$$\frac{\beta_{[s-1-j]}}{1-q_{[s-1-j]}}+\frac{\beta_j}{1-q_j}=1.$$

Taking into account (5.1), the equality $\dfrac{\beta_{[s-1-j]}}{1-q_{[s-1-j]}}+\dfrac{\beta_j}{1-q_j}=1$ will exist when $q_i=q_{[s-1-i]}$ for any $i\in A_s$.

Then, the derivative $I'(x_0)$ does not exist at the points $x_0=\Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n(j)}$, $0\neq j\neq s-1$ for $q_i\neq q_{[s-1-i]}$ for all $i\in A_s$.

3) Let the frequencies $\nu_0(x_0),\nu_1(x_0),\ldots,\nu_{s-1}(x_0)$ of digits $0,1,\ldots,s-1$ exist in the Q_s-representation of the number x_0.

Suppose that the derivative $I'(x_0)$ exists at the point $x_0=\Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\ldots}$.

Consider the case when s is an even number. Suppose that for formula (5.3) is executed

$$I'(x)=\lim_{n\to\infty}\left(\left(\frac{q_0}{q_{s-1}}\right)^{\nu_{s-1}(x)\to\nu_0(x)}\cdot\left(\frac{q_1}{q_{s-2}}\right)^{\nu_{s-2}(x)\to\nu_1(x)}\cdot\ldots\cdot\left(\frac{q_{\frac{s}{2}-1}}{q_{\frac{s}{2}}}\right)^{\nu_{\frac{s}{2}}(x)\to\nu_{\frac{s}{2}-1}(x)}\right)^n=\infty.$$

Then from the condition $(\nu_{s-1}(x_0)-\nu_0(x_0))(q_0-q_{s-1})<0$ it follows that the $\left(\dfrac{q_0}{q_{s-1}}\right)^{\nu_{s-1}(x_0)-\nu_0(x_0)}>1$. From the condition $(\nu_{s-2}(x_0)-\nu_1(x_0))(q_1-q_{s-2})<0$ it follows that the $\left(\dfrac{q_1}{q_{s-2}}\right)^{\nu_{s-2}(x_0)-\nu_1(x_0)}>1$.

Analogously, from the condition $(\nu_{\frac{s}{2}}(x_0)-\nu_{\frac{s}{2}-1}(x_0))(q_{\frac{s}{2}-1}-q_{\frac{s}{2}})<0$ it follows that the $\left(\dfrac{q_{\frac{s}{2}-1}}{q_{\frac{s}{2}}}\right)^{\nu_{\frac{s}{2}}(x_0)-\nu_{\frac{s}{2}-1}(x_0)}>1$.

So, if there exist frequency $\nu_0(x_0), \nu_1(x_0), \ldots, \nu_{s-1}(x_0)$ of digits $0, 1, \ldots, s-1$ in the Q_s-representation of the number x_0, s is an even number, and

$$\begin{cases} (\nu_{s-1}(x_0) - \nu_0(x_0))(q_0 - q_{s-1}) < 0, \\ (\nu_{s-2}(x_0) - \nu_1(x_0))(q_1 - q_{s-2}) < 0, \\ \ldots, \\ (\nu_{\frac{s}{2}}(x_0) - \nu_{\frac{s}{2}-1}(x_0))(q_{\frac{s}{2}-1} - q_{\frac{s}{2}}) < 0, \end{cases}$$

then the finite derivative $I'(x_0)$ does not exist at the point x_0.

Let s be an odd number, then Eq. (5.4) holds. Then

$$I'(x) = \lim_{n\to\infty} \left(\left(\frac{q_0}{q_{s-1}}\right)^{\nu_{s-1}(x)-\nu_0(x)} \cdot \left(\frac{q_1}{q_{s-2}}\right)^{\nu_{s-2}(x)-\nu_1(x)} \cdot \ldots \cdot \left(\frac{q_{\frac{s-3}{2}}}{q_{\frac{s+1}{2}}}\right)^{\nu_{\frac{s+1}{2}}(x)-\nu_{\frac{s-3}{2}}(x)} \right)^n = \infty.$$

Using analogous considerations, it can be argued that the finite derivative $I'(x_0)$ does not exist at the point x_0, with odd s if

$$\begin{cases} (\nu_{s-1}(x_0) - \nu_0(x_0))(q_0 - q_{s-1}) < 0, \\ (\nu_{s-2}(x_0) - \nu_1(x_0))(q_1 - q_{s-2}) < 0, \\ \ldots, \\ (\nu_{\frac{s+1}{2}}(x_0) - \nu_{\frac{s-3}{2}}(x_0))(q_{\frac{s-3}{2}} - q_{\frac{s+1}{2}}) < 0. \end{cases}$$

The theorem is proved.

Corollary 5.3 *If s is odd number, then the digit $\dfrac{s-1}{2}$ retains its frequency $\nu_{\frac{s-1}{2}}(x)$ in the Q_s-representation. That is, the frequency $\nu_{\frac{s-1}{2}}(x)$ of the digit $\dfrac{s-1}{2}$ in the Q_s-representation of the argument is equal to the frequency $\nu_{\frac{s-1}{2}}(y)$ of the digit $\dfrac{s-1}{2}$ in the value of the function $y = I(x)$.*

It was proved in [26] that in the case $q_i \neq q_{s-1-i}$, where $i \in A_s$, the inversor I is a purely singular function (different from the constant continuous function of bounded variation, which is almost everywhere (in difference Lebesgue's measure) is zero.)

5.5 Fractal Properties of the Inversor

Let (M, ρ) be a metric space, E is some limited subset of it, $d(E)$ is the diameter E, $h(t)$ is a continuous increasing function given on a real semiaxis $t \geq 0$ such that $h(0) = 0$ (the class of such functions is denoted by H_0), F_M is a subset of the space M such that $\forall\ E \subset M$ and $\forall\ \varepsilon > 0$ there exists no-more-than countable ε-covering $\{G_i\}$, $G_i \in F_M$, the set E (that is, there exists $\{G_i\}$, $G_i \in F_M$, $E \subset \bigcup\limits_i G_i$, $d(G_i) \leq \varepsilon$). For given E, h and any ε, we define the function

$$m_h^{\varepsilon}(E) = \inf_{d(G_i)\leq\varepsilon} \left\{ \sum_i h[d(G_i)] : E \subset \bigcup_i G_i \right\},$$

where the lower bound is taken for all possible, no-more-than countable ε-coverings $\{G_i\}$, $G_i \in F_M$ of the set E.

Definition 5.3 ([14]) The number

$$H_h(E) = \lim_{\varepsilon\downarrow 0} m_h^{\varepsilon}(E) = \sup_{\varepsilon>0} m_h^{\varepsilon}(E)$$

is called the exterior Hausdorff's h-measure of the set E. In this case, the function h is called the measuring, and the exterior measure $m_h^{\varepsilon}(E)$ is an approximate exterior of the order ε.

The function $h \in H_0$ is called the Hausdorff's dimensionally function of the set E if $0 < H_h(E) < \infty$. For all convex function $h \in H_0$ such that $\dfrac{h(t)}{t} \to \infty$, if $t \to 0$, there exists a set $E \subset R^1$ for which h is a dimensionally function.

If $0 < H_\alpha(E) < \infty$, then the number α is called the *Hausdorff dimension of the set* E.

Definition 5.4 ([14]) The number $\alpha_0 = \alpha_0(E)$, determined by the equation

$$\alpha_0(E) = \sup\{\alpha : H_\alpha(E) \neq 0\} = \inf\{\alpha : H_\alpha(E) = 0\}$$

called the *Hausdorff-Besicovitch dimension of the set* E.

The Hausdorff-Besicovitch dimension of the set $E \subset M$ is determined by the behavior of $H_\alpha(E)$ not as a function of a dependent E, but as functions of α.

Theorem 5.3 *The inversor I saves the Hausdorff-Besicovitch dimension, that is, the set and its image have the same dimension then and only if $q_i = q_{s-1-i}$ for all $i \in A_s$.*

If $q_i \neq q_{s-1-i}$ and H is the set of points x in which there is no finite derivative $I'(x)$, then its Hausdorff-Besicovitch dimension satisfies the inequality

$$\alpha_0(H) \geq \frac{\ln \frac{1}{s}}{\ln q_0 q_1 \dots q_{[s-1]}}, \quad s \text{ is an even number},$$

$$\alpha_0(H) \geq \frac{\ln \frac{1}{s}}{\ln q_0 q_1 \dots q_{[\frac{s-3}{2}]} q_{[\frac{s+1}{2}]} \dots q_{[s-1]}}, \quad s \text{ is an odd number}.$$

Proof Let us now prove the first part of the theorem.

For $q_i = q_{s-1-i}$ has the place $I(x) = 1 - x$, then in this case the inversor I saves the Hausdorff-Besicovitch dimension.

Let's consider the Besicovitch-Egglestone set [14]

$$E[\nu_0, \nu_1, \dots, \nu_{s-1}] = \{x : \nu_0^{Q_s}(x) = \nu_0, \nu_1^{Q_s}(x) = \nu_1, \dots, \nu_{s-1}^{Q_s}(x) = \nu_{s-1}\}$$

for $\nu_i \neq \nu_{s-1-i}$. It was proved in [14] that Hausdorff-Besicovitch dimension is equal to

$$\alpha_0(E) = \frac{\ln \nu_0^{\nu_0} \nu_1^{\nu_1} \dots \nu_{s-1}^{\nu_{s-1}}}{\ln q_0^{\nu_0} q_1^{\nu_1} \dots q_{s-1}^{\nu_{s-1}}}.$$

The image $E[\nu_0, \nu_1, \dots, \nu_{s-1}]$ under the action of the inversor I is the set $E'[\nu_{s-1}, \nu_{s-2}, \dots, \nu_0]$, whose dimension

$$\alpha_0(E') = \frac{\ln \nu_{s-1}^{\nu_{s-1}} \nu_{s-2}^{\nu_{s-2}} \dots \nu_0^{\nu_0}}{\ln q_0^{\nu_{s-1}} q_1^{\nu_{s-2}} \dots q_{s-1}^{\nu_0}}.$$

$\alpha_0(E) = \alpha_0(E')$, when $\ln q_0^{\nu_0} q_1^{\nu_1} \dots q_{s-1}^{\nu_{s-1}} = \ln q_0^{\nu_{s-1}} q_1^{\nu_{s-2}} \dots q_{s-1}^{\nu_0}$, the last equivalently

$$1 = \frac{q_0^{\nu_0} q_1^{\nu_1} \dots q_{s-1}^{\nu_{s-1}}}{q_0^{\nu_{s-1}} q_1^{\nu_{s-2}} \dots q_{s-1}^{\nu_0}} = \left(\frac{q_0}{q_{s-1}}\right)^{\nu_0 - \nu_{s-1}} \cdot \left(\frac{q_1}{q_{s-2}}\right)^{\nu_1 - \nu_{s-2}} \cdot \dots \left(\frac{q_{\frac{s}{2}-1}}{q_{\frac{s}{2}}}\right)^{\nu_{\frac{s}{2}-1} - \nu_{\frac{s}{2}}},$$

where $s = 2t$;

$$1 = \frac{q_0^{\nu_0} q_1^{\nu_1} \dots q_{s-1}^{\nu_{s-1}}}{q_0^{\nu_{s-1}} q_1^{\nu_{s-2}} \dots q_{s-1}^{\nu_0}} = \left(\frac{q_0}{q_{s-1}}\right)^{\nu_0 - \nu_{s-1}} \cdot \left(\frac{q_1}{q_{s-2}}\right)^{\nu_1 - \nu_{s-2}} \cdot \dots \left(\frac{q_{\frac{s-3}{2}}}{q_{\frac{s+1}{2}}}\right)^{\nu_{\frac{s-3}{2}} - \nu_{\frac{s+1}{2}}},$$

where $s = 2t + 1$.

From the last equations implies that the inversor I saves the Hausdorff-Besicovitch dimension then and only if $q_i = q_{s-1-i}$ for all $i \in A_s$.

Let $q_i > q_{s-1-i}$. According to Theorem 5.2, the Besicovitch-Egglestone set $E\left[\frac{1}{s}-\varepsilon;\frac{1}{s}-\varepsilon;\ldots;\frac{1}{s}+\varepsilon;\frac{1}{s}+\varepsilon\right]$ belongs to set H of non-differentiable of the function I for all $0 < \varepsilon < \frac{1}{s}$. Then, in accordance with the properties of the monotonity and stable countable of Hausdorff-Besicovitch dimension [14], we have

$$\alpha_0(H) \geq \sup_{\varepsilon} \alpha_0\left(E\left[\frac{1}{s}-\varepsilon;\frac{1}{s}-\varepsilon;\ldots;\frac{1}{s}+\varepsilon;\frac{1}{s}+\varepsilon\right]\right)$$

$$= \sup_{\varepsilon} \frac{\ln\left(\frac{1}{s}-\varepsilon\right)^{\left(\frac{1}{s}-\varepsilon\right)}\cdot\ldots\cdot\left(\frac{1}{s}+\varepsilon\right)^{\left(\frac{1}{s}+\varepsilon\right)}}{\ln q_0^{\left(\frac{1}{s}-\varepsilon\right)}\ldots q_{s-1}^{\left(\frac{1}{s}+\varepsilon\right)}},$$

$$\alpha_0(H) \geq \frac{\ln\frac{1}{s}}{\ln q_0 q_1\ldots q_{s-1}}, \quad s=2t,$$

$$\alpha_0(H) \geq \frac{\ln\frac{1}{s}}{\ln q_0 q_1\ldots q_{[\frac{s-3}{2}]}q_{[\frac{s+1}{2}]}\cdots q_{[s-1]}}, \quad s=2t+1.$$

If $q_i < q_{s-1-i}$, then, according to Theorem 5.2, the Besicovitch-Egglestone set $E\left[\frac{1}{s}+\varepsilon;\frac{1}{s}+\varepsilon;\ldots;\frac{1}{s}-\varepsilon;\frac{1}{s}-\varepsilon\right]$ belongs to the set H, where the function I is non-differentiable for all $0 < \varepsilon < \frac{1}{s}$ and

$$\alpha_0(H) \geq \sup_{\varepsilon} \alpha_0\left(E\left[\frac{1}{s}+\varepsilon;\frac{1}{s}+\varepsilon;\ldots;\frac{1}{s}-\varepsilon;\frac{1}{s}-\varepsilon\right]\right)$$

$$= \sup_{\varepsilon} \frac{\ln\left(\frac{1}{s}+\varepsilon\right)^{\left(\frac{1}{s}+\varepsilon\right)}\cdot\ldots\cdot\left(\frac{1}{s}-\varepsilon\right)^{\left(\frac{1}{s}-\varepsilon\right)}}{\ln q_0^{\left(\frac{1}{s}+\varepsilon\right)}\ldots q_{s-1}^{\left(\frac{1}{s}-\varepsilon\right)}},$$

$$\alpha_0(H) \geq \frac{\ln\frac{1}{s}}{\ln q_0 q_1\ldots q_{s-1}}, \quad s=2t,$$

$$\alpha_0(H) \geq \frac{\ln\frac{1}{s}}{\ln q_0 q_1\ldots q_{[\frac{s-3}{2}]}q_{[\frac{s+1}{2}]}\cdots q_{[s-1]}}, \quad s=2t+1.$$

What it had to prove.

Lemma 5.2 *The graph* $\Gamma_I = \{(x, I(x)) : x \in [0, 1]\}$ *of the function* I *is a self-affine set, namely*

$$\Gamma_I = \bigcup_{i=0}^{s-1} \phi_i(\Gamma_I) \equiv \phi(\Gamma_I),$$

where ϕ_i *are affine transformations, such that*

$$\phi_i : \begin{cases} x' = q_i x + \beta_i, \\ y' = -q_{[s-1-i]} y + \beta_{[s-i]}, \end{cases}$$

$i \in A_s$, $\beta_s = 1$.

Proof Let $\Gamma \equiv \phi_0(\Gamma_I) \cup \phi_1(\Gamma_I) \cup \ldots \cup \phi_{s-1}(\Gamma_I)$. Let's show $\Gamma \subset \Gamma_I$. For an arbitrary point $M(x'_M, y'_M) \in \Gamma$ there is an i such that $M \in \phi_i(\Gamma_I)$, that is ϕ_i : $\begin{cases} x' = q_i x + \beta_i, \\ y' = -q_{[s-1-i]} y + \beta_{[s-i]}. \end{cases}$

It's easy to see that $y'_M = I(x'_M)$.

Let's show $\Gamma_I \subset \Gamma$. Let $M(x, f(x)) \in \Gamma_I$, then $x_M = \Delta^{Q_s}_{\alpha_1\alpha_2\ldots\alpha_n\ldots} = q_i x + \beta_i$, that is, $f(x) = y$ and $M \in \Gamma$.

Then $\Gamma = \Gamma_I$ and the graph Γ_I of the function I is a self-affine set.

Corollary 5.4 *All the levels (the set of level* y_0 *of the function* f *is called the set* $f^{-1}(y_0) = \{x : f(x) = y_0\}$*) functions of* I *are finite.*

5.6 Conclusion

Functions with complex local structure and fractal properties, in particular singular and "piecewise singular" functions, are an actual object of modern research. But their general theory is poorly developed, its development is realized mainly through individual theories (individual functions and family of functions that depends on the parameters). This problem, first of all, is related to the search for effective means of their assignment and study. In recent years, different systems of real-time encoding with the finite and infinite, constant and changing alphabets are increasingly used to solve this problem. These include the classical s-adic representation of real numbers, and its generalization is the polybase Q_s-representation of numbers. This is the representation we used to determining (construct) a function depends on $s - 1$ parameters, with a non-trivial a set of features of a differentiated nature. Conditions for non-differentiability and singularity were found for her, and self-similar and fractal properties were investigated.

References

1. Agadzhanov, A.N.: Singular functions that do not have intervals of monotonicity in problems of finite control of distributed systems. Rep. Acad. Sci. **454**(5), 503–506 (2014) (in Russian)
2. Albeverio, S., Baranovskyi, O., Kondratiev, Y., Pratsiovytyi, M.: On one class of functions related to Ostrogradsky series and containing singular and nowhere monotonic functions. Sci. J. Natl. Pedagogical Dragomanov Univ. Serya 1. Phys. Math. Sci. **15**, 35–55 (2013) (National Pedagogical Dragomanov University, Kiev)
3. Bernstein, D.: Algorithmic Definitions of Singular Functions. Davidson College, Davidson (2013)
4. Freilich, G.: Increasing continuous singular functions. Am. Math. Mon. **80**, 918–919 (1973)
5. Gelbaum, R.B., Olmsted, J.M.H.: Counterexamples in Analysis. HoldenDay, San Francisco (1964)
6. Hewitt, E., Stromberg, K.: Real and Abstract Analysis. Springer, New York (1965)
7. Kalpazidou, S., Knopfmacher, A., Knopfmacher, J.: L*ü*roth-type alternating series representations for real numbers. Acta Arith. **55**, 311–322 (1990)
8. Kapustyan, O.V., Kapustyan, O.A., Sukretna, A.V.: Approximate bounded synthesis for one weakly nonlinear boundary-value problem. Nonlinear Oscil. **12**(3), 297–304 (2009)
9. Kawamura, K.: On the set of points where lebesgues singular function has the derivative zero. Proc. Jpn. Acad. **87**(A), 162–166 (2011)
10. Kyoungsoo, P., Jeronymo, P.P., Armando Duarte, C., Paulino, G.H.: Integration of singular enrichment functions in the generalized/extended finite element method for three-dimensional problems. Int. J. Numer. Methods Eng. **78**, 1220–1257 (2009)
11. Luroth, J.: *Ü*eber eine eindeutige Entwickelung von Zahlen in eine unendliche Reihe. Math. Ann. **21**, 411–423 (1883)
12. Massopust, P.R.: Fractal Functions, Fractal Surfaces, and Wavelets, 1st edn., 383 pp. Academic, Cambridge (1995)
13. Mironovsky, L.A., Petrova, X.Y.: Singular functions of a nonlinear pendulum on finite time intervals. In: Conference: Physics and Control, Proceedings. 2003 International Conference, vol. 4 (2003)
14. Pratsiovytyi, M.V.: The fractal approach in the research of singular distributions, 296 pp. (1998). View of the National Pedagogical Dragomanov University, Kyiv (in Ukrainian)
15. Pratsiovytyi, M.V., Hetman, B.I.: Engel's series and their application. Sci. J. Natl. Pedagogical Dragomanov Univ. Serya 1. Phys. Math. Sci. **7**, 105–116 (2006) (National Pedagogical Dragomanov University, Kiev (in Ukrainian))
16. Pratsiovytyi, M.V., Skrypnyk, S.V.: Q_2-representation for fractional part of real number and the inversor of its digits. Sci. J. Natl. Pedagogical Dragomanov Univ. Serya 1. Phys. Math. Sci. **15**, 134–143 (2013) (National Pedagogical Dragomanov University, Kiev (in Ukrainian))
17. Pratsiovytyi, M.V., Zamrii, I.V.: Continuous functions preserving digit 1 in the Q_3-representation of number. Bukovinsky Math. J. **3**(3–4), 142–159 (2015). Chernivtsi: Chernivtsi National University (in Ukrainian)
18. Pratsovyta, I.M., Zadniprianyi, M.V.: Schedules of numbers in the Sylvester series and their application. Sci. J. Natl. Pedagogical Dragomanov Univ. Serya 1. Phys. Math. Sci. **10**, 73–87 (2009) (National Pedagogical Dragomanov University, Kiev (in Ukrainian))
19. Ricardo, E., Fulling, S.A.: How singular functions define distributions. J. Phys. A Math. Gen. **35**(13), 3079–3089 (2002)
20. Riesco, A., Rodriguez-Hortala, J.: Singular and plural functions for functional logic programming: detailed proofs. Technical report SIC-9/11, Dpto. Sistemas Informaticos y Computacion, Universidad Complutense de Madrid (2011)
21. Riesz, F., Nagy, B.Sz.: Functional Analysis. Ungar, New York (1965)

22. Salem, R.: On some singular monotonic function which are strictly increasing. Trans. Am. Math. Soc. **53**(3), 427–439 (1943)
23. Takacs, L.: An increasing continuous singular function. Am. Math. Mon. **85**, 35–37 (1978)
24. Turbin, A.F., Pratsiovytyi, M.V.: Fractal Sets, Functions, Distributions, 208 pp. Naukova dumka, Kiev (1992) (in Russian)
25. Wen, L.: An approach to construct the singular monotone functions by using markov chains. Taiwan. J. Math. **2**(3), 361–368 (1998)
26. Zamrii, I.V.: Lebesgue structure and properties of the inversor of digits of Q_s-representation for fractional part of real number. Sci. Educ. New Dimens. Nat. Tech. Sci. **16**(148), 47–49 (2017). Budapest
27. Zamrii, I.V., Pratsiovytyi, M.V.: The singularity of the inversor of digits of Q_3-representation of the fractional part of the real number, its fractal and integral properties. Nonlinear oscil. **18**(1), 55–70 (2015). ISSN 1562-3076, Institute of Mathematics, National Academy of Sciences of Ukraine
28. Zhiqiang, C., Seokchan, K., Sangdong, K., Sooryun, K.: A finite element method using singular functions for Poisson equations: mixed boundary conditions. Comput. Methods Appl. Mech. Eng. **195**, 2635–2648 (2006)
29. Zhykharieva, Yu.I., Pratsovytyi, M.V.: Representation of numbers by the applicable Liurot's series: the basis of the metric theories. Sci. J. Natl. Pedagogical Dragomanov Univ. Serya 1. Phys. Math. Sci. **9**, 200–211 (2008) (National Pedagogical Dragomanov University, Kiev (in Ukrainian))

Chapter 6
Almost Sure Asymptotic Properties of Solutions of a Class of Non-homogeneous Stochastic Differential Equations

Oleg I. Klesov and Olena A. Tymoshenko

Abstract We study non-homogeneous stochastic differential equation with separation of stochastic and deterministic variables. We express the asymptotic behavior of solutions of such equations in terms of that for the corresponding ordinary differential equation. The general results are discussed for some particular equations, mainly in the field of mathematics of finance.

6.1 Introduction

Stochastic differential equations are one of the effective models of stochastic processes that are used in many fields of science including of insurance and financial mathematics, economics, control theory and many others. Linear stochastic differential equations

$$dX(t) = a(t, X(t))dt + b(t, X(t))dw(t), \qquad t \geq 0, \tag{6.1}$$

describe quite well various natural and engineering phenomena. Here a and b are two nonrandom functions and w is a Wiener process. A partial case of Eq. (6.1) is presented by the homogeneous stochastic differential equation perturbed by a Wiener process

$$dX(t) = g\,(X(t))\,dt + \sigma\,(X(t))\,dw(t). \tag{6.2}$$

Stochastic differential equations often have an economic interpretation that really makes research on its solutions particularly interesting for economists. The Vašiček

O. I. Klesov (✉) · O. A. Tymoshenko
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Department of Mathematical Analysis and Probability Theory, Kyiv, Ukraine
e-mail: klesov@matan.kpi.ua; klesov@matan.kpi.edu; zorot@ukr.net

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_6

equation [33]

$$dX(t) = \alpha\,(\beta - X(t))\,dt + \sigma\,dw(t). \tag{6.3}$$

is used in finance for the valuation of interest rate derivatives, and has also been adapted for credit markets. Obviously it is a member of the class of homogeneous stochastic differential equations. In Eq. (6.3), $w(\cdot)$ is a standard Wiener process; β is a long term mean level, α characterizes the velocity at which such trajectories will regroup in time, and σ determines the volatility of the interest rate. Many other stochastic differential equations of this type are widely used in mathematics of finance (see, for example [4, 5, 14, 15, 26]).

The following stochastic differential equation

$$dX(t) = g(X(t))\phi(t)dt + \sigma(X(t))\theta(t)dw(t) \tag{6.4}$$

is studied in the current paper. Regarding the generality, this equation occupies a place between homogeneous equation (6.2) and general linear equation (6.1).

Among other popular models of type (6.4) are the Rendleman–Bartter model [27]

$$dX(t) = X(t)\phi(t)dt + X(t)\theta(t)dw(t) \tag{6.5}$$

and the Hull–White model [18]

$$dX(t) = (\alpha(t) + \beta(t)X(t))\,dt + \theta(t)dw(t). \tag{6.6}$$

Model (6.5) is considered in finance to investigate the evolution of interest rates. Model (6.6) is still popular in the financial markets today, since it is possible to value many derivatives dependent solely on a single bond analytically when working in the Hull–White model.

We are mainly interested in studying the asymptotic behavior of solutions of stochastic differential equations. Close problems are studied in [7, 32]. In doing so, we follow an idea due to Gihman and Skorohod [17] to describe this behavior in terms of the behavior of a solution of the corresponding ordinary differential equation.

Gihman and Skorohod [17] dealt with the asymptotic behavior of solutions of homogeneous stochastic differential equation (6.2).

Another approach to solving this problem is presented in the paper Keller, Kersting, and Rösler [19]. The same problem was later investigated by Kersting [20] for multidimensional stochastic differential equations. Sufficient conditions for the so-called ψ-asymptotic behaviour of solutions stochastic differential equation (6.2) are presented in [12]. Closer investigations are done by Samoĭlenko and Stanzhitskiĭ [28].

Asymptotic behavior of solutions of differential equation with state-independent perturbation

$$dX(t) = g(X(t))\,dt + \theta(t)\,dw(t) \tag{6.7}$$

being a particular case of Eq. (6.4) was investigated in [1, 2].

Asymptotic properties of solutions of stochastic differential equations have also been studied by D'Anna et al. [16], Mitsui [25], Strauss and Yorke [30] to mention a few. Properties of solutions of stochastic differential equations are described by these authors in terms of properties of the corresponding deterministic differential equation.

6.2 Setting of the Problem

We consider the following Cauchy problem for non-homogeneous stochastic differential equation (6.4)

$$\begin{aligned} &dX(t) = g(X(t))\phi(t)dt + \sigma(X(t))\theta(t)dw(t), \qquad t \geq 0, \\ &X(0) = b > 0. \end{aligned} \tag{6.8}$$

Here $w(\cdot)$ is a standard Wiener process, $\theta(\cdot)$ is a continuous function, $g(\cdot)$, $\phi(\cdot)$ and $\sigma(\cdot)$ are continuous positive functions such that a unique and continuous solution $X(\cdot)$ of Eq. (6.1) exists. Clearly (6.1) generalizes the homogeneous equation (6.2). The corresponding ordinary differential equation

$$d\mu(t) = g(\mu(t))\phi(t)\,dt \tag{6.9}$$

separates the variables, so that Eq. (6.1) is called a *stochastic differential equation with separated variables*. Another name for (6.1) is *stochastic differential equation with time-dependent coefficients*. Applications for some particular equations arising in the field of mathematics of finance are also discussed. The same problem for Eq. (6.2) has been solved in [17]. Another set of sufficient conditions is proposed in [19].

Suppose that a continuous stochastic process $X(t) = (X(\omega, t), \omega \in \Omega, t \geq 0)$ is defined on a complete probability space $\{\Omega, \mathscr{F}, \mathbf{P}\}$.

The following two functions

$$\Phi(t) = \int_0^t \phi(u)du, \qquad t \geq 0, \tag{6.10}$$

and

$$\Theta(t) = \int_0^t \theta^2(s)\,ds, \qquad t > 0, \tag{6.11}$$

are involved in the statement of the main result below. A crucial role in our approach is played by the following condition

$$\lim_{t\to\infty} \Phi(t) = \infty. \tag{6.12}$$

The results obtained in this paper are valid only for those solutions $X(t)$ for which

$$\lim_{t\to\infty} X(t) = \infty \qquad \text{a.s.} \tag{6.13}$$

Some sufficient conditions for (6.13) are obtained in [17] in the case of homogeneous equation (6.2). Our conditions guarantee that there is no explosion of a solution during a finite time. This problem, in general, is studied by Taniguchi [31].

6.3 Main Result

The main result of the paper is the following one. A preliminary version has been obtained in [23].

Theorem 6.1 *Assume that $w(\cdot)$ is a Wiener process and $\sigma(\cdot)$ is a continuous positive bounded function with*

$$\sup_{x\in\mathbb{R}} \sigma(x) < \infty. \tag{6.14}$$

Further let $g(\cdot)$, $\phi(\cdot)$, and $\theta(\cdot)$ be continuous positive functions. In addition, let $g(\cdot)$ be such that, for each $T > 0$, there exists a positive constant K for which

$$|g(x) - g(y)| \le K|x - y| \tag{6.15}$$

and

$$g^2(x) \le K^2(1 + x^2) \tag{6.16}$$

for all $t \in [0, T]$ and for all $x, y \in \mathbb{R}$. Moreover let conditions (6.12) and (6.13). If

$$\sum_{k=0}^{\infty} \frac{\Theta(2^{k+1})}{\Phi^2(2^k)} < \infty, \tag{6.17}$$

then

$$\lim_{t\to\infty} \frac{1}{\Phi(t)} \int_0^t \sigma(X(s))\theta(s)dw(s) = 0 \qquad a.s. \tag{6.18}$$

Remark 6.1 Using the technique developed in [21] and bounds on the moments of solutions of stochastic differential equations one can treat the case of an unbounded coefficient of diffusion $\sigma(\cdot)$ as well.

Remark 6.2 In the case of homogeneous equation (6.2),

$$\phi(t) \equiv 1, \qquad \theta(t) \equiv 1,$$

whence

$$\Phi(t) = t, \qquad \Theta(t) = t.$$

and condition (6.17) follows.

Proof First, note that conditions (6.15) and (6.16) imply that, for every $T > 0$, a continuous solution of equation (6.1) exists in the interval $[0, T]$. Indeed the following conditions

$$|a(t, x) - a(t, y)| \le K|x - y|, \qquad |b(t, x) - b(t, y)| \le K|x - y|$$

for some constant $K > 0$ and all $t \in [0, T]$ and $x, y \in \mathbb{R}$; and

$$a^2(t, x) + b^2(t, x) \le K(1 + x^2)$$

for some constant $K > 0$ and all $t \in [0, T]$ and $x, y \in \mathbb{R}$ are sufficient for the problem (6.1) to have a continuous solution in the interval $[0, T]$. Since $a(t, x) = g(x)\phi(t)$ and $b(t, x) = \theta(t)\sigma(x)$ in our case, the above conditions simplify. Moreover, $\phi(\cdot)$ and $\theta(\cdot)$ are bounded in every interval $[0, T]$, since they are continuous. As $\sigma(\cdot)$ is bounded in view of (6.14), these conditions reduce to (6.15) and (6.16).

Next, we introduce, for all $k \ge 0$ and $\varepsilon > 0$, the following two events:

$$B_k = \left\{ \sup_{2^k \le t \le 2^{k+1}} \frac{1}{\Phi(t)} \left| \int_0^t \sigma(X(s))\theta(s)dw(s) \right| > \varepsilon \right\}$$

and

$$C_k = \left\{ \sup_{2^k \le t \le 2^{k+1}} \frac{1}{\Phi(2^k)} \left| \int_0^t \sigma(X(s))\theta(s)dw(s) \right| > \varepsilon \right\}.$$

Since $\sigma(\cdot)$ is bounded and continuous, $X(\cdot)$ is continuous almost surely, and $\theta(\cdot)$ is continuous, the integral

$$\int_0^t \sigma^2(X(s))\theta^2(s)\,ds$$

exists and is finite almost surely for all $t > 0$. Moreover

$$\int_0^t \mathbf{E}\left[\sigma^2(X(s))\theta^2(s)\right] ds < \infty.$$

By the Chebyshev–Markov inequality, for all $\varepsilon > 0$

$$\begin{aligned}\mathbf{P}(C_k) &= \mathbf{P}\left\{\sup_{2^k\le t\le 2^{k+1}}\left|\int_0^t \sigma(X(s))\theta(s)\,dw(s)\right| > \varepsilon\Phi(2^k)\right\}\\ &\le \frac{1}{\Phi^2(2^k)\varepsilon^2}\,\mathbf{E}\left[\sup_{2^k\le t\le 2^{k+1}}\left|\int_0^t \sigma(X(s))\theta(s)\,dw(s)\right|^2\right]\\ &\le \frac{1}{\Phi^2(2^k)\varepsilon^2}\,\mathbf{E}\left[\sup_{t\le 2^{k+1}}\left|\int_0^t \sigma(X(s))\theta(s)\,dw(s)\right|^2\right].\end{aligned}$$

According to Theorem 1 of §3 in [17], for all $k \ge 0$

$$\begin{aligned}\mathbf{E}\left[\sup_{t\le 2^{k+1}}\left|\int_0^t \sigma(X(s))\theta(s)\,dw(s)\right|^2\right] &\le 4\int_0^{2^{k+1}} \left(\mathbf{E}\left[\sigma(X(s))\theta(s)\right]\right)^2 ds\\ &\le 4M^2\int_0^{2^{k+1}} \theta^2(s)\,ds,\end{aligned}$$

where

$$M = \sup_{x\in\mathbb{R}}|\sigma(x)|.$$

Note that $\Phi(\cdot)$ is an increasing function, whence $B_k \subset C_k$, $k \ge 0$. Therefore

$$\mathbf{P}(B_k) \le \mathbf{P}(C_k) \le \frac{4M^2}{\Phi^2(2^k)\varepsilon^2}\cdot\left(\int_0^{2^{k+1}}\theta^2(s)\,ds\right) = \frac{4M^2}{\varepsilon^2}\cdot\frac{\Theta(2^{k+1})}{\Phi^2(2^k)}.$$

Now, by condition (6.17),

$$\sum_{k=0}^{\infty}\mathbf{P}(B_k) < \infty,$$

whence the Borel–Cantelli lemma implies

$$\mathbf{P}(B_k \text{ i.o.}) = 0$$

for any $\varepsilon > 0$. Here "i.o." abbreviates "infinitely often". This means that if $\varepsilon > 0$ is given, then almost surely a number k_0 exists such that

$$\sup_{2^k \le t \le 2^{k+1}} \frac{1}{\Phi(t)} \left| \int_0^t \sigma(X(s))\theta(s)\, dw(s) \right| \le \varepsilon, \qquad k \ge k_0.$$

Moreover,

$$\frac{1}{\Phi(t)} \left| \int_0^t \sigma(X(s))\theta(s)\, dw(s) \right| \le \varepsilon, \qquad t \ge k_0.$$

This completes the proof, since $\varepsilon > 0$ is arbitrary. □

6.3.1 *Some Sufficient Conditions for (6.17)*

Below we discuss some simple sufficient conditions for assumption (6.17). These conditions are expressed in terms of growth conditions imposed on the functions $\phi(\cdot)$ and $\theta(\cdot)$. First, we prove a result where we use growth conditions imposed on the integrals of functions $\phi(\cdot)$ and $\theta^2(\cdot)$, namely

$$\liminf_{t\to\infty} \frac{\Phi(t)}{t^\alpha} > 0 \tag{6.19}$$

and

$$\limsup_{t\to\infty} \frac{\Theta(t)}{t^\beta} < \infty \tag{6.20}$$

Lemma 6.1 *If conditions* (6.19) *and* (6.20) *are satisfied with* $\beta < 2\alpha$, *then condition* (6.17) *holds.*

Proof By conditions (6.19) and (6.20), there exists a constant $M > 0$ such that

$$\frac{t^\alpha}{\Phi(t)} < M \qquad \text{for all} \quad t > 0,$$

$$\frac{\Theta(t)}{t^\beta} < M \qquad \text{for all} \quad t > 0.$$

Now we estimate the general term of series (6.17):

$$\frac{\Theta(2^{k+1})}{\Phi^2(2^k)} \le M \cdot \frac{2^{(k+1)\beta}}{\Phi^2(2^k)} \le M^3 \cdot \frac{2^{(k+1)\beta}}{2^{2k\alpha}} = \frac{M^3 2^\beta}{2^{k(2\alpha-\beta)}}.$$

This means that

$$\sum_{k=0}^{\infty} \frac{\Theta(2^{k+1})}{\Phi^2(2^k)} \le M^3 2^\beta \sum_{k=0}^{\infty} 2^{-k(2\alpha-\beta)}.$$

The series on the right-hand side converges if $\alpha > \beta/2$.

Remark 6.3 If

$$\lim_{u\to\infty} \inf \phi(u) u^\gamma > 0 \tag{6.21}$$

for some $\gamma < 1$, then condition (6.19) holds for $\alpha = 1 - \gamma$. Indeed, according to condition (6.21), there exist two numbers $\varepsilon > 0$ and $u_0 \ge 0$ such that

$$\phi(u) > \varepsilon u^{-\gamma}, \qquad u > u_0.$$

Without loss of generality, we will assume that $u_0 = 0$. So for $t \ge 1$ we have

$$\int_0^t \phi(u)\, du \ge \varepsilon \int_0^t u^{-\gamma}\, du = \frac{\varepsilon}{1-\alpha} \cdot t^{1-\gamma}$$

if $\gamma < 1$. Therefore condition (6.19) follows for $\alpha = 1 - \gamma$.

Remark 6.4 Similarly to Remark 6.3, if

$$\limsup_{t\to\infty} \frac{\theta(t)}{t^\delta} < \infty, \tag{6.22}$$

then condition (6.20) holds with $\beta = 1 + \delta$.

Now Remarks 6.3 and 6.4 together with Lemma 6.1 imply the following result.

Proposition 6.1 *Condition* (6.17) *holds if* (6.21) *and* (6.22) *are satisfied with*

$$\gamma < \min\left\{1, \frac{1-\delta}{2}\right\}.$$

6.3.1.1 The Case of Regularly Varying Coefficients

Let $\mathscr{RV}_\gamma$ denote the set of regularly varying functions of index γ (see [3] or [29] for definitions and properties of regularly varying functions). Assume that $\phi(\cdot) \in \mathscr{RV}_\alpha$

and $\theta(\cdot) \in \mathscr{RV}_\beta$. According to Karamata's Theorem

$$\Phi(\cdot) \in \mathscr{RV}_{\alpha+1}, \qquad \Theta(\cdot) \in \mathscr{RV}_{2\beta+1}.$$

Therefore condition (6.17) holds if $\alpha > \beta$. The case of $\alpha = \beta$ is more involved but condition (6.17) still holds. Indeed, let $\phi(t) = t^\alpha \ell_1(t)$ and $\theta(t) = t^\alpha \ell_2(t)$, where $\ell_1(\cdot)$ and $\ell_2(\cdot)$ are some slowly varying functions. Then, by Karamata's theorem,

$$\Phi(t) = t^{\alpha+1}\ell_1'(t), \qquad \Theta(t) = t^{2\alpha+1}\ell_2'(t),$$

where

$$\ell_1'(t) = \frac{1}{t^{\alpha+1}} \int_0^t \phi(s)\, ds \qquad \text{and} \qquad \ell_2'(t) = \frac{1}{t^{2\alpha+1}} \int_0^t \theta^2(s)\, ds$$

are slowly varying functions, that is $\ell_1'(\cdot), \ell_2'(\cdot) \in \mathscr{SV}$. Here $\mathscr{SV} = \mathscr{RV}_0$ denotes the set of all slowly varying functions. Hence

$$\frac{\Theta(2^{k+1})}{\Phi^2(2^k)} = \frac{\ell_3(2^k)}{2^k} \qquad \text{with} \qquad \ell_3(t) = 2^{2\alpha+1} \frac{\ell_2'(2t)}{(\ell_1'(t))^2} \in \mathscr{SV}.$$

Since $\ell_3(t) = o(t^{1/2})$ as $t \to \infty$, we obtain (6.17). However (6.17) fails if $\alpha < \beta$.

One can generalize the case of regularly varying functions discussed above. One possible extension is related to the so-called functions with non-degenerate groups of regular points (see [13] or [9] for the origin of the notion). A measurable positive function $f(\cdot)$ is said to have a regular scale point $\lambda > 0$ if the limit

$$\lim_{t\to\infty} \frac{f(\lambda t)}{f(t)}$$

exists. Any function has at least one regular scale point, namely $\lambda = 1$. The set of regular scale points $G_f(\cdot)$ of any function $f(\cdot)$ is always a multiplicative group. We say that $G_f(\cdot)$ is non-degenerate if there are at least two elements in it. In such a case, the corresponding function $f(\cdot)$ is said to have a non-degenerate group of regular scale points. Note that every positive real number is a regular scale point for every regularly varying function. However there are many other functions whose groups of regular scale points are non-degenerate but "thinner" than the whole positive semi-axes. In fact, for every multiplicative semigroup, there exists a function $f(\cdot)$ for which this group coincides with the set of regular scale points for $f(\cdot)$ (see [9]). Of course, there are functions whose group of regular scale points is degenerate.

For the current paper, it is important that some classes of functions with non-degenerate group of regular scale points allow a nice asymptotic of integrals, similar to that given by the Karamata Theorem (see [6]). The asymptotic expression for the integral in such a case involves some *periodic* components as well as power and

slowly varying functions. Therefore, one can easily check condition (6.17) in such a case.

6.3.2 Sharpness of Theorem 6.1

Just for the sake of demonstration, we provide below a corollary of Theorem 6.1. This result also follows from the law of the logarithm for the Wiener process. Our aim in stating Corollary 6.1 is to show how close Theorem 6.1 is to an optimal result for the Wiener process.

Corollary 6.1 *Assume that $w(\cdot)$ is a Wiener processes, then for any $\varepsilon > 0$*

$$\lim_{t\to\infty} \frac{w(t)}{\sqrt{t}(\log t)^{\frac{1}{2}+\varepsilon}} = 0 \quad \text{a.s.}$$

Proof Consider the equation

$$dX(t) = \phi(t)\,dt + dw(t), \tag{6.23}$$

which coincides with Eq. (6.1) if $g(x) = 1$ for all $x \in \mathbb{R}$ and $\sigma(t) = 1$ and $\theta(t) = 1$ for all $t > 0$. Note that $X(t) = X(0) + \Phi(t) + w(t)$ and the result follows if

$$\lim_{t\to\infty} \frac{w(t)}{\Phi(t)} = 0 \qquad \text{a.s.}$$

For $\phi(\cdot)$ defined by

$$\phi(t) = \frac{(\log t)^{\varepsilon+(1/2)} + (1+2\varepsilon)(\log t)^{\varepsilon-(1/2)}}{2\sqrt{t}}, \qquad t > 0,$$

for some $\varepsilon > 0$ the latter result follows from the law of the iterated logarithm. However we want to prove it independently via Theorem 6.1.

First we prove condition (6.13) in the case under consideration. The function

$$\widetilde{g}(t,x) = -\int_0^t \frac{b'_t(t,x)}{b^2(t,y)}\,dy + \frac{a(t,x)}{b(t,x)} - \frac{1}{2}b'_x(t,x)$$

defined for the general linear stochastic differential equation (6.1) reduces to $\phi(t)$ for all t and x in the case of equation (6.23). Thus

$$u(t) = \sup_{x\in\mathbb{R}} \widetilde{g}(t,x) = \phi(t).$$

Note that

$$\Phi(t) = \sqrt{t}(\log t)^{\frac{1}{2}+\varepsilon}, \qquad t > 0.$$

Since

$$\int_0^T u(t)\,dt = \Phi(T) \asymp T^{1/2}(\log T)^{(1/2)+\varepsilon}, \qquad T \to \infty,$$

by Karamata's theorem,

$$\lim_{T\to\infty} \frac{1}{\sqrt{2T\log\log T}} \int_0^T u(t)\,dt = \infty.$$

Therefore Theorem 2.1 of [22] implies condition (6.13).

Condition (6.17) in our case reads as follows

$$\sum_{k=0}^{\infty} \frac{2^{k+1}}{\Phi^2(2^k)} < \infty.$$

Since

$$\sum_{k=0}^{\infty} \frac{2^{k+1}}{\Phi^2(2^k)} \asymp 2\sum_{k=1}^{\infty} \frac{1}{k^{1+2\varepsilon}}$$

and $\varepsilon > 0$, the series on the right hand side converges. Hence condition (6.17) holds and Corollary 6.1 follows from Theorem 6.1. □

Remark 6.5 The same reasoning but with

$$\Phi(t) = (t\log t)^{1/2}(\log\log t)^{(1/2)+\varepsilon}, \qquad t > 1,$$

proves that

$$\lim_{t\to\infty} \frac{w(t)}{(t\log t)^{1/2}(\log\log t)^{\frac{1}{2}+\varepsilon}} = 0 \qquad \text{a.s.}$$

6.4 Some Examples

Below are some examples of applications of Theorem 6.1 to several particular stochastic differential equations.

6.4.1 Population Growth Model

Consider the Cauchy problem

$$dX(t) = \phi(t)X(t)dt + \beta X(t)dw(t), \qquad t \geq 0; \tag{6.24}$$
$$X(0) = 1.$$

A solution of problem (6.24) describes the growth of a population with unit initial size (see [26]), where $X(\cdot)$ is the size of population at time t; $\phi(\cdot)$ is relative growth rate of the population that depends on time; $w(\cdot)$ is a Wiener process; $\beta \in (0; +\infty)$. Let $\phi(\cdot)$ be a positive continuous function.

Clearly, this is a particular case of problem (6.8) corresponding to $g(x) = x$, $\theta(t) \equiv \beta$, and $\sigma(x) = x$, however Theorem 6.1 is not applied directly here, since $\sigma(x) = x$ in problem (6.24) and condition (6.14) does not hold. So one needs a way around this problem.

Theorem 6.2 *Let $X(\cdot)$ be a solution of problem (6.24). Assume that*

$$\lim_{t\to\infty} \frac{\Phi(t)}{t} > \frac{1}{2}\beta^2 \tag{6.25}$$

where $\Phi(\cdot)$ is defined by (6.10) and denote the left-hand side of (6.25) by K. Then

$$\lim_{t\to\infty} \frac{\ln X(t)}{\Phi(t)} = 1 - \frac{\beta^2}{2K} \qquad \text{a. s.}$$

The right-hand side equals 1 if $K = \infty$.

Proof First we check whether condition (6.13) holds. The solution of problem (6.24) is given by

$$X(t) = \exp\left\{\left(\Phi(t) - \frac{1}{2}\beta^2 t\right) + \beta w(t)\right\}.$$

Now we obtain from the strong law of large numbers, $w(t) = o(t)$ a.s., that

$$\lim_{t\to\infty} X(t) = \lim_{t\to\infty} \exp\left\{\left(\Phi(t) - \frac{1}{2}\beta^2 t\right) + \beta w(t)\right\}$$
$$= \lim_{t\to\infty} \exp\left\{t\left(\frac{\Phi(t)}{t} - \frac{1}{2}\beta^2 + \beta\frac{w(t)}{t}\right)\right\} = \infty$$

almost surely.

Next we apply the Itô formula for $\ln X(t)$ and arrive at the equation

$$d\ln X(t) = \left(\phi(t) - \frac{1}{2}\beta^2\right)dt + \beta dw(t). \tag{6.26}$$

Clearly, all assumptions of Theorem 6.1 are satisfied for Eq. (6.26) and thus we have

$$\lim_{t\to\infty} \frac{1}{\Phi(t)} \int_0^t \sigma(X(s))\theta(s)\, dw(s) = \lim_{t\to\infty} \frac{1}{\Phi(t)} \int_0^t \beta\, dw(s) = 0 \qquad \text{a.s.}$$

Finally, using the closed form of a solution of equation (6.26) and condition (6.25), we get

$$\lim_{t\to\infty} \frac{\ln X(t)}{\Phi(t)} = \lim_{t\to\infty} \frac{1}{\Phi(t)} \left(\int_0^t (\phi(s) - \frac{1}{2}\beta^2)\, ds + \int_0^t \beta\, dw(s)\right) =$$

$$= \lim_{t\to\infty} \frac{1}{\Phi(t)} \left(\Phi(t) - \frac{1}{2}\beta^2 t\right) + \lim_{t\to\infty} \frac{1}{\Phi(t)} \int_0^t \beta\, dw(s) = 1 - \frac{\beta^2}{2K}$$

almost surely. □

Remark 6.6 The result of Theorem 6.2 is typical in the following sense. Consider the Cauchy problem of the ordinary differential equation corresponding to (6.24):

$$\begin{aligned} &d\mu(t) = \phi(t)\mu(t)dt, \qquad t \geq 0; \\ &\mu(0) = 1. \end{aligned} \tag{6.27}$$

Then its solution $\mu(\cdot)$ is such that

$$\ln \mu(t) = \Phi(t).$$

Therefore the result of Theorem 6.2 is equivalent to

$$\lim_{t\to\infty} \frac{\ln X(t)}{\ln \mu(t)} = 1 - \frac{\beta^2}{2K} \qquad \text{a.s.}$$

In other words, denoting $\psi(\cdot) = \log(\cdot)$, the solution of stochastic problem (6.24) is ψ-equivalent almost surely to the solution of deterministic problem (6.27). This kind of asymptotic equivalence is studied in [12] in more detail (also see [13]).

6.4.2 Rendleman–Bartter Model

Consider the Cauchy problem for the Rendleman–Bartter Model (6.5)

$$dX(t) = X(t)\phi(t)\,dt + X(t)\theta(t)\,dw(t), \qquad t \geq 0; \tag{6.28}$$
$$X(0) = b > 0.$$

This is a particular case of problem (6.8) with $g(x) = x$ and $\sigma(x) = x$. Here $\phi(\cdot)$ represents an expected instantaneous rate of change in the interest rate, $\theta(\cdot)$ is a volatility parameter, and $w(\cdot)$ is a Wiener process.

Like the population growth model discussed in Sect. 6.4.1 the function $\sigma(\cdot)$ is unbounded in the Rendleman–Bartter model, as well. To overcome this difficulty we use again the Itô formula for $\ln X(t)$ in (6.28) and obtain the following equation

$$d\ln X(t) = \left(\phi(t) - \frac{1}{2}\theta^2(t)\right) dt + \theta(t)\,dw(t)$$

which can be treated with the help of Theorem 6.1.

Theorem 6.3 *Let $\phi(\cdot)$ and $\theta(\cdot)$ be continuous functions and let $X(\cdot)$ be a solution of Cauchy problem* (6.28)*. Assume that*

$$\lim_{t\to\infty} \frac{1}{\Phi(t)} \int_0^t \theta^2(s)\,ds = K, \qquad K \in [0;\infty), \tag{6.29}$$

and

$$\sum_{k=0}^{\infty} \frac{\Phi(2^{k+1})}{\Phi^2(2^k)} < \infty. \tag{6.30}$$

If conditions (6.12) *and* (6.13) *are satisfied, then*

$$\lim_{t\to\infty} \frac{\ln X(t)}{\Phi(t)} = 1 - \frac{1}{2}K.$$

Proof Note that (6.17) follows from (6.29) and (6.30). Thus all assumptions of Theorem 6.1 hold, hence

$$\lim_{t\to\infty} \frac{1}{\Phi(t)} \int_0^t \sigma\theta(s)\,dw(s) = 0 \qquad \text{a.s.}$$

Therefore

$$\lim_{t\to\infty}\frac{\ln X(t)}{\Phi(t)}=\lim_{t\to\infty}\frac{1}{\Phi(t)}\left(\int_0^t(\phi(s)-\frac{1}{2}\theta^2(s))\,ds+\int_0^t\sigma\theta(s)\,dw(s)\right)$$
$$=\lim_{t\to\infty}\left(1-\frac{1}{2\Phi(t)}\int_0^t\theta^2(s)\,ds\right)+\lim_{t\to\infty}\frac{1}{\Phi(t)}\int_0^t\sigma\theta(s)\,dw(s)$$
$$=1-\frac{1}{2}K\qquad\text{a.s.}$$

6.4.3 *Asymptotic Behavior of Solutions of Stochastic Differential Equation (6.7)*

We consider Eq. (6.7) which obviously is a member of the class of Eq. (6.4) corresponding to the case of $\phi(t)\equiv 1$, $\sigma(x)\equiv 1$. Condition (6.17) in this case can be rewritten as follows

$$\sum_{k=0}^{\infty}\frac{\Theta(2^{k+1})}{\Phi^2(2^k)}=\frac{1}{4}\sum_{k=1}^{\infty}\frac{1}{2^{2k}}\int_0^{2^k}\sigma\theta^2(s)\,ds<\infty.$$

It can be simplified further, since

$$\sum_{k=1}^{\infty}\frac{1}{2^{2k}}\int_1^{2^k}\theta^2(s)\,ds=\sum_{k=1}^{\infty}\frac{1}{2^{2k}}\sum_{i=1}^{k}\int_{2^{i-1}}^{2^i}\theta^2(s)\,ds=\sum_{i=1}^{\infty}\int_{2^{i-1}}^{2^i}\theta^2(s)\,ds\sum_{k=i}^{\infty}\frac{1}{2^{2k}}$$
$$=\frac{4}{3}\sum_{i=1}^{\infty}\frac{1}{2^{2i}}\int_{2^{i-1}}^{2^i}\theta^2(s)\,ds\asymp\sum_{i=1}^{\infty}\int_{2^{i-1}}^{2^i}\frac{\theta^2(s)}{s^2}\,ds$$
$$=\int_1^{\infty}\frac{\theta^2(s)}{s^2}\,ds.$$

Theorem 6.1 reads as follows for the Cauchy problem of Eq. (6.7):

$$dX(t)=g(X(t))\,dt+\theta(t)\,dw(t),\qquad t\geq 0;\tag{6.31}$$
$$X(0)=b>0.$$

Corollary 6.2 *Let $g(\cdot)$ and $\theta(\cdot)$ be continuous positive functions and conditions (6.15) and (6.16) hold. We further assume that a solution of problem (6.31)*

satisfies condition (6.13)*. If*

$$\int_1^\infty \frac{\theta^2(s)}{s^2}\,ds < \infty,$$

then

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \theta(s)\,dw(s) = 0 \qquad a.s. \tag{6.32}$$

This result can be called the *weighted strong law of large numbers for the Wiener process*. The so-called (moment, in probability, almost sure) stability can also be obtained for solutions of problem (6.31) (see [24], also see [1, 2]).

Note that (6.32) does not involve the solution $X(\cdot)$ at all. On the other hand, limit result (6.32) can be used to obtain a precise asymptotic behavior. Following the technique developed in [13] one can conclude from (6.32) that, under some additional assumption imposed on the function $g(\cdot)$,

$$\lim_{t\to\infty} \frac{X(t)}{\mu(t)} = 1 \qquad \text{a.s.}$$

where $\mu(\cdot)$ is a solution of the Cauchy problem for the following ordinary differential equation

$$\begin{aligned} &d\mu(t) = g(\mu(t))dt, \qquad t \ge 0;\\ &\mu(0) = b. \end{aligned}$$

The additional condition mentioned above is

$$G(u) = \int_0^u \frac{du}{g(u)} \to \infty \qquad \text{as} \qquad x \to \infty,$$

$$G \text{ preserves the asymptotic equivalence.} \tag{6.33}$$

Functions with property (6.33) are called *pseudo-regularly varying* in [13] (more detail on relationships between pseudo-regularly varying functions and limit properties of solutions of stochastic differential equations is given in [13], also see [8, 10, 11]).

Acknowledgements Supported by the grants from Ministry of Education and Science of Ukraine (projects N 2105 ϕ and M/68-2018).

References

1. Appleby, A.D., Cheng, J.: On the asymptotic stability of a class of perturbed ordinary differential equations with weak asymptotic mean reversion. Electronic J. Qualitative Theory Differ. Equ. Proc. 9th Coll. **1**, 1–36 (2011)
2. Appleby, A.D., Cheng, J.: Rodkina, A.: Characterisation of the asymptotic behaviour of scalar linear differential equations with respect to a fading stochastic perturbation. Discrete Contin. Dyn. Syst. Suppl. **2011**, 79–90 (2011)
3. Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular Variation. Cambridge University Press, Cambridge (1987)
4. Black, F., Karasinski, P.: Bond and option pricing when short rates are lognormal. Financ. Anal. J. **47**, 52–59 (1991)
5. Black, F., Derman, E., Toy, W.: A one-factor model of interest rates and its application to treasury bond options. Financ. Anal. J. **46**, 24–32 (1990)
6. Buldygin, V.V., Pavlenkov, V.V.: Karamata theorem for regularly log-periodic functions. Ukr. Math. J. **64**, 1635–1657 (2013)
7. Buldygin, V.V., Tymoshenko, O.A.: On the exact order of growth of solutions of stochastic differential equations with time-dependent coefficients. Theory Stoch. Process. **16**, 12–22 (2010)
8. Buldygin, V.V., Klesov, O.I., Steinebach, J.G.: On some properties of asymptotically quasi-inverse functions and their applications. I. Theory Probab. Math. Stat. **70**, 9–25 (2003)
9. Buldygin, V.V., Klesov, O.I., Steinebach, J.G.: On factorization representations for Avakumović–Karamata functions with nondegenerate groups of regular points. Anal. Math. **30**, 161–192 (2004)
10. Buldygin, V.V., Klesov, O.I., Steinebach, J.G.: The PRV property of functions and the asymptotic behavior of solutions of stochastic differential equations. Theory Probab. Math. Stat. **72**, 63–78 (2004)
11. Buldygin, V.V., Klesov, O.I., Steinebach, J.G.: On some properties of asymptotically quasi-inverse functions and their applications. II. Theory Probab. Math. Stat. **71**, 63–78 (2004)
12. Buldygin, V.V., Klesov, O.I., Steinebach, J.G., Tymoshenko, O.A.: On the φ-asymptotic behavior of solutions of stochastic differential equations. Theory Stoch. Process. **14**, 11–30 (2008)
13. Buldygin, V.V., Indlekofer, K.-H., Klesov, O.I., Steinebach, J.G.: Pseudo-Regularly Varying Functions and Generalized Renewal Processes. Springer, Berlin (2018)
14. Chen, L.: Stochastic mean and stochastic volatility – a three-factor model of the term structure of interest rates and its application to the pricing of interest rate derivatives. Financ. Mark. Inst. Instrum. **5**, 1–88 (1996)
15. Cox, J.C., Ingersoll, J.E., Ross, S.A.: A theory of the term structure of interest rates. Econometrica **53**, 385–407 (1985)
16. D'Anna, A., Maio, A., Moauro, V.: Global stability properties by means of limiting equations. Nonlinear Anal. **4**(2), 407–410 (1980)
17. Gikhman, I.I., Skorokhod, A.V.: Stochastic Differential Equations. Springer, Berlin (1972)
18. Hull, J., White, A.: Pricing interest-rate derivative securities. Rev. Financ. Stud. **3**, 573–592 (1990)
19. Keller, G., Kersting, G., Rösler, U.: On the asymptotic behavior of solutions of stochastic differential equations. Z. Wahrsch. Geb. **68**(2), 163–184 (1984)
20. Kersting, G.: Asymptotic properties of solutions of multidimensional stochastic differential equations. Probab.Theory Relat. Fields **88**, 187–211 (1982)
21. Klesov, O.I.: Limit Theorems for Multi-Indexed Sums of Random Variables. Springer, Berlin (2014)
22. Klesov, O.I., Tymoshenko, O.A.: Unbounded solutions of stochastic differential equations with time-dependent coefficients. Ann. Univ. Sci. Budapest Sect. Comput. **41**, 25–35 (2013)

23. Klesov, O.I., Siren 'ka, I.I., Tymoshenko, O.A.: Strong law of large numbers for solutions of non-autonomous stochastic differential equations. Naukovi Visti NTUU KPI **4**, 100–106 (2017)
24. Mao, X.: Stochastic Differential Equations and Applications, 2nd edn. Woodhead Publishing, Cambridge (2010)
25. Mitsui, T.: Stability analysis of numerical solution of stochastic differential equations. Res. Inst. Math. Sci. Kyoto Univ. **850**, 124–138 (1995)
26. Øksendal, B.K.: Stochastic Differential Equations: An Introduction with Applications. Springer, Berlin (2003)
27. Rendleman, R., Bartter, B.: The pricing of options on debt securities. J. Financ. Quant. Anal. **15**, 11–24 (1980)
28. Samoĭlenko, A.M., Stanzhytskyi, O.M.: Qualitative and Asymptotic Analysis of Differential Equations with Random Perturbations. World Scientific Publishing, Hackensack, NJ (2011)
29. Seneta, E.: Regularly Varying Functions. Springer, Berlin (1976)
30. Strauss, A., Yorke, J.A.: On asymptotically autonomous differential equations. Math. Syst. Theory **1**, 175–182 (1967)
31. Taniguchi, T.: On sufficient conditions for nonexplosion of solutions to stochastic differential equations, J. Math. Anal. Appl. **153**, 549–561 (1990)
32. Tymoshenko, O.A.: Generalization of asymptotic behavior of non-autonomous stochastic differential equations, Naukovi Visti NTUU KPI **4**, 100–106 (2016)
33. Vašiček, O.: An equilibrium characterization of the term structure. J. Financ. Econ. **5**, 177–188 (1977)

Part II
Solid Mechanics

Chapter 7
Procedure of the Galerkin Representation in Transversely Isotropic Elasticity

Dimitri V. Georgievskii

Abstract An algorithm for splitting an equilibrium displacement equation system with bulk forces for a transversely isotropic linearly-elastic medium is described that leads to three uncoupled equations with certain canonical fourth-order differential operators in the three components of the displacement vector. It is shown that, in the special case of isotropy, the proposed algorithm is mathematically equivalent to the Galerkin representation, well-known in the theory of elasticity.

7.1 The Classic Galerkin Representation in Isotropic Elasticity

As is known, the expressions of a displacement vector $\mathbf{u} = u_i \mathbf{e_i}$ $(i = 1, 2, 3)$ in terms of vectors complying with equations that seem to be more simple than the Lamé equations

$$(\lambda + \mu)\text{grad}\,\text{div}\,\mathbf{u} + \mu\Delta\mathbf{u} + \rho\mathbf{F} = 0 \tag{7.1}$$

in isotropic elasticity (λ, μ are Lamé constants; ρ is a density; $\mathbf{F}$ is a mass force) are called representations of solutions of a boundary-value problem in elasticity theory. The Galerkin representation is one of these classic representations which reduce the Lamé operator to the biharmonic operator $\Delta\Delta$. Let us remind an essence of the Galerkin representation in isotropic elasticity.

By applying the operator div for two hands of (7.1) we receive

$$\Delta\text{div}\,\mathbf{u} = -\frac{\rho\,\text{div}\,\mathbf{F}}{\lambda + 2\mu}. \tag{7.2}$$

D. V. Georgievskii (✉)
Lomonosov Moscow State University, Moscow, Russian Federation
e-mail: georgiev@mech.math.msu.su

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_7

Then applying the operator Δ for two hands of (7.1) and taking into account the relation (7.2) we derive the system of non-coupled not uniform biharmonic equations

$$\mu\Delta\Delta\mathbf{u} = \frac{\lambda+\mu}{\lambda+2\mu}\rho\,\mathrm{grad}\,\mathrm{div}\,\mathbf{F} - \rho\Delta\mathbf{F}, \tag{7.3}$$

which may be written in operator form by means of the Galerkin representation

$$\Delta\Delta\mathbf{u} = \check{M}\cdot\mathbf{F}, \tag{7.4}$$

with the Galerkin tensor differential operator $\check{M}$:

$$\check{M} = \frac{(1+\nu)\rho}{(1-\nu)E}\mathrm{grad}\,\mathrm{div} - \frac{2}{E}(1+\nu)\rho\Delta, \tag{7.5}$$

where E is Young modulus, ν is Poisson ratio.

Solution of the system (7.4) is sought in the form

$$\mathbf{u} = \check{M}\cdot\Gamma, \tag{7.6}$$

where Γ is the Galerkin vector. Using permutability of two linear differential operators $\check{M}$ and Δ we derive the following sufficient conditions for compliance of the system (7.4):

$$\Delta\Delta\Gamma = \mathbf{F}. \tag{7.7}$$

Reduction of the Lamé equations system (7.1) to biharmonic equation (7.7) makes an essence of the Galerkin representation procedure in isotropic elasticity. The system (7.7) with respect to vector Γ is appeared to be simpler than another biharmonic system (7.4) with respect to vector $\mathbf{u}$ because a class of smoothness of mass forces in (7.7) is more broad than in (7.4).

In the capacity of known example we choose a mass force $\mathbf{F}$ in form of point loading at the origin of coordinates in unbounded three-dimensional elastic space: $\rho\mathbf{F} = \mathbf{P}\delta(\mathbf{x})$ where $\mathbf{P}$ is a value with dimension of force. It is naturally to find the solution of the following from (7.7) equations

$$\Delta\Delta\Gamma = \frac{\mathbf{P}}{\rho}\,\delta(\mathbf{x}) \tag{7.8}$$

in the form

$$\Gamma = \frac{Cr}{\rho}\,\mathbf{P}, \quad r = |\mathbf{x}|, \quad C = \mathrm{const}. \tag{7.9}$$

As $\Delta r = 2/r$ and $\Delta(1/r) = -4\pi\delta(\mathbf{x})$ in R^3 than $C = -1/(8\pi)$, so

$$\Gamma_i = -\frac{P_i r}{8\pi\rho}, \quad u_i = \frac{P_j(1+\nu)}{8\pi E(1-\nu)}\Big(\frac{x_i x_j}{r^3} + (3-4\nu)\frac{\delta_{ij}}{r}\Big). \tag{7.10}$$

The small Latin subscripts change from 1 to 3 while the large Latin subscripts encountered later from 1 to 2. Summation is carried out over a doubly repeated index.

Thus, on the basis of the solution (7.10) of the problem on point force action in unbounded elastic space (the Kelvin problem) one can explicitly write all components $U_i^{(k)}(\mathbf{x}, \boldsymbol{\xi})$ of the Kelvin displacement tensor:

$$U_i^{(k)}(\mathbf{x}, \boldsymbol{\xi}) = \frac{P(1+\nu)}{8\pi(1-\nu)E}\Big[\frac{(x_i - \xi_i)(x_k - \xi_k)}{r^3(\mathbf{x}, \boldsymbol{\xi})} + (3-4\nu)\frac{\delta_{ik}}{r(\mathbf{x}, \boldsymbol{\xi})}\Big], \tag{7.11}$$

where $r(\mathbf{x}, \boldsymbol{\xi}) = |\mathbf{x} - \boldsymbol{\xi}|$, $P = |\mathbf{P}|$.

7.2 Splitting of the System of Displacement Equations in Anisotropic Elasticity

The equilibrium equations in terms of displacements in anisotropic elasticity have the form

$$C_{ijkl}u_{k,lj} + X_i = 0, \tag{7.12}$$

where C_{ijkl} and X_i are the components of the elastic moduli tensor $\mathbf{C}^{(4)}$ and volume force vector $\mathbf{X}(\mathbf{x})$. We differentiate both sides of (7.12) with respect to x_m and x_n and multiply by the constant components D_{pnim} of the fourth rank tensor $\mathbf{D}^{(4)}$, from which we require that: (a) it has the same symmetry as $\mathbf{C}^{(4)}$, that is, $D_{pnim} = D_{npim} = D_{pnmi} = D_{impn}$; (b) physical dimension of its components coincides with dimension of elastic compliances.

Hence,

$$D_{pnmi}C_{ijkl}u_{k,ljmn} + D_{pnim}X_{i,mn} = 0. \tag{7.13}$$

The idea of choosing the tensor $\mathbf{D}^{(4)}$ using a known tensor $\mathbf{C}^{(4)}$ is involved in reducing equations (7.13) to the form

$$\check{L}_{(p)}u_p + D_{pnim}X_{i,mn} + Y_p = 0, \tag{7.14}$$

where $\check{L}_{(p)}$ are certain scalar canonical fourth-order operator; Y_p are known functions of coordinates. Because of increase in the order, system (7.14) unlike (7.13)

consists of three split equations with respect to u_1, u_2 and u_3 that are not connected with one another.

In case of elastic moduli tensor for isotropic medium $\mathbf{C}^{(4)}$ with the components

$$C_{ijkl} = \frac{E\nu}{(1-2\nu)(1+\nu)}\,\delta_{ij}\delta_{kl} + \frac{E}{2(1+\nu)}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \tag{7.15}$$

a tensor with components

$$D_{pnmi} = -\frac{(1+\nu)(3-2\nu)}{E(1-\nu)}\,\delta_{pn}\delta_{mi} + \frac{2(1+\nu)}{E}(\delta_{pm}\delta_{ni} + \delta_{pi}\delta_{nm}) \tag{7.16}$$

may be taken as the tensor $\mathbf{D}^{(4)}$. Equalities in (7.13) are then written in the following way:

$$u_{p,kkmm} + \frac{1+\nu}{E}\left(2X_{p,kk} - \frac{1}{1-\nu}X_{k,kp}\right) = 0. \tag{7.17}$$

Comparing equalities (7.17) and (7.14) we see that in isotropic medium $\check{L}_{(1)} = \check{L}_{(2)} = \check{L}_{(3)} = \Delta^2$ and $Y_p = 0$. It should be noted that the tensor $\mathbf{D}^{(4)}$ is not identical to the elastic compliance tensor $\mathbf{J}^{(4)}$ with the components

$$J_{pnmi} = -\frac{\nu}{E}\,\delta_{pn}\delta_{mi} + \frac{1+\nu}{2E}(\delta_{pm}\delta_{ni} + \delta_{pi}\delta_{nm}). \tag{7.18}$$

7.3 Transversely Isotropic Medium

Let us consider the case of transversely isotropic elastic medium with the rotational axes of symmetry x_3 [1, 2] (such materials are sometimes called transtropic [3]). This medium is defined by five material constants $\lambda_1, \ldots, \lambda_5$:

$$C_{ijkl} = \lambda_1\gamma_{ij}\gamma_{kl} + \lambda_2(\gamma_{ik}\gamma_{jl} + \gamma_{il}\gamma_{jk}) + \lambda_3(\gamma_{ij}\delta_{3k}\delta_{3l} + \gamma_{kl}\delta_{3i}\delta_{3j}) +$$

$$+ \lambda_4\delta_{3i}\delta_{3j}\delta_{3k}\delta_{3l} + \lambda_5(\gamma_{ik}\delta_{3j}\delta_{3l} + \gamma_{jk}\delta_{3i}\delta_{3l} + \gamma_{il}\delta_{3j}\delta_{3k} + \gamma_{jl}\delta_{3i}\delta_{3k}), \tag{7.19}$$

$$\gamma_{ij} = \delta_{1i}\delta_{1j} + \delta_{2i}\delta_{2j}. \tag{7.20}$$

It is conveniently to introduce the notation

$$\xi_1 = \lambda_1 + \lambda_2, \quad \xi_2 = \lambda_1 + 2\lambda_2, \quad \xi_3 = \lambda_3 + \lambda_5, \quad \xi_4 = \lambda_3 + \lambda_4 + \lambda_5.$$

We seek [4] the tensor $\mathbf{D}^{(4)}$ in the class of tensors that are similar by structure to $\mathbf{C}^{(4)}$:

$$D_{pnmi} = d_1\gamma_{pn}\gamma_{mi} + d_2(\gamma_{pm}\gamma_{ni} + \gamma_{pi}\gamma_{nm}) + d_3(\gamma_{pn}\delta_{3m}\delta_{3i} + \gamma_{mi}\delta_{3p}\delta_{3n}) +$$

$$d_4\delta_{3p}\delta_{3n}\delta_{3m}\delta_{3i} + d_5(\gamma_{pm}\delta_{3n}\delta_{3i} + \gamma_{nm}\delta_{3p}\delta_{3i} + \gamma_{pi}\delta_{3n}\delta_{3m} + \gamma_{ni}\delta_{3p}\delta_{3m}). \quad (7.21)$$

The Case $p = 3$ Let us substitute expressions (7.19) and (7.21) into Eq. (7.13) and initially write one of them when $p = 3$. The first term on the left-hand side of it has the form

$$D_{3nmi}C_{ijkl}u_{k,ljmn} = [\xi_2 d_3 + (\xi_2 + \xi_3)d_5]u_{K,KLL3} + [\lambda_5 d_3 + \xi_3 d_4 + \lambda_5 d_5]u_{K,K333} +$$

$$+ \lambda_5 d_5 u_{3,KKLL} + [\xi_3 d_3 + \lambda_5 d_4 + \xi_4 d_5]u_{3,KK33} + \lambda_4 d_4 u_{3,3333}. \quad (7.22)$$

The following obvious properties of the symbols γ_{ij} (7.13) are used:

$$\gamma_{ij}\gamma_{jk} = \gamma_{ik}, \quad \gamma_{kl}(*)_{kl} = (*)_{KK}. \quad (7.23)$$

We require that the normalization condition $\lambda_5 d_5 = 1$ and the constraint between the constants:

$$d_3 = -\frac{\xi_2 + \xi_3}{\xi_2}d_5, \quad d_4 = \frac{\lambda_5}{\xi_2}d_5 \quad (7.24)$$

are satisfied. The coefficients by $u_{K,KLL3}$ and $u_{K,K333}$ in (7.22) become equal to zero.

From equalities (7.13) and (7.22) we obtain the required equation for the displacement component u_3:

$$u_{3,KKLL} + B_1 u_{3,KK33} + B_2 u_{3,3333} + D_{3nim}X_{i,mn} = 0 \quad (7.25)$$

with the coefficients

$$B_1 = \frac{1}{\lambda_5}\left(\lambda_4 - \lambda_3\frac{\xi_3 + \lambda_5}{\xi_2}\right), \quad B_2 = \frac{\lambda_4}{\xi_2}. \quad (7.26)$$

Comparing equalities (7.14) and (7.25) we see that $Y_3 = 0$ as well as the canonical operator $\check{L}_{(3)}$ is

$$\check{L}_{(3)} = \frac{\partial^4}{\partial x_K \partial x_K \partial x_L \partial x_L} + B_1\frac{\partial^4}{\partial x_K \partial x_K \partial^2 x_3} + B_2\frac{\partial^4}{\partial^4 x_3}. \quad (7.27)$$

As in isotropic elasticity, the solution of (7.25) can be sought in the form

$$u_3 = D_{3nim}\Gamma_{i,mn} = (d_3 + d_5)\Gamma_{K,K3} + d_5\Gamma_{3,KK} + d_4\Gamma_{3,33}, \tag{7.28}$$

where $\Gamma_i(\mathbf{x})$ are the components of the unknown Galerkin vector. To satisfy equality (7.25), it is sufficient to require that

$$\check{L}_{(3)}\Gamma_i \equiv \Gamma_{i,KKLL} + B_1\Gamma_{i,KK33} + B_2\Gamma_{i,3333} + X_i = 0. \tag{7.29}$$

The Case p=P In a similar way to (7.22) we now write out the first term on the left-hand side of (7.13) (after substituting expressions (7.19) and (7.21) into it) for values $p = 1$ and $p = 2$:

$$D_{Pnmi}C_{ijkl}u_{k,ljmn} = [\xi_2 d_1 + (\xi_1 + \xi_2)d_2]u_{K,KLLP} +$$

$$+[\lambda_5 d_1 + \lambda_5 d_2 + \xi_3 d_3 + (\xi_1 + \xi_3)d_5]u_{K,K33P} +$$

$$+(\xi_3 d_1 + 2\xi_3 d_2 + \lambda_5 d_3 + \lambda_5 d_5)u_{3,LL3P} +$$

$$+ (\lambda_4 d_3 + \xi_4 d_5)u_{3,333P} + \lambda_2 d_2 u_{P,KKLL} + (\lambda_5 d_2 + \lambda_2 d_5)u_{P,LL33} + \lambda_5 d_5 u_{P,3333}. \tag{7.30}$$

In addition to constraints (7.24), we require that the constraints

$$d_1 = -\Big(1 + \frac{\xi_1}{\xi_2}\Big)\xi_5 d_5, \quad d_2 = \xi_5 d_5, \quad \xi_5 = \frac{\xi_1\xi_2 - \xi_3^2}{\xi_1\lambda_5} \tag{7.31}$$

are satisfied. Then the coefficients by $u_{K,KLLP}$ and $u_{K,K33P}$ in equality (7.30) vanish.

From equalities (7.13) and (7.30), we obtain two equations for u_P $(P = 1, 2)$

$$B_3 u_{P,KKLL} + B_4 u_{P,LL33} + u_{P,3333} + B_5 u_{3,LL3P} + B_6 u_{3,333P} + D_{Pnim}X_{i,mn} = 0 \tag{7.32}$$

with the coefficients

$$B_3 = \frac{\lambda_2}{\lambda_5}\xi_5, \quad B_4 = \frac{\lambda_2}{\lambda_5} + \xi_5, \quad B_5 = \Big(\frac{\lambda_2}{\lambda_5}\xi_5 - 1\Big)\frac{\xi_3}{\xi_2}, \quad B_6 = \Big(1 - \frac{\lambda_4}{\xi_2}\Big)\frac{\xi_3}{\lambda_5}. \tag{7.33}$$

Comparing (7.14) and (7.32), we write out the operators $\check{L}_{(P)}$ and the components Y_P that are known from the results for the case $p = 3$:

$$\check{L}_{(P)} = B_3 \frac{\partial^4}{\partial x_K \partial x_K \partial x_L \partial x_L} + B_4 \frac{\partial^4}{\partial x_K \partial x_K \partial^2 x_3} + \frac{\partial^4}{\partial^4 x_3}, \tag{7.34}$$

$$Y_P = (B_5 u_{3,LL} + B_6 u_{3,33})_{,3P}.$$

The operators $\check{L}_{(P)}$ (7.34) and $\check{L}_{(3)}$ (7.27) only differ in the values of the coefficients.

Now let us represent the solutions u_P of two equations (7.32) and the known functions $u_{3,P}$ in the form

$$u_P = D_{Pnim} \Gamma^*_{i,mn} = \eta_1 \Gamma^*_{K,KP} + d_2 \Gamma^*_{P,KK} + \eta_3 \Gamma^*_{3,3P} + d_5 \Gamma^*_{P,33}, \tag{7.35}$$

$$u_{3,P} = D_{Pnim} \Phi_{i,mn3} = \eta_1 \Phi_{K,K3P} + d_2 \Phi_{P,KK3} + \eta_3 \Phi_{3,33P} + d_5 \Phi_{P,333}, \tag{7.36}$$

where $\eta_1 = d_1 + d_3$, $\eta_3 = d_3 + d_5$, and Γ^*_i are the components of a further Galerkin vector that differ from Γ_i. The known functions

$$\Phi_1 = \Phi_2 = 0, \quad \Phi_3 = \frac{1}{\eta_3} \int \int u_3 \, dx_3 dx_3 \tag{7.37}$$

can be taken as the functions Φ_i. Substituting (7.35) and (7.36) into (7.32), we derive the sufficient conditions for which these equations are satisfied:

$$B_3 \Gamma^*_{i,KKLL} + B_4 \Gamma^*_{i,LL33} + \Gamma^*_{i,3333} + B_5 \Phi_{i,LL33} + B_6 \Phi_{i,3333} + X_i = 0. \tag{7.38}$$

By virtue of (7.37), conditions (7.38) can also be written as follows:

$$B_3 \Gamma^*_{P,KKLL} + B_4 \Gamma^*_{P,LL33} + \Gamma^*_{P,3333} + X_P = 0. \tag{7.39}$$

$$B_3 \Gamma^*_{3,KKLL} + B_4 \Gamma^*_{3,LL33} + \Gamma^*_{3,3333} + \frac{B_5}{\eta_3} u_{3,KK} + \frac{B_6}{\eta_3} u_{3,33} + X_3 = 0. \tag{7.40}$$

Thus, in relations (7.24) and (7.31), where the equality $d_5 = 1/\lambda_5$ has to be taken into account, all the material constants $d_1, \ldots, d_5$ of the tensor $\mathbf{D}^{(4)}$ participating in the extended Galerkin representation have been found. The equilibrium equations in terms of displacements for a transversely isotropic medium are reduced to three equations (7.25) and (7.32) that are not connected to one another, containing the fourth-order operators $\check{L}_{(3)}$ (7.27) and $\check{L}_{(P)}$ (7.34). By introducing two Galerkin vectors with the components Γ_i (7.28) and Γ^*_i (7.35), the equations can be written in the simpler form (7.29), (7.39), (7.40). The volume (mass) forces occurring in the initial equations (7.12), rather than its derivatives with respect to coordinates, are the discontinuities. This is important if the loads are concentrated at one point as well as possess some similar singular character [5].

As a test of all mentioned relations we take the limit of an anisotropic elastic medium with Lamé constants λ and μ. Putting

$$\lambda_1 = \lambda_3 = \lambda, \quad \lambda_2 = \lambda_5 = \mu, \quad \lambda_4 = \lambda + 2\mu, \tag{7.41}$$

in accordance with the adopted notation we obtain

$$\xi_1 = \xi_3 = \lambda + \mu, \quad \xi_2 = \lambda + 2\mu, \quad \xi_4 = 2\lambda + 3\mu, \quad \xi_5 = 1. \tag{7.42}$$

After simplification we will have

$$d_1 = d_3 = -\frac{2\lambda + 3\mu}{(\lambda + 2\mu)\mu} = -\frac{(1+\nu)(3-2\nu)}{E(1-\nu)}, \quad d_2 = d_5 = \frac{1}{\mu} = \frac{2(1+\nu)}{E},$$

$$d_4 = d_1 + 2d_2 = \frac{1}{\lambda + 2\mu},$$

$$B_1 = B_4 = 2, \quad B_2 = B_3 = 1, \quad B_5 = B_6 = 0, \quad \check{L}_{(i)} = \Delta^2, \quad i = 1, 2, 3.$$

The components D_{pnmi} are identical to the components determined using (7.16) and three equilibrium equations (7.25) and (7.32) are identical to three equations (7.17) for an isotropic medium.

References

1. Lekhnitskii, S.G.: Theory of Elasticity of an Anisotropic Body. Holden-Day, San-Francisco (1963)
2. Pobedria, B.E.: Numerical Methods in Theory of Elasticity and Plasticity. Moscow State University Publication, Moscow (1995) (in Russian)
3. Ashkenazi, E.K., Ganov, E.V.: Anisotropy of Structural Materials. Handbook. Mashinostroenie, Leningrad (1972) (in Russian)
4. Georgievskii, D.V.: An extended Galerkin representation for a transversely isotropic linearly elastic medium. J. Appl. Math. Mech. **79**(6), 618–621 (2015)
5. Georgievskii, D.V.: The Galerkin tensor operator, reduction to tetraharmonic equations, and their fundamental solutions. Dokl. Phys. **60**(8), 364–367 (2015)

Chapter 8
Symmetries and Fundamental Solutions of Displacement Equations for a Transversely Isotropic Elastic Medium

Alexander V. Aksenov

Abstract A fourth-order linear elliptic partial differential equation describing the displacements of a transversely isotropic linear elastic medium is considered. Its symmetries and the symmetries of an inhomogeneous equation with a delta function on the right-hand side are found. The latter symmetries are used to construct an invariant fundamental solution of the original equation in terms of elementary functions.

8.1 Introduction and the Main Result

In [1], the system of displacement equilibrium equations for a transversely isotropic linear elastic medium is reduced to a system of three linear inhomogeneous equations for three displacement components. The homogeneous equations are associated with canonical linear partial differential equations of the fourth order. These canonical equations are a generalization of the biharmonic equation describing the displacements of an isotropic linear elastic medium. To find the displacements of a transversely isotropic linear elastic medium subjected to a given body force, we need to know fundamental solutions of the canonical equations.

Note that reductions of the system of displacement equations in 3D elasticity to systems of higher order equations based on operators that are more suitable for a numerical-analytical study than the Lamé operator are called representations of the solution of the elasticity problem and are described in the classical theory. Specifically, a reduction to tetraharmonic equations was discussed in [2].

Fundamental solutions of linear partial differential equations are frequently invariant under transformations admitted by the original equation [3]. Below, a fundamental solution is constructed using the algorithm from [4] proposed for finding fundamental solutions of linear partial differential equations. The algorithm

A. V. Aksenov (✉)
Lomonosov Moscow State University, Moscow, Russian Federation

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_8

makes use of the symmetries admitted by a linear partial differential equation with a delta function on its right-hand side. Let us briefly describe the main result of this work. Consider the p th-order linear partial differential equation

$$Lu \equiv \sum_{\alpha=1}^{p} A_\alpha(x) D^\alpha u = 0\,, \qquad x \in R^m\,. \tag{8.1}$$

Here, the standard notation is used: $\alpha = (\alpha_1, \ldots, \alpha_m)$ is a multi-index with nonnegative integer components, $\alpha = \alpha_1 + \cdots + \alpha_m$, and

$$D^\alpha \equiv \left(\frac{\partial}{\partial x^1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x^m}\right)^{\alpha_m}.$$

The fundamental solutions of Eq. (8.1) are solutions of the equation

$$Lu = \delta(x - x_0)\,. \tag{8.2}$$

It was shown in [5] that Eq. (8.1) with $p \geq 2$ and $m \geq 2$ can only admit symmetry operators of the form

$$Z = \sum_{i=1}^{m} \xi^i(x)\,\frac{\partial}{\partial x^i} + \eta(x,u)\,\frac{\partial \eta}{\partial u}, \qquad \frac{\partial^2 \eta}{\partial u^2} = 0\,.$$

The basic Lie algebra of symmetry operators of Eq. (8.1) regarded as a vector space is a direct sum of two subalgebras: one consisting of operators of the form

$$X = \sum_{i=1}^{m} \xi^i(x)\,\frac{\partial}{\partial x^i} + \zeta(x)\,u\,\frac{\partial}{\partial u} \tag{8.3}$$

and the infinite-dimensional subalgebra generated by the operators

$$X_\infty = \varphi(x)\,\frac{\partial}{\partial u}\,, \tag{8.4}$$

where $\varphi(x)$ is an arbitrary solution of Eq. (8.1). Note that operators (8.4) are symmetry operators of Eq. (8.2). In what follows, we consider only symmetry operators of form (8.3). Let $\underset{p}{X}$ denote an extension of order p of symmetry operator (8.3).

Proposition 8.1 *The infinitesimal operator given by (8.3) is a symmetry operator of Eq. (8.1) if and only if there exists a function $\lambda = \lambda(x)$ satisfying the identity*

$$\underset{p}{X}(Lu) \equiv \lambda(x)\,Lu \tag{8.5}$$

for any function $u = u(x)$ from the domain of Eq. (8.1).

Theorem 8.1 *The Lie algebra of symmetry operators of Eq. (8.2) is a subalgebra of the Lie algebra of symmetry operators of Eq. (8.1) and is defined by the relations*

$$\begin{aligned} &\xi^i(x_0) = 0, \qquad i = 1, \dots, m\,, \\ &\lambda(x_0) + \sum_{i=1}^{m} \frac{\partial \xi^i(x_0)}{\partial x^i} = 0\,. \end{aligned} \tag{8.6}$$

Let us describe an algorithm for finding fundamental solutions by applying symmetries [4]:

1. Find a general symmetry operator of Eq. (8.1) and the corresponding function $\lambda(x)$ satisfying identity (8.5).
2. Use this operator and relations (8.6) to obtain the basis for the Lie algebra of symmetry operators of Eq. (8.2).
3. Construct invariant fundamental solutions with the help of the symmetries of Eq. (8.2).
4. Obtain new fundamental solutions from the known ones with the help of the symmetries of Eq. (8.2) (production of solutions).

Remark 8.1 To find generalized invariant fundamental solutions, we need to search for invariants in the class of generalized functions.

Example 8.1 Consider a two-dimensional biharmonic equation

$$\Delta\Delta u \equiv \frac{\partial^4 u}{\partial x^4} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} = 0\,. \tag{8.7}$$

Fundamental solutions of the two-dimensional biharmonic equation satisfy the equation

$$\frac{\partial^4 u}{\partial x^4} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} = \delta(x, y)\,. \tag{8.8}$$

The finite part of the basis of the Lie algebra of symmetry operators of Eq. (8.7) is given by [6]

$$\begin{aligned} &X_1 = \frac{\partial}{\partial x}\,, \qquad X_2 = \frac{\partial}{\partial y}\,, \qquad X_3 = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\,, \\ &X_4 = y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\,, \qquad X_5 = (x^2 - y^2)\frac{\partial}{\partial x} + 2xy\frac{\partial}{\partial y} + 2xu\frac{\partial}{\partial u}\,, \\ &X_6 = 2xy\frac{\partial}{\partial x} + (y^2 - x^2)\frac{\partial}{\partial y} + 2yu\frac{\partial}{\partial u}\,, \qquad X_7 = u\frac{\partial}{\partial u}\,. \end{aligned}$$

To find the symmetry operators admitted by Eq. (8.8) we write the general form of the symmetry operator admitted by Eq. (8.7) as $X = \sum a_i X_i \quad (i = 1, \ldots, 7)$ or

$$X = \big[a_1 + a_3 x + a_4 y + a_5(x^2 - y^2) + 2a_6 xy\big]\frac{\partial}{\partial x} + \\ + \big[a_2 + a_3 y - a_4 x + 2a_5 xy + a_6(y^2 - x^2)\big]\frac{\partial}{\partial y} + \\ + (2a_5 x + 2a_6 y + a_7)u\frac{\partial}{\partial u}\,, \tag{8.9}$$

where a_i $(i = 1, \ldots, 7)$ are arbitrary constants. Symmetry operator (8.9) corresponds to the function $\lambda(x, y)$

$$\lambda = a_7 - 4a_3 - 6a_5 x - 6a_6 y\,.$$

Then using theorem 1 we find

$$a_1 = a_2 = 0\,, \qquad a_7 - 2a_3 = 0\,.$$

Proposition 8.2 *Equation (8.8) admits the following basis of Lie algebra symmetry operators*

$$Y_1 = y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\,, \qquad Y_2 = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + 2u\frac{\partial}{\partial u}\,,$$

$$Y_3 = (x^2 - y^2)\frac{\partial}{\partial x} + 2xy\frac{\partial}{\partial y} + 2xu\frac{\partial}{\partial u}\,,$$

$$Y_4 = 2xy\frac{\partial}{\partial x} + (y^2 - x^2)\frac{\partial}{\partial y} + 2yu\frac{\partial}{\partial u}\,.$$

The fundamental solution of the two-dimensional biharmonic equation is known to be [7]

$$u = \frac{x^2 + y^2}{16\pi}\ln(x^2 + y^2)\,. \tag{8.10}$$

Solution (8.10) is invariant under the one-parameter group of rotations corresponding to the symmetry operator Y_1. The symmetry operator Y_2 generates a one-parameter group of inhomogeneous scaling transformations $x' = e^a x$, $t' = e^a t$, $u' = e^{2a} u$, where a is a group parameter. Under the action of this one-parameter group the fundamental solution of (8.10) is transformed into fundamental solution

$$u = \frac{x^2 + y^2}{16\pi}\big[\ln(x^2 + y^2) + 2a\big]\,.$$

Consider the symmetry operator Y_3. The symmetry operator Y_3 generates a one-parameter transformation group

$$\begin{aligned} x' &= \frac{x - a(x^2 + y^2)}{1 - 2ax + a^2(x^2 + y^2)}, \\ y' &= \frac{y}{1 - 2ax + a^2(x^2 + y^2)}, \\ u' &= \frac{u}{1 - 2ax + a^2(x^2 + y^2)}, \end{aligned} \tag{8.11}$$

where a is a group parameter. Under the action of this one-parameter group the fundamental solution of (8.10) is transformed into nontrivial fundamental solution

$$u = \frac{[x - a(x^2 + y^2)]^2 + y^2}{16\pi[1 - 2ax + a^2(x^2 + y^2)]} \cdot \ln\left[\frac{\left(x - a(x^2 + y^2)\right)^2 + y^2}{\left(1 - 2ax + a^2(x^2 + y^2)\right)^2}\right]. \tag{8.12}$$

Similarly, we can consider the symmetry operator Y_4. The symmetry operator Y_4 corresponds to the one-parameter transformation group

$$\begin{aligned} x' &= \frac{x}{1 - 2ay + a^2(x^2 + y^2)}, \\ y' &= \frac{y - a(x^2 + y^2)}{1 - 2ay + a^2(x^2 + y^2)}, \\ u' &= \frac{u}{1 - 2ay + a^2(x^2 + y^2)}, \end{aligned} \tag{8.13}$$

where a is a group parameter. Under the action of this one-parameter group the fundamental solution of (8.10) is transformed into nontrivial fundamental solution

$$u = \frac{1}{16\pi} \cdot \frac{x^2 + \left(y - a(x^2 + y^2)\right)^2}{1 - 2ay + a^2(x^2 + y^2)} \cdot \ln\left[\frac{x^2 + \left(y - a(x^2 + y^2)\right)^2}{\left(1 - 2ay + a^2(x^2 + y^2)\right)^2}\right]. \tag{8.14}$$

Remark 8.2 We can consider the composition of transformations (8.11) and (8.13). Then instead of one-parameter families of fundamental solutions (8.12) and (8.14) one can obtain a nontrivial two-parameter family of fundamental solutions.

The main result of this paper is the construction, in terms of elementary functions, of an invariant fundamental solution to the equation of a transversely isotropic linear elastic medium.

8.2 The Basic Equations

Consider the following fourth-order linear differential equations, which were introduced in [1]:

$$\begin{aligned} &L_1 u \equiv u_{xxxx} + 2u_{xxyy} + u_{yyyy} + B_1(u_{xxzz} + u_{yyzz}) + B_2 u_{zzzz} = 0\,, \\ &L_2 u \equiv B_3(u_{xxxx} + 2u_{xxyy} + u_{yyyy}) + B_4(u_{xxzz} + u_{yyzz}) + u_{zzzz} = 0\,. \end{aligned} \tag{8.15}$$

Here B_1, B_2, B_3, B_4, and are positive constants characterizing a linear elastic medium. The fundamental solutions of Eq. (8.15) are solutions of the equations

$$L_1 u = \delta(x)\delta(y)\delta(z)\,, \qquad L_2 u = \delta(x)\delta(y)\delta(z)\,. \tag{8.16}$$

Let us show that Eqs. (8.15) and (8.16) can be reduced to identical equations by changing variables. For this purpose, in the equations for the differential operator L_1, we pass to the new variables

$$\bar{z} = \frac{z}{\sqrt[4]{B_2}}\,, \qquad \bar{u} = \sqrt[4]{B_2}u\,.$$

After omitting the bars over the new variables, the corresponding equations (8.15) and (8.16) become

$$L_3 u \equiv u_{xxxx} + 2u_{xxyy} + u_{yyyy} + b(u_{xxzz} + u_{yyzz}) + u_{zzzz} = 0\,, \tag{8.17}$$

$$L_3 u = \delta(x)\delta(y)\delta(z)\,. \tag{8.18}$$

Here, $b = B_1/\sqrt{B_2}$. Similarly, by changing to the variables

$$\bar{x} = \frac{x}{\sqrt[4]{B_3}}\,, \qquad \bar{y} = \frac{y}{\sqrt[4]{B_3}}\,, \qquad \bar{u} = \sqrt{B_3}u$$

the equations for the differential operator L_2 are reduced to Eq. (8.17) and (8.18) with $b = B_4/\sqrt{B_3}$.

Assume that Eq. (8.17) is elliptic. Then it must hold that $b \geq 2$. In what follows, Eq. (8.17) is considered the basic equation. The axisymmetric solutions of Eq. (8.17) satisfy the equation

$$L_4 u \equiv u_{rrrr} + b u_{rrzz} + u_{zzzz} + \frac{2}{r}u_{rrr} + \frac{b}{r}u_{rzz} - \frac{1}{r^2}u_{rr} + \frac{1}{r^3}u_r = 0\,, \tag{8.19}$$

while the axisymmetric fundamental solutions (or axisymmetric solutions of Eq. (8.18)) satisfy the equation

$$r L_4 u = \frac{1}{\pi}\,\delta(r)\delta(z)\,. \tag{8.20}$$

Here $r = \sqrt{x^2 + y^2}$, and

$$\int_0^\infty \delta(r)dr = \frac{1}{2}\,.$$

Equation (8.20) can be rewritten in conservative form as

$$\left(ru_{rrr} + bru_{rzz} + u_{rr} - \frac{1}{r}u_r\right)_r + \left(ru_{zzz}\right)_z = \frac{1}{\pi}\,\delta(r)\delta(z)\,. \tag{8.21}$$

8.3 Symmetries of the Basic Equations

The symmetries of Eq. (8.17) can be found using the symmetry-finding algorithm from [3].

Proposition 8.3 *Equation (8.17) with an arbitrary parameter admits the following basis of the Lie algebra of symmetry operators:*

$$X_1 = \frac{\partial}{\partial x}\,, \qquad X_2 = \frac{\partial}{\partial y}\,, \qquad X_3 = \frac{\partial}{\partial z}\,,$$

$$X_4 = y\,\frac{\partial}{\partial x} - x\,\frac{\partial}{\partial y}\,, \qquad X_5 = x\,\frac{\partial}{\partial x} + y\,\frac{\partial}{\partial y} + z\,\frac{\partial}{\partial z}\,,$$

$$X_6 = u\,\frac{\partial}{\partial u}\,, \qquad X_\infty = \varphi(x, y, z)\,\frac{\partial}{\partial u}\,.$$

For $b = 2$, the basis of the Lie algebra is supplemented with the symmetry operators

$$X_7 = z\,\frac{\partial}{\partial x} - x\,\frac{\partial}{\partial z}\,, \qquad X_8 = z\,\frac{\partial}{\partial y} - y\,\frac{\partial}{\partial z}\,,$$

$$X_9 = (x^2 - y^2 - z^2)\,\frac{\partial}{\partial x} + 2xy\,\frac{\partial}{\partial y} + 2xz\,\frac{\partial}{\partial z} + xu\,\frac{\partial}{\partial u}\,,$$

$$X_{10} = 2xy\,\frac{\partial}{\partial x} + (y^2 - x^2 - z^2)\,\frac{\partial}{\partial y} + 2yz\,\frac{\partial}{\partial z} + yu\,\frac{\partial}{\partial u}\,,$$

$$X_{11} = 2xz\,\frac{\partial}{\partial x} + 2yz\,\frac{\partial}{\partial y} + (z^2 - x^2 - y^2)\,\frac{\partial}{\partial z} + zu\,\frac{\partial}{\partial u}\,.$$

Here, $u = \varphi(x, y, z)$ is an arbitrary solution of Eq. (8.17).

To find the symmetries of Eq. (8.18) we use the results of [4]. Using the finite-dimensional part of the Lie algebra of symmetry operators of Eq. (8.17), we consider

the general symmetry operator

$$X = \sum_{i=1}^{6} a_i X_i \,.$$

Here, a_i $(i = 1, \ldots, 6)$ are arbitrary constants.

Proposition 8.4 *It is true that*

$$\underset{4}{X} L_3 u = (a_6 - 4a_5) L_3 u \,.$$

Then, using Theorem 8.1, we find that

$$a_1 = a_2 = a_3 = 0\,, \qquad a_5 - a_6 = 0\,.$$

Proposition 8.5 *For an arbitrary parameter b, Eq. (8.18) admits the following basis of the Lie algebra of symmetry operators:*

$$Y_1 = y\,\frac{\partial}{\partial x} - x\,\frac{\partial}{\partial y}\,, \qquad Y_2 = x\,\frac{\partial}{\partial x} + y\,\frac{\partial}{\partial y} + z\,\frac{\partial}{\partial z} + u\,\frac{\partial}{\partial u}\,. \tag{8.22}$$

Remark 8.3 It can also be shown that Eq. (8.18) admits symmetry operators (8.22), the symmetry operator

$$Y_3 = z\,\frac{\partial}{\partial x} - x\,\frac{\partial}{\partial z}\,, \tag{8.23}$$

and the symmetry operators X_8, X_9, X_{10}, and X_{11}.

8.4 Fundamental Solution

Let us find a solution of Eq. (8.17) that is invariant under symmetry operators (8.22). The invariants of the admitted transformation group are $J_1 = r^2/z^2 = \tau$ and $J_2 = u/z$. Then an invariant solution is sought in the form

$$u = z f(\tau)\,. \tag{8.24}$$

Substituting (8.24) into Eq. (8.17) (or Eq. (8.19)), we obtain the fourth-order ordinary differential equation

$$\begin{aligned} &4\tau^2(\tau^2 + b\tau + 1)\frac{d^4 f}{d\,\tau^4} + 2\tau(14\tau^2 + 11b\tau + 8)\frac{d^3 f}{d\,\tau^3} + \\ &+ (39\tau^2 + 22b\tau + 8)\frac{d^2 f}{d\,\tau^2} + 2(3\tau + b)\frac{d\,f}{d\,\tau} = 0\,. \end{aligned} \tag{8.25}$$

Proposition 8.6 *The ordinary differential equation (8.25) has the following fundamental set of solutions:*

$$f_1 = 1\,,$$

$$f_2 = \sqrt{\frac{\tau}{a}+1} - \operatorname{arcoth}\sqrt{\frac{\tau}{a}+1} + \sqrt{a\tau+1} - \operatorname{arcoth}\sqrt{a\tau+1}\,,$$

$$f_3 = \frac{a}{a^2-1}\left(\sqrt{\frac{\tau}{a}+1} - \operatorname{arcoth}\sqrt{\frac{\tau}{a}+1} - \sqrt{a\tau+1} + \operatorname{arcoth}\sqrt{a\tau+1}\right),$$

$$f_4 = \frac{a}{a^2-1}\left(\sqrt{\frac{\tau}{a}+1}\operatorname{arcoth}\sqrt{\frac{\tau}{a}+1} - \frac{1}{2}\operatorname{arcoth}^2\sqrt{\frac{\tau}{a}+1} - \right.$$
$$\left. - \sqrt{a\tau+1}\operatorname{arcoth}\sqrt{a\tau+1} + \frac{1}{2}\operatorname{arcoth}^2\sqrt{a\tau+1}\right),$$

where the parameter a satisfies the relation $b = a + 1/a$.

Consider the general solution of Eq. (8.25):

$$f = \sum_{i=1}^{4} c_i f_i\,, \tag{8.26}$$

where c_i $(i = 1, \ldots, 4)$ are arbitrary constants. Among solutions (8.26), we find ones that take finite values, together with their first derivatives, at $\tau = 0$.

Proposition 8.7 *As $\tau \to 0$,we obtain*

$$f = \left[c_2 + \frac{a\ln a}{2(a^2-1)}\,c_4\right]\ln\tau + O(1)\,,$$

$$\frac{df}{d\tau} = \left[c_2 + \frac{a\ln a}{2(a^2-1)}\,c_4\right]\frac{1}{\tau} + \frac{c_4}{8}\ln\tau + O(1)\,.$$

It follows that $c_2 = 0$ and $c_4 = 0$. Assume also that $c_1 = 0$. As a result, we obtain the following one parameter family of solutions to Eq. (8.25):

$$f = \frac{c_3 a}{a^2-1}\left(\sqrt{\frac{\tau}{a}+1} - \sqrt{a\tau+1} - \operatorname{arcoth}\sqrt{\frac{\tau}{a}+1} + \operatorname{arcoth}\sqrt{a\tau+1}\right).$$

Then, using (8.24) yields a one-parameter family of solutions to Eq. (8.17) (or to Eq. (8.19)):

$$u = \frac{c_3 a}{a^2-1}\left(\sqrt{\frac{r^2}{a}+z^2} - \sqrt{ar^2+z^2} - z\cdot \operatorname{arcoth}\frac{\sqrt{\frac{r^2}{a}+z^2}}{z} + \right. \\ \left. + z\cdot \operatorname{arcoth}\frac{\sqrt{ar^2+z^2}}{z}\right). \tag{8.27}$$

Let us show that solutions (8.27) contain a fundamental one. For this purpose, both sides of Eq. (8.21) are integrated over the rectangular domain $\Pi = \{0 \leqslant r \leqslant r_0,\ -z_1 \leqslant z \leqslant z_2,\ r_0 > 0,\ z_1 > 0,\ z_2 > 0\}$. By using the Stokes formula, the integral on the left-hand side can be written in terms of an integral along the boundary of Π. Then solution (8.27) is substituted into the resulting integrand. Finally, we find that $c_3 = 1/(4\pi)$.

Below is the main result of this work.

Theorem 8.2 *The fundamental solution of Eq. (8.17) can be written as*

$$u_f = \frac{a}{4\pi(a^2-1)}\left[\sqrt{\frac{r^2}{a}+z^2} - \sqrt{ar^2+z^2} + \right. \\ \left. + \frac{z}{2}\ln\frac{\left(\sqrt{\frac{r^2}{a}+z^2}-z\right)\left(\sqrt{ar^2+z^2}+z\right)}{\left(\sqrt{\frac{r^2}{a}+z^2}+z\right)\left(\sqrt{ar^2+z^2}-z\right)}\right]. \tag{8.28}$$

Remark 8.4 For $a = 1$ (or $b = 2$), the fundamental solution (8.28) becomes

$$u_{\mathrm{f}} = -\frac{1}{8\pi}\sqrt{r^2+z^2}$$

and coincides with the fundamental solution of the three-dimensional biharmonic equation [7].

Remark 8.5 When $b = 2$, the construction of a fundamental solution based on symmetries is especially effective. Specifically, the solution of Eq. (8.17) invariant under symmetry operators (8.22) and (8.23) is immediately determined up to a multiplicative constant and is given by

$$u = c\sqrt{r^2+z^2}\,. \tag{8.29}$$

Proceeding as described above, we find that (8.29) is a fundamental solution with $c = -1/(8\pi)$.

8.5 Conclusion

The main result of this work is the construction (in terms of elementary functions) of an invariant fundamental solution to the equation of a transversely isotropic linear elastic medium. To conclude, we note that the symmetry approach can also be effectively used to construct fundamental solutions of linear partial differential equations with variable coefficients and for higher-order equations.

References

1. Georgievskii, D.V.: An extended Galerkin representation for a transversely isotropic linearly elastic medium. J. Appl. Math. Mech. **79**(6), 618–621 (2015)
2. Georgievskii, D.V.: The Galerkin tensor operator, reduction to tetraharmonic equations, and their fundamental solutions. Dokl. Phys. **60**(8), 364–367 (2015)
3. Ovsiannikov, L.V.: Group Analysis of Differential Equations. Academic, New York (1982)
4. Aksenov, A.V.: Symmetries of linear partial differential equations and fundamental solutions. Dokl. Math. **51**(3), 329–331 (1995)
5. Bluman, G.: Simplifying the form of Lie groups admitted by a given differential equation. J. Math. Anal. Appl. **145**(1), 52–62 (1990)
6. Bluman, G., Anco, S.: Symmetry and Integration Methods for Differential Equations. Springer, Berlin (2002)
7. Vladimirov, V.S.: Generalized Functions in Mathematical Physics. Mir, Moscow (1979)

Chapter 9
Modification of Hydrodynamic and Acoustic Fields Generated by a Cavity with Fluid Suction

Volodymyr G. Basovsky, Iryna M. Gorban, and Olha V. Khomenko

Abstract The hybrid numerical technique coupled with the vortex method for simulation of viscous incompressible flow and the Ffowcs William-Hawkings acoustic analogy is applied to the investigation of hydrodynamic and acoustic fields generated by a two-dimensional open cylindrical cavity. The problem is considered for a thin laminar boundary layer before the cavity and with the Reynolds number of $Re = 2 \cdot 10^4$, based on the cavity chord. The obtained results indicate that the cavity flow oscillates in the shear-layer mode and radiates a dipole in the far acoustic field so that the sound intensity in the backward direction is higher than in the forward direction. The effectiveness of controlling of the flow oscillations by applying steady suction through the rear cavity wall is studied. The results show that the suction allows us to localize the vortical flow inside the cavity when saving the mode of self-sustained oscillations in the shear layer. The vortices generated in the shear layer do not hit the trailing edge now but are absorbed by the suction causing the rise of pressure fluctuations in the vicinity of suction point. As a result, the obtained levels of radiated sound are much higher than in the uncontrolled cavity flow. The obtained positive effect of the suction on the cavity flow is that it suppresses the pressure fluctuations on the wall portion behind the cavity that leads to stabilization of the attached boundary layer.

9.1 Introduction

Cavity flows are found in various engineering devices and means of transport, including aircrafts, submarines and ground vehicles. The cavity may be either a design component or installed for creation of a special near-wall flow pattern. When

V. G. Basovsky · I. M. Gorban (✉)
Institute of Hydromechanics, National Academy of Sciences of Ukraine, Kyiv, Ukraine
e-mail: basovsky@ukr.net

O. V. Khomenko
Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
e-mail: olgkhomenko@ukr.net

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_9

the purpose of a flow control is intensification of heat and mass transfer, cavities are arranged on the surface for additional turbulization in the flow [1]. On the other hand, lift enhancement requires localization of a large vortex inside the cavity as in the Kasper wing [2].

Numerous technical applications of cavities have stimulated a large number of studies devoted to cavity flows. Note that canonical problem in this area is defined by a rectangular cavity located on the flat wall. Firstly, the cavities were classified into deep and shallow according to the ratio of cavity length to its depth [3]. Then shallow cavities were separated into open, transitional and closed basing on the pressure distribution along the cavity floor [4]. In the majority of works, flows near open shallow cavities have been investigated because those are the most relevant in practical applications. Gharib and Roshko [5] revealed that the flow regime in such a cavity is determined by the ratio between the cavity span L and the displacement thickness θ of the oncoming boundary layer. It was found that under a wide range of these parameters open cavity flows operate in the regime of self-sustained oscillations, which can be considered as one of the most important sources of the noise generated by the flow over a cavity.

The mechanism of self-sustained oscillations in a cavity flow has been described in details by Rockwell and Naudascher [6], who denoted its dependence on the Mach number. In compressible flows, there is a feedback loop between hydrodynamic and acoustic disturbances when the shear layer instability generates the pressure wave near the rear cavity edge and this wave propagates upstream amplifying disturbances in the shear layer. The frequency of the shear layer instability is defined by the semi-empirical formula proposed by Rossiter [7]. At low Mach numbers ($M < 0.2$), the acoustic wavelength is much greater than the cavity span, so the self-sustaining oscillation mechanism has to be considered as purely hydrodynamic. In this case, the oscillation frequency is governed by the convective velocity of vortices in the shear layer [8].

The typical flow characteristics for the open shallow cavities operating in self-sustained oscillation mode include free-stream flow separation at the cavity leading edge, a shear layer developed between the free-stream flow and the flow inside the cavity, impingement of the shear layer on the cavity aft wall and the shear layer disintegration into separate vortex structures. The processes produce high dynamic loads on the construction, which can lead to undesirable phenomena such as structural vibration and fatigue. In addition, the flow under consideration generates noise in the surrounding space that consists of intense discrete and broadband components.

The different techniques have been proposed to reduce the cavity flow unsteadiness and suppress acoustic resonances excited by impinging shear flow. The detailed summary of studies devoted to dynamics and control of open cavity flows can be found in reviews [9, 10]. The emphasis in these reviews is made on experimental investigations of open- and closed-loop suppression techniques developed in recent years. Numerical modeling has been also applied to provide a basis for techniques used for cavity flow control. Shao and Li [11] presented two passive control schemes where the cavity either with recessed leading edge step or with sloping trailing edge

wall is considered. To demonstrate the effects of the control, numerical simulation of the cavity flow was carried out by the LES-method. The obtained results showed the decreasing of resonant Strouhal numbers and the reduction of overall sound pressure levels in both cases. The control technique proposed by Suponitsky et al. [12] is based on applying simultaneous steady injection and suction through the front and rear cavity walls. The large eddy simulation technique coupled with the Lighthill-Curle acoustic analogy was used to get flow characteristics and estimate the effectiveness of control. The major effect of the control was the reduction of the reverse flow inside the cavity to the levels at which the absolute instability of flow is impossible.

Note that the majority of research deals with the rectangular cavity and either supersonic or compressible subsonic cavity flows because of their relevance to aeronautical applications. Much less attention has been given to low Mach number cavity flows and especially to non-rectangular cavities, although they are also widely implied in various engineering devices. In the present work, the incompressible fluid flow grazing over a shallow cylindrical cavity is numerically studied in the shear-layer mode. The developing flow field and its associated sound are derived and modification of those performed with a help of steady suction of fluid is discussed.

The use of suction is considered in the framework of the overall control strategy for near-wall flows, which involves cavities for creation of stable vortex systems near a body surface. Artificial generation and sustentation of the coherent vortices in near-wall flows is known to be an effective way to reduce surface friction and intensify heat and mass transfer [13, 14]. Taking into account the non-stationary nature of cavity flow, as mentioned above, the suction is assumed to facilitate localization of the vortex sheet generated at the leading edge inside the cavity.

The nonlinear theoretical model of the vortical flow control in a cross cavity has been developed in [15]. Being based on the Ringleb theory of the trapped vortex [14], the model describes the dynamic behavior of a single vortex inside the cavity in the system that consists of the uniform free stream and fluid suction. Analysis of the topology of this flow allowed us to get the optimal parameters of suction device, location and strength, with different cavity geometries. Note that optimal control in this case implies the trapped vortex laying at the stable critical point. It was also found that the region of vortex stability in shallow cavities is wider than in deep ones therefore the latter are more promising for near-wall flow control.

In the present work, numerical simulation of the two-dimensional viscous incompressible flow in the shallow cylindrical cavity is carried out by the vortex method [16], which belongs to the high-resolution Lagrangian-type schemes developed as fast alternative to direct numerical simulations. The vortex numerical schemes have been successfully used for calculation of various flows that take place in natural environment and technical applications [17]. The version of the vortex method used in this work as well as its advantages and restrictions have been described in details in papers [18, 19]. To derive the acoustic field, the hydrodynamic solution is coupled with the Ffowcs William-Hawkings equation, which is a development of the Lighthill acoustic analogy for bounded flows [20, 21].

The first part of the paper focuses on study of flow patterns and acoustic pressure generated by the cylindrical cavity with an aspect ratio of length to maximum depth of $L/D \approx 5$. The problem is considered for a thin laminar boundary layer on the wall portion before the cavity and with the Reynolds number $Re = 2 \cdot 10^4$, based on the cavity length L. The characteristics of the near hydrodynamic field show that the cavity operates in the shear-layer mode, when the self-sustained oscillations of flow are caused by collision of the large-scale vortex structures developed in the shear layer with the cavity trailing edge. The flow is the source of the dipole radiation with unequal lobes in the far acoustic field so that the sound intensity in the backward direction is higher than in the forward direction.

In the second part of the paper, modification of hydrodynamic and acoustic fields with the help of a steady suction through the rear cavity wall is considered. It was found that the applied control technique, which was used allows us to localize the vortical flow inside the cavity when saving the mode of self-sustained oscillations in the shear layer. The vortex structures developed in the shear layer do not hit the trailing edge now but are absorbed by the suction that causes considerable fluctuations as the mean pressure as the root-mean-square deviation of pressure in the vicinity of suction point. As a result, the levels of radiated sound are much higher than in the uncontrolled cavity flow. The obtained positive effect of suction on the cavity flow is that it suppresses the pressure fluctuations on the wall portion behind the cavity that leads to stabilization of the attached boundary layer.

9.2 Problem Statement and Numerical Procedure

Hydrodynamic and acoustic fields generated by a two-dimensional flow of viscous fluid in the region bounded by a wall with an embedded cylindrical cavity are studied. The problem is considered at uniform flow velocity U_0 and naturally laminar boundary layer before the cavity. The geometry of interest and coordinate systems are presented in Fig. 9.1.

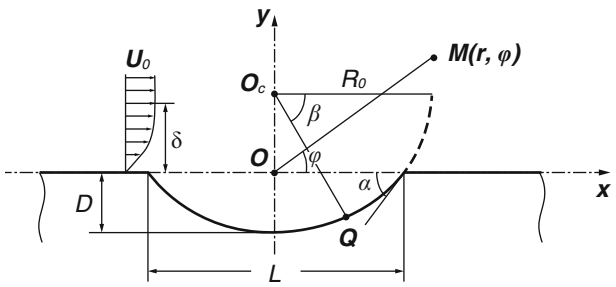


Fig. 9.1 Schemes of the flat wall with a cylindrical cavity and the flow

The axis Ox of the Cartesian coordinate system is directed along the wall, vertical axis Oy passes through the cavity center and the origin of the polar coordinate system coincides with that of the Cartesian system (r, φ). The angle φ of the polar system is measured counter-clockwise from the positive direction of x-axis. The point $M(r, \varphi)$ in Fig. 9.1 indicates the listener position.

The cavity geometry is described by its chord L and the angle α between the axis Ox and the tangent to the cavity at the intersection point. The shallow cavity, which is a subject of this research, is a part of the circle the center of which $O_c(x_c, y_c)$ is located above the wall. The fluid suction to be used in the second part of the current research is modeled by a hydrodynamic sink of constant strength Q placed on the cavity rear wall. Its coordinates are uniquely defined by the angle β (Fig. 9.1): $x_q = R_0 \sin\beta$, $y_q = R_0 \cos\beta + y_c$, where R_0 is the radius of the circle.

The problem under consideration is fully characterized by the following parameters: the ratio of the cavity length to its maximum depth, L/D; the initial boundary layer thickness at the cavity leading edge, δ; the Reynolds number Re based on the cavity length L, free-stream velocity U_0 and the kinematic viscosity of the ambient flow ν; the Mach number of the free-stream, $M = U_0/c_0$, where c_0 is the speed of sound in the ambient medium. The velocity U_0 is supposed to be much less than c_0, so the problem deals with a very low Mach number and the model of incompressible fluid can be used for simulation of both hydrodynamic and acoustic fields.

Note that all geometrical lengths are normalized with the cavity length L, velocities with U_0, physical times with L/U_0 and hydrodynamic pressure is specified by the dynamic coefficient $\rho_0 U_0^2/2$, where ρ_0 is the ambient fluid density.

Since a flow of very low Mach number is considered, there is no a reverse effect of the acoustic field on the hydrodynamic process. So, the hybrid numerical technique combining an independent evaluation of near hydrodynamic field with a certain acoustic analogy can be applied. In the present work, the viscous incompressible flow past the cylindrical cavity is simulated by the vortex method and the Ffowcs William-Hawkings analogy is used to derive the sound field.

9.2.1 Hydrodynamic Calculations

The flow evolution in the region under consideration is governed by the continuity equation and the Navier–Stokes equations with the non-leaking condition on the whole solid boundary. The no-slip condition is assumed to be satisfied on the cavity boundary and on the flat wall portions located directly before and beyond the cavity. This approach allows one to obtain the specified boundary layer thickness δ_{99} just above the cavity leading edge and constrain an influence of the boundary layer emerging along the back of flat plate to the cavity flow.

The vortex numerical scheme used for simulation of the flow field has been described in details in papers [18, 19] therefore its generalities are only considered here. In this approach, the vorticity $\omega = \mathbf{k} \cdot \nabla \times \mathbf{U}$, where $\mathbf{U}$ is the velocity vector and $\mathbf{k}$ is the unit vector out of the page, is considered as the primary variable. Its

transfer in the flow field is examined according to the vorticity transport equation:

$$\frac{\partial \omega}{\partial t} + (\mathbf{U} \cdot \nabla)\omega = \frac{1}{Re} \triangle \omega, \tag{9.1}$$

Equation (9.1) is solved using a fractional step procedure (see Cottet and Koumoutsakos [16]) when that is split into convective and diffusive parts, which are solved separately. In the present realization of the vortex method, the spatial derivative in the diffusion equation is approximated by the finite-difference scheme on the orthogonal grid put on the calculation domain. The convection of vorticity is simulated by the finite volume method, which controls vorticity flows across boundaries of elementary volumes. The volumes are connected with node points of the orthogonal grid. The vorticity is assumed to distribute evenly inside each volume. Note that the cavity boundary is approximated by the step function in this approach. To integrate the process in time, the explicit scheme of the second order with correction of all variables after each operator performed is applied.

The velocity field $\mathbf{U}(\mathbf{r}, t)$ is recovered from the vorticity field with help of the Biot–Savart integral. To build the integral, the fundamental solution of the Laplacian for a vortex in the region under consideration must be found. The conformal mapping of the flow field in the physical plane $z(x, y)$ into an upper half-plane of the auxiliary plane $\zeta(\xi, \eta)$ is performed to ensure this solution. The vortex function satisfying the non-leaking condition of the solid wall in the canonical plane is constructed with applying the well-known technique of mirror images:

$$G(\zeta, \zeta_v) = \frac{1}{2\pi i}[\ln(\zeta - \zeta_v) - \ln(\zeta - \overline{\zeta}_v)], \tag{9.2}$$

where $\zeta, \overline{\zeta}_v$ are the complex coordinate of a vortex and its image in the canonical plane, respectively.

The function that realizes the necessary mapping is the following [22]:

$$\zeta(z) = \frac{L\gamma}{2}\left[1 + \left(\frac{z - L/2}{z + L/2}\right)^{\gamma}\right] \Bigg/ \left[1 - \left(\frac{z - L/2}{z + L/2}\right)^{\gamma}\right], \quad \gamma = \frac{\alpha}{\pi - \alpha}. \tag{9.3}$$

Taking into account that the vorticity in any field point is conserved when conformal mapping and solution (9.2) annihilates the effect of the horizontal wall, which the flow boundary is reflected in, one obtains the Biot–Savart integral in the following form:

$$\mathbf{U}(\mathbf{r}, t) = \left[U_0 + \frac{Q}{\pi}\frac{1}{\zeta(r) - \zeta(r_q)} + \int_S \int \omega(\mathbf{r}', t)\mathbf{k} \times \nabla G(\zeta(\mathbf{r}), \zeta(\mathbf{r}'))ds(\mathbf{r}')\right]\frac{d\zeta}{dr}, \tag{9.4}$$

where S is the flow domain, $\mathbf{r}$ is the radius-vectors of the field points. Note that the second term in square brackets represents an influence of the sink located on the cavity boundary.

To derive the boundary condition for the vorticity on a solid wall, the *Lighthill's* vorticity creation mechanism, which attaches the vortex sheet to the wall, is applied. The vortex sheet is assumed to compensate the spurious slip on the wall that appears due to vorticity flow modifications. Taking into account the velocity jump across the vortex sheet, we obtain the following relation between the spurious slip and the vortex sheet strength:

$$\mathbf{U}_\tau^0 + \frac{\gamma}{2} = 0, \tag{9.5}$$

where γ is the strength of the adjoined sheet and $\mathbf{U}_\tau^0$ is the tangential velocity in the wall points calculated from (9.4).

As soon as the vortex sheet at a solid wall is obtained, it has to be transferred to the flow using either the *Neuman* or *Dirichlet-type* boundary condition for the vorticity. In this work, we follow Wu [23] who divided the strength of the vortex sheet by the distance from the wall to the first mesh point in the computational domain and then get the following equation for the wall vorticity ω_0 from (9.5):

$$\omega_0 = -\frac{2\mathbf{U}_\tau^0}{\Delta s}, \tag{9.6}$$

where Δs is the grid spacing perpendicularly to the wall.

The vorticity created on smooth parts of the flow boundary enters the fluid through a mechanism of viscous diffusion. To simulate the flow separation on the sharp cavity edges, the *Kutta-Joukowskii* condition is applied. In the numerical scheme, the condition is realized by convective transferring the vorticity from the sharp edge to the surrounding flow.

The formulation of the Navier-Stokes equations with vorticity and velocity variables permits us to decouple purely kinematical problem from the pressure problem that simplifies significantly numerical modeling of fluid flows. The viscous flow equations connecting pressure, velocity and vorticity were obtained by Lamb [24]. Then one can retrieve the pressure coefficient $\overline{C_p} = 2(p - p_0)/\rho U_0^2$ by direct integration of those relative to the field coordinates (see [19]). It is convenient for the present flow configuration to integrate the second equation of the Lamb system top-down along the vertical direction. Note that the integration must start sufficiently far from the source of perturbations of the fluid field.

9.2.2 Far Acoustic Field

In the present work, the Ffowcs William-Hawkings (FW-H) analogy for the computation of the acoustic pressure from the obtained hydrodynamic field is applied. The approach deals with most of the practical cases where the problem of aerodynamic sound involves moving or fixed boundaries. The classical FW-H equation is written in the time domain in the coordinate system connected rigidly with a body [20]. The body is believed to move in stationary environment. The monomial terms in the right part of the equation, which describe monopoly and dipole sound sources located on the body boundary, include the Dirac function of the argument depending both on time and on spatial coordinates. This does not allow us to apply the Fourier transformation directly to the FW-H equation. But for the important practical case of a uniform rectilinear body motion, one can pass to the fixed coordinate system using the Galilean transformation. Note that system is still connected with the body, which is at rest now. After the Galilean transformation, the wave FW-H equation becomes the convective wave equation [20]:

$$\left\{\frac{\partial^2}{\partial t^2}+U_iU_j\frac{\partial^2}{\partial y_i\partial y_j}+2U_j\frac{\partial^2}{\partial y_j\partial t}-c_0^2\frac{\partial^2}{\partial y_i^2}\right\}[\rho' H(f)]= \\ =\frac{\partial}{\partial t}[Q\delta(f)]-\frac{\partial}{\partial y_i}[F_i\delta(f)]+\frac{\partial^2}{\partial y_i\partial y_j}[T_{ij}H(f)], \tag{9.7}$$

where monopoly, dipole and quadrupole terms in the right-hand side are respectively written as:

$$Q(\mathbf{y},t)=(\rho u_i-\rho_0U_i)\,n_i,$$

$$F_i(\mathbf{y},t)=\left(p\delta_{ij}+\rho\,(u_i-2U_i)\,u_j+\rho_0U_iU_j\right)n_j, \tag{9.8}$$

$$T_{ij}(\mathbf{y},t)=\rho u_iu_j+p\delta_{ij}-c_0^2\rho'\delta_{ij}.$$

Here t is the time and $\mathbf{y}$ is the radius-vector of a point in the Cartesian coordinate system, where axis are denoted as y_1, y_2; c_0 is the sound velocity; $\rho=\rho_0+\rho'$, $p=p_0+p'$, $u_i=U_i+u_i'$ are the total density, pressure and velocity; U_i are the velocity components of the oncoming flow. Note that the parameters with subscript "0" in (9.8) characterize the undisturbed flow and a prime is used to denote a perturbation quantity. The numerical indices in formulae (9.7), (9.8) are used instead of the letter indexation of coordinates for convenience in summation. Since we deal with a two-dimensional problem, the indices i, j take the value 1 or 2. It must be also noted that in further consideration the viscous part of the Lighthill stress tensor T_{ij} will be neglected and only the term $p\delta_{ij}$ is considered.

In the present statement, the function f is only a function of the spatial coordinates, so: $f=f(\mathbf{y})$. It is introduced in such a way that the equation $f(\mathbf{y})=0$

defines a closed control surface outside which the acoustic field is considered. The region $f(\mathbf{y}) > 0$ lies outside the surface and $f(\mathbf{y}) < 0$ elsewhere. The vector $n_j = \partial f/\partial y_j$ designates the outer normal to the control surface. The Heaviside function $H(f)$ is introduced in such a way that $H = 1$ for $f \geq 0$ and $H = 0$ for $f < 0$. Its derivative $H'(f) = \delta(f)$ is the Dirac delta-function and δ_{ij} in (9.8) is the Kronecker symbol. It must be noted that the control surface includes the solid boundary of a flow.

For a low Mach number and an uniform rectilinear flow, Eq. (9.7) can be transformed into the frequency domain with the help of the Fourier transformation. Assuming isotropic acoustically ideal medium in the far field, where $\rho' \ll \rho_0$, $p' \ll p_0$, $p' = c_0^2\rho'$ one obtains after some simplifications the following formula for the acoustic pressure [21]:

$$p'(\mathbf{y}_o, \Omega) = -\oint\limits_{f=0} i\Omega\,\widehat{Q}(\mathbf{y}, \Omega)G(\mathbf{y}_o, \mathbf{y}, \Omega)\,dl - \oint\limits_{f=0} \widehat{F}_i(\mathbf{y}, \Omega)\frac{\partial G(\mathbf{y}_o, \mathbf{y}, \Omega)}{\partial y_i}\,dl - \\ - \int\limits_{f>0} \widehat{T}_{ij}(\mathbf{y}, \Omega)\frac{\partial^2 G(\mathbf{y}_o, \mathbf{y}, \Omega)}{\partial y_i \partial y_j}\,d\mathbf{y}, \tag{9.9}$$

where $\widehat{Q}(\mathbf{y}, \Omega)$, $\widehat{F}_i(\mathbf{y}, \Omega)$ and $\widehat{T}_{ij}(\mathbf{y}, \Omega)$ are the Fourier transformation of monopoly, dipole and quadrupole terms of the right part of Eq. (9.7); G is the Green function; Ω is the angular frequency; $\mathbf{y} = (y_1, y_2)$, $\mathbf{y_0} = (y_{01}, y_{02})$ are the radius-vectors of the point in the hydrodynamic field and the listener position, respectively.

The Green function in (9.2.2) takes into account the convective effects and has the following form:

$$G(\mathbf{y}_o, \mathbf{y}, \Omega) = \frac{\mathrm{i}}{4\beta}\exp\left(\frac{\mathrm{iMkr}_1}{\beta^2}\right)\cdot H_0^2\left(\frac{k}{\beta^2}\sqrt{r_1^2 + \beta^2 r_2^2}\right), \tag{9.10}$$

where $r_1 = (y_{01} - y_1)$, $r_2 = (y_{02} - y_2)$, H_0^2 is the Hankel function of the second kind of order zero, $k = \Omega/c_0$ is the wave number, M is the Mach number, $\beta = \sqrt{1 - M^2}$ is the Prandt-Glauert factor, $\mathrm{i} = \sqrt{-1}$ is the imaginary unit.

The transition into the frequency domain is a very important step when solving the FW-H equation for two-dimensional problems because this allows one to estimate the extension of sound sources across the flow. It is impossible in the time domain where the Green function is expressed by the Heaviside function and one has to integrate over time on a half-indefinite interval, which is actually impossible. To turn back into the time domain after computing the sound field, the inverse Fourier transformation must be performed.

Sound field calculations can be greatly simplified if one takes into account two factors. The first is conditioned by the fact that the sound wave in the far field

is cylindrical and the amplitudes of pressure pulsations are proportional to $M^{3/2}$, $M^{5/2}$, $M^{7/2}$ for monopoly, dipole and quadrupole sound sources respectively. Consequently, at low Mach numbers the amplitude of sound from quadrupole sources is much smaller than that of others and one can neglect the last double integral in Eq. (9.2.2). The second simplification follows from the problem statement, because the hydrodynamic data used in (9.2.2) are estimated on the impermeable surface, where $u_i = 0$, and the uniform rectilinear external flow is considered. Only the monomial $p\delta_{ij}n_{ij}$ in (9.8) depends on time. Therefore, to determine the sound field, it is sufficient to calculate the second contour integral in (9.2.2), which describes the dipole source of sound and is conditioned by the unsteady force acting on the body.

9.2.3 Details of the Numerical Scheme

With the described hybrid numerical technique that combines the vortex method and the Ffowcs William-Hawkings analogy, the flow patterns and generated acoustic field above the open shallow cavity with $L/D \approx 5$ were derived at $Re = 2 \cdot 10^4$. The structure of the calculation domain and the boundary conditions involved are presented in Fig. 9.2.

Since the only source of vorticity in the region is a solid surface where the no-slip condition is satisfied, the flow is assumed to be irrotational at the inlet cross-section and at the upper boundary. At the output section, where the gradients of hydrodynamic parameters are small enough, the so-called soft boundary condition $\partial^2\omega/\partial t^2$ is set. The width of the calculation region was $2L$ and the length of the wall portion beyond the cavity was $25L$ at that the length of the no-slip segment, where $\omega = \omega_0$, was $5L$ (Fig. 9.2). As test calculations have shown, a further enlargement of the computational domain has no appreciable influence on the flow characteristics.

The non-dimensional length l_0 of the wall portion before the cavity is a very important parameter in the present study because it controls the properties of the oncoming boundary layer before the zone of flow separation. Since the boundary layer is laminar, the famous Blasius solution is available to estimate near-wall flow parameters [25]. But Larsson et al. [26] have shown that the upstream boundary

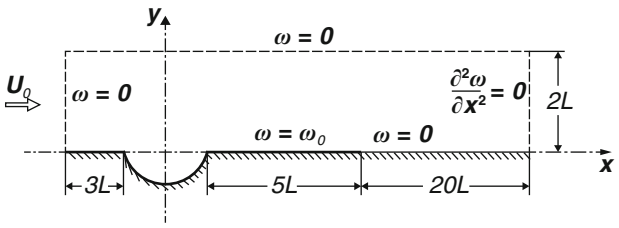


Fig. 9.2 Sketch of the computational domain

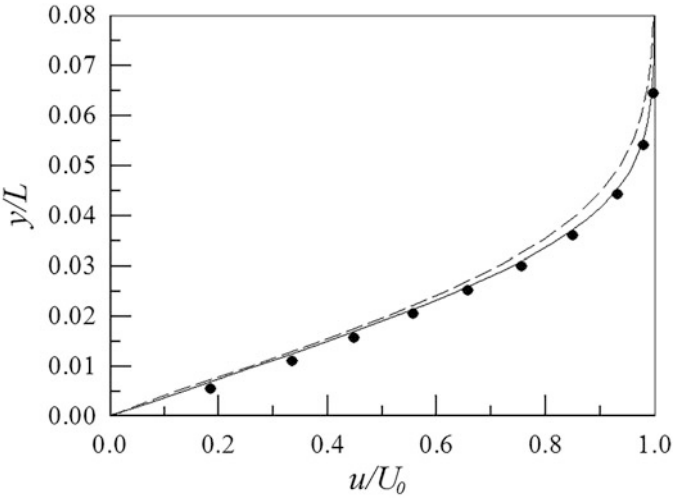


Fig. 9.3 Average velocity profiles at the leading edge of the cavity: solid line—Blasius solution, dash line and circles—numerical solutions for the wall with a cavity and flat plate, respectively

layer is essentially affected by the presence of the cavity; as a result, the actual parameters of the boundary layer above the cavity leading edge may be greater than in Blasius solution. Figure 9.3 illustrates the average velocity profiles just above the cavity leading edge obtained from a Blasius solution for the flat plate (solid line) and in the present calculations (dash line). Note that we set the length of l_0 to obtain a thin enough boundary layer before the cavity, $l_0 = 3$. The Blasius solution ensures the boundary layer thickness $\delta_{99} \approx 0.061$ for this l_0. At the same time, the value of δ_{99} obtained in the present simulation is about 0.073. To estimate an error of the numerical scheme, the simulation of the boundary layer developing along the flat wall was carried out. The average velocity profile obtained in the calculation is depicted by circles in Fig. 9.3. One can see that the numerical results are close to an exact solution, which indicates the good accuracy of the numerical scheme.

The value of the boundary layer momentum thickness θ_{99}, which is important for cavity flow classification, is deduced from the parameter δ_{99}, assuming a ratio $\delta/\theta = 8$ [25]. This gives the relation $L/\theta \approx 110$ above the cavity leading edge and we expect the cavity oscillating in the shear-layer mode [5].

The uniform grid with the square elements of size $h = 0.005$ was used and the normalized time step Δt was chosen from the Courant–Friedrichs–Lewy condition:

$$\frac{\max\{u, v\}\Delta t}{h} \leq 1.$$

The maximum value of the local flow velocity $\mathbf{U}(u, v)$ obviously depends on the flowed surface curvature and the Reynolds number. As preliminary calculations

have shown, the numerical stability for the present geometry and flow conditions is achieved at $\Delta t = 0.5h$.

The pressure fluctuations on the flow boundary obtained by the hydrodynamic modeling at the successive instances were the input data of the acoustic problem, which was characterized by the Mach number $M = 0.2$. As it has been mentioned, the far acoustic field is only generated by dipole sources of sound. The integration contour $f(\mathbf{y}) = 0$ in Eq. (9.2.2) coincides with the outside of the flow boundary and the curvilinear integral is calculated in quadrature by the trapezium method.

9.3 Results

9.3.1 Natural Flow in Open Cylindrical Cavity

In this subsection, the results of simulation of the natural flow, without suction, in the cavity under consideration are analyzed. The flow regime in the shallow rectangular cavity is known to depend on the momentum thickness θ of the boundary layer formed on the wall before the cavity. Depending on its magnitude, the flow in the rectangular cavity oscillates either in the shear-layer mode or in the wake mode, which are characterized by very different values of the average drop coefficient $\overline{C}_D$, which describes the cavity drag resulting from a pressure difference on rear and front walls. It is well known that $\overline{C}_D$ is much higher in the wake mode than in the shear-layer mode, 0.3 against 0.01 [9]. So, to define the oscillation mode in the cavity flow under consideration, we first consider the drop coefficient obtained in the calculations. For a cylindrical cavity, it is normalized on the maximum cavity depth D and reduced to the unit length along the cavity axis. The time evolution of the instantaneous drop coefficient C_D is presented in Fig. 9.4, where the time segment covers four statistically stationary fluctuation periods. It follows from these results that $C_D = 0.0096$ and the non-dimensional period of oscillations $T \cdot (U_0/L)$

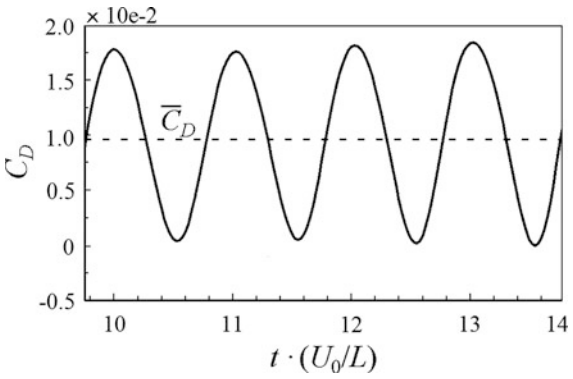


Fig. 9.4 Time history of the cavity drop coefficient C_D

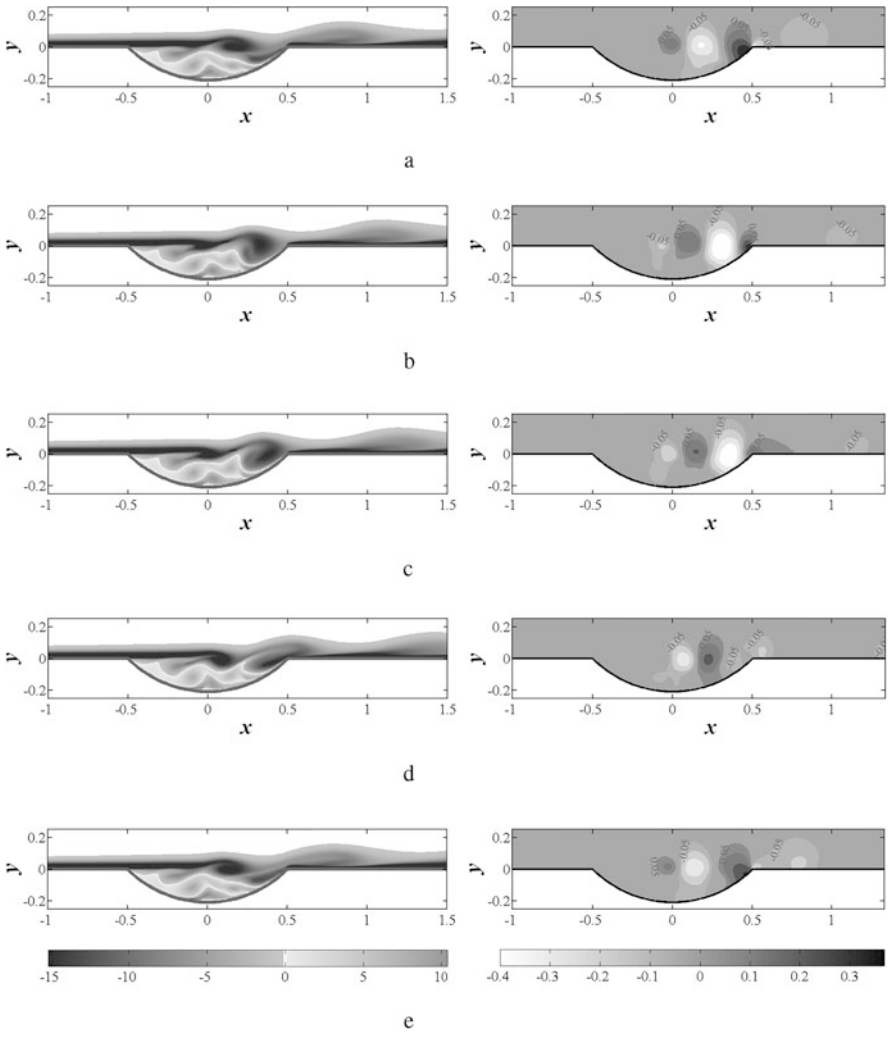


Fig. 9.5 Instantaneous vorticity field (on the left) and pressure field (on the right) at five times during one cycle. (**a**) $t = 0$. (**b**) $t = T/4$. (**c**) $t = T/2$. (**d**) $t = 3T/4$. (**e**) $t \to T$

is about 1. It follows that the frequency of oscillations is of $f = 1/T = 1$ that ensures $St = fL/U_0 = 1$, where St is the Strouhal number of flow.

The conclusion is confirmed by the results presented in Fig. 9.5, where the instantaneous fields of vorticity (on the left) and pressure (on the right) over one period of the well-established self-sustained oscillations are depicted. Note that the time instance $t = 0$ corresponds here to the maximum value of the drop coefficient C_D. The pictures demonstrate the classical behavior of the cavity flow in the regime of shear-layer oscillations.

As the flow separates at the upstream edge of the cavity, the shear layer develops between the free stream and the flow inside the cavity. Further the layer loses stability and clockwise large-scale vortical structures are generated there and move to the trailing edge. The snapshots of the vorticity field (Fig. 9.5, left) show that two vortical structures can be identified in the cavity shear layer at any time. In Fig. 9.5a, the first vortex is located approximately in the middle of the cavity and the second vortex is born near the leading edge. These vortices travel downstream in the next pictures, growing with convection. As the first vortex impinges on the trailing edge (Fig. 9.5d), the second vortex locates at the cavity center. After the impingement, part of the vortex spills over the cavity and is convected downstream, increasing the thickness of the reattached boundary layer. The other component is swept downwards into the cavity amplifying the recirculation zone. The process is accompanied by pressure fluctuations of large amplitude near the output edge, as a result, the drop coefficient decreases to its minimum value. In Fig. 9.5e, the impinging vortex is finally divided and the vortex generated at the leading edge in the first picture travels to the trailing edge to sustain the vortex impingement process.

At the same time, the secondary recirculation flow is developing inside the cavity. It consists of two counter-clockwise vortical structures born near the cavity lateral walls and the clockwise vortex located near the bottom. During the oscillation cycle, the flow configuration in the cavity changes according to the vortex dynamics in the shear-layer. This fact points out that the secondary vortices are one of the elements of the hydrodynamic feedback in the self-sustaining oscillation mechanism. But the main cause of a feedback phenomenon connects with the velocity production in the shear layer according to the Biot-Savart law. As a result, the oscillation frequency is governed by the convective velocity of the vortices in the shear layer. Since the length of acoustic wave exceeds considerably the cavity span at a low Mach number, only hydrodynamic mechanism produces self-sustained oscillations of the cavity flow in this case.

The fact that the flow field clearly oscillates in a shear-layer mode is confirmed by the instantaneous pressure contours presented in Fig. 9.5 (on the right). Here the negative pressure zones are the markers of the vortical structures in the shear-layer. Two vortices are clearly visible in Fig. 9.5b, c and only one vortex is present explicitly in other pressure snapshots. This is due to the fact that the second vortex is either already broken or is still in its developing. It is also seen that pronounced areas of positive pressure are emerging between the vortices.

To determine the convective velocity of vortex structures in the shear layer, let us consider the cross correlation coefficient of fluctuating pressure $R_{12}(\tau)$ between the points located on the cavity chord. Here τ is the time delay between the peaks of the pressure function in the considered points. It is clearly seen in Fig. 9.5 that the shear layer vortices are formed not directly nearby the leading edge but at some distance downstream from it. So, the functions $R_{12}(\tau)$ between the reference point $x = 0$ and the points located at $x > 0$ are only analyzed. The cross correlation coefficient $R_{12}(\tau)$ for some points of x-axis is shown in Fig. 9.6a. Since the direction of travel of the vortices is obvious it is enough to consider this coefficient only for positive values of τ. To liquidate the phase coupling of the pressure functions

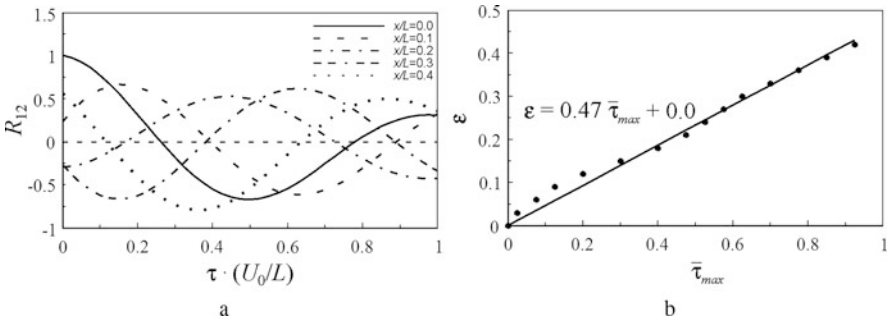


Fig. 9.6 Cross correlation coefficients R_{12} of fluctuating pressure along the cavity chord at various x—locations—(**a**) and time delay ε of maximum correlation coefficients at various separation distances—(**b**)

at two points, the abscissa τ_{max} of the first maximum of $R_{12}(\tau)$ for $\tau \geq 0$ is considered as a real time delay. The non-dimension separation distance ε between each sampling point and the reference point as well as the non-dimension time delay $\bar{\tau}_{max}$ are determined. The function $\varepsilon(\bar{\tau}_{max})$ is depicted in Fig. 9.6b by the bold points for each location. After fitting a linear curve through all the data points and the point of coordinates $(0, 0)$, the convection velocity of vortices in the shear layer U_c normalized with the flow velocity U_0 is estimated as the angular slope of the approximating straight line. The obtained value of the non-dimensional velocity of the vortices is $U_c = 0.47$ that is in close agreement with the experimental data of Özsoy et al. [27], where $U_c \approx 0.5$ has been found.

It is obvious that the frequency of collisions of the shear-layer vortices with the cavity trailing edge depends on the number of vortices, the distance between them and the velocity U_c. Due to absence of the acoustic feedback in the flow under consideration, the frequency of collisions defines the frequency of self-sustained oscillations of shear layer, known as the Strouhal number of cavity flow. It can be determined from the Rossiter formula adapted to incompressible fluid [28]:

$$St = fL/U_0 = U_c(n - \alpha), \quad n = 1, 2, \ldots \tag{9.11}$$

where n is the number of vortices and $\alpha = 0.25$ is the empiric constant interpreted as a phase delay of the vortices in the shear layer. Using $St = 1$, which has been obtained on the base of drop pressure fluctuations (Fig. 9.4) and $n = 2$ as follows from snapshots of Fig. 9.5, one calculates the value of $U_c = 0.57$. There is some difference between this value and the convective velocity derived in Fig. 9.6b because the Rossiter formula considers the vortices running the entire cavity chord. At the same time, we have clarified in Fig. 9.5 that the shear-layer vortices are born not near the leading edge of cavity that is a consequence of the specified geometry.

Periodical vortex processes occurring in the shear layer above the cavity generate pressure fluctuations on its surface to be characterized by the time-mean-pressure coefficient $\overline{C}_p$ and the root-mean-square deviation of pressure $C_{p,rms}$. In Fig. 9.7,

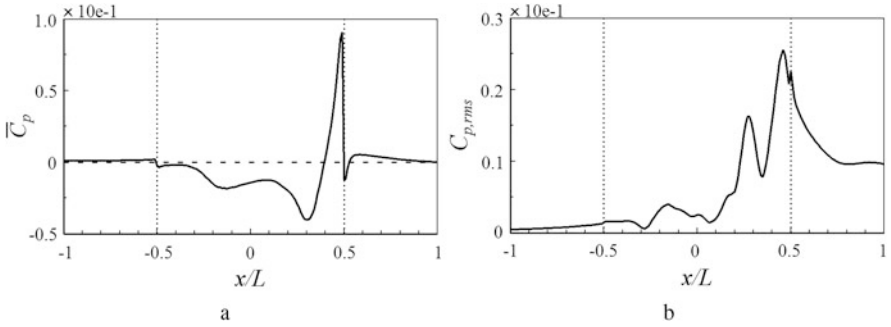


Fig. 9.7 Mean pressure coefficient $\overline{C}_p$—(**a**) and root-mean-square deviation of pressure $C_{p,rms}$—(**b**) on the cavity boundary and on the wall just before and beyond the cavity

these characteristics calculated on the cavity floor and on the wall portions located directly before and beyond the cavity are plotted. Note that the points of cavity boundary are projected to x-axis and closed between dashed vertical lines. The curves reflect the dynamics of flow in the cavity during the period of oscillations. It is seen that both the value and the amplitude of coefficient $\overline{C}_p$ are small enough in the vicinity of the cavity leading edge and they grow significantly on the rear cavity wall, where sharp jumps in $\overline{C}_p$ are observed. It is clear that those are caused by non-stationary behavior of the cavity shear layer in this flow region. The moderate increase of $\overline{C}_p$ on the wall just beyond the cavity is connected with the movement of the vortex structures which have left the cavity.

The root-square deviation of pressure from its mean value characterized by the coefficient $C_{p,rms}$ is also the largest in the aft part of cavity where the shear-layer vortices hit the wall. The fluctuations of $C_{p,rms}$ in other cavity parts are seen caused by the change of configuration of the secondary vortices in the cavity during the oscillation cycle. On the wall before the cavity, the coefficient $C_{p,rms}$ is small and it gradually decays beyond the cavity trailing edge.

Thus it has been found that self-sustained oscillations of shear layer are the reason of substantial pressure fluctuations in the closed hydrodynamic field, which, in turn, radiate an acoustic wave in the far field. The frequency of fluctuations of the sound pressure in the far field fixed point is equal to the Strouhal number of flow. So, the directivity characteristics of the sound field are only considered for the amplitude of pressure fluctuations p'_{amp}. The calculated directivity chart for the sound pressure levels is depicted in Fig. 9.8, in which the amplitude of pressure fluctuations is normalized with the dynamic pressure: $\overline{p}'_{amp}(r,\varphi) = p'_{amp}(r,\varphi)/(\rho_0 U_0^2/2)$. This picture demonstrates that the cavity flow is a source of dipole radiation with unequal lobes. The sound intensity in the backward direction is seen to be higher than that in the forward direction. The maximum value of the sound pressure is observed at $\varphi \approx 140°$. The inverse proportion between the amplitude of acoustic pressure $\overline{p}'_{amp}(r,\varphi)$ and the square root from radius r at any fixed angle φ follows from these

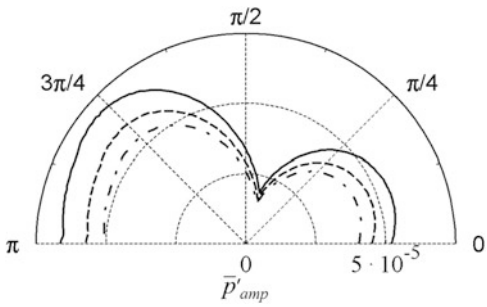


Fig. 9.8 Directivity chart for the overall sound pressure level: solid line—$r/L = 60$, dash line—$r/L = 80$, dash-dotted line—$r/L = 100$

results. This fact points out that the present cavity flow radiates a cylindrical sound wave in the far field.

9.3.2 Cavity Flow with Fluid Suction

In this section, an influence of the fluid suction on the cavity flow and radiated noise is examined. The fluid suction may be used to localize the vortical flow inside a cavity and reduce the level of fluctuations of hydrodynamic parameters in the attached boundary layer. The problem arises, for example, when cavities are applied for creation of a stable vortical pattern in the near-wall flow. The theoretical model of near-wall flow control that uses a vortex trapped in a cylindrical cavity and suction of fluid is developed in [15]. It is based on the analysis of the flow topology in the domain. The parameters of the control device are set by us to obtain the trapped vortex located in a stable critical point. To estimate feasible advantages and disadvantages of this control technique, the numerical simulation of hydrodynamic and acoustic fields generated by the cylindrical cavity with fluid suction is carried out.

The suction is modeled by the hydrodynamic sink, whose strength and location are determined with applying the theoretical model [15]. The sink is located on the rear cavity edge as seen in Fig. 9.1. In this study, the angle β characterizing the sink position is equal to 55° and the non-dimensional strength is set $\overline{Q} = Q/U_0 L = 0.08$.

It is revealed in the present simulation that the cavity flow with suction operates in the regime of self-sustained oscillations, as in the previous case. The instantaneous fields of vorticity (on the left) and pressure (on the right) over one period of the oscillations are represented in Fig. 9.9. One can see generation, development and movement of vortex structures in the shear layer. The difference is that the vortices do not impinge on the traveling edge but are attracted by the sink point located on the aft cavity wall. The second mode of self-sustained oscillations is seen to be realized here because two vortices exist in the shear layer at any time. The suction absorbs the vortex structures so that they have almost no effect on the attached

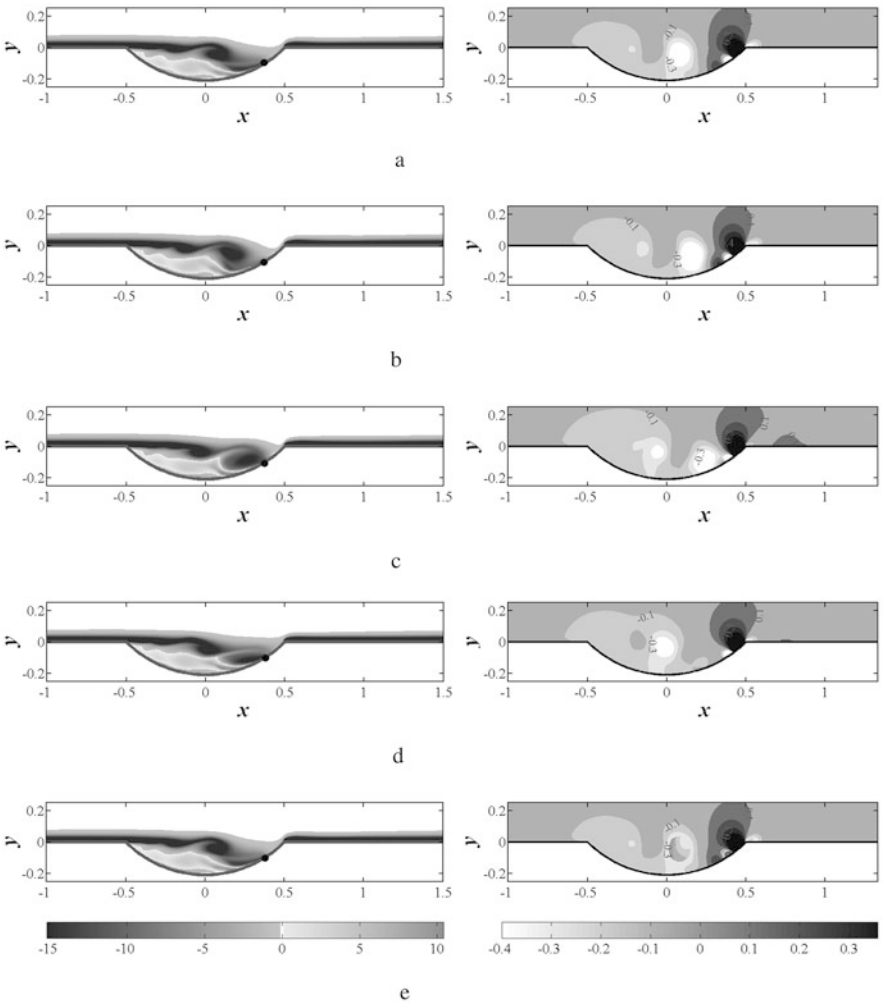


Fig. 9.9 Instantaneous vorticity field (on the left) and pressure field (on the right) at five times during one cycle in the cavity with suction. (**a**) $t = 0$. (**b**) $t = T/4$. (**c**) $t = T/2$. (**d**) $t = 3T/4$. (**e**) $t \to T$

boundary layer beyond the cavity. It is seen on the pressure pictures that there is always a large-scale zone of positive pressure near the trailing edge that prevents the vorticity concentrated in the cavity from emission. The secondary flow in the cavity is weak because of the limited space in this case.

The periodicity of the shear-layer processes is confirmed by the time history of the drag coefficient C_D presented in Fig. 9.10a. It was found that the non-dimensional period of the oscillations T increased in comparison with the natural

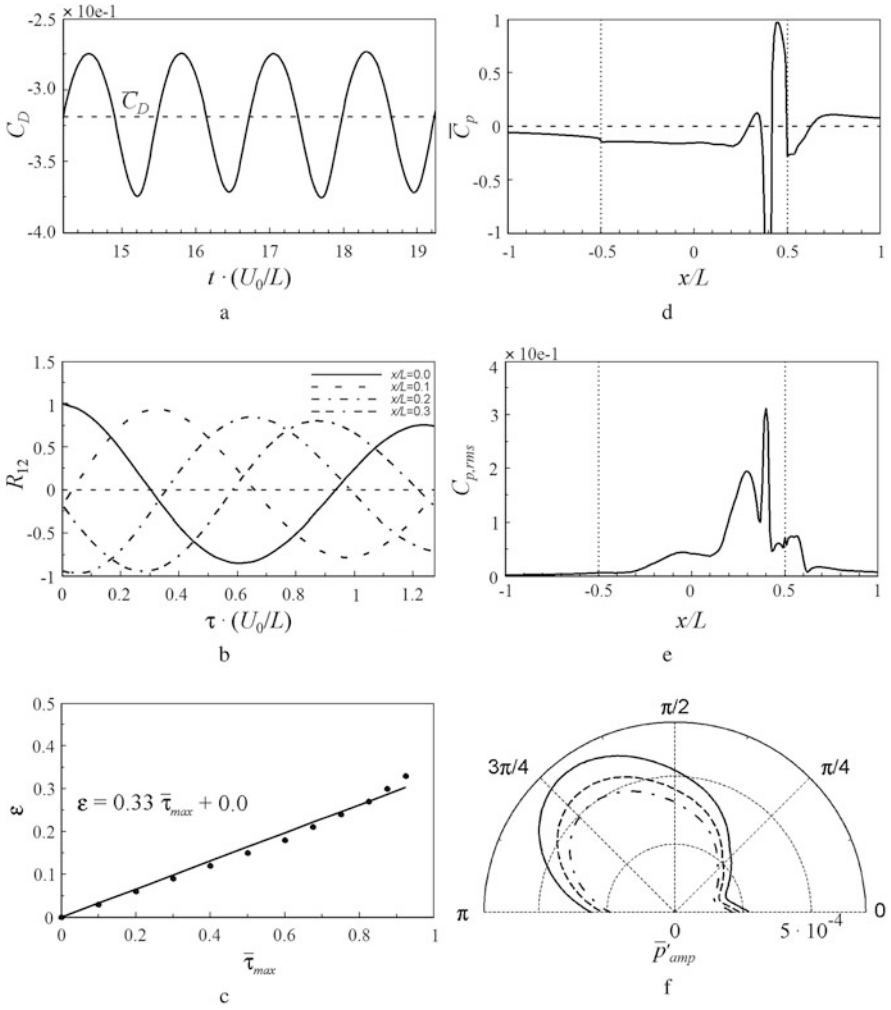


Fig. 9.10 Characteristics of near hydrodynamic and far acoustic fields for the cavity flow with fluid suction: (**a**)—time history of the cavity drop coefficient C_D; (**b**)—cross correlation coefficients R_{12} of fluctuating pressure along the cavity chord at various x—locations; (**c**)—time delay ε of maximum correlation coefficients at various separation distances; (**d**)—mean pressure coefficient $\overline{C}_p$ on the cavity boundary and on the wall just before and beyond the cavity; (**e**)—root-mean-square deviation of pressure $C_{p,rms}$ on the cavity boundary and on the wall just before and beyond the cavity; (**f**)—directivity chart for the overall sound pressure level: solid line—$r/L = 60$, dash line—$r/L = 80$, dash-dotted line $r/L = 100$

cavity flow. Here $T = 1.257$ against $T = 1$, when the suction is absent. So, the Strouhal number of the flow under consideration is equal to 0.795. It is significant that the coefficient C_D is always below zero and its average value $\overline{C}_D = -0.319$ that makes the obtained flow regime similar to the wake mode.

Since the control action increases the period of cavity flow oscillations one can expect that the convection velocity of vortices U_c will drop. It is seen in Fig. 9.9 that the vortices travel along the line that is slightly inclined to the cavity chord. Therefore, the change of pressure at the points on the cavity chord occurs synchronously with the displacement of the vortices in the shear layer. So, to estimate the convection velocity, the cross correlation coefficient of fluctuating pressure $R_{12}(\tau)$ can be evaluated between the points located on the cavity chord. We consider $x = 0$ as the reference point and $0.1 \leq x \leq 0.3$ as the sampling points. The corresponding curves are presented in Fig. 9.10b and the dependence $\varepsilon(\tau)$ which illustrates the convection velocity calculation is depicted in Fig. 9.10c. It follows from Fig. 9.10c, $U_c = 0.33$.

Figure 9.10d, e demonstrate the effect of suction on the pressure fluctuations estimated on the cavity surface. The time-mean-pressure coefficient $\overline{C}_p$ demonstrates the most important changes on the aft wall. The minimum of $\overline{C}_p$ is achieved in the vicinity of the suction point. This value is obtained to be $\overline{C}_p \approx -13$. As in the uncontrolled case, the function $\overline{C}_p(x)$ has its maximum value just before the cavity trailing edge. Note the range $-1 \leq \overline{C}_p \leq 1$ in Fig. 9.10d is chosen to show distinctly the behavior of this coefficient along the cavity surface. The zone of the highest values of the root-mean-square deviation of pressure $C_{p,rms}$ shifts from the trailing edge, as it has been in the uncontrolled cavity, to the vicinity of the suction point. Taking into account the results presented in Fig. 9.10d, e one can conclude that the absolute values of pressure characteristics grow as compared with the uncontrolled case everywhere except the region located beyond the trailing edge. Therefore it can be deduced that the fluid suction does indeed reduce the pulsations of the hydrodynamic parameters on the wall behind the cavity, thus stabilizing the attached boundary layer.

Since the far acoustic field is the linear reflection of fluctuating pressure close to the wall, one can expect that essential quantitative changes observed in the fluctuating pressure levels should modify dramatically the directivity chart when the fluid suction takes place. Comparing the plots of the root-mean-square deviation of pressure $C_{p,rms}$ derived in the natural cavity flow and in the cavity with fluid suction (Figs. 9.7b and 9.10e) we obtain that in the last case the maximum value of $C_{p,rms}$ is twelve times higher. Besides, it is displaced from the trailing edge into the cavity. These two factors have the largest effect on the configuration of the directivity chart, which is depicted in Fig. 9.10e for the same values of radius r as in the previous case (Fig. 9.8). It is seen that the maximum of acoustic pressure here is nine times higher than the corresponding value for the natural cavity flow. This chart also has a dipole configuration, but it is less pronounced. The backward lobe of the chart extends and the lobe in the forward direction decreases considerably. In this case, the directivity chart rotates counter-clockwise by about 10°, and the maximum of the sound energy is propagated in the direction $\varphi = 130°$.

9.4 Conclusion

The hybrid numerical technique coupled with the vortex method for simulation of viscous incompressible flow and the Ffowcs William-Hawkings acoustic analogy is developed. It is applied to the investigation of hydrodynamic and acoustic fields generated by a two-dimensional cylindrical cavity with an aspect ratio of $L/D \approx 5$. The problem is considered for a thin laminar boundary layer before the cavity and with the Reynolds number of $Re_L = 2 \cdot 10^4$. The Mach number was set to 0.2. The natural cavity flow and the flow with fluid suction through the cavity aft wall were studied and obtained results were compared to clarify advantages and disadvantages of such control.

The present simulation revealed that the given flow conditions result in the shear-layer regime above the cavity without suction. This is characterized by the self-sustained oscillations of cavity flow which are caused by dynamics of large-scale vortex structures in the shear layer. The second mode of oscillations is realized because no more than two vortices exist in the shear layer at any time. The calculated frequency of oscillations is in a good agreement with the analytical solution and the convective velocity of vortices is slightly lower in comparison with the known data for a rectangular cavity which is consequence of the specific flow geometry. The directivity chart for the sound pressure levels demonstrates that the present cavity flow is a source of dipole radiation with unequal lobes. The sound intensity in the backward direction is higher than that in the forward direction. The maximum value of the sound pressure is observed at $\varphi \approx 140°$.

A steady suction through the rear cavity wall allows to localize the vortical flow inside the cavity while maintaining the mode of self-sustained oscillations in the shear layer. The vortices generated in the shear layer do not hit the trailing edge here but are absorbed by the suction that causes the considerable fluctuations as the mean pressure as the root-mean-square deviation of pressure in the vicinity of suction point. As a result, the obtained levels of radiated sound are much higher than in the uncontrolled cavity flow. The obtained positive effect of the suction on the cavity flow is that it suppresses the pressure fluctuations on the wall portion behind the cavity that leads to stabilization of the attached boundary layer.

References

1. Voskoboinick, V., Kornev, N., Turnow, J.: Study of near wall coherent flow structures on dimpled surfaces using unsteady pressure measurements. Flow Turbul. Combust. **90**(2), 86–99 (2013)
2. Gregorio, F., Fraioli, G.: Flow control on a high thickness airfoil by a trapped vortex cavity. In: Proceedings of 14th International Symposium on Applications of Laser Techniques to Fluid Mechanics, Lisbon, Portugal, 7–10 July, p. 112 (2008)
3. East, L.F.: Aerodynamically induced resonance in rectangular cavities. J. Sound Vib. **3**(3), 277–287 (1996)

4. Stallings, R. Jr., Wilcox, F. Jr.: Experimental cavity pressure distributions at supersonic speeds. Technical Paper 2683, NASA, June 1987
5. Gharib, M., Roshko, A. The effect of flow oscillations on cavity drag. J. Fluid Mech. **177**, 501–530 (1987)
6. Rockwell, D., Naudascher, E.: Review: self-sustaining oscillations of flow past cavities. J. Fluids Eng. **100**, 152–165 (1978)
7. Rossiter, J.E.: Wind tunnel experiments on the flow over rectangular cavities at subsonic and transonic speeds. ARC Reports and Memoranda, 3438 (1964)
8. Lin, J.-C., Rockwell, D. Organized oscillations of initially turbulent flow past a cavity. AIAA J. **39**(6), 1139–1151 (2001)
9. Rowley, C.W., Williams, D.R.: Dynamics and control of high-Reynolds-number flow over open cavities. Annu. Rev. Fluid Mech. **38**, 251276 (2006)
10. Cattafesta, L.N., Song, Q., Williams, D.R., et al.: Active control of flow-induced cavity oscillations. Prog. Aerosp. Sci. **44**(7), 7479502 (2008)
11. Shao, W., Li, J.: Subsonic flow over open cavities. Part 2: passive control methods. In: Proceedings of ASME Turbo Expo 2016: Turbomachinery Technical Conference and Exposition, GT2016, , Seoul 13–17 June 2016
12. Suponitsky, V., Avital, E., Gaster, M.: On three-dimensionality and control of incompressible cavity flow. Phys. Fluids. **17**(10), 104103 (2005)
13. Mkhitaryan, A.M., Lukashuk, S.A., Trubenok, V.D., Fridland, V.Y.: Influence of spoilers on the aerodynamic characteristics of a wing and a solid of revolution [in Russian]. Naukova Dumka, Kyiv, 254–263 (1966)
14. Ringleb F.O.: Two-dimensional flow with standing vortex in ducts and diffusers. Trans. ASME. J. Basic Eng. **10**, 921–927 (1960)
15. Gorban, I.M., Khomenko, O.V.: Active near-wall flow control via a cross groove with suction. In: Zgurovsky, M.Z., Sadovnichiy, V.A. (eds.) Studies in Systems, Decision and Control. Continuous and Distributed Systems II: Theory and Applications, vol. 30, pp. 353–367. Springer, Berlin (2015)
16. Cottet, G.-H., Koumoutsakos, P.: Vortex Methods: Theory and Practice. Cambridge University Press, London (2000)
17. Vortex methods.: Proceeds of the 1-st International Conference of Vortex Motions. Kamemoto, K., Tsutahara, M. (eds.). World Scientific, Singapore (2000)
18. Gorban, V., Gorban, I.: Vortical flow structure near a square prism: numerical model and algorithms of control [in Ukrainian]. J. Appl. Hydromech. **7**, 8–26 (2005)
19. Gorban, I.M., Khomenko, O.V.: Flow control near a square prism with the help of frontal flat plates. In: Zgurovsky, M.Z., Sadovnichiy, V.A. (eds.) Studies in Systems, Decision and Control. Advances in Dynamical Systems and Control, vol. 69, pp. 327–350. Springer, Berlin (2016)
20. Lockard, D.P.: An efficient, two-dimensional implementation of the Ffowcs Williams and Hawkings equation. J. Sound Vib. **229**(4), 897–911 (2000)
21. Guo, Y.P.: Application of the Ffowcs Williams-Hawkings equation to two-dimensional problems. J. Fluid Mech. **403**, 201–221 (2000)
22. Filchakov, P.F: Approximate methods of conformal mappings [in Russian]. K. Naukova Dumka, Kiev (1964)
23. Wu, J.C.: Numerical boundary conditions for viscous flow problems. AIAA J. **14**(8), 1042–1049 (1976)
24. Lamb, G.: Hydromechanics. Cambridge University Press, London (1916)
25. Schlichting, H.: Boundary-Layer Theory. McGraw-Hill, New York (1979)
26. Larsson, J., et al.: Aeroacoustic investigation of an open cavity at low Mach number. AIAA J. **42**(12), 2462–2473 (2004)
27. Ozsoy, E., Rambaud, P., Stitou, A., Riethmuller, M.L.: Vortex characteristics in laminar cavity flow at very low Mach number. Exp. Fluid. **38**, 133–145 (2005)
28. Gloerfelt, X., Bogey, C., Bailly, C., Juvé D.: Aerodynamic noise induced by laminar and turbulent boundary layers over rectangular cavities. AIAA Paper 2002–2476 (2002)

Chapter 10
Numerical Modeling of the Wing Aerodynamics at Angle-of-Attack at Low Reynolds Numbers

Iryna M. Gorban and Oleksiy G. Lebid

Abstract Flows over symmetrical airfoils are numerically investigated for Reynolds number of 500. The high-resolution vortex method is used for the computations. The effects of both airfoil thickness and angle-of-attack (AoA) on non-linear wake and aerodynamic loads are examined. When increasing AoA from 0° to 60°, a flow regime in the airfoil wake was found to change from stationary to multiperiodic one through the Hopf bifurcation and period-doubling bifurcation. The highest lift-drag ratio of the airfoil is achieved in the stationary regime, when AoA< 15°. With further increase in the angle-of-attack, the airfoil performance drops due to increment in the drag force. The obtained results show that a thinner airfoil has better hydrodynamic characteristics but the effect of thickness is considerable in the stationary regime only. The analysis of pressure fields shows that negative pressure zones form not only in the airfoil frontal part, as at large Reynolds numbers, but near the trailing edge that is due to effect of boundary layer. The intensity of those grows with increasing an angle-of-attack.

10.1 Introduction

Interest in aerodynamics of low Reynolds numbers is kept up by the performance of mechanical systems operating with relatively small speeds such as unmanned aerial vehicles (UAVs) and wind turbines. Besides, UAVs move at large enough angles of attack during a maneuver, especially when landing that is followed by the sudden drop of its lift force. So, the study of such flight regimes is important for the further improvement of UAV technology. As for wind turbines, those work under significant non-stationary loads due to rapid changes in direction and speed of air

I. M. Gorban (✉)
Institute of Hydromechanics, National Academy of Sciences of Ukraine, Kyiv, Ukraine

O. G. Lebid
Institute of Telecommunications and Global Information Space, National Academy of Sciences of Ukraine, Kyiv, Ukraine

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_10

flows that requires evaluating the forces acting on the wing in wide range of both angles of attack and Reynolds numbers.

Flows around aerodynamic shapes at low Reynolds numbers are known to have a complex nature conditioned by boundary layer separation, flow reattachment and unsteady shedding of vortices. When increasing the angle of attack, the phenomena become stronger that affects greatly on the airfoil efficiency.

The creation of slow-speed flying apparatus, which operate in a wide range of angles of attack, has stimulated researches dealing with unsteady aerodynamics of wing at a small Reynolds number. In most of them, biomechanical analogies of a wing like flexible or flapping configurations were considered [1–3]. At subcritical angles of attack, birds and insects were found to enhance the lift force owing to generation of a stable leading-edge vortex.

On the other hand, there is a lack of papers devoted to translational movement of a rigid wing at low Reynolds numbers. Most of the research effort in this point is directed toward the understanding of the flow field over flat-plate airfoils, whose separation points are fixed at the plate edges. In paper [4], both two- and three-dimensional numerical simulations of the flow over an impulsively started flat plate at a chord Reynolds number of 300 were carried out. A number of simulations were performed with varied aspect ratio AR, angle of attack α ($\alpha = 0°$–$60°$) and plan-form geometry. It was found that aspect ratio influences significantly on the wake patterns and forces experienced by the plate. In three-dimensional flows, leading-edge vortices are evolved into hairpin vortices that interact with the tip vortices. These interactions weaken non-stationary effects; as a result, the highest loads on the plate develop at large aspect ratios when the airfoil flow is two-dimensional.

The fundamental aspects of the flow separated behind an inclined plate were numerically analyzed by Zhang et al. [5] at $Re \in [100, 850]$ and $\alpha \in [0°, 45°]$. It was found that the vortical flow patterns behind the plate depend on correlation between the angle of attack and the Reynolds number. The transition from steady to chaotic flow is realized through the Hopf bifurcation, period-doubling bifurcations and various incommensurate bifurcations. Yang et al. [6] captured numerically the shedding characteristics of the wake behind a flat plate at $\alpha = 30°$ and $Re = 750$. Two different problem statements were considered, in which either the plate or the incoming flow were inclined relative to the Cartesian grid. The central conclusion of the research lies in the fact that the statements are quite coincident.

In papers [7, 8], a flow near the NACA-0012 airfoil was calculated by Lattice-Boltzmann Equation solver based on the kinetic theory of fluid. The 2D and 3D statements as well as different Reynolds numbers and angles of attack were considered. It was found that the drag coefficient C_D at $\alpha = 0°$ and $Re = 500$ compares very well with that obtained by a Navier-Stokes equation-based finite difference method.

Kuns and Kroo [9] explored numerically flows near two-dimensional airfoils at $Re < 10^4$ and $\alpha \in [0°, 10°]$ under the assumption that the flow field is fully laminar. Variations in the thickness, camber, and leading/trailing edge shape were considered. The main conclusion of this study affirms that airfoil flows at low Reynolds numbers are viscously dominated that leads to very large increase of the

drag coefficient. So, flight at these Reynolds numbers is much less efficient than at higher Re, although the lift of wing increases due to generating the intensive bubble at the low pressure side.

Experimental investigations of the problem are limited by the aerodynamic regimes, when $\alpha \leq 15°$. The researches of Mueller and Batillt [10] accentuated on the airfoil boundary layer and the leading-edge separation bubble. The dynamics of the bubble and its influence on airfoil loads for a chord Reynolds number range of $4 \cdot 10^4$–$4 \cdot 10^5$ were investigated. The rapid rise in lift with increasing the angle of attack was observed at $Re = 4 \cdot 10^4$. At the same time, it was shown that a negative lift can be produced at $Re = 1.3 \cdot 10^5$ and $\alpha \leq 4°$ as a result of laminar separation on upper and lower surfaces downstream of midchord.

Systematic studies of aerodynamic characteristics of different shape airfoils were examined by Sunada et al. at $Re = 4 \cdot 10^3$ [11, 12]. In this case, airfoils with good aerodynamic performance were found to be thinner than airfoils for higher Reynolds numbers and have a sharp leading edge and a camber of about five percent. However, a thinner wing has lower rigidity; as a result, it cannot support the favorable distribution of the surface pressure for a long time. It was shown that by analogy with the insects wings the corrugation of surface can be used to eliminate this disadvantage.

It follows from the above studies that the airfoil flows at low Reynolds numbers essentially depend on this parameter. So, the successful operation of wing systems at these conditions demands additional investigations of aerodynamic flows in the specified range of Reynolds number.

In this work, numerical simulations of viscous two-dimensional flows around aerodynamic profiles are performed with applying the vortex method [13], which belongs to high-resolution Lagrangiantype schemes developed as fast alternative to direct numerical simulations. The vortex numerical schemes have been successfully used for calculation of various flows that take place in natural environment and technical applications [14]. Those are based on the idea of transition in a mathematical model from natural variables, pressure and velocity, to vorticity and focus on its creation, transport and diffusion. Advantages of the vortex schemes are the absence of pressure in the equations; automatic enforcing boundary conditions at infinity; an inherent economy in the number of computational elements since vorticity is usually concentrated near bodies and in their wakes; a direct physical interpretation of the results and internal ability of the algorithm to be parallelized.

At first the algorithm was tested by the example of the airfoil flow over NACA0012 profile at $Re = 500$ and $\alpha = 0°$. The calculated surface pressure as well as drag coefficient are close to its values obtained by high order numerical schemes. Once the numerical results have been validated, a number of simulations are carried out for NACA0008 and NACA0018 profiles, to estimate an influence of the thickness alongside other factors.

In this simulation, the vortical flow patterns behind the airfoils and the corresponding lift and drag coefficients are calculated at $Re = 500$ when the flow near an airfoil is still laminar. The fundamental features of the airfoil flow in the wide range of the angle of attack α, from 0° to 60°, are derived. The rise of α is shown to lead

to changing the vortical flow pattern in the wake from stationary to multiperiodic through the Hopf bifurcation and period-doubling bifurcation. Calculations of the airfoil dynamic characteristics corresponding to these regimes indicate that the highest ratio of the lift to drag is achieved with the stationary flow. When the angle of attack increases, the aerodynamic performance drops to the values that are less than one due to a significant increment in the drag.

An analysis of the pressure fields near the airfoils shows that the lift force is conditioned by the dynamics of the separation bubble generated on the upper side. At $\alpha < 15°$, the lift is provided by the vortex sheet leaving the trailing edge and the leading-edge separation bubble. When increasing α, unsteady effects prevail in the wake, as a result, vortex shedding occurs. This leads to both oscillations of the lift and the change in the mechanism of its generation. At $\alpha > 20°$, the lift depends mainly on the strength and size of the recirculation bubble generated on the upper surface. The results of calculations demonstrate the formation of negative pressure zones not only in the airfoil frontal part, as at large Reynolds numbers, but also near the trailing edge, which is due to the influence of viscosity. When increasing an angle of attack, the strength of the rear recirculation zone growths.

It follows from the comparison of dynamic characteristics of NACA0008 and NACA0018 airfoils, the thinner profile has better performance but the flow around the thick one is more regular. In general, the results obtained show that operating wing systems at low Reynolds numbers is significantly different from traditional aerodynamic regimes due to the domination of viscous effects in the flow.

10.2 Problem Statement and Method Description

A two-dimensional flow of incompressible Newtonian fluid of velocity V_0 around the symmetrical airfoil of chord d is studied. The geometry of the problem and coordinate system are depicted in Fig.10.1.

Two control parameters of the problem are the inclined angle α (angle of attack) with respect to the free-stream flow and the Reynolds number $Re = V_0 d/\nu$, where ν is the kinematic viscosity of fluid. The flow under consideration is described by the Navier-Stokes equations:

$$\nabla \mathbf{V} = 0, \tag{10.1}$$

$$\frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V} \cdot \nabla)\mathbf{V} = -\nabla p + \frac{1}{Re}\nabla^2 \mathbf{V}, \tag{10.2}$$

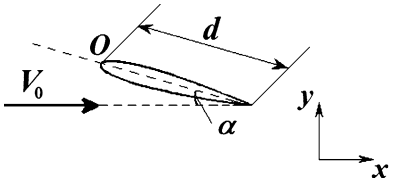


Fig. 10.1 The geometry of interest

where $\mathbf{V}(u, v)$ is the velocity vector, p is the non-dimensional pressure and t is the time scale. Note that all geometrical lengths are normalized with d, velocities with V_0, physical times with d/V_0 and pressure is specified by the dynamic coefficient $\rho V_0^2/2$, where ρ is the fluid density.

On the airfoil, the slipping condition must be satisfied

$$\mathbf{V} \cdot \mathbf{n}|_L = 0, \tag{10.3}$$

$$\mathbf{V} \cdot \boldsymbol{\tau}|_L = 0, \tag{10.4}$$

where L denotes the airfoil contour, $\mathbf{n}$, $\boldsymbol{\tau}$ are the normal and tangential unit vectors to L.

It should be noted that the application of the model of incompressible fluid to simulation of the flows at the subsonic velocities that corresponds to the Mach number less than 0.2 is quite justified and does not introduce a significant error in the calculations [15].

The governing equations are solved by the vortex method, which considers vorticity $\omega = \mathbf{k} \cdot \nabla \times \mathbf{V}$ as the primary variable (here $\mathbf{k}$ is the unit vector out of the page) and examines its translation in the flow field according to the vorticity transport equation (see Cottet and Koumoutsakos [13]).

$$\frac{\partial \omega}{\partial t} + (\mathbf{V} \cdot \nabla)\omega = \frac{1}{Re}\Delta\omega. \tag{10.5}$$

To satisfy boundary conditions in the vorticity equation, the Lighthill creation mechanism describing generation of vorticity at the solid boundary is used. It attaches a vortex sheet on the airfoil surface to cancel the spurious slip arising due to transformations in the vorticity field. Then the Biot–Savart integral that recovers the velocity field from the vorticity takes the following form:

$$\begin{aligned}\mathbf{V}(\mathbf{r}, t) = &\int_S \int \omega(\mathbf{r}', t)\mathbf{k} \times \nabla G(\mathbf{r}, \mathbf{r}')ds(\mathbf{r}') + \\ &\int_L \gamma(\mathbf{r}', t)\mathbf{k} \times \nabla G(\mathbf{r}, \mathbf{r}')dl(\mathbf{r}') + V_0,\end{aligned} \tag{10.6}$$

where S is the flow domain, $\mathbf{r}$ is the radius-vectors of the field points, ω is the vorticity, γ is the strength of bound vortex sheet, G is the fundamental solution of the Laplacian: $G(\mathbf{r}, \mathbf{r}') = \frac{1}{2\pi} ln|\mathbf{r} - \mathbf{r}'|$.

The strength γ is determined from no-slip boundary condition (10.4). Taking into account the tangential velocity jump on the vortex sheet, which is equal to $\gamma/2$, and

equality $\mathbf{n}(\mathbf{r}) \times \boldsymbol{\tau}(\mathbf{r}) = \mathbf{k}$ one obtains from (10.4) the following equation relative to γ:

$$\int\limits_L \gamma(\mathbf{r}', t) \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial n} dl(\mathbf{r}') - \frac{\gamma(\mathbf{r})}{2} = \boldsymbol{\tau}(\mathbf{r}) \cdot \left(V_0 + \int\limits_S \int \omega(\mathbf{r}', t) \mathbf{k} \times \nabla G(\mathbf{r}, \mathbf{r}') ds(\mathbf{r}') \right), \tag{10.7}$$

where $\mathbf{r} \in L$.

Note that in compliance with the conception of *linked boundary conditions* substantiated by Shiels [16], satisfaction of the tangential velocity condition means that no-through boundary condition (10.3) is also satisfied since only a translation motion of the boundary enclosed body is considered.

Equation (10.7) that defines the strength of bound vortex sheet is a Fredholm equation of the second kind. It is a singular one and additionally it admits a non-unique solution. The uniqueness of the solution of equation (10.7) is ensured with Kelvins circulation theorem that relates the vortex sheet strength with the change of the flow circulation. The theorem requires the conservation of the total circulation in the flow domain and it is expressed by the equation:

$$\int\limits_L \gamma(\mathbf{r}', t) dl(\mathbf{r}') + \int\limits_S \int \omega(\mathbf{r}', t) ds(\mathbf{r}') = 0. \tag{10.8}$$

After a vortex sheet on the boundary is obtained to cancel the slip velocity, it has to be transferred to the fluid domain. This is achieved by solving a diffusion equation in respect to the vorticity with a Neuman-type boundary condition that connects the vortex sheet strength with the vorticity flux. This condition has been formulated in paper [17], which is classical for the vortex methodology. In accordance with it, the total flux of vorticity to be emitted into the flow for the small time step δt is connected with the sheet strength by the expression:

$$\frac{\partial \omega_0}{\partial n} = -\frac{\gamma}{\nu \delta t}, \tag{10.9}$$

where ω_0 is the surface vorticity.

10.3 Numerical Methodology

To realize numerically the mathematical model described in Eqs. (10.5)–(10.9), we put a uniform orthogonal grid $(i\Delta x, j\Delta y)$ on the flow domain, where Δx, Δy are small spatial scales, $i = 1, 2, \ldots, N_x$, $j = 1, 2, \ldots, N_y$. We introduce also the

volume cells $Q(x, y) = \{\xi, \eta : |\xi - x| < \Delta x/2, \ \ |\eta - y| < \Delta y/2\}$ around the mesh points of the grid and assume the vorticity occupying the cell Q_{ij} is converted into a vortex particle of circulation Γ_{ij}. Then the continuous vorticity field is replaced in full with a finite system of the discrete vortices, whose circulation is determined by the expression:

$$\Gamma_{ij} = \int\limits_{Q_{ij}} \int \omega(x, y) dx dy \approx \omega_{ij} \Delta x \Delta y. \tag{10.10}$$

Note that the present realization of the vortex method supposes a uniform distribution of vorticity inside the cell Q_{ij}.

Equation (10.5) is solved on the grid introduced using a fractional step procedure when that is split into convective and diffusive parts, which are integrated separately. The algorithm and its applications to modeling bluff body flows have been described in details in papers [18, 19]; so, only the principal features of the numerical scheme are considered here.

The viscous diffusion equation is integrated by the finite-difference method on the orthogonal grid with Γ_{ij} in the mesh points. If the spatial derivative of this equation is approximated by the scheme of the second order and a simple Euler explicit scheme for temporal discretization is used for updating Γ_{ij}, the resulting finite difference scheme will take the following form:

$$\frac{\Gamma_{ij}^{t+\Delta t} - \Gamma_{ij}^{t}}{\Delta t} = \frac{1}{Re} \left(\frac{\Gamma_{i+1j}^{t} - 2\Gamma_{ij}^{t} + \Gamma_{i-1j}^{t}}{(\Delta x)^2} + \frac{\Gamma_{ij+1}^{t} - 2\Gamma_{ij}^{t} + \Gamma_{ij-1}^{t}}{(\Delta y)^2} \right), \tag{10.11}$$

where Δt is the time step.

In the vortex methods, the vorticity convection is usually simulated by a displacement of vortex elements along fluid particle trajectories. It means the vortices move with their induced velocities while the vortex circulation does not change during this sub step. In our approach, we calculate numerical fluxes of vorticity through boundaries of grid cells Q_{ij} according to the convection equation that leads to changing the circulation of the vortices fixed in the mesh points of the given orthogonal grid. Taking into account that the normal vectors to the boundaries of the vorticity element Q_{ij} coincide with the directions of coordinate axes, one can write the vorticity balance inside Q_{ij} during the time step Δt as follows:

$$\begin{aligned} &\frac{\omega_{ij}^{t+\Delta t} - \omega_{ij}^{t}}{\Delta t} \Delta x \Delta y \approx (\omega_{i-1j}^{t} u_{i-1j}^{t} - \omega_{i+1j}^{t} u_{i+1j}^{t}) \Delta y + \\ &(\omega_{ij-1}^{t} v_{ij-1}^{t} - \omega_{ij+1}^{t} v_{ij+1}^{t}) \Delta x - \omega_{ij}^{t} (|u_{ij}^{t}| \Delta y + |v_{ij}^{t}| \Delta x). \end{aligned} \tag{10.12}$$

From (10.12), the evolutionary equations for the vortex circulation are derived:

$$\Gamma_{ij}^{t+\Delta t} = \Gamma_{ij}^{t} + \left[\left(\Gamma_{i-1j}^{t} u_{i-1j}^{t} - \Gamma_{i+1j}^{t} u_{i+1j}^{t}\right)/\Delta x + \left(\Gamma_{ij-1}^{t} v_{ij-1}^{t} - \Gamma_{ij+1}^{t} v_{ij+1}^{t}\right)/\Delta y - \Gamma_{ij}^{t}\left(|u_{ij}^{t}|/\Delta x + |v_{ij}^{t}|/\Delta y\right)\right]\Delta t. \tag{10.13}$$

Scheme (10.13) is of the second order in space and of the first order in time. Note also it is a non-dissipative and has improved dispersion properties compared to classical linear schemes.

The airfoil contour L is discretized into N_p boundary elements (panels) and each of them assumes a constant value of vortex strength γ_k, $k = 1, 2, \ldots, N_p$. In the computation scheme, the panel is replaced by a single point vortex, whose circulation Γ_k^p is equal to that of the panel: $\Gamma_k^p = \gamma_k \Delta l_k$, where Δl_k is the length of k-th panel. The vortices are assumed to be located in the middle of the panels and then integrals in expressions (10.6)–(10.8) can be replaced by sums based on trapezoidal quadrature. Formula (10.6) for determination of the velocity in the flow domain takes the following form:

$$\mathbf{V}(\mathbf{r}, t) = \frac{1}{2\pi}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\Gamma_{ij}\frac{\mathbf{k}\times(\mathbf{r}-\mathbf{r}_{ij})}{|\mathbf{r}-\mathbf{r}_{ij}|^2} + \frac{1}{2\pi}\sum_{k=1}^{N_p}\Gamma_k^p\frac{\mathbf{k}\times(\mathbf{r}-\mathbf{r}_k^p)}{|\mathbf{r}-\mathbf{r}_k^p|^2} + V_0 \tag{10.14}$$

where $\mathbf{r}_{ij}$ and $\mathbf{r}_k^p$ are the radius-vectors of free and bound vortices, respectively.

Singular integral equation (10.9) together with condition (10.10) are reduced to a system of linear algebraic equations with respect to the circulations of bound vortices Γ_k^p:

$$\sum_{k=1,k\neq m}^{N_p}\Gamma_k^p\frac{\mathbf{n}(\mathbf{r}_m^p)\cdot(\mathbf{r}_m^p-\mathbf{r}_k^p)}{|\mathbf{r}_m^p-\mathbf{r}_k^p|^2} - \frac{\pi\Gamma_m^p}{\Delta l_m} + R = \boldsymbol{\tau}(\mathbf{r}_m^p)\cdot\left[V_0 + \sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\Gamma_{ij}\frac{\mathbf{k}\times(\mathbf{r}_m^p-\mathbf{r}_{ij})}{|\mathbf{r}_m^p-\mathbf{r}_{ij}|^2}\right], \quad m = 1, 2, \ldots, N_p, \tag{10.15}$$

$$\sum_{k=1}^{N_p}\Gamma_k^p = \sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\Gamma_{ij}. \tag{10.16}$$

Here $\mathbf{n}(\mathbf{r}_m^p)$, $\boldsymbol{\tau}(\mathbf{r}_m^p)$ are the normal and tangent unit vectors to the contour L at the m-th panel, R is the regularizing variable, which is introduced because of the obtained system is overdetermined. Equation (10.15) are written for the control points $\mathbf{r}_m^p$,

those are located in the middle of the panels and coincide with the bound vortices. Such an approach to constructing a discrete analogue of Eq. (10.7) has been applied in paper [20], in which the correctness of the quadrature formula for the principal value of the considered singular integral is substantiated.

The calculated vortex sheet $\gamma_k = \Gamma_k^p / \Delta l_k, k = 1, 2, \ldots, N_p$, must be distributed to neighbor vortices of the fluid domain in accordance with the diffusion equation and Neumann boundary condition (10.9). In this work, we apply the solution of the diffusion equation developed by Ploumhans and Winckelmans [21] since it allows accurate calculation of the vortical flows past bluff bodies of general geometry including those with great surface curvature.

To simulate the inertial flow separation in the trailing edge of airfoil, the Kutta-Joukowski condition is used. This is satisfied by appropriate discretization of the airfoil contour, when the vortex is displaced in the sharp edge and it is transferred to the flow. Note that the technique has been successfully used by Belotserkovsky et al. [22] for simulation of viscous flows around bluff bodies with sharp edges.

The velocity-vorticity formulation of the Navier-Stokes equations allows decouple purely kinematical problem from the pressure problem that simplifies significantly numerical modeling of fluid flows. To recover the pressure field from the vorticity and velocity, we integrate the Navier-Stokes equations in the Lamb representation on the base orthogonal grid (see [19]). The coefficients of the forces acting on the airfoil are calculated from the pressure distribution on its surface:

$$C_D = \int_L \bar{p} n_x \, dx, \quad C_L = \int_L \bar{p} n_y \, dy, \tag{10.17}$$

where C_D, C_L are the coefficients of drag and lift, respectively, n_x, n_y are the components of the internal normal to the airfoil and $\bar{p} = 2(p - p_\infty)/\rho V_0^2$ is the pressure coefficient.

10.4 Results

With the vortex method described above, simulations of the fluid flows past an inclined symmetrical NACA profiles were carried out at $Re = 500$ and $0° \leq \alpha \leq 60°$. The geometry of such a profile depends on its thickness only and is given by the following function:

$$y(x) = \pm \frac{t}{0.2}(0.2969\sqrt{x} - 0.1260x - 0.3516x^2 + 0.2843x^3 - 0.1015x^4) \tag{10.18}$$

where t is the airfoil thickness.

We discuss the evolution of the vortical flow and pressure fields generating around 8% and 18% airfoils as well as their dynamic characteristics.

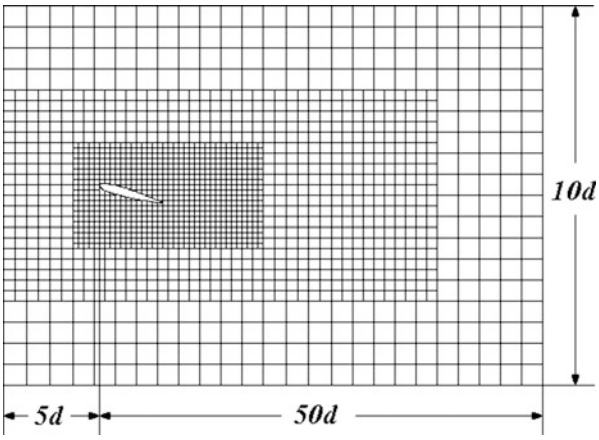


Fig. 10.2 Sketch of the computational grid

10.4.1 Discretization Details

In this study, we adopt the three-level rectangular grid with a constant cell size at each level as presented in Fig.10.2.

The grid spacing in the domain adjoining the airfoil is chosen as $\Delta x = \Delta y = 0.005d$ and the cell size of each next grid is doubled compared with the previous. The dimensionless width of the calculation region is 10 and the lengths of upstream and wake regions are 5 and 50, respectively (Fig.10.2). As test calculations have shown, a further enlargement of the computational domain has no appreciable influence on kinematic and dynamic flow characteristics to be derived.

The explicit scheme for integration in time is employed and that is why the normalized time step Δt has to be chosen from the Courant–Friedrichs–Lewy condition:

$$\frac{\max(u, v)\Delta t}{\min(\Delta x, \Delta y)} \leq 1$$

The maximum value of the local flow velocity $V(u, v)$ obviously depends on the flowed surface curvature, angle of attack and Reynolds number. As calculations have shown, the numerical stability for the present flow conditions can be achieved at $\Delta t = 0.5$, where h is the finest grid spacing.

The airfoil contour is sampled with the following equation [22]:

$$x_k = 1 - sin^2\left(\pi \frac{k-1}{N_p}\right), \qquad k = 1, 2, \ldots, N_p, \tag{10.19}$$

where x_k is the abscissa of the control point at zero angle of attack. Its ordinate is obtained from function (10.18). Note the present calculations are performed at $N_p = 200$.

The developed numerical scheme was earlier applied by us for modeling separation flows past bluff bodies and in body systems at moderate Reynolds numbers. In particular, the results of the detailed calculations for a square prism at $10^2 \leq Re \leq 10^3$ are presented in [19]. Those demonstrate a good agreement of the Strouhal number and averaged hydrodynamic loads with related numerical and experimental studies. A small difference (up to 10%) is only observed for the amplitudes of non-stationary forces.

We performed also the test calculations of the viscous flow around the NACA0012 profile at $Re = 500$ and zero angle of attack. The obtained averaged fields of vorticity and pressure around the profile are presented in Fig. 10.3a, b. The results indicate that the main contribution into the wing drag in this flow configuration is provided by the surface friction, so, the net force acting on the wing is here calculated from the vorticity field with applying the impulse theorem [13]. The obtained value $C_D = 0.175$ compares very well with $C_D = 0.1761$ reported in paper [7].

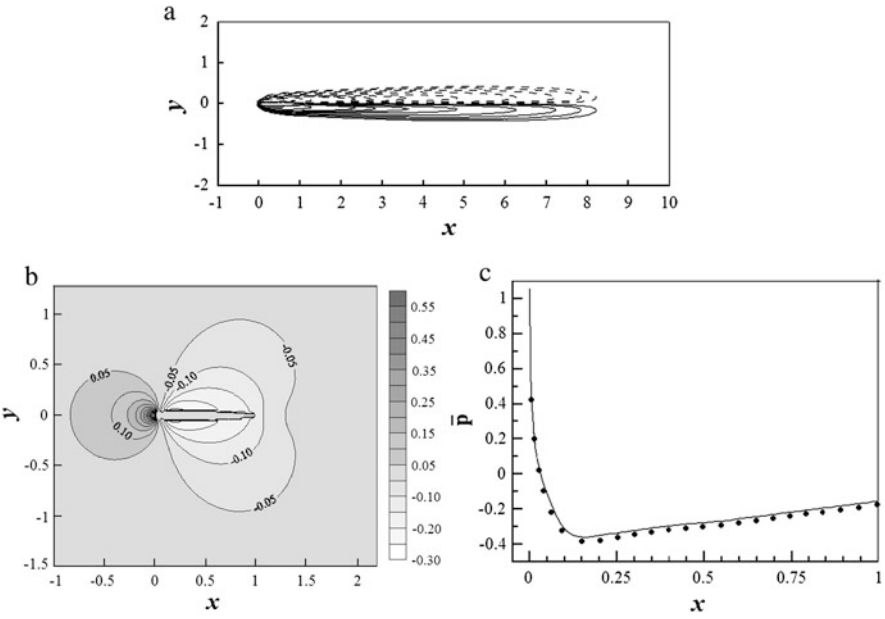


Fig. 10.3 NACA0012 profile at $\alpha = 0°$ and $Re = 500$: (**a**) vorticity contours, (**b**) pressure contours, (**c**) surface pressure coefficient. The solid line corresponds to the present simulations, markers illustrate the data from [8]

Figure 10.3c illustrates the pressure coefficient along the profile and its comparison with related data from paper [8]. Remember that calculations in [7, 8] are based on the PowerFLOW technique, which is in a good concurrency to finite difference methods used for integrating the Navier-Stokes equations. The above results indicate that the presented version of the vortex method is able to predict correctly a viscous flow around an airfoil in the considered range of the Reynolds number.

10.4.2 Vortical Flow Patterns and Frequency Analysis

In this section, the characteristics of the vortex field around an airfoil obtained in the present simulation are analyzed. The calculations were performed at $Re = 500$ and $\alpha \in [0°, 60°]$. Two symmetrical profiles NACA0008 and NACA0018 were considered to evaluate the effect of thickness along with other factors. Figure 10.4 illustrates the patterns of near wake past 8%-profile corresponding to different incidence, at $\alpha = 15°, 20°, 40°, 60°$. Here the instantaneous vorticity contours are presented and the solid lines denote the counter-clockwise rotation, the dash lines mean clockwise rotation. Figure 10.5 shows the Fourier power spectra of the velocity at the point located in the airfoil wake at $\alpha = 20°$ and $\alpha = 40°$. Note the frequencies in Fig. 10.5 are normalized by V_0/d.

It is seen in Fig. 10.4 that the vortical flow pattern past the airfoil changes from stationary to multiperiodic one in the considered range of angles of attack. In the stationary regime, when $\alpha \leq 15°$ (Fig. 10.4a), the airfoil wake is composed of the opposite regular vortex sheets. The clockwise sheet separates from the trailing edge and the counter-clockwise sheet generates in the front face. When increasing the angle of attack to $\alpha = 20°$, the vortical flow turns into periodic state via a Hopf bifurcation. It is seen in Fig. 10.4b that the regular von Karman vortex street is formed past the airfoil in this case. The regularity of the process is confirmed by the power spectrum of velocity which contains peaks at the primary frequency f_1 and its harmonic frequency $2f_1$ (Fig. 10.5a). The vortices separating from the front face and in the trailing edge are approximately of the same scale and almost do not interact with each other. It is obtained in the calculations, the Hopf bifurcation in the wake of NACA0008 profile occurs at $\alpha \approx 18°$, while a regular vortex street is observed before $\alpha \approx 30°$.

When α is further increased to 40°, significant complication of the vortical flow in the airfoil wake is observed. In this state, both the scale and the strength of the vortex structures grow, in addition, the opposite vortices intensively interact (Fig. 10.4c—$\alpha = 40°$). The vortex street is relatively stable only at a short distance from the profile, approximately to $x \approx 12d$. Further the opposite vortices push off from each other; as a result, the wake broadens greatly. The corresponding power spectrum (Fig. 10.5b) contains a series of peaks at the primary frequency f_1 and its harmonic frequencies $2f_1$, $3f_1$ etc. It indicates that the wake is still regular enough but the presence of low periodic phenomena indicates a strong interaction between the vortices.

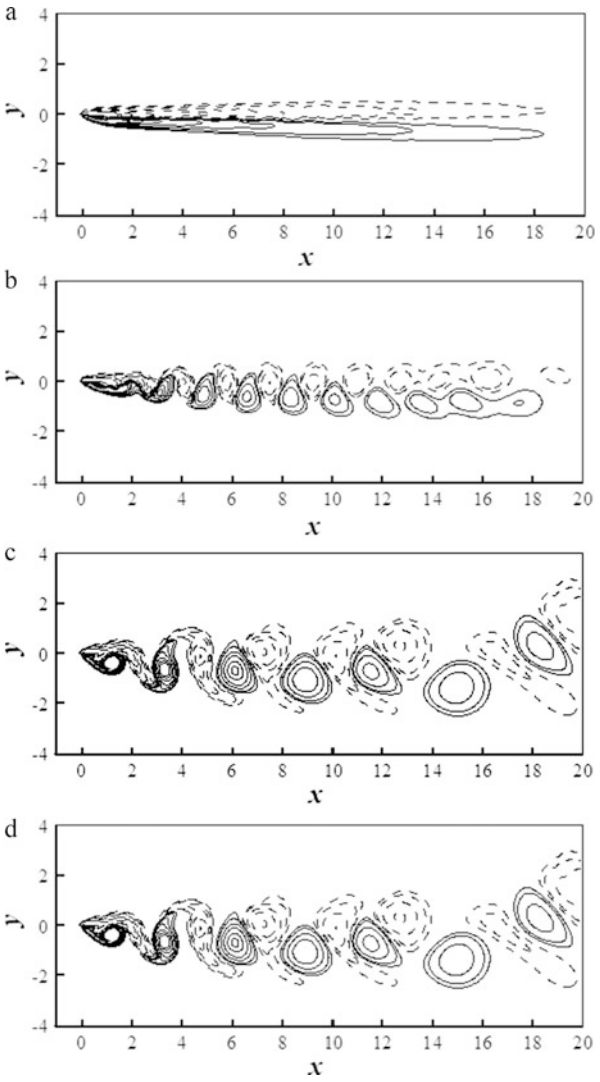


Fig. 10.4 Instantaneous vorticity fields in the wake of NACA0008 profile at different angles of attack: $|\omega_{min}| = 0.25$, $|\omega_{max}| = 4$; $t = 50$. (**a**) $\alpha = 15°$. (**b**) $\alpha = 20°$. (**c**) $\alpha = 40°$. (**d**) $\alpha = 60°$

The wake pattern derived at $\alpha = 60°$ (Fig. 10.4d) is known as period-doubling state [23], when opposite vortices form stable vortex pairs without merging. The clockwise vortex separated from the front face being split into two parts where the foot is attracted to the counter-clockwise vortex generated in the trailing edge. This leads to violation of the basic spatial structure of the vortex street but not to chaotic states in the wake. The power spectrum of velocity in this case is found to be characterized by occurrence of subharmonic frequency $f_1/2$, which corresponds to the period of vortex pairing. We have identified also that the transformation of the wake to the period-doubling state begins at $\alpha \approx 45°$.

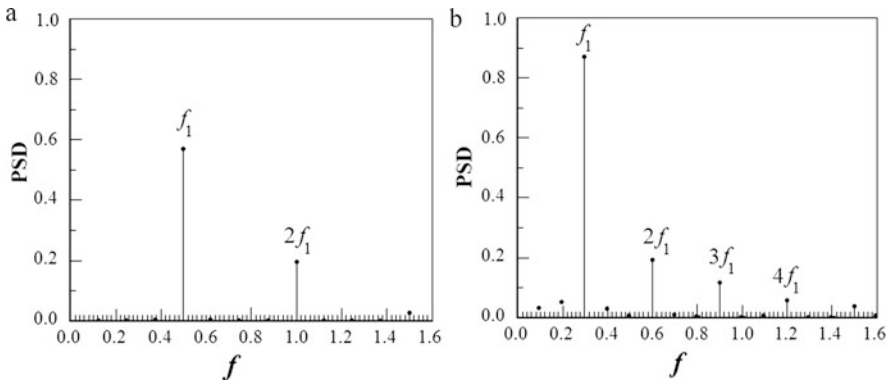


Fig. 10.5 Power spectrum density of the velocity component in the y- direction in the near wake of NACA0008 profile at different angles of attack: (**a**) $\alpha = 20°$, (**b**) $\alpha = 40°$

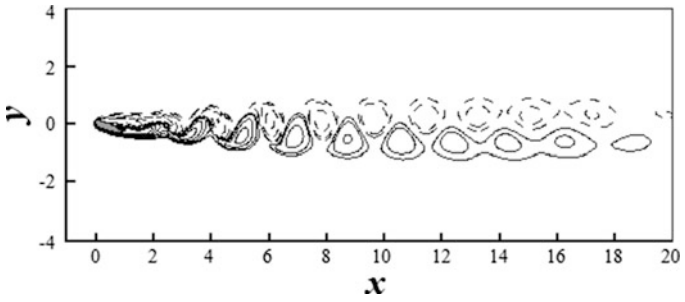


Fig. 10.6 Instantaneous vorticity in the wake of NACA0018 profile at $\alpha = 15°$, $t = 50$

It is revealed in the simulation that an increase of the thickness of profile leads to destabilization of the wake in the sense that transitional phenomena occur earlier relative to the angle of attack. This fact is confirmed by Fig. 10.6, where the pattern of vorticity past NACA0018 profile at $\alpha = 15°$ is presented. One can see here that the vortex sheets have disintegrated into the vortex street in contrast to the results presented in Fig. 10.4a.

The dependence of the main frequency f_1, or Strouhal number St on the angle of attack is presented in Fig. 10.7. The frequency characterizes shedding of large-scale vortices from the airfoil. The results also reflect the influence of the profile thickness because the two curves in this figure correspond to NACA0008 and NACA0018 profiles. The obtained vortex shedding frequency is seen to decrease from 0.52 to 0.2 that indicates a significant augmentation in the scale and intensity of the vortex structures. The effect of profile thickness is significant at smaller angles of attack from the predetermined range and it disappears with transition to period-doubling state.

In summary, the results presented show the dominance of viscous effects in the airfoil flow at the Reynolds number under consideration. Those include the viscous

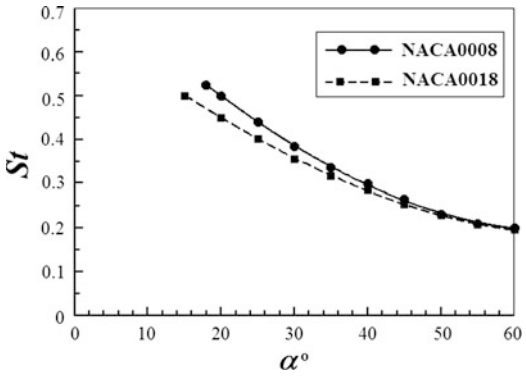


Fig. 10.7 Strouhal number St against the angle of attack α

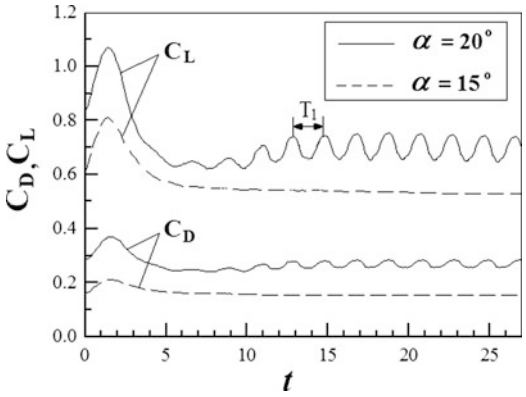


Fig. 10.8 Lift and drag history of NACA0008 profile at $\alpha = 15°$ and $\alpha = 20°$

separation of the boundary layer in the front part and intense interaction between the opposite sheets which leads to its disintegration and establishment of a certain wake pattern depending on the angle of attack.

10.4.3 Forces

In this section, the dynamic characteristics of the profiles under consideration are analyzed. It is obvious that the time history of force coefficients correlates with patterns of vortex shedding by the profile. At the start of the process, a sharp peak in both force coefficients, C_D and C_L, is observed in the entire range of an angle of attack. In the stationary regime, both coefficients reach their steady-state values after the peak. It is seen in Fig. 10.8, where the time history of C_D and C_L coefficients for NACA0008 profile in the stationary regime, at $\alpha = 15°$, is presented by the dotted line.

The pressure field near the profile indicates on the existence of a stable separation bubble attached to the upper surface at this regime. When an angle of attack

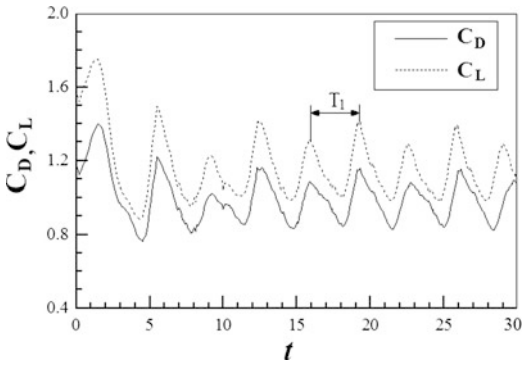


Fig. 10.9 Lift and drag history of NACA0008 profile at $\alpha = 40°$

is further increased, the flow on the upper surface of the airfoil separates and phenomenon known as stall developes. The wake modification at increase of an angle of attack causes changes in time evolution of airfoil forces. At $\alpha = 20°$, there is a broad bump in C_L and C_D before the regular oscillations (Fig. 10.8—solid line) that is a consequence of the gradual development of the separation bubble into a large leading edge vortex. In contrast, at an angle of attack of $\alpha = 40°$, the leading vortex develops rapidly and it is quite large, so its separation causes force oscillations that are comparable to action of the trailing edge vortex (Fig. 10.9). The curves in Figs. 10.8 and 10.9 show increase of both force coefficients when raising an angle of attack but the drag force grows much faster than the lift force and those are approximately equal to one another at $\alpha = 45°$.

Amplitudes of force coefficients also increase with an angle of attack due to strengthening the vortices generated by the airfoil. At $\alpha < 45°$, the magnitude of force oscillations is more substantial in the time history of C_L than of C_D, but after 45° oscillations in drag are more noticeable. It would be mentioned that the forces reach a maximum value around $t = 1.5$ for 8%—profile and when the profile is thickened the value shifts towards a smaller time. Note also that the period T_1 in Fig. 10.8 corresponds to the Strouhal number from Fig. 10.7, which has been calculated from the velocity oscillations in the wake.

Figure 10.10 illustrates the flow dynamics around NACA0008 profile at $\alpha = 40°$ by the way of distribution of vorticity (left) and pressure (right). The picture in Fig. 10.10a describes the fields at the time instance when the lift force coefficient amounts to its maximum value; on the contrary, Fig. 10.10b corresponds to the lift force minimum. The leading vortex in Fig. 10.10a is developed enough but it is still attached to the upper surface of airfoil. This ensures the intense rarefaction over the airfoil. At this instance, the trailing vortex is separating that causes the widening of positive pressure region at the lower surface. As a result, the difference between the upper and lower pressure increases and the airfoil lift achieves its maximum value. When the leading vortex is separating (Fig. 10.10b) the pressure over the airfoil grows and the lift force drops to the minimum value. So, at high angles of attack the lift generation mainly depends on the dynamics of the leading vortex. Besides, the presented results demonstrate the formation of negative pressure zones not only

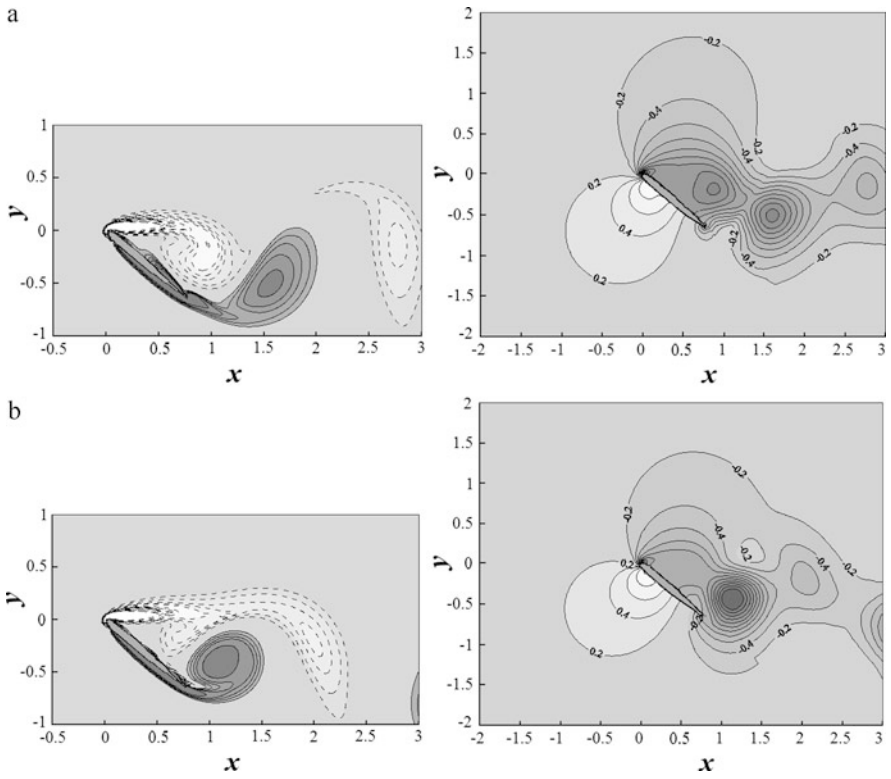


Fig. 10.10 Instantaneous vorticity (on the left) and pressure (on the right) fields around NACA0008 profile at $\alpha = 40°$ (**a**) at maximum lift (**b**) at minimum lift

in the airfoil frontal part, as at large Reynolds numbers, but also near the trailing edge, that is due to the influence of viscosity. When an angle of attack increases, the strength of the rear recirculation zone growths.

General conclusions about the efficiency of wing systems in the given ranges of angles of attack and Reynolds number come to light from the analysis of the time-average loads. The calculated time-average drag $\overline{C}_D$ and lift $\overline{C}_L$ coefficients as well as the lift-drag ratio $\overline{C}_D/\overline{C}_L$ versus the angle of attack α for the profiles under consideration are presented in Figs. 10.11 and 10.12, respectively. One can see three distinctive areas as for behavior of the airfoil forces. In the first area, the lift gradient is much higher than that of the drag, so the lift-drag ratio grows to its maximal value. For a NACA0008 profile, the value is about three and it is reached at $\alpha \approx 12°$. The efficiency of the NACA0018 profile is lower because its maximum performance is approximately equal to two. So, the thinner wing has better hydrodynamic characteristics. The conclusion complies with the data

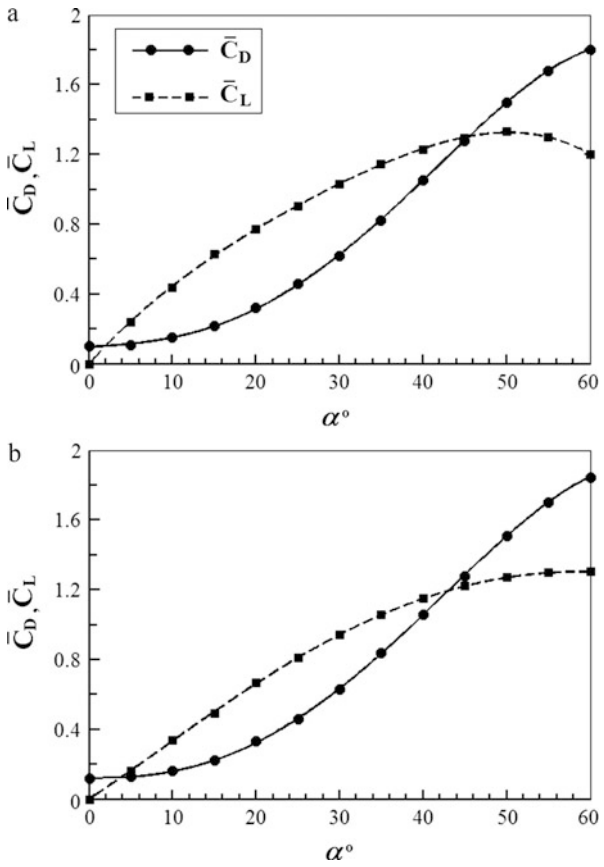


Fig. 10.11 Time-averaged force coefficients $\overline{C}_D$ and $\overline{C}_L$ against angle of attack α. (**a**) NACA0008 profile. (**b**) NACA0018 profile

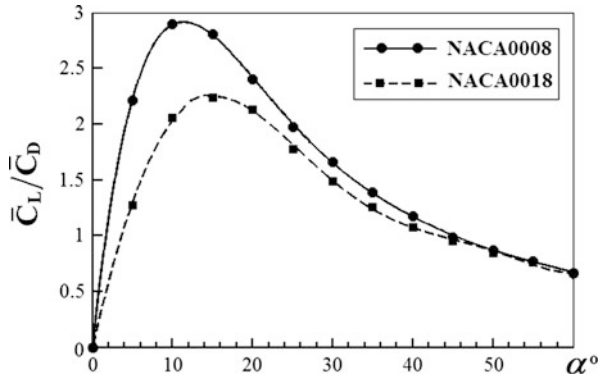


Fig. 10.12 Aerodynamic performance of NACA0008 and NACA0018 profiles against angle of attack α

of experimental study [12], in which the dynamic characteristics of the wings of different airfoil shapes have been measured at $Re = 4 \cdot 10^3$.

In the second area of the above-mentioned the ratio $\overline{C}_D/\overline{C}_L$ drops with increasing an angle of attack but it still remains higher than unit. For both considered profiles, the lift-drag ratio approaches a unit at $\alpha \approx 45°$. When $45° < \alpha \leq 60°$, the drag increases rapidly with an angle of attack, at the same time, the lift remains near a constant value and it even decreases slightly for a NACA0008 profile. So, the aerodynamic performance drops from 1 to 0.75 in this area.

The results in Fig. 10.11 show that the lift coefficient of NACA0008 profile grows greatly as compared with that for NACA0018 profile at small angles of attack from the range under consideration. Those are just the angles at which an effect of the profile thickness on the aerodynamic performance is maximal. Further an influence of viscosity and nonstationarity on the airfoil flow exceeds geometrical effects; accordingly the loads on both profiles are almost equal.

The estimates obtained in the present simulation are important for understanding the flow evolution and loads developing when wing systems operate at a low Reynolds number and in the wide range of angles of attack.

10.5 Conclusion

In this paper, the vortex numerical scheme is implemented to simulation of a two-dimensional viscous flow around wing profiles. The scheme was shown to calculate correctly non-stationary fields of vorticity and pressure as well as aerodynamics characteristics of wings at Reynolds numbers lying in the range from 10^2 to 10^3.

The developed technique is used for modeling a vortex flow past symmetrical NACA0008 and NACA0018 profiles at $Re = 500$ and angles of attack from 0° to 60°. The obtained results indicate the domination of viscous effects in the airfoil flow in this case. The rise of α is shown to lead to changing the vortical flow pattern in the wake from stationary to multiperiodic through the Hopf bifurcation and period-doubling bifurcation.

Calculations of the airfoil dynamic characteristics corresponding to these regimes revealed that the highest lift-drag ratio is achieved with the stationary flow, at $\alpha < 15°$. When an angle of attack increases, the aerodynamic performance drops to values less than one, which is due to a significant increment in the drag.

As follows from the pressure fields, the lift generation is mainly conditioned by the dynamics of the separation bubble on the upper surface of profile, as in the case of large Reynolds numbers.

A comparison of dynamic characteristics of NACA0008 and NACA0018 profiles revealed that the thinner profile has better performance but the flow around the thick one is more regular. The effect of the profile thickness on the force coefficients is significant only in the stationary flow regime.

In general, the results obtained in this work show that operation of wing systems at low Reynolds numbers is significantly different from traditional aerodynamic regimes due to the domination of viscous effects in the flow.

References

1. Ifju, P., Stanford, B., Sytsma M., Albertani, R.: Analysis of a flexible wing micro air vehicle. AIAA Paper 3311 (2006)
2. Dickinson, M.H., Geotz, K.G.: Unsteady aerodynamic performance of model wings at low Reynolds numbers. J. Exp. Biol. **174**, 45–64 (1993)
3. Wang, Z.J.: Vortex shedding and frequency selection in flapping flight. J. Fluid Mech. **410**, 323–341 (2000)
4. Taira, K., Colonius, T.: Three-dimensional flows around low-aspect-ratio flat-plate wings at low Reynolds numbers. J. Fluid Mech. **623**, 187–207 (2009)
5. Zhang, J., Liu, N.-S., Lu, X.-Y.: Route to a chaotic state in fluid flow past an inclined flat plate. J. Phys. Rev. **79**, 045306: 1–4 (2009)
6. Yang, D., Pettersen, B., Andersson, H.I., Narasimhamurthy, V.D.: Numerical simulation of flow past a rectangular flat plate. In: Proceeds of V European Conference on Computer Fluid Dynamics, Lisbon, 14–17 June (2010)
7. Yu, D., Mei, R., Shyy, W. A multi-block lattice Boltzmann method for viscous fluid flows. Int. J. Numer. Methods Fluids **39**, 99–120 (2002)
8. Lockard, D.P., Luo, L.-S., Milder, S.D., Singer, B.A. Evaluation of power FLOW for aerodynamics applications. J. Stat. Phys. **107**, 423–478 (2002)
9. Kunz, P.J., Kroo, I.: Analysis, design and testing of airfoils for use at ultra-low Reynolds numbers. In: Mueller, T.J. (ed.) Univ. of Notr-Dam, Proceeds of the Conference on Fixed, Flapping and Rotary Vehicles at Very Low Reynolds Numbers, pp. 349–372 (2000)
10. Mueller, T.J., Batillt, S.M.: Experimental studies of separation on a two-dimensional airfoil at low Reynolds numbers. AIAA J. **20**(4), 457–463 (1982)
11. Sunada, S., Sakaguchi, A., Kawachi, K.: Airfoil section characteristics at a low Reynolds number. J. Fluids Eng. **119**, 129–135 (1997)
12. Sunada S., Kawachi, K.: Comparison of wing characteristics at an ultralow Reynolds number. J. Aircraft **39**, 331–338 (2002)
13. Cottet, G.-H., Koumoutsakos, P. Vortex Methods: Theory and Practice. Cambridge University Press, London (2000)
14. Vortex methods. In: Kamemoto, Tsutahara (eds.) Proceeds of the 1-st International Conference of Vortex Motions. World Scientific, Singapore (2000)
15. Chornyy, G.G.: Gas Dynamics. Nauka, Moscow (1988) (in Russian)
16. Shiels, D.: Simulation of controlled bluff body flow with a viscous vortex method, PhD thesis. California Institute of Technology (1998)
17. Koumoutsakos, P., Leonard, A., Pepin, F.: Boundary conditions for viscous vortex methods. J. Comput. Phys. **113**, 52–61 (1994)
18. Gorban, V., Gorban, I.: Vortical flow structure near a square prism: numerical model and algorithms of control. J Appl. Hydromech. **7**, 8–26 (2005) (in Ukrainian)
19. Gorban, I.M., Khomenko, O.V.: Flow control near a square prism with the help of frontal flat plates. In: Zgurovsky, M.Z., Sadovnichiy, V.A. (eds.) Studies in Systems, Decision and Control. Advances in Dynamical Systems and Control, vol. 69, pp. 327–350. Springer, Berlin (2016)
20. Kempka, S.N., Glass, M.W., Peery, J.S., Strickland, J.H.: Accuracy consideration for implementing velocity boundary conditions in vorticity formulations. SANDIA Reports N. SAND. 96-0583, UC-700 (1996)

21. Ploumhans, P., Winckelmans, G.S.: Vortex methods for high-resolution simulations of viscous flow past bluff bodies of general geometry. J. Comput. Phys. **165**, 354–406 (2000)
22. Belotserkovsky, S.M., Kotovsky, V.N., Nisht, M.I., Fedorov, R.M. Mathematical simulation of two-dimensional parallel separation flows near bodies. Nauka, Moscow (1988) (in Russian)
23. Karniadakis, G.E., Triantafyllou, G.S. Three-dimensional dynamics and transition to turbulence in the wake of bluff objects. J. Fluid Mech. **238**, 1–30 (1992)

Chapter 11
Strong Solutions of the Thin Film Equation in Spherical Geometry

Roman M. Taranets

Abstract We study existence and long-time behaviour of strong solutions for the thin film equation using a priori estimates in a weighted Sobolev space. This equation can be classified as a doubly degenerate fourth-order parabolic and it models coating flow on the outer surface of a sphere. It is shown that the strong solution asymptotically decays to the flat profile.

11.1 Introduction

In this paper, we study the following doubly degenerate fourth-order parabolic equation

$$u_t + \Big((1-x^2)|u|^n((1-x^2)u_x)_{xx}\Big)_x = 0 \text{ in } Q_T, \tag{11.1}$$

where $Q_T = \Omega \times (0,T)$, $n > 0$, $T > 0$, and $\Omega = (-1,1)$. This equation describes the dynamics of a thin viscous liquid film on the outer surface of a solid sphere. More general dynamics of the liquid film for the cases when the draining of the film due to gravity were balanced by centrifugal forces arising from the rotation of the sphere about a vertical axis and by capillary forces due to surface tension was considered in [12]. In addition, Marangoni effects due to temperature gradients were taken into account in [13]. The spherical model without the surface tension and Marangoni effects was studied in [15, 17].

In [12], the authors derived the following equation for no-slip regime in dimensionless form

$$h_t + \tfrac{1}{\sin\theta}(h^3 \sin\theta\, J)_\theta = 0,$$
$$J := a\sin\theta + b\sin\theta\cos\theta + c[2h + \tfrac{1}{\sin\theta}(\sin\theta\, h_\theta)_\theta]_\theta,$$

R. M. Taranets (✉)
Institute of Applied Mathematics and Mechanics of the NASU, Sloviansk, Ukraine

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_11

where $h(\theta, t)$ represent the thickness of the thin film, $\theta \in (0, \pi)$ is the polar angle in spherical coordinates, with t denoting time; the dimensionless parameters a, b and c describe the effects of gravity, rotation and surface tension, respectively. After the change of variable $x = -\cos\theta$, this equation can be written in the form:

$$h_t + [h^3(1-x^2)(a - bx + c(2h + ((1-x^2)h_x)_x)_x)]_x = 0, \tag{11.2}$$

where $x \in (-1, 1)$. As a result, Eq. (11.1) for $n = 3$ is a particular case of (11.2) for no-slip regime. On the other hand, (11.1) for $n < 3$ generalises (11.2) with $a = b = 0$ for different slip regimes, for example, like weak or partial wetting.

In contrast to the classical thin film equation:

$$u_t + \left(|u|^n u_{xxx}\right)_x = 0, \tag{11.3}$$

which describes the behaviour of a thin viscous film on a flat surface under the effect of surface tension, Eq. (11.1) is not yet well analysed. To the best of our knowledge, there is only one analytical result [14], where the authors proved existence of weak solutions in a weighted Sobolev space. In 1990, Bernis and Friedman [3] constructed non-negative weak solutions of the equation (11.3) when $n \geqslant 1$, and it was also shown that for $n \geqslant 4$, with a positive initial condition, there exists a unique positive classical solution. In 1994, Bertozzi et al. [4] generalised this positivity property for the case $n \geqslant \frac{7}{2}$. In 1995, Beretta et al. [2] proved the existence of non-negative weak solutions for the equation (11.3) if $n > 0$, and the existence of strong ones for $0 < n < 3$. Also, they could show that this positivity-preserving property holds for almost every time t in the case $n \geqslant 2$. A similar result on a cylindrical surface was obtained in [9, 10]. Regarding the long-time behaviour, Carrillo and Toscani [8] proved the convergence to a self-similar solution for equation (11.3) with $n = 1$ and Carlen and Ulusoy [7] gave an upper bound on the distance from the self-similar solution. A similar result on a cylindrical surface was obtained in [1, 5].

In the present article, we obtain the existence of weak solutions in a wider weighted classes of functions than it was done in [14]. Moreover, we show the existence of non-negative strong solutions and we also prove that this solution decays asymptotically to the flat profile. Note that (11.1) loses its parabolicity not only at $u = 0$ (as in (11.3)) but also at $x = \pm 1$. For this reason, it is natural to seek solution in a Sobolev space with weight $1 - x^2$. For example, it is the well-known that the non-negative steady state of Eq. (11.3) for $x \in (-1, 1)$ has the form

$$u_s(x) = c_1(1 - x^2) + c_2, \text{ where } c_i \geqslant 0.$$

On the other hand, the Eq. (11.1) has the following non-negative steady state

$$u_s(x) = (c_1 + c_2)\ln(1 + x) + (c_1 - c_2)\ln(1 - x) + c_3,$$

where $0 \leqslant |c_2| \leqslant -c_1$, $c_3 \geqslant -(c_1 + c_2)\ln(1 + \frac{c_2}{c_1}) + (c_1 - c_2)\ln(1 - \frac{c_2}{c_1})$, hence $u_s(x) \to +\infty$ as $x \to \pm 1$.

11.2 Existence of Strong Solutions

We study the following thin film equation

$$u_t + \left((1-x^2)|u|^n \left((1-x^2)u_x\right)_{xx}\right)_x = 0 \text{ in } Q_T \tag{11.4}$$

with the no-flux boundary conditions

$$(1-x^2)u_x = (1-x^2)\left((1-x^2)u_x\right)_{xx} = 0 \text{ at } x = \pm 1,\ t > 0, \tag{11.5}$$

and the initial condition

$$u(x,0) = u_0(x) \geqslant 0. \tag{11.6}$$

Here $n > 0$, $Q_T = \Omega \times (0,T)$, $\Omega := (-1,1)$, and $T > 0$. Integrating the Eq. (11.4) by using boundary conditions (11.5), we obtain the mass conservation property

$$\int_\Omega u(x,t)dx = \int_\Omega u_0(x)dx =: M > 0. \tag{11.7}$$

Consider initial data $u_0(x) \geqslant 0$ for all $x \in \bar{\Omega}$ satisfying

$$\int_\Omega \{u_0^2(x) + (1-x^2)u_{0,x}^2(x)\}dx < \infty. \tag{11.8}$$

Definition 11.1 (Weak Solution) Let $n > 0$. A function u is a weak solution of the problem (11.4)–(11.6) with initial data u_0 satisfying (11.8) if $u(x,t)$ has the following properties

$$(1-x^2)^{\frac{\beta}{2}} u \in C_{x,t}^{\frac{\alpha}{2},\frac{\alpha}{8}}(\bar{Q}_T),\ 0 < \alpha < \beta \leqslant \tfrac{2}{n},$$

$$u_t \in L^2(0,T;(H^1(\Omega))^*),\ (1-x^2)^{\frac{1}{2}} u_x \in L^\infty(0,T;L^2(\Omega)),$$

$$(1-x^2)^{\frac{1}{2}}|u|^{\frac{n}{2}}((1-x^2)u_x)_{xx} \in L^2(P),$$

u satisfies (11.4) in the following sense:

$$\int_0^T \langle u_t, \phi\rangle\, dt - \iint_P (1-x^2)|u|^n((1-x^2)u_x)_{xx}\phi_x\, dxdt = 0$$

for all $\phi \in L^2(0,T;H^1(\Omega))$, where $P := \bar{Q}_T \setminus \{\{u=0\} \cup \{t=0\}\}$,

$$(1-x^2)^{\frac{1}{2}} u_x(.,t) \to (1-x^2)^{\frac{1}{2}} u_{0,x}(.) \text{ strongly in } L^2(\Omega) \text{ as } t \to 0,$$

and boundary conditions (11.5) hold at all points of the lateral boundary, where $\{u \neq 0\}$.

Let us denote by

$$0 \leqslant G_0(z) := \begin{cases} \frac{z^{2-n}-A^{2-n}}{(n-1)(n-2)} - \frac{A^{1-n}}{1-n}(z-A) \text{ if } n \neq 1,2, \\ z\ln z - z(\ln A + 1) + A \text{ if } n = 1, \\ \ln(\frac{A}{z}) + \frac{z}{A} - 1 \text{ if } n = 2, \end{cases} \tag{11.9}$$

where $A \geqslant 0$ if $n \in (1,2)$ and $A > 0$ if else. Next, we establish existence of a more regular solution u of (11.4) than a weak solution in the sense of Definition 11.1. Besides, we show that L^1-norm of this strong solution u decays to $\frac{M}{|\Omega|}$.

Theorem 11.1 (Strong Solution) *Assume that $n \geqslant 1$ and initial data u_0 satisfies $\int\limits_\Omega G_0(u_0)\,dx < +\infty$ then the problem (11.4)–(11.6) has a non-negative weak solution, u, in the sense of Definition 11.1, such that*

$$(1-x^2)u_x \in L^2(0,T;H^1(\Omega)),\ (1-x^2)^{\frac{\gamma}{2}} u_x \in L^2(Q_T),\ \gamma \in (0,1],$$

$$u \in L^\infty(0,T;L^2(\Omega)),\ (1-x^2)^{\frac{\mu}{2}} u \in L^2(Q_T),\ \mu \in (-1,\beta].$$

Moreover, if $n \in [1,2)$ then there exist positive constants A, B depending on initial data and n such that

$$\|u - \tfrac{M}{|\Omega|}\|_{L^1(\Omega)} \leqslant \tfrac{A}{1+Bt} \to 0 \text{ as } t \to +\infty.$$

11.3 Proof of Theorem 11.1

11.3.1 Regularised Problems

Equation (11.4) is doubly degenerate when $u = 0$ and $x = \pm 1$. For this reason, for any $\epsilon > 0$ and $\delta > 0$ we consider two-parametric regularised equations

$$u_{\epsilon\delta,t} + \left[(1-x^2+\delta)(|u_{\epsilon\delta}|^n + \epsilon)\left((1-x^2+\delta)u_{\epsilon\delta,x}\right)_{xx}\right]_x = 0 \text{ in } Q_T \tag{11.10}$$

with boundary conditions

$$u_{\epsilon\delta,x} = \left((1-x^2+\delta)u_{\epsilon\delta,x}\right)_{xx} = 0 \text{ at } x = \pm 1, \tag{11.11}$$

and initial data

$$u_{\epsilon\delta}(x,0) = u_{0,\epsilon\delta}(x) \in C^{4+\gamma}(\bar{\Omega}),\ \gamma > 0, \tag{11.12}$$

where

$$u_{0,\epsilon\delta}(x) \geqslant u_{0\delta}(x) + \epsilon^{\theta},\ \ \theta \in (0, \tfrac{1}{2(n-1)}), \tag{11.13}$$

$$u_{0,\epsilon\delta} \to u_{0\delta} \text{ strongly in } H^1(\Omega) \text{ as } \epsilon \to 0, \tag{11.14}$$

$$(1-x^2+\delta)^{\frac{1}{2}} u_{0x,\delta} \to (1-x^2)^{\frac{1}{2}} u_{0,x} \text{ strongly in } L^2(\Omega) \text{ as } \delta \to 0. \tag{11.15}$$

The parameters $\epsilon > 0$ and $\delta > 0$ in (11.10) make the problem regular up to the boundary (i.e. uniformly parabolic). The existence of a local in time solution to (11.10) is guaranteed by the Schauder estimates in [11]. Now suppose that $u_{\epsilon\delta}$ is a solution of equation (11.10) and that it is continuously differentiable with respect to the time variable and fourth order continuously differentiable with respect to the spatial variable.

11.3.2 Existence of Weak Solutions

In order to get an *a priori* estimation for $u_{\epsilon\delta}$, we multiply both sides of Eq. (11.10) by $-[(1-x^2+\delta)u_{\epsilon\delta,x}]_x$ and integrate over Ω by (11.11). This gives us

$$\tfrac{1}{2}\tfrac{d}{dt}\int_{\Omega}(1-x^2+\delta)u_{\epsilon\delta,x}^2\,dx + \int_{\Omega}(1-x^2+\delta)(|u_{\epsilon\delta}|^n+\epsilon)[(1-x^2+\delta)u_{\epsilon\delta,x}]_{xx}^2\,dx = 0. \tag{11.16}$$

Integrating (11.16) in time, we get

$$\tfrac{1}{2}\int_{\Omega}(1-x^2+\delta)u_{\epsilon\delta,x}^2\,dx + \iint_{Q_T}(1-x^2+\delta)(|u_{\epsilon\delta}|^n+\epsilon)[(1-x^2+\delta)u_{\epsilon\delta,x}]_{xx}^2\,dxdt = \tfrac{1}{2}\int_{\Omega}(1-x^2+\delta)u_{0x,\epsilon\delta}^2\,dx. \tag{11.17}$$

By (11.15) we have

$$\int\limits_{\Omega} (1 - x^2 + \delta) u_{\epsilon\delta,x}^2 \, dx \leqslant C_0, \tag{11.18}$$

where $C_0 > 0$ is independent of ϵ and δ. From (11.18) and (11.17) it follows that

$$\{u_{\epsilon\delta}\}_{\epsilon>0} \text{ is uniformly bounded in } L^{\infty}(0, T; H^1(\Omega)), \tag{11.19}$$

$$\{(1 - x^2 + \delta)^{\frac{1}{2}} (|u_{\epsilon\delta}|^n + \epsilon)^{\frac{1}{2}} [(1 - x^2 + \delta) u_{\epsilon\delta,x}]_{xx}\}_{\epsilon,\, \delta>0} \text{ is u. b. in } L^2(Q_T). \tag{11.20}$$

By (11.19) and (11.20), using the same method as [3], we can prove that solutions $u_{\epsilon\delta}$ have uniformly (in ϵ) bounded $C_{x,t}^{1/2,1/8}$-norms. By the Arzelà-Ascoli theorem, this equicontinuous property, together with the uniformly boundedness shows that every sequence $\{u_{\epsilon\delta}\}_{\epsilon>0}$ has a subsequence such that

$$u_{\epsilon\delta} \to u_{\delta} \text{ uniformly in } Q_T \text{ as } \epsilon \to 0. \tag{11.21}$$

As a result, we obtain a solution u_{δ} of the problem (11.10)–(11.12) with $\epsilon = 0$ in the sense of [3, Theorem 3.1, pp. 185–186].

Next, we show that the family of solutions $\{u_{\delta}\}_{\delta>0}$ is uniformly bounded in some weighted space. Using the conservation of mass property

$$\int\limits_{\Omega} u_{\delta}(x, t) dx = M_{\delta} > 0, \tag{11.22}$$

we arrive at

$$|u_{\delta} - \tfrac{M_{\delta}}{|\Omega|}| = \Big| \int\limits_{x_0}^{x} u_x \, dx \Big| \leqslant \Big(\int\limits_{\Omega} (1 - x^2) u_x^2 \, dx \Big)^{\frac{1}{2}} \Big(\int\limits_{x_0}^{x} \tfrac{dx}{1 - x^2} \Big)^{\frac{1}{2}}. \tag{11.23}$$

Multiplying (11.23) by $(1 - x^2)^{\frac{\beta}{2}}$, where $\beta > 0$, by (11.18) we deduce that

$$(1 - x^2)^{\frac{\beta}{2}} |u_{\delta} - \tfrac{M_{\delta}}{|\Omega|}| \leqslant (\tfrac{C_0}{2})^{\frac{1}{2}} \Big((1 - x^2)^{\beta} \ln(\tfrac{(1+x)(1-x_0)}{(1-x)(1+x_0)}) \Big)^{\frac{1}{2}} \leqslant C_1 \tag{11.24}$$

for all $x \in \bar{\Omega}$, where $C_1 > 0$ is independent of $\delta > 0$. From (11.24) we find that

$$\{(1 - x^2)^{\frac{\beta}{2}} u_{\delta}\}_{\delta>0} \text{ is u. b. in } Q_T \text{ for any } \beta > 0. \tag{11.25}$$

In particular, by (11.18) we get

$$(1-x^2)^{\frac{\beta}{2}}|u_\delta(x_1,t)-u_\delta(x_2,t)| \leqslant C_2|x_1-x_2|^{\frac{\alpha}{2}} \quad \forall x_1,x_2 \in \Omega,\ \alpha \in (0,\beta). \tag{11.26}$$

By (11.20), (11.25) and (11.26) with $\beta \in (0, \frac{2}{n}]$, using the same method as [3, Lemma 2.1, p. 183], we can prove similarly that

$$(1-x^2)^{\frac{\beta}{2}}|u_\delta(x,t_1)-u_\delta(x,t_2)| \leqslant C_3|t_1-t_2|^{\frac{\alpha}{8}} \quad \forall t_1,t_2 \in (0,T). \tag{11.27}$$

The inequalities (11.26) and (11.27) show the uniform (in δ) boundedness of a sequence $\{(1-x^2)^{\frac{\beta}{2}}u_\delta\}_{\delta>0}$ in the $C^{\frac{\alpha}{2},\frac{\alpha}{8}}_{x,t}$-norm. By the Arzelà-Ascoli theorem, this a priori bound together with (11.25) shows that as $\delta \to 0$, every sequence $\{(1-x^2)^{\frac{\beta}{2}}u_\delta\}_{\delta>0}$ has a subsequence $\{(1-x^2)^{\frac{\beta}{2}}u_{\delta_k}\}_{\delta_k>0}$ such that

$$(1-x^2)^{\frac{\beta}{2}}u_{\delta_k} \to (1-x^2)^{\frac{\beta}{2}}u \text{ uniformly in } \bar{Q}_T \text{ as } \delta_k \to 0. \tag{11.28}$$

Following the idea of proof [3, Theorem 3.1], we obtain a solution u of the problem (11.10)–(11.12) in the sense of Definition 11.1.

11.3.3 Existence of Strong Solutions

Let us denote by $G_\epsilon(z)$ the following function

$$G_\epsilon(z) \geqslant 0\ \forall z \in \mathbb{R},\ G''_\epsilon(z) = \tfrac{1}{|s|^n+\epsilon}.$$

Now we multiply Eq. (11.10) by $G'_\epsilon(u_{\epsilon\delta})$ and integrate over Ω to get

$$\tfrac{d}{dt}\int_\Omega G_\epsilon(u_{\epsilon\delta}(x,t))\,dx + \int_\Omega [(1-x^2+\delta)u_{\epsilon\delta,x}]^2_x\,dx = 0. \tag{11.29}$$

After integration in time, Eq. (11.29) becomes

$$\int_\Omega G_\epsilon(u_{\epsilon\delta}(x,T))\,dx + \iint_{Q_T}[(1-x^2+\delta)u_{\epsilon\delta,x}]^2_x\,dxdt = \int_\Omega G_\epsilon(u_{0,\epsilon\delta}(x))\,dx. \tag{11.30}$$

We compute

$$G''_0(z) - G''_\epsilon(z) = \tfrac{\epsilon}{|z|^n(|z|^n+\epsilon)},$$

and consequently

$$G_0(z) - G_\epsilon(z) = \epsilon \int_A^z \int_A^v \frac{ds\,dv}{|s|^n(|s|^n+\epsilon)},$$

where A is some positive constant. As $u_{0,\epsilon\delta}(x)$ is bounded then by (11.13) it follows that

$$|G_0(u_{0,\epsilon\delta}(x)) - G_\epsilon(u_{0,\epsilon\delta}(x))| \leqslant C\,\epsilon^{1-2\theta(n-1)} \to 0 \text{ as } \epsilon \to 0,$$

and therefore, due to (11.14), we have

$$\int_\Omega G_\epsilon(u_{0,\epsilon}(x))\,dx \to \int_\Omega G_0(u_{0\delta}(x))\,dx \text{ as } \epsilon \to 0. \tag{11.31}$$

As a result, by (11.30), (11.31) we deduce that

$$\int_\Omega G_\epsilon(u_{\epsilon\delta}(x,T))\,dx \leqslant C_4, \tag{11.32}$$

$$\{(1-x^2+\delta)u_{\epsilon\delta,x}\}_{\epsilon,\,\delta>0} \text{ is u. b. in } L^2(0,T;H^1(\Omega)), \tag{11.33}$$

where $C_4 > 0$ is independent of ϵ and δ. Similar to [3, Theorem 4.1, p. 190], using (11.19) and (11.32), we can show that the limit solution u_δ is non-negative if $n \in [1,4)$ and strictly positive if $n \geqslant 4$. Next, letting $\delta \to 0$, we get a non-negative strong solution.

11.3.4 Asymptotic Behaviour

First of all, note that the energy functional $\mathscr{E}_0(u(t)) := \frac{1}{2}\int_\Omega (1-x^2)u_x^2\,dx$ is decaying (by (11.16) with $\epsilon = \delta = 0$), bounded from below and lower semi-continuous it must have a minimizer, $u_{min}(x)$, which is continuous on Ω. Taking into account the mass conservation, we find that $u_{\min}(x) = \frac{M}{|\Omega|}$ and

$$\mathscr{E}_0(u(t)) \to 0 \text{ as } t \to +\infty.$$

Next, we will use the mass conservation property (11.22) and the following interpolation inequality.

Lemma 11.1 ([6]) *Let p, q, r, α, β, γ, σ and θ be real numbers satisfying $p,\ q \geqslant 1$, $r > 0$, $0 \leqslant \theta \leqslant 1$, $\gamma = \theta\sigma + (1-\theta)\beta$, $\frac{1}{p} + \frac{\alpha}{n} > 0$, $\frac{1}{q} + \frac{\beta}{n} > 0$ and $\frac{1}{r} + \frac{\gamma}{n} > 0$.*

There exists a positive constant C such that the following inequality holds for all $v \in C_0^\infty(\mathbb{R}^n)$ $(n \geqslant 1)$

$$\||x|^\gamma v\|_{L^r} \leqslant C\||x|^\alpha |Dv|\|_{L^p}^\theta \||x|^\beta v\|_{L^q}^{1-\theta}$$

if and only if

$$\tfrac{1}{r} + \tfrac{\gamma}{n} = \theta\left(\tfrac{1}{p} + \tfrac{\alpha-1}{n}\right) + (1-\theta)\left(\tfrac{1}{q} + \tfrac{\beta}{n}\right)$$

and

$$\begin{cases} 0 \leqslant \alpha - \sigma \ \ \textit{if} \ \ a > 0 \\ \alpha - \sigma \leqslant 1 \ \ \textit{if} \ \ a > 0 \ \ \textit{and} \ \ \tfrac{1}{p} + \tfrac{\alpha-1}{n} = \tfrac{1}{r} + \tfrac{\gamma}{n}. \end{cases}$$

Applying Lemma 11.1 to $v = u_\delta - \frac{M_\delta}{|\Omega|}$ with $\Omega = (-1, 1)$, $\gamma = \beta = 0$, $\alpha = \frac{1}{2}$, $r = p = 2$, $q = 1$, $n = 1$, and $\theta = \frac{1}{2}$, we have

$$\|u_\delta - \tfrac{M_\delta}{|\Omega|}\|_2 \leqslant C_N \|(1-x^2)^{\frac{1}{2}} u_{\delta,x}\|_2^\theta \|u_\delta - \tfrac{M_\delta}{|\Omega|}\|_1^{1-\theta},$$

whence for $u_\delta \geqslant 0$ we deduce that

$$\int_\Omega (u_\delta - \tfrac{M_\delta}{|\Omega|})^2\, dx \leqslant 2M_\delta C_N^2 \Big(\int_\Omega (1-x^2) u_{\delta,x}^2\, dx\Big)^{\frac{1}{2}}. \tag{11.34}$$

Next, we will use the following Hardy's inequality

$$\int_{-1}^{1} (1-x^2)^{-1} v^2(x)\, dx \leqslant C_H \int_{-1}^{1} v_x^2(x)\, dx \tag{11.35}$$

for all $v \in H^1(-1, 1)$ such that $v(\pm 1) = 0$, where $C_H = 4K \approx 7.028$. Really, using integration by parts and Cauchy inequality, we have

$$\int_{-1}^{1} (1-x^2)^{-1} v^2(x)\, dx = v^2(x) \ln(\tfrac{1+x}{1-x})\Big|_{-1}^{1} - 2\int_{-1}^{1} v(x) v_x(x) \ln(\tfrac{1+x}{1-x})\, dx \leqslant$$

$$2\Big(\int_{-1}^{1} (1-x^2)^{-1} v^2(x)\, dx\Big)^{\frac{1}{2}} \Big(\int_{-1}^{1} (1-x^2)(\ln(\tfrac{1+x}{1-x}))^2 v_x^2(x)\, dx\Big)^{\frac{1}{2}}.$$

As $v(x) \in C^{\frac{1}{2}}[-1,1]$ then $v^2(x)\ln(\frac{1+x}{1-x}) = 0$ at $x = \pm 1$. From here we find that

$$\int\limits_{-1}^{1} (1-x^2)^{-1} v^2(x)\,dx \leqslant 2\Big(\int\limits_{-1}^{1} (1-x^2)^{-1} v^2(x)\,dx\Big)^{\frac{1}{2}} \Big(K\int\limits_{-1}^{1} v_x^2(x)\,dx\Big)^{\frac{1}{2}},$$

where

$$K := \max_{[-1,1]} (1-x^2)(\ln(\tfrac{1+x}{1-x}))^2 \approx 1.757,$$

whence it follows (11.35).

Applying (11.35) to $v = (1-x^2)u_x$, we obtain that

$$\int\limits_{\Omega} (1-x^2)u_x^2\,dx \leqslant C_H \int\limits_{\Omega} [(1-x^2)u_x]_x^2\,dx,$$

whence, due to (11.34), we find that

$$\tfrac{1}{4C_H M_\delta^2 C_N^4}\Big(\int\limits_{\Omega} (u_\delta - \tfrac{M_\delta}{|\Omega|})^2\,dx\Big)^2 \leqslant \int\limits_{\Omega} [(1-x^2)u_x]_x^2\,dx. \tag{11.36}$$

Assume that $n \in [1,2)$. Taking $A = \frac{M_\delta}{|\Omega|}$ in the definition of $G_0(z)$, we have

$$0 \leqslant G_0(z) \leqslant C_n(z - \tfrac{M_\delta}{|\Omega|})^2 \text{ for all } z \geqslant 0,$$

where C_n depends on n only. As a result, by (11.36)

$$\tfrac{1}{4C_H M_\delta^2 C_N^4 C_n^2}\Big(\int\limits_{\Omega} G_0(u_\delta)\,dx\Big)^2 \leqslant \int\limits_{\Omega} [(1-x^2)u_x]_x^2\,dx. \tag{11.37}$$

Taking $\delta \to 0$ in (11.30), due to (11.37), we arrive at

$$\int\limits_{\Omega} G_0(u)\,dx + B_0\int\limits_0^t \Big(\int\limits_{\Omega} G_0(u_\delta)\,dx\Big)^2 ds \leqslant A_0 := \int\limits_{\Omega} G_0(u_0)\,dx, \tag{11.38}$$

where $B_0 := \frac{1}{4C_H M^2 C_N^4 C_n^2}$. From (11.38) by comparing to the solution $y(t)$ of the problem for ODE

$$y'(t) + B_0 y^2(t) = 0, \quad y(0) = A_0,$$

we get

$$0 \leqslant \int\limits_{\Omega} G_0(u)\,dx \leqslant \tfrac{A_0}{1+A_0B_0t} \to 0 \text{ as } t \to +\infty. \tag{11.39}$$

As a result, applying Csiszár-Kullback inequality [16], by (11.39) we obtain

$$\|u - \tfrac{M}{|\Omega|}\|_{L^1(\Omega)} \leqslant \tfrac{A_0}{1+A_0B_0t} \to 0 \text{ as } t \to +\infty$$

provided $n \in [1, 2)$. This proves Theorem 11.1 completely. □

Acknowledgements This paper is supported by Ministry of Education and Science of Ukraine, grant number is 0118U003138.

References

1. Badali, D., Chugunova, M., Pelinovsky, D.E., Pollack, S.: Regularized shock solutions in coating flows with small surface tension. Phys. Fluids **23**(9), 093103-1–093103-8 (2011)
2. Beretta, E., Bertsch, M., Dal Passo, R.: Nonnegative solutions of a fourth-order nonlinear degenerate parabolic equation. Arch. Ration. Mech. Anal. **129**(2), 175–200 (1995)
3. Bernis, F., Friedman, A.: Higher order nonlinear degenerate parabolic equations. J. Differ. Equ. **83**(1), 179–206 (1990)
4. Bertozzi, A. L., et al.: Singularities and Similarities in Interface Flows. Trends and Perspectives in Applied Mathematics, pp. 155–208. Springer, New York (1994)
5. Burchard, A., Chugunova, M., Stephens, B.K.: Convergence to equilibrium for a thin-film equation on a cylindrical surface. Commun. Partial Differ. Equ. **37**(4), 585–609 (2012)
6. Caffarelli, L., Kohn, R., Nirenberg, L.: First order interpolation inequalities with weights. Compos. Math. **53**(3), 259–275 (1984)
7. Carlen, E.A., Ulusoy, S.: Asymptotic equipartition and long time behavior of solutions of a thin-film equation, J. Differ. Equ. **241**(2), 279–292 (2007)
8. Carrillo, J.A., Toscani, G.: Long-time asymptotics for strong solutions of the thin film equation. Commun. Math. Phys. **225**(3), 551–571 (2002)
9. Chugunova M., Taranets, R.M.: Qualitative analysis of coating flows on a rotating horizontal cylinder. Int. J. Differ. Equ. **2012**, Article ID 570283, 30 pp. (2012)
10. Chugunova, M., Pugh, M.C., Taranets, R.M.: Nonnegative solutions for a long-wave unstable thin film equation with convection. SIAM J. Math. Anal. **42**(4), 1826–1853 (2010)
11. Friedman, A.: Interior estimates for parabolic systems of partial differential equations. J. Math. Mech. **7**(3), 393–417 (1958)
12. Kang, D., Nadim, A., Chugunova, M.: Dynamics and equilibria of thin viscous coating films on a rotating sphere, J. Fluid Mech. **791**, 495–518 (2016)
13. Kang, D., Nadim, A., Chugunova, M.: Marangoni effects on a thin liquid film coating a sphere with axial or radial thermal gradients. Phys. Fluids **29**, 072106-1–072106-15 (2017)
14. Kang, D., Sangsawang, T., Zhang, J.: Weak solution of a doubly degenerate parabolic equation (2017). arXiv:1610.06303v2
15. Takagi, D., Huppert, H.E.: Flow and instability of thin films on a cylinder and sphere, J. Fluid Mech. **647**, 221–238 (2010)

16. Unterreiter, A., Arnold, A., Markowich, P., Toscani, G.: On generalized Csiszár-Kullback inequalities. Monatshefte für Mathematik **131**(3), 235–253 (2000)
17. Wilson, S.K.: The onset of steady Marangoni convection in a spherical geometry. J. Eng. Math. **28**, 427–445 (1994)

Part III
Dynamics of Differential and Difference Equations and Applications

Chapter 12
Sequence Spaces with Variable Exponents for Lattice Systems with Nonlinear Diffusion

Xiaoying Han, Peter E. Kloeden, and Jacson Simsen

Abstract Motivated by the study of lattice dynamical systems, i.e., infinite dimensional systems of ordinary differential equations, with nonlinear and state dependent diffusion, a new sequence space with variable exponents is introduced. In particular, given an exponent sequence $\boldsymbol{p} = (p_i)_{i\in\mathbb{Z}}$, a discrete Musielak-Orlicz space of real valued bi-infinite sequences $\ell_{\boldsymbol{p}}$ is defined and equipped with a norm $\|\cdot\|_{\boldsymbol{p}}$ induced by a semi-modular $\rho(\cdot)$. Properties of $\|\cdot\|_{\boldsymbol{p}}$ and $\rho(\cdot)$, as well as properties of the space $(\ell_{\boldsymbol{p}}, \|\cdot\|_{\boldsymbol{p}})$ are discussed in greater detail. While these properties largely facilitate dynamical analysis of a much wider class of lattice systems, this work is a step towards the construction of an integral mathematical framework for the study of lattice models with complicated diffusion structures.

12.1 Introduction

The simplest lattice dynamical system (LDS) has the form

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \nu(u_{i-1} - 2u_i + u_{i+1}) + f(u_i), \quad i \in \mathbb{Z}, \tag{12.1}$$

X. Han
Department of Mathematics and Statistics, Auburn University, Auburn, AL, USA
e-mail: xzh0003@auburn.edu

P. E. Kloeden (✉)
School of Mathematics and Statistics, Huazhong University of Science & Technology, Wuhan, Hubei, China
e-mail: kloeden@math.uni-frankfurt.de

J. Simsen
Instituto de Matemática e Computação, Universidade Federal de Itajubá, Bairro Pinheirinho, Itajubá - MG, Brazil
e-mail: jacson@unifei.edu.br

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_12

where $u_i \in \mathbb{R}$ for each $i \in \mathbb{Z}$ and $f : \mathbb{R} \to \mathbb{R}$ satisfies proper conditions. The LDS (12.1) can be viewed as the spatial discretization of the reaction-diffusion equation $\frac{\partial u}{\partial t} = \nu \Delta u + f(u)$ on a one-dimensional domain, where the Laplacian operator Δ is discretized by using a finite difference quotient to obtain the leading operator

$$(Au)_i := u_{i-1} - 2u_i + u_{i+1}. \tag{12.2}$$

Extensive studies have been done to investigate long term dynamics for variations of the system (12.1) in the forcing term (see, e.g., [1, 2, 8, 16, 17] and references therein), but mostly with the same leading operator A as defined in (12.2). Such systems can be formulated as a deterministic or stochastic ordinary differential equation in the Hilbert space ℓ^2 of real valued square summable bi-infinite sequences with the inner product and the norm

$$(\boldsymbol{u}, \boldsymbol{v}) := \sum_{i\in\mathbb{Z}} u_i v_i, \quad \|\boldsymbol{u}\|_2 := \left(\sum_{i\in\mathbb{Z}} u_i^2\right)^{\frac{1}{2}} \quad \text{for } \boldsymbol{u} = (u_i)_{i\in\mathbb{Z}}, \boldsymbol{v} = (v_i)_{i\in\mathbb{Z}} \in \ell^2.$$

Notice that the operator A defined in (12.2) describes the simplest linear tri-diagonal interconnection structure that allows only linear and uniform diffusion within the nearest neighborhood. This excludes numerous applications where the diffusion does not follow a linear or a uniform structure such as cell dynamics (see, e.g., [4, 15]). To model nonlinear diffusion, a lattice dynamical system with the discretized p-Laplacian operator

$$(\Gamma u)_i := \left|D^+ u_i\right|^{p-2} D^+ u_i - \left|D^- u_i\right|^{p-2} D^- u_i \tag{12.3}$$

where $D^+ u_i := u_{i+1} - u_i$ and $D^- u_i := u_i - u_{i-1}$, was proposed and studied in [7] in the space ℓ^p of p-times summable bi-infinite sequences with norm

$$\|\boldsymbol{u}\|_p := \left(\sum_{i\in\mathbb{Z}} u_i^p\right)^{1/p}, \quad \text{for } \boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} \in \ell^p.$$

To further model nonlinear and state dependent diffusive structures, the $p(x)$-Laplacian operator $\operatorname{div}\left(|\nabla u|^{p(x)-2}\nabla u\right)$ had been used in the continuum context (see, e.g., [9, 10]). Thereby partial differential equations with the $p(x)$-Laplacian on a bounded smooth domain $\Omega \in \mathbb{R}^n$ were studied in the function space $L^{p(\cdot)}$ defined by

$$L^{p(\cdot)}(\Omega) := \left\{\boldsymbol{u} : \Omega \to \mathbb{R} : u \text{ is measurable}, \int_\Omega |\boldsymbol{u}(x)|^{p(x)} dx < \infty\right\}, \tag{12.4}$$

with the exponent function $p(\cdot) \in \mathscr{C}(\bar{\Omega})$ satisfying

$$1 < \min_{x\in\overline{\Omega}} p(x) \le \max_{x\in\overline{\Omega}} p(x).$$

On the contrary, lattice dynamical systems with nonlinear and state dependent diffusion operators have never been studied in the past. A major obstacle is that the most important coercive properties of the diffusion operator to investigate behavior of solutions may not hold in the classical sequence spaces such as ℓ^2 or ℓ^p when the operator is state dependent and nonlinear. New spaces and techniques are required to study the dynamics of such systems.

The goal of this work is to construct a new sequence space with variable exponents which is an analog to the function space $L^{p(\cdot)}$, that allows the exponent p in ℓ^p to vary with respect to the state. In particular, the spatially uniform exponent p is generalized to state dependent exponents described by an infinite sequence $\boldsymbol{p} := (p_i)_{i\in\mathbb{Z}}$. Such spaces show strong relevance to lattice dynamical systems with nonlinear and state dependent diffusion, but have never been studied systematically in the past.

The rest of this paper is organized as follows. In Sect. 12.2 we formulate the new sequence space $\ell_{\boldsymbol{p}}$ and define a semi-modular $\rho(\cdot)$ and a norm $\|\cdot\|_{\boldsymbol{p}}$ on it. In Sect. 12.3 we investigate properties of $\rho(\cdot)$ and $\|\cdot\|_{\boldsymbol{p}}$, which are essential to study dynamics of lattice models in the space $\ell_{\boldsymbol{p}}$. In Sect. 12.4 we show that the space $\ell_{\boldsymbol{p}}$ equipped with the norm $\|\cdot\|_{\boldsymbol{p}}$ is a separable and reflexive Banach space. In addition, a weak compact embedding theorem and a Hölder-like inequality are established. Some closing remarks are provided in Sect. 12.5.

12.2 Formulation of Sequence Spaces with Variable Exponents

The formulation of $\ell_{\boldsymbol{p}}$ arises from the function space $L^{p(\cdot)}$ defined in (12.4). More precisely, the exponent function $p(x)$ will be descritized into a real valued bi-infinite sequences $\boldsymbol{p} = (p_i)_{i\in\mathbb{Z}}$, with $p_i = p(i\,\Delta x)$ and Δx is the spatial scaling. Define

$$p^- := \inf_{i\in\mathbb{Z}} p_i, \qquad p^+ := \sup_{i\in\mathbb{Z}} p_i.$$

It is assumed throughout the paper that

(P0) $\quad 1 < p^- \le p^+ < \infty.$

Given such an exponent sequence $\boldsymbol{p}$, define the discrete Musielak-Orlicz space of real valued bi-infinite sequences $\ell_{\boldsymbol{p}}$ as

$$\ell_{\boldsymbol{p}} := \left\{ \boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} : \sum_{i\in\mathbb{Z}} |u_i|^{p_i} < \infty \right\}. \tag{12.5}$$

Such spaces were first considered by Orlicz [14] in 1931, and later appeared scatteredly in textbooks and a few papers (see, e.g., [12, 13]).

The new $\ell_{\boldsymbol{p}}$ space can be regarded as a discretised counterpart of $L^{p(\cdot)}$, but does not share the same properties. For example, $\ell^2 \subset \ell_{\boldsymbol{p}}$ while $L^{p(\cdot)}(\Omega) \subset L^2(\Omega)$ on any bounded domain Ω when $p^- \geq 2$. In fact, the space $L^{p(\cdot)}$ is defined on a bounded domain Ω but the space $\ell_{\boldsymbol{p}}$ is defined on an infinite lattice, which is essentially an unbounded domain. As a result, the properties of $\ell_{\boldsymbol{p}}$ ought to be studied by techniques different from those used for $L^{p(\cdot)}$.

We first show that the space $\ell_{\boldsymbol{p}}$ is a linear space (see Lemma 12.1 below). For any $\boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{p}}$ define the mapping $\rho : \ell_{\boldsymbol{p}} \to \mathbb{R}^+$ by

$$\rho(\boldsymbol{u}) := \sum_{i\in\mathbb{Z}} |u_i|^{p_i}, \tag{12.6}$$

and the mapping $\|\cdot\|_{\boldsymbol{p}} : \ell_{\boldsymbol{p}} \to \mathbb{R}^+$ by

$$\|\boldsymbol{u}\|_{\boldsymbol{p}} := \inf\left\{\lambda > 0 : \rho\left(\frac{\boldsymbol{u}}{\lambda}\right) \leq 1\right\}. \tag{12.7}$$

We will then show that $\|\cdot\|_{\boldsymbol{p}}$ is a norm on the linear space $\ell_{\boldsymbol{p}}$ (see Lemma 12.2 below).

Lemma 12.1 *The space $\ell_{\boldsymbol{p}}$ defined in (12.5) is a linear space.*

Proof First of all it is obvious that the zero sequence $\mathbf{0}$ is in $\ell_{\boldsymbol{p}}$. We next show that $\ell_{\boldsymbol{p}}$ is closed under component-wise scalar multiplication and addition.

1. (Scalar multiplication) For any scalar $\alpha \in \mathbb{R}$, define

$$\zeta(\alpha) := \begin{cases} p^-, & |\alpha| \leq 1, \\ p^+, & |\alpha| > 1. \end{cases} \tag{12.8}$$

   Then $|\alpha|^{p_i} \leq |\alpha|^{\zeta(\alpha)}$ for all $i \in \mathbb{Z}$. Thus for any $\boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{p}}$,

$$\sum_{i\in\mathbb{Z}} |\alpha u_i|^{p_i} = \sum_{i\in\mathbb{Z}} |\alpha|^{p_i} |u_i|^{p_i} \leq |\alpha|^{\zeta(\alpha)} \sum_{i\in\mathbb{Z}} |u_i|^{p_i} < \infty,$$

   which implies that $\alpha\boldsymbol{u} \in \ell_{\boldsymbol{p}}$.
2. (Addition) First note that for any $p \geq 1$ the function $x \mapsto |x|^p$ is convex on $\mathbb{R}^+$. Then for any $\boldsymbol{u} = (u_i)_{i\in\mathbb{Z}}$, $\boldsymbol{v} = (v_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{p}}$,

$$\begin{aligned} |u_i + v_i|^{p_i} &= \left|\frac{1}{2}(2u_i) + \frac{1}{2}(2v_i)\right|^{p_i} \leq \frac{1}{2}|2u_i|^{p_i} + \frac{1}{2}|2v_i|^{p_i} \\ &\leq 2^{p_i-1}|u_i|^{p_i} + 2^{p_i-1}|v_i|^{p_i}, \end{aligned}$$

which implies that

$$\sum_{i\in\mathbb{Z}} |u_i + v_i|^{p_i} \le 2^{p^+-1}\left(\sum_{i\in\mathbb{Z}} |u_i|^{p_i} + \sum_{i\in\mathbb{Z}} |v_i|^{p_i}\right) < \infty, \tag{12.9}$$

i.e., $\boldsymbol{u}+\boldsymbol{v} \in \ell_{\boldsymbol{p}}$. □

From Part II of the above proof we can deduce that ρ is a convex mapping. For later reference we give an explicit statement and a direct proof below.

Corollary 12.1 *The mapping $\rho : \ell_{\boldsymbol{p}} \to \mathbb{R}^+$ is convex.*

Proof For any $\boldsymbol{u}, \boldsymbol{v} \in \ell_{\boldsymbol{p}}$ and $\theta \in [0,1]$, by Lemma 12.1, $\theta\boldsymbol{u}, (1-\theta)\boldsymbol{v}, \theta\boldsymbol{u}+(1-\theta)\boldsymbol{v} \in \ell_{\boldsymbol{p}}$. The mapping $x \mapsto |x|^p$ is convex for $p \ge 1$, so for each $i \in \mathbb{Z}$

$$|\theta u_i + (1-\theta)v_i|^{p_i} \le \theta|u_i|^{p_i} + (1-\theta)|v_i|^{p_i},$$

and thus

$$\begin{aligned}\rho(\theta\boldsymbol{u}+(1-\theta)\boldsymbol{v}) &= \sum_{i\in\mathbb{Z}} |\theta u_i + (1-\theta)v_i|^{p_i} \\ &\le \theta\sum_{i\in\mathbb{Z}} |u_i|^{p_i} + (1-\theta)\sum_{i\in\mathbb{Z}} |v_i|^{p_i} = \theta\rho(\boldsymbol{u}) + (1-\theta)\rho(\boldsymbol{v}),\end{aligned}$$

i.e., ρ is convex. □

Lemma 12.2 *$\|\cdot\|_{\boldsymbol{p}}$ defined in (12.7) is a norm on the linear space $\ell_{\boldsymbol{p}}$.*

Proof First observe that $\|\boldsymbol{u}\|_{\boldsymbol{p}} < \infty$ for all $\boldsymbol{u} \in \ell_{\boldsymbol{p}}$. In fact, for each given $\boldsymbol{u} \in \ell_{\boldsymbol{p}}$ the set $\{\lambda > 0 : \rho(\boldsymbol{u}/\lambda) \le 1\}$ is bounded below by zero, and hence the completeness property of real numbers guaranties the existence of

$$\inf\left\{\lambda > 0 : \rho\left(\frac{\boldsymbol{u}}{\lambda}\right) \le 1\right\} \in \mathbb{R}.$$

We next show the positive definiteness, scalar multiplication and triangular inequality for $\|\cdot\|_{\boldsymbol{p}}$.

1. (Positive definiteness) If $\boldsymbol{u} = \boldsymbol{0}$, then $u_i = 0$ for each $i \in \mathbb{Z}$, and thus $u_i/\lambda = 0$ for each $i \in \mathbb{Z}$. As a result $\rho(\boldsymbol{u}/\lambda) = 0$ for each $\lambda > 0$, from which it follows that $\|\boldsymbol{0}\|_{\boldsymbol{p}} = 0$.

   On the other hand, if $\|\boldsymbol{u}\|_{\boldsymbol{p}} = 0$, then there is a positive sequence $\lambda_n \to 0^+$ such that $\rho(\boldsymbol{u}/\lambda_n) \le 1$. Suppose (for contradiction) that $\boldsymbol{u} \ne \boldsymbol{0}$, then there exists an $\iota \in \mathbb{Z}$ such that $u_\iota \ne 0$. It then follows that $|u_\iota|^{p_\iota}/\lambda_n \le \rho(\boldsymbol{u}/\lambda_n) \le 1$, and hence $0 < |u_\iota|^{p_\iota} \le \lambda_n \to 0^+$ as $n \to \infty$, which contradicts with $u_\iota \ne 0$.

2. (Scalar multiplication) For any $\boldsymbol{u} \in \ell_{\boldsymbol{p}}$ and $\alpha \in \mathbb{R}$, $\alpha\boldsymbol{u} \in \ell_{\boldsymbol{p}}$ and in addition

$$\|\alpha\boldsymbol{u}\|_{\boldsymbol{p}} = \inf\left\{\lambda > 0 : \rho\left(\frac{\alpha\boldsymbol{u}}{\lambda}\right) \leq 1\right\} = \inf\left\{\lambda = \mu|\alpha| > 0 : \rho\left(\frac{\boldsymbol{u}}{\mu}\right) \leq 1\right\}$$
$$= |\alpha| \inf\left\{\mu > 0 : \rho\left(\frac{\boldsymbol{u}}{\mu}\right) \leq 1\right\} = |\alpha| \cdot \|\boldsymbol{u}\|_{\boldsymbol{p}}\,.$$

3. (Triangular inequality) For any $\boldsymbol{u}, \boldsymbol{v} \in \ell_{\boldsymbol{p}}$, $\boldsymbol{u} + \boldsymbol{v} \in \ell_{\boldsymbol{p}}$. Given an arbitrary $\varepsilon > 0$, by properties of infimum there exist $\lambda_u, \lambda_v \in \{\lambda > 0 : \rho(\boldsymbol{u}/\lambda) \leq 1\}$ such that

$$\lambda_u < \|\boldsymbol{u}\|_{\boldsymbol{p}} + \varepsilon \quad \text{and} \quad \lambda_v < \|\boldsymbol{v}\|_{\boldsymbol{p}} + \varepsilon.$$

with $\rho\left(\frac{\boldsymbol{u}}{\lambda_u}\right) \leq 1$ and $\rho\left(\frac{\boldsymbol{v}}{\lambda_v}\right) \leq 1$. Let $\theta := \frac{\lambda_u}{\lambda_u+\lambda_v}$, then by the convexity of ρ we have

$$\rho\left(\frac{\boldsymbol{u}+\boldsymbol{v}}{\lambda_u+\lambda_v}\right) \leq \theta\rho\left(\frac{\boldsymbol{u}}{\lambda_u}\right) + (1-\theta)\rho\left(\frac{\boldsymbol{v}}{\lambda_v}\right) \leq 1.$$

It then follows from the definition of $\|\cdot\|_{\boldsymbol{p}}$ that

$$\|\boldsymbol{u}+\boldsymbol{v}\|_{\boldsymbol{p}} \leq \lambda_u + \lambda_v < \|\boldsymbol{u}\|_{\boldsymbol{p}} + \|\boldsymbol{v}\|_{\boldsymbol{p}} + 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, then

$$\|\boldsymbol{u}+\boldsymbol{v}\|_{\boldsymbol{p}} \leq \|\boldsymbol{u}\|_{\boldsymbol{p}} + \|\boldsymbol{v}\|_{\boldsymbol{p}}.$$

□

Now define the *sequence space with variable exponents* by

$$\mathscr{P} := (\ell_{\boldsymbol{p}}, \|\cdot\|_{\boldsymbol{p}}).$$

While properties for L^p and $L^{p(\cdot)}$ spaces have been well documented (see, e.g., [5, 6, 11]), properties of the space $\mathscr{P}$ have only appeared scatteredly in different sources. One of main aims of this work is to establish properties of $\mathscr{P}$ that do not exist in the literature. They are closely related to the properties of the semi-modular ρ and norm $\|\cdot\|_{\boldsymbol{p}}$, which are constructed in the next Section.

12.3 Properties of ρ and $\|\cdot\|_{\boldsymbol{p}}$

In this section we discuss important properties of the semi-modular ρ and the norm $\|\cdot\|_{\boldsymbol{p}}$, as well as relations between them.

Lemma 12.3 *For every $\boldsymbol{u} \in \ell_{\boldsymbol{p}}$ and $\alpha \in \mathbb{R}$, the mapping ρ defined in (12.6) satisfies*

(i) $\rho(\alpha \boldsymbol{u}) \le |\alpha| \rho(\boldsymbol{u})$ *if* $|\alpha| \le 1$ *and* $\rho(\alpha \boldsymbol{u}) \ge |\alpha| \rho(\boldsymbol{u})$ *if* $|\alpha| \ge 1$*;*
(ii) $\rho(\boldsymbol{u}) \le \alpha \rho(\boldsymbol{u}) \le \alpha^{p^-} \rho(\boldsymbol{u}) \le \rho(\alpha \boldsymbol{u}) \le \alpha^{p^+} \rho(\boldsymbol{u})$ *if* $\alpha \ge 1$, *and* $\rho(\boldsymbol{u}) \ge \alpha \rho(\boldsymbol{u}) \ge \alpha^{p^-} \rho(\boldsymbol{u}) \ge \rho(\alpha \boldsymbol{u}) \ge \alpha^{p^+} \rho(\boldsymbol{u})$ *if* $0 < \alpha < 1$*;*
(iii) the mapping $\alpha \mapsto \rho(\alpha \boldsymbol{u})$ is continuous;
(iv) the mapping $\alpha \mapsto \rho(\alpha \boldsymbol{u})$ increasing for $\alpha > 0$ and decreasing for $\alpha < 0$.

Proof

(i) For $|\alpha| \le 1$, $|\alpha|^{p_i} \le |\alpha|^{p^-} \le |\alpha|$, and thus

$$\rho(\alpha \boldsymbol{u}) = \sum_{i \in \mathbb{Z}} |\alpha|^{p_i} |u_i|^{p_i} \le \sum_{i \in \mathbb{Z}} |\alpha|^{p^-} |u_i|^{p_i} \le |\alpha| \sum_{i \in \mathbb{Z}} |u_i|^{p_i} = |\alpha| \rho(\boldsymbol{u}).$$

For $|\alpha| \ge 1$, by convexity of ρ,

$$\rho(\boldsymbol{u}) = \rho\left(\frac{1}{|\alpha|} |\alpha| \boldsymbol{u} + \left(1 - \frac{1}{|\alpha|}\right) \boldsymbol{0} \right) \le \frac{1}{|\alpha|} \rho(|\alpha| \boldsymbol{u}).$$

It then follows immediately that $\rho(\alpha \boldsymbol{u}) = \rho(|\alpha| \boldsymbol{u}) \ge |\alpha| \rho(\boldsymbol{u})$.

(ii) If $\alpha \ge 1$,

$$|u_i|^{p_i} \le \alpha |u_i|^{pi} \le \alpha^{p^-} |u_i|^{p_i} \le |\alpha u_i|^{p_i} \le \alpha^{p^+} |u_i|^{p_i}$$

and if $0 < \alpha < 1$

$$|u_i|^{p_i} \ge \alpha |u_i|^{pi} \ge \alpha^{p^-} |u_i|^{p_i} \ge |\alpha u_i|^{p_i} \ge \alpha^{p^+} |u_i|^{p_i}.$$

Summing all the inequalities for $i \in \mathbb{Z}$ gives the result.

(iii) For any $\alpha_0 \in \mathbb{R}$, $\boldsymbol{u} \in \ell_{\boldsymbol{p}}$, due to the continuity of $\alpha \mapsto |\alpha|^{p_i}$ for every $\varepsilon > 0$ there exists $\delta_i > 0$ such that

$$\left| |\alpha|^{p_i} - |\alpha_0|^{p_i} \right| < \frac{\varepsilon}{\rho(\boldsymbol{u})}, \quad \text{for all} \quad |\alpha - \alpha_0| < \delta_i.$$

Let $\delta := \inf_{i \in \mathbb{Z}} \delta_i$, then $\delta > 0$ and for every $|\alpha - \alpha_0| < \delta$ we have

$$|\rho(\alpha \boldsymbol{u}) - \rho(\alpha_0 \boldsymbol{u})| \le \sum_{i \in \mathbb{Z}} \left| |\alpha|^{p_i} - |\alpha_0|^{p_i} \right| \cdot |u_i|^{p_i} < \frac{\varepsilon}{\rho(\boldsymbol{u})} \sum_{i \in \mathbb{Z}} |u_i|^{p_i} = \varepsilon,$$

which implies the continuity of the mapping $\alpha \mapsto \rho(\alpha \boldsymbol{u})$.

(iv) For any $0 < \alpha_1 < \alpha_2$, $0 < \alpha_1 / \alpha_2 < 1$ and thus

$$\rho(\alpha_1 \boldsymbol{u}) = \rho\left(\frac{\alpha_1}{\alpha_2} \alpha_2 \boldsymbol{u} \right) \le \frac{\alpha_1}{\alpha_2} \rho(\alpha_2 \boldsymbol{u}) < \rho(\alpha_2 \boldsymbol{u}).$$

Similarly for $\alpha_1 < \alpha_2 < 0$, $\alpha_1/\alpha_2 > 1$ and thus

$$\rho(\alpha_1 \boldsymbol{u}) = \rho\left(\frac{\alpha_1}{\alpha_2}\alpha_2 \boldsymbol{u}\right) \geq \frac{\alpha_1}{\alpha_2}\rho(\alpha_2 \boldsymbol{u}) > \rho(\alpha_2 \boldsymbol{u}).$$

□

Theorem 12.1 (Unit Ball Theorem) *For every* $\boldsymbol{u} \in \ell_p$,

(i) $\rho(\boldsymbol{u}) < 1$ *if and only if* $\|\boldsymbol{u}\|_p < 1$;
(ii) $\rho(\boldsymbol{u}) \leq 1$ *if and only if* $\|\boldsymbol{u}\|_p \leq 1$;
(iii) $\rho(\boldsymbol{u}) \geq 1$ *if and only if* $\|\boldsymbol{u}\|_p \geq 1$;
(iv) $\|\boldsymbol{u}\|_p = 1$ *if and only if* $\rho(\boldsymbol{u}) = 1$.

Proof

(i) On the one hand, for $\|\boldsymbol{u}\|_p < 1$, there exists a $\alpha > 1$ with $\rho(\alpha \boldsymbol{u}) \leq 1$. Then by Lemma 12.3-(i),

$$\rho(\boldsymbol{u}) \leq \frac{1}{\alpha}\rho(\alpha \boldsymbol{u}) \leq \frac{1}{\alpha} < 1.$$

On the other hand, for $\rho(\boldsymbol{u}) < 1$, by Lemma 12.3-(ii) there exists $\alpha > 1$ such that $\rho(\alpha \boldsymbol{u}) < 1$. Hence $\|\alpha \boldsymbol{u}\|_p < 1$ and $\|\boldsymbol{u}\|_p \leq \frac{1}{\alpha} < 1$.

(ii) On the one hand if $\rho(\boldsymbol{u}) \leq 1$, then $\|\boldsymbol{u}\|_p \leq 1$ by the definition of the norm. On the other hand, if $\|\boldsymbol{u}\|_p \leq 1$, then for each $\alpha > 1$ there exists $\lambda_\alpha \in \{\lambda > 0 : \rho(\boldsymbol{u}/\lambda) \leq 1\}$ such that $\lambda_\alpha < \alpha$. From Lemma 12.3 -(iii) we know that the function $\rho(\alpha \boldsymbol{u})$ is increasing in α, and hence

$$\rho(\boldsymbol{u}/\alpha) \leq \rho(\boldsymbol{u}/\lambda_\alpha) \leq 1.$$

Letting $\alpha \to 1+$ gives immediately $\rho(\boldsymbol{u}) \leq 1$.

(iii) The assertion follows immediately from (i).

(iv) The assertion follows immediately from parts (i) and (ii). □

Lemma 12.4 *For every* $\boldsymbol{u} \in \ell_p$,

$$\rho(\boldsymbol{u}) \leq \|\boldsymbol{u}\|_p \ \text{ if } \ \|\boldsymbol{u}\|_p \leq 1 \quad \text{and} \quad \rho(\boldsymbol{u}) \geq \|\boldsymbol{u}\|_p \ \text{ if } \ \|\boldsymbol{u}\|_p \geq 1.$$

Proof The argument is obvious if $\boldsymbol{u} = \boldsymbol{0}$. First assume that $\|\boldsymbol{u}\|_p \in (0, 1]$. Then by Lemma 12.3-(i),

$$\frac{\rho(\boldsymbol{u})}{\|\boldsymbol{u}\|_p} \leq \rho\left(\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|_p}\right).$$

On the other hand, by Theorem 12.1-(iv), $\rho(\boldsymbol{u}/\|\boldsymbol{u}\|_p) = 1$ because $\|\boldsymbol{u}/\|\boldsymbol{u}\|_p\|_p = 1$. It then follows immediately that $\rho(\boldsymbol{u})/\|\boldsymbol{u}\|_p \leq 1$, i.e., $\rho(\boldsymbol{u}) \leq \|\boldsymbol{u}\|_p$.

The proof of the case for $\|\boldsymbol{u}\|_{\boldsymbol{p}} \geq 1$ is analogous. □

Corollary 12.2 *Let* $\Lambda := \{\boldsymbol{u} \in \ell_{\boldsymbol{p}} : \rho(\boldsymbol{u}) \leq 1\}$. *Then* $\Lambda = \overline{\mathscr{B}(0,1)}$, *the unit closed ball centered at zero, with respect to the norm* $\|\cdot\|_{\boldsymbol{p}}$.

Proof For any $\boldsymbol{u} \in \Lambda$, $\rho(\boldsymbol{u}) \leq 1$. It follows immediately from Lemma 12.3 that $\|\boldsymbol{u}\|_{\boldsymbol{p}} \leq 1$, i.e., $\boldsymbol{u} \in \overline{\mathscr{B}(0,1)}$. Therefore, $\Lambda \subset \overline{B(0,1)}$. On the other hand, for any $\boldsymbol{u} \in \overline{\mathscr{B}(0,1)}$, $\|\boldsymbol{u}\|_{\boldsymbol{p}} \leq 1$. Then by Lemma 12.4, $\rho(\boldsymbol{u}) \leq 1$ and thus $\boldsymbol{u} \in \Lambda$. Therefore $\overline{B(0,1)} \subset \Lambda$ which concludes that $\Lambda = \overline{\mathscr{B}(0,1)}$. □

Lemma 12.5 *Let* $\boldsymbol{u}^n \in \ell_{\boldsymbol{p}}$ *for* $n \in \mathbb{N}$. *Then* $\|\boldsymbol{u}^n\|_{\boldsymbol{p}} \to 0$ *as* $n \to \infty$ *if and only if* $\rho(\alpha \boldsymbol{u}^n) \to 0$ *as* $n \to \infty$ *for every* $\alpha > 0$.

Proof Assume that $\|\boldsymbol{u}^n\|_{\boldsymbol{p}} \to 0$ as $n \to \infty$. Then for every $\alpha > 0$, we have

$$\lim_{n\to\infty} \|K\alpha \boldsymbol{u}^n\|_{\boldsymbol{p}} = \lim_{n\to\infty} |K||\alpha|\|\boldsymbol{u}^n\|_{\boldsymbol{p}} = 0, \quad \forall K > 0.$$

Hence for any $K > 1$ there exists $N_0 = N_0(K) > 0$ such that $\|K\alpha \boldsymbol{u}^n\|_{\boldsymbol{p}} < 1$ for all $n > N_0$. As a consequence

$$\|K\alpha \boldsymbol{u}^n\|_{\boldsymbol{p}} = \inf\left\{\lambda > 0 : \rho\left(\frac{K\alpha \boldsymbol{u}^n}{\lambda}\right) \leq 1\right\} < 1, \text{ for all } n > N_0(K).$$

By Lemma 12.3-(i),

$$\inf\left\{\lambda > 0 : \frac{1}{\lambda}\rho(K\alpha \boldsymbol{u}^n) \leq \rho\left(K\alpha \frac{\boldsymbol{u}^n}{\lambda}\right) \leq 1\right\} < 1, \text{ for all } n > N_0(K),$$

i.e., $\rho(K\alpha \boldsymbol{u}^n) \leq \lambda < 1$ for all $n > N_0(K)$. As a result, for any $K > 1$ we have

$$\rho(\alpha \boldsymbol{u}^n) \leq \frac{1}{K}\rho(K\alpha \boldsymbol{u}^n) < \frac{1}{K}, \text{ for all } n > N_0(K),$$

which implies that $\rho(\alpha \boldsymbol{u}^n) \to 0$ as $n \to \infty$.

Now assume that if $\rho(\alpha \boldsymbol{u}^n) \to 0$ as $n \to \infty$ for every $\alpha > 0$. Then for any $\alpha > 0$ there exists $N_0 = N_0(\alpha)$ such that $\rho(\alpha \boldsymbol{u}^n) \leq 1$ for $n > N_0$. In particular, by the definition of the norm, $\|\boldsymbol{u}^n\|_{\boldsymbol{p}} \leq 1/\alpha$ for $n > N_0(\alpha)$. Since $\alpha > 0$ is arbitrary, $\|\boldsymbol{u}^n\|_{\boldsymbol{p}} \to 0$ as $n \to \infty$. □

Lemma 12.6 *The mapping* $\rho : \ell_{\boldsymbol{p}} \to \mathbb{R}^+$ *is lower semi continuous, i.e.,*

$$\rho(\hat{\boldsymbol{u}}) \leq \liminf_{n\to\infty} \rho(\boldsymbol{u}^n) \quad \textit{for all} \quad \boldsymbol{u}^n \to \hat{\boldsymbol{u}} \ \textit{in} \ \mathscr{P}.$$

Proof Let $\boldsymbol{u}^n, \hat{\boldsymbol{u}} \in \ell_{\boldsymbol{p}}$ with $\|\boldsymbol{u}^n - \hat{\boldsymbol{u}}\|_{\boldsymbol{p}} \to 0$. By Lemma 12.5

$$\lim_{n\to\infty} \rho(\alpha(\hat{\boldsymbol{u}} - \boldsymbol{u}^n)) = \lim_{n\to\infty} \rho(\alpha(\boldsymbol{u}^n - \hat{\boldsymbol{u}})) = 0 \quad \forall \alpha > 0.$$

Let $\varepsilon \in (0, \frac{1}{2})$. Then by the convexity of ρ we have

$$\begin{aligned}\rho((1-\varepsilon)\hat{\boldsymbol{u}}) &= \rho\left(\frac{1}{2}\hat{\boldsymbol{u}} + \frac{1-2\varepsilon}{2}\left(\hat{\boldsymbol{u}} - \boldsymbol{u}^n\right) + \frac{1-2\varepsilon}{2}\boldsymbol{u}^n\right) \\ &\leq \frac{1}{2}\rho(\hat{\boldsymbol{u}}) + \frac{1}{2}\rho\Big((1-2\varepsilon)(\hat{\boldsymbol{u}} - \boldsymbol{u}^n) + (1-2\varepsilon)\boldsymbol{u}^n\Big) \\ &\leq \frac{1}{2}\rho(\hat{\boldsymbol{u}}) + \frac{2\varepsilon}{2}\rho\left(\frac{1-2\varepsilon}{2\varepsilon}(\hat{\boldsymbol{u}} - \boldsymbol{u}^n)\right) + \frac{1-2\varepsilon}{2}\rho\left(\boldsymbol{u}^n\right).\end{aligned}$$

Taking the limit of the above inequality as $n \to \infty$ gives

$$\rho((1-\varepsilon)\hat{\boldsymbol{u}}) \leq \frac{1}{2}\rho(\hat{\boldsymbol{u}}) + \frac{1-2\varepsilon}{2}\liminf_{n\to\infty}\rho\left(\boldsymbol{u}^n\right). \tag{12.10}$$

Then taking the limit of (12.10) as $\varepsilon \to 0+$ to obtain

$$\rho(\hat{\boldsymbol{u}}) \leq \frac{1}{2}\rho(\hat{\boldsymbol{u}}) + \frac{1}{2}\liminf_{n\to\infty}\rho\left(\boldsymbol{u}^n\right).$$

Finally, since $\rho(\hat{\boldsymbol{u}}) < \infty$, this gives $\rho(\hat{\boldsymbol{u}}) \leq \liminf_{n\to\infty}\rho\left(\boldsymbol{u}^n\right)$. □

The following lemma provide a direct relationship between the norm $\|\cdot\|_{\boldsymbol{p}}$ and the semi-modular ρ.

Lemma 12.7 *Let $\boldsymbol{u} \in \ell_{\boldsymbol{p}} - \{\mathbf{0}\}$. Then $\|\boldsymbol{u}\|_{\boldsymbol{p}} = a$ if and only if $\rho(\boldsymbol{u}/a) = 1$.*

Proof For convenience, write

$$I_{\boldsymbol{u}} := \{\lambda > 0 : \rho(\boldsymbol{u}/\lambda) \leq 1\}.$$

($\Rightarrow$) When $\|\boldsymbol{u}\|_{\boldsymbol{p}} = a$, $I_{\boldsymbol{u}} = [a, \infty)$ and $\rho(\boldsymbol{u}/a) \leq 1$. Suppose (for contradiction) that $\rho(\boldsymbol{u}/a) < 1$. Then for $\lambda > 0$, the function $\lambda \mapsto \rho(\boldsymbol{u}/\lambda)$ is continuous and decreasing as a direct consequence of Lemma 12.3. So, there exists $\delta > 0$ such that $\rho(\boldsymbol{u}/\lambda) < 1$ for $\lambda \in (a-\delta, a+\delta)$. This implies that $a - \frac{\delta}{2} \in I_{\boldsymbol{u}}$, which is a contradiction. Therefore, $\rho(\boldsymbol{u}/a) = 1$.
($\Leftarrow$) When $\rho(\boldsymbol{u}/a) = 1$, $a \in I_{\boldsymbol{u}}$ and thus $\|\boldsymbol{u}\|_{\boldsymbol{p}} \leq a$. Suppose (for contradiction) that $\|\boldsymbol{u}\|_{\boldsymbol{p}} < a$. Then by the properties of infimum, there exists $\lambda_0 \in I_{\boldsymbol{u}}$ such that $\|\boldsymbol{u}\|_{\boldsymbol{p}} \leq \lambda_0 < a$. This implies that $\rho(\boldsymbol{u}/a) < \rho(\boldsymbol{u}/\lambda_0) \leq 1$ by contradiction. Therefore, $\|\boldsymbol{u}\|_{\boldsymbol{p}} = a$. □

In the rest of this section we construct several inequalities on $\|\cdot\|_{\boldsymbol{p}}$ which are critical for studying lattice systems that can be formulated as an ordinary differential equation or an evolution equation on $\mathscr{P}$.

Theorem 12.2 *For any* $\boldsymbol{u} \in \mathscr{P}$,

$$\min\left\{\rho(\boldsymbol{u})^{1/p^-}, \rho(\boldsymbol{u})^{1/p^+}\right\} \le \|\boldsymbol{u}\|_{\boldsymbol{p}} \le \max\left\{\rho(\boldsymbol{u})^{1/p^-}, \rho(\boldsymbol{u})^{1/p^+}\right\}. \tag{12.11}$$

Proof Assume that $\rho(\boldsymbol{u}) < 1$. Then $1/p^+ < 1/p^- \le 1$, and thus $\rho(\boldsymbol{u})^{1/p^-} \le \rho(\boldsymbol{u})^{1/p^+}$. In this case the inequality (12.11) becomes

$$\rho(\boldsymbol{u})^{1/p^-} \le \|\boldsymbol{u}\|_{\boldsymbol{p}} \le \rho(\boldsymbol{u})^{1/p^+}. \tag{12.12}$$

Since $p_i/p^+ \le 1$ for all $i \in \mathbb{Z}$ and $\rho(\boldsymbol{u}) < 1$, then $\sum_{i\in\mathbb{Z}} |u_i|^{p_i} = \rho(\boldsymbol{u}) \le \rho(\boldsymbol{u})^{p_i/p^+}$. Hence $\rho(\boldsymbol{u})^{-p_i/p^+} \le \rho(\boldsymbol{u})^{-1}$ for all $i \in \mathbb{Z}$, and thus

$$\rho\left(\frac{\boldsymbol{u}}{\rho(\boldsymbol{u})^{1/p^+}}\right) = \sum_{i\in\mathbb{Z}} \left|\frac{u_i}{\rho(\boldsymbol{u})^{1/p^+}}\right|^{p_i} = \sum_{i\in\mathbb{Z}} \frac{|u_i|^{p_i}}{\rho(\boldsymbol{u})^{p_i/p^+}} \le \frac{\sum_{i\in\mathbb{Z}} |u_i|^{p_i}}{\rho(\boldsymbol{u})} = 1$$

By the Unit Ball Theorem this means that $\left\|\frac{\boldsymbol{u}}{\rho(\boldsymbol{u})^{1/p^+}}\right\|_{\boldsymbol{p}} \le 1$. Finally, by the homogeneity property of a norm,

$$\frac{1}{\rho(\boldsymbol{u})^{1/p^+}} \|\boldsymbol{u}\|_{\boldsymbol{p}} = \left\|\frac{\boldsymbol{u}}{\rho(\boldsymbol{u})^{1/p^+}}\right\|_{\boldsymbol{p}} \le 1,$$

i.e., $\|\boldsymbol{u}\|_{\boldsymbol{p}} \le \rho(\boldsymbol{u})^{1/p^+}$, which is the upper inequality in (12.12).

Similarly, since $p_i/p^- \ge 1$ for all $i \in \mathbb{Z}$ and $\rho(\boldsymbol{u}) < 1$, then $\rho(\boldsymbol{u})^{p_i/p^-} \le \rho(\boldsymbol{u})$. Hence $\rho(\boldsymbol{u})^{-1} \le \rho(\boldsymbol{u})^{-p_i/p^-}$ or all $i \in \mathbb{Z}$, so

$$1 = \frac{\sum_{i\in\mathbb{Z}} |u_i|^{p_i}}{\rho(\boldsymbol{u})} \le \sum_{i\in\mathbb{Z}} \frac{|u_i|^{p_i}}{\rho(\boldsymbol{u})^{p_i/p^-}} = \sum_{i\in\mathbb{Z}} \left|\frac{u_i}{\rho(\boldsymbol{u})^{1/p^-}}\right|^{p_i} = \rho\left(\frac{\boldsymbol{u}}{\rho(\boldsymbol{u})^{1/p^-}}\right).$$

By the Unit Ball Theorem again, $1 \le \left\|\frac{\boldsymbol{u}}{\rho(\boldsymbol{u})^{1/p^+}}\right\|_{\boldsymbol{p}}$ and the homogeneity property of a norm

$$1 \le \left\|\frac{\boldsymbol{u}}{\rho(\boldsymbol{u})^{1/p^-}}\right\|_{\boldsymbol{p}} = \frac{1}{\rho(\boldsymbol{u})^{1/p^-}} \|\boldsymbol{u}\|_{\boldsymbol{p}},$$

i.e., $\rho(\boldsymbol{u})^{1/p^+} \le \|\boldsymbol{u}\|_{\boldsymbol{p}}$, which is the lower inequality in (12.12).

The case $\rho(\boldsymbol{u}) > 1$ can be shown by a similar process. □

Corollary 12.3 *For any* $\boldsymbol{u} \in \mathscr{P}$, $\|\boldsymbol{u}\|_{\boldsymbol{p}}^{p^-} \le \rho(\boldsymbol{u}) \le \|\boldsymbol{u}\|_{\boldsymbol{p}}^{p^+}$ *if* $\|\boldsymbol{u}\|_{\boldsymbol{p}} > 1$; *and* $\|\boldsymbol{u}\|_{\boldsymbol{p}}^{p^+} \le \rho(\boldsymbol{u}) \le \|\boldsymbol{u}\|_{\boldsymbol{p}}^{p^-}$ *if* $\|\boldsymbol{u}\|_{\boldsymbol{p}} \le 1$. *In other words,*

$$\min\left\{\|\boldsymbol{u}\|_{\boldsymbol{p}}^{p^-}, \|\boldsymbol{u}\|_{\boldsymbol{p}}^{p^+}\right\} \le \rho(\boldsymbol{u}) \le \max\left\{\|\boldsymbol{u}\|_{\boldsymbol{p}}^{p^-}, \|\boldsymbol{u}\|_{\boldsymbol{p}}^{p^+}\right\}.$$

By using the Corollary 12.3 above Lemma 12.5 can be improved to the following important result.

Theorem 12.3 *Let* $\{\boldsymbol{u}^n\}$ *be a sequence in* $\ell_{\boldsymbol{p}}$ *and let* $\boldsymbol{u} \in \ell_{\boldsymbol{p}}$. *Then*

$$\lim_{n\to\infty} \rho(\boldsymbol{u}^n - \boldsymbol{u}) = 0 \text{ if and only if } \lim_{n\to\infty} \|\boldsymbol{u}^n - \boldsymbol{u}\|_{\boldsymbol{p}} = 0.$$

Proof ($\Rightarrow$) Assume $\lim_{n\to\infty} \rho(\boldsymbol{u}^n - \boldsymbol{u}) = 0$. Given any $\varepsilon \in (0,1)$ there exists $N_1 = N_1(\varepsilon) \in \mathbb{N}$ such that $\rho(\boldsymbol{u}^n - \boldsymbol{u}) < \varepsilon^{p^+} < \varepsilon < 1$ for all $n \ge N_1$. Then by Theorem 12.1 (i), we have $\|\boldsymbol{u}^n - \boldsymbol{u}\|_{\boldsymbol{p}} < 1$ for all $n \ge N_1$ and by Corollary 12.3, $\|\boldsymbol{u}^n - \boldsymbol{u}\|_{\boldsymbol{p}}^{p^+} \le \rho(\boldsymbol{u}^n - \boldsymbol{u}) < \varepsilon^{p^+}$ for all $n \ge N_1$. Therefore, $\lim_{n\to\infty} \|\boldsymbol{u}^n - \boldsymbol{u}\|_{\boldsymbol{p}} = 0$.

($\Leftarrow$) Assume $\lim_{n\to\infty} \|\boldsymbol{u}^n - \boldsymbol{u}\|_{\boldsymbol{p}} = 0$. Given any $\varepsilon \in (0,1)$ there exists $N_2 = N_2(\varepsilon) \in \mathbb{N}$ such that $\|\boldsymbol{u}^n - \boldsymbol{u}\|_{\boldsymbol{p}} < \varepsilon < 1$ for all $n \ge N_2$. Then by Corollary 12.3, $\rho(\boldsymbol{u}^n - \boldsymbol{u}) \le \|\boldsymbol{u}^n - \boldsymbol{u}\|_{\boldsymbol{p}}^{p^-} < \varepsilon^{p^-} < \varepsilon$ for all $n \ge N_2$, which implies that $\lim_{n\to\infty} \rho(\boldsymbol{u}^n - \boldsymbol{u}) = 0$. □

12.4 Properties of the Space $\mathscr{P}$

In this section we discuss properties of the space $\ell_{\boldsymbol{p}}$ equipped with the norm $\|\cdot\|_{\boldsymbol{p}}$. Due to Lemmas 12.1 and 12.2 the space $\mathscr{P}$ is a linear normed space. Indeed, $\mathscr{P}$ is a Banach space (see Theorem 12.4 below) and moreover, $\mathscr{P}$ is separable (see Theorem 12.5 below) and reflexive (see Theorem 12.7 below).

Theorem 12.4 *The space* $\mathscr{P}$ *is a Banach space.*

Proof With Lemmas 12.1 and 12.2 it remains to prove the completeness of $\mathscr{P}$. To this end, let $\{\boldsymbol{u}^n\}_{n\in\mathbb{N}}$ be a Cauchy sequence on $\mathscr{P}$. Then given any $\varepsilon \in (0,1)$ there exists an $N_1(\varepsilon) \in \mathbb{N}$ such that $\|\boldsymbol{u}^n - \boldsymbol{u}^m\|_{\boldsymbol{p}} \le \varepsilon < 1$ for all $n, m \ge N_1(\varepsilon)$. By Lemma 12.4,

$$\rho\left(\boldsymbol{u}^n - \boldsymbol{u}^m\right) \le \|\boldsymbol{u}^n - \boldsymbol{u}^m\|_{\boldsymbol{p}} \le \varepsilon, \qquad \forall n, m \ge N_1(\varepsilon).$$

Thus for each $i \in \mathbb{Z}$,

$$\left|u_i^n - u_i^m\right| \le \varepsilon^{1/p_i} \qquad \text{for } n, m \ge N_1(\varepsilon), \tag{12.13}$$

which implies that the sequence $\{u_i^n\}$ is a Cauchy sequence in $\mathbb{R}$ for each $i \in \mathbb{Z}$ and thus has a convergent subsquence $\{u_i^{n_j}\}$ such that $\lim_{n_j \to \infty} u_i^{n_j} = \hat{u}_i$ for each $i \in \mathbb{Z}$.

Define the bi-infinite sequence $\hat{\boldsymbol{u}} := (\hat{u}_i)_{i \in \mathbb{Z}}$. Then by a diagonal subsequence argument, there exists a subsequence $\{\boldsymbol{u}^{n_j}\}_{j \in \mathbb{N}}$ in $\ell_{\boldsymbol{p}}$ of the sequence $\{\boldsymbol{u}^n\}_{n \in \mathbb{N}}$ that converges componentwise to $\hat{\boldsymbol{u}}$. Replacing n and m in (12.13) by n_j, n_{j+k}, i.e., using elements $u_i^{n_j}$, $u_i^{n_{j+k}}$ of this diagonal subsequence, and taking the limit as $k \to \infty$ gives

$$\left| u_i^{n_j} - \hat{u}_i \right| \le \varepsilon^{1/p_i} \qquad \text{for } n_j \ge N_1(\varepsilon),\ i \in \mathbb{Z}.$$

The rest of the proof follows from that of [5, Theorem 2.3.13]. Let $\alpha > 0$ and $\varepsilon \in (0, 1)$. Since $\{\boldsymbol{u}^n\}_{n \in \mathbb{N}}$ is a Cauchy sequence in $\mathscr{P}$ there exists $N_2 = N_2(\alpha, \varepsilon) \in \mathbb{N}$ such that $\|\alpha(\boldsymbol{u}^n - \boldsymbol{u}^m)\|_{\boldsymbol{p}} < \varepsilon$ for all $n, m \ge N_2$. Lemma 12.4 implies that $\rho\left(\alpha(\boldsymbol{u}^n - \boldsymbol{u}^m)\right) < \varepsilon$ and by Fatou's Lemma we obtain that

$$\begin{aligned}
\rho\left(\alpha(\boldsymbol{u}^{n_j} - \hat{\boldsymbol{u}})\right) &= \sum_{i \in \mathbb{Z}} \lim_{k \to \infty} \left| \alpha(u_i^{n_j} - u_i^{n_k}) \right|^{p_i} \\
&\le \lim_{k \to \infty} \sum_{i \in \mathbb{Z}} \left| \alpha(u_i^{n_j} - u_i^{n_k}) \right|^{p_i} = \lim_{k \to \infty} \rho\left(\alpha(\boldsymbol{u}^{n_j} - \boldsymbol{u}^{n_k})\right) \le \varepsilon,
\end{aligned}$$

which implies that $\rho\left(\alpha(\boldsymbol{u}^{n_j} - \hat{\boldsymbol{u}})\right) \to 0$ as $j \to \infty$ for all $\alpha > 0$. Then by Lemma 12.5, $\|\boldsymbol{u}^{n_j} - \hat{\boldsymbol{u}}\|_{\boldsymbol{p}} \to 0$ as $j \to \infty$ and thus the Cauchy sequence $\boldsymbol{u}^n$ converges in $\mathscr{P}$.

It remains to show that $\hat{\boldsymbol{u}} \in \ell_{\boldsymbol{p}}$. Note that by inequality (12.9)

$$|\hat{u}_i|^{p_i} \le 2^{p^+ - 1} |\hat{u}_i - u_i^{n_j}|^{p_i} + 2^{p^+ - 1} |u_i^{n_j}|^{p_i}, \quad \forall n_j > \max\{N_1, N_2\}$$

for each $i \in Z$. Therefore

$$\begin{aligned}
\rho\left(\hat{\boldsymbol{u}}\right) = \sum_{i \in \mathbb{Z}} |\hat{u}_i|^{p_i} &\le 2^{p^+ - 1} \sum_{i \in \mathbb{Z}} |\hat{u}_i - u_i^{n_j}|^{p_i} + 2^{p^+ - 1} \sum_{i \in \mathbb{Z}} |u_i^{n_j}|^{p_i} \\
&\le 2^{p^+ - 1} \rho\left(\boldsymbol{u}^{n_j} - \hat{\boldsymbol{u}}\right) + 2^{p^+ - 1} \rho\left(\boldsymbol{u}^{n_j}\right) \\
&\le 2^{p^+ - 1} \varepsilon + 2^{p^+ - 1} \rho\left(\boldsymbol{u}^{n_j}\right) < \infty
\end{aligned}$$

for $j > \max\{N_1, N_2\}$. It follows from $\boldsymbol{u}^{n_j} \in \ell_{\boldsymbol{p}}$ that $\hat{\boldsymbol{u}} \in \ell_{\boldsymbol{p}}$. □

Theorem 12.5 *The Banach space $\mathscr{P}$ is separable.*

Proof Let $\boldsymbol{e}^n = (e_i^n)_{i \in \mathbb{Z}}$ be the element in $\ell_{\boldsymbol{p}}$ with $e_i^n = \delta_{ni}$ where δ is the Kronecker delta. Then $\boldsymbol{e}^n$ forms a Schauder base for $\mathscr{P}$ and the desired assertion follows immediately. □

To show that $\mathscr{P}$ is reflexive, we now introduce the dual space of $\ell_{\boldsymbol{p}}$. More precisely, given a sequence $\boldsymbol{p} = (p_i)_{i\in\mathbb{Z}}$, let $\boldsymbol{q} = (q_i)_{i\in\mathbb{Z}}$ be the bi-infinite sequence such that

$$\frac{1}{p_i} + \frac{1}{q_i} = 1, \quad \forall\, i \in \mathbb{Z}.$$

Define the discrete Musielak-Orlicz space $\ell_{\boldsymbol{q}}$ by

$$\ell_{\boldsymbol{q}} := \left\{ \boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} : \sum_{i\in\mathbb{Z}} |u_i|^{q_i} < \infty \right\}.$$

Similarly to (12.6) and (12.7) for any $\boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{q}}$ define

$$\eta(\boldsymbol{u}) := \sum_{i\in\mathbb{Z}} |u_i|^{q_i} \quad \text{and} \quad \|\boldsymbol{u}\|_{\boldsymbol{q}} := \inf\left\{ \lambda > 0 : \eta\left(\frac{\boldsymbol{u}}{\lambda}\right) \le 1 \right\}. \tag{12.14}$$

The following theorem presents a Hölder-like inequality.

Theorem 12.6 *For any $\boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{p}}$ and $\boldsymbol{v} = (v_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{q}}$ it holds*

$$\sum_{i\in\mathbb{Z}} |u_i v_i| \le \left(\frac{1}{p^-} + \frac{1}{q^-} \right) \|\boldsymbol{u}\|_{\boldsymbol{p}} \|\boldsymbol{v}\|_{\boldsymbol{q}},$$

where $q^- := \inf_{i\in\mathbb{Z}} q_i$.

Proof Let $\|\boldsymbol{u}\|_{\boldsymbol{p}} = a$ and $\|\boldsymbol{u}\|_{\boldsymbol{q}} = b$. By Young's inequality and Lemma 12.7 we obtain

$$\begin{aligned}
\frac{1}{ab} \sum_{i\in\mathbb{Z}} |u_i v_i| &\le \sum_{i\in\mathbb{Z}} \frac{1}{p_i} \left|\frac{u_i}{a}\right|^{p_i} + \sum_{i\in\mathbb{Z}} \frac{1}{q_i} \left|\frac{v_i}{b}\right|^{q_i} \\
&\le \frac{1}{p^-} \sum_{i\in\mathbb{Z}} \left|\frac{u_i}{a}\right|^{p_i} + \frac{1}{q^-} \sum_{i\in\mathbb{Z}} \left|\frac{v_i}{b}\right|^{q_i} \\
&= \frac{1}{p^-} \rho(\boldsymbol{u}/a) + \frac{1}{q^-} \eta(\boldsymbol{v}/b) = \frac{1}{p^-} + \frac{1}{q^-},
\end{aligned}$$

which completes the proof. □

Denote by $\mathscr{Q} := (\ell_{\boldsymbol{q}}, \|\cdot\|_{\boldsymbol{q}})$. We next show that $\mathscr{P}$ is reflexive with the dual space $\mathscr{Q}$.

Lemma 12.8 *$\ell_{\boldsymbol{p}}$ is reflexive provided $p^- \ge 2$.*

Proof We will prove that $\ell_{\boldsymbol{p}}$ is uniformly convex, which implies its reflexivity. Given any $\varepsilon > 0$ let $\boldsymbol{u}, \boldsymbol{v} \in \ell_{\boldsymbol{p}}$ be such that $\|\boldsymbol{u}\|_{\boldsymbol{p}} \le 1$, $\|\boldsymbol{v}\|_{\boldsymbol{p}} \le 1$ and $\|\boldsymbol{u} - \boldsymbol{v}\|_{\boldsymbol{p}} > \varepsilon$.

Since $p_i \geq p^- \geq 2$ for all $i \in \mathbb{Z}$, Clarkson's inequality gives

$$\left|\frac{u_i + v_i}{2}\right|^{p_i} + \left|\frac{u_i - v_i}{2}\right|^{p_i} \leq \frac{1}{2}\left(|u_i|^{p_i} + |v_i|^{p_i}\right), \ \forall\, i \in \mathbb{Z}.$$

Summing the above inequality over $i \in \mathbb{Z}$ and using Theorem 12.1 (ii) we have

$$\rho\left(\frac{\boldsymbol{u}+\boldsymbol{v}}{2}\right) + \rho\left(\frac{\boldsymbol{u}-\boldsymbol{v}}{2}\right) \leq \frac{1}{2}\rho(\boldsymbol{u}) + \frac{1}{2}\rho(\boldsymbol{v}) \leq 1,$$

which implies that $\rho(\frac{\boldsymbol{u}+\boldsymbol{v}}{2}) \leq 1$ and $\rho(\frac{\boldsymbol{u}-\boldsymbol{v}}{2}) \leq 1$.

Using Theorem 12.1 again we obtain $\|\frac{\boldsymbol{u}+\boldsymbol{v}}{2}\|_{\boldsymbol{p}} \leq 1$ and $\|\frac{\boldsymbol{u}-\boldsymbol{v}}{2}\|_{\boldsymbol{p}} \leq 1$. It then follows immediately from Corollary 12.3 (ii) that

$$\left\|\frac{\boldsymbol{u}+\boldsymbol{v}}{2}\right\|_{\boldsymbol{p}}^{p^+} + \left\|\frac{\boldsymbol{u}-\boldsymbol{v}}{2}\right\|_{\boldsymbol{p}}^{p^+} \leq 1.$$

Since $\|\boldsymbol{u}-\boldsymbol{v}\|_{\boldsymbol{p}} > \varepsilon$, $\left\|\frac{\boldsymbol{u}+\boldsymbol{v}}{2}\right\|_{\boldsymbol{p}} < \left[1 - (\frac{\varepsilon}{2})^{p^+}\right]^{1/p^+} < 1$, which implies that $\ell_{\boldsymbol{p}}$ is uniformly convex and completes the proof.

To prove the duality $\ell_{\boldsymbol{p}}^* \equiv \ell_{\boldsymbol{q}}$ for $p^- \geq 2$ the following lemma is needed.

Lemma 12.9 *Let $T : \ell_{\boldsymbol{q}} \to \ell_{\boldsymbol{p}}^*$ be the linear operator defined by*

$$\boldsymbol{v} \mapsto \langle T\boldsymbol{v}, \boldsymbol{u}\rangle_{\ell_{\boldsymbol{p}}^*, \ell_{\boldsymbol{p}}} := \sum_{i\in\mathbb{Z}} u_i v_i, \quad \forall\, \boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{p}},\ \boldsymbol{v} = (v_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{q}}.$$

Then T is injective.

Proof For $\boldsymbol{v} = (v_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{q}}$ with $\|\boldsymbol{v}\|_{\boldsymbol{q}} = a$, let

$$\hat{\boldsymbol{v}} = \left(\hat{v}_i\right)_{i\in\mathbb{Z}} \quad \text{with} \quad \hat{v}_i = \left|\frac{v_i}{a}\right|^{q_i - 1} \cdot \operatorname{sign} v_i \quad \text{and} \quad \operatorname{sign} v_i := \begin{cases} 1, & v_i > 0, \\ 0, & v_i = 0, \\ -1, & v_i < 0. \end{cases}$$

Then $\hat{\boldsymbol{v}} \in \ell_{\boldsymbol{p}}$ and $\|\hat{\boldsymbol{v}}\|_{\boldsymbol{p}} = 1$. In fact, by Lemma 12.7, $\eta(\boldsymbol{v}/a) = 1$ and hence

$$\rho(\hat{\boldsymbol{v}}) = \sum_{i\in\mathbb{Z}} \left|\hat{v}_i\right|^{p_i} = \sum_{i\in\mathbb{Z}} \left|\frac{v_i}{a}\right|^{q_i} = \eta\left(\frac{\boldsymbol{v}}{a}\right) = 1.$$

By Lemma 12.7 again, $\|\hat{\boldsymbol{v}}\|_{\boldsymbol{p}} = 1$. On the other hand,

$$\langle T\boldsymbol{v}, \hat{\boldsymbol{v}}\rangle_{\ell_{\boldsymbol{p}}^*,\ell_{\boldsymbol{p}}} = \sum_{i\in\mathbb{Z}} v_i \left|\frac{v_i}{a}\right|^{q_i-1} \operatorname{sign} v_i = \sum_{i\in\mathbb{Z}} \left|\frac{v_i}{a}\right|^{q_i} a = a\eta\left(\frac{\boldsymbol{v}}{a}\right) = a.$$

Therefore

$$\|T\boldsymbol{v}\|_{\ell_{\boldsymbol{p}}^*} = \sup_{\|\boldsymbol{u}\|_{\boldsymbol{p}}\le 1} \langle T\boldsymbol{v}, \boldsymbol{u}\rangle_{\ell_{\boldsymbol{p}}^*,\ell_{\boldsymbol{p}}} \ge \langle T\boldsymbol{v}, \hat{\boldsymbol{v}}\rangle_{\ell_{\boldsymbol{p}}^*,\ell_{\boldsymbol{p}}} = a = \|\boldsymbol{v}\|_{\boldsymbol{q}}\,, \tag{12.15}$$

which implies that $\boldsymbol{v}^{(1)} = \boldsymbol{v}^{(2)}$ if $T\boldsymbol{v}^{(1)} = T\boldsymbol{v}^{(2)}$ and completes the proof. □

Lemma 12.10 *Assume that (i) $p^- \ge 2$ or (ii) $1 < p^- \le p_i \le p^+ \le 2$. Then $\mathscr{Q}$ is the dual space of $\mathscr{P}$ with $\ell_{\boldsymbol{p}}^* \equiv \ell_{\boldsymbol{q}}$.*

Proof

(i) First notice that $1 < q_i \le 2$ when $2 \le p^- \le p_i \le p^+ < \infty$, because

$$1 < 1 + \frac{1}{p^+ - 1} \le q_i = 1 + \frac{1}{p_i - 1} \le 1 + \frac{1}{p^- - 1} \le 2.$$

Let T be the linear injection defined in Lemma 12.9 and let $E := T(\ell_{\boldsymbol{q}})$. Then E is a vectorial subspace of $\ell_{\boldsymbol{p}}^*$. Due to the Hölder inequality in Theorem 12.6, $\|T\boldsymbol{v}\|_{\ell_{\boldsymbol{p}}^*} \le C\|\boldsymbol{v}\|_{\boldsymbol{q}}$ for some constant $C > 0$. Together with (12.15) we obtain $\|\boldsymbol{v}\|_{\boldsymbol{q}} \le \|T\boldsymbol{v}\|_{\ell_{\boldsymbol{p}}^*} \le C\|\boldsymbol{v}\|_{\boldsymbol{q}}$. It then follows that E is closed, because $\ell_{\boldsymbol{q}}$ is a Banach space.

We next show that T is an isomorphism. To this end, we only need to show that E is dense in $\ell_{\boldsymbol{p}}^*$. Let $\boldsymbol{u} \in \ell_{\boldsymbol{p}}^{**} = \ell_{\boldsymbol{p}}$ (by Lemma 12.8) be such that

$$\langle T\boldsymbol{v}, \boldsymbol{u}\rangle_{\ell_{\boldsymbol{p}}^*,\ell_{\boldsymbol{p}}} = 0 \quad \text{for all } \boldsymbol{v} \in \ell_{\boldsymbol{q}}.$$

Let $\hat{\boldsymbol{v}} = \left(\hat{v}_i\right)_{i\in\mathbb{Z}}$ with $\hat{v}_i = |u_i|^{p_i-2}u_i$. Then

$$\eta(\hat{\boldsymbol{v}}) = \sum_{i\in\mathbb{Z}} \left||u_i|^{p_i-2}u_i\right|^{q_i} = \sum_{i\in\mathbb{Z}} |u_i|^{p_i} = \rho(\boldsymbol{u}) < \infty.$$

Thus $\hat{\boldsymbol{v}} \in \ell_{\boldsymbol{q}}$ and should satisfy $\langle T\hat{\boldsymbol{v}}, \boldsymbol{u}\rangle_{\ell_{\boldsymbol{p}}^*,\ell_{\boldsymbol{p}}} = 0$. On the other hand,

$$\langle T\hat{\boldsymbol{v}}, \boldsymbol{u}\rangle_{\ell_{\boldsymbol{p}}^*,\ell_{\boldsymbol{p}}} = \sum_{i\in\mathbb{Z}} |u_i|^{p_i-2}u_i^2 = \rho(\boldsymbol{u}).$$

Hence $\rho(\boldsymbol{u}) = 0$, which implies that $\boldsymbol{u} = 0$. By Corollary 1.8 of [3], E is dense in $\ell_{\boldsymbol{p}}$. As a result, T is an isomorphism and therefore $\ell_{\boldsymbol{p}}^* \equiv \ell_{\boldsymbol{q}}$.

(ii) If $1 < p^- \le p_i \le p^+ \le 2$, then $q_i \ge 2$ for all i, which implies that $q^- \ge 2$. Hence by Lemma 12.8, $\ell_{\boldsymbol{q}}$ is reflexive, i.e., $\ell_{\boldsymbol{q}} = \ell_{\boldsymbol{q}}^{**}$. On the other hand by part (i) we have $\ell_{\boldsymbol{q}}^* = \ell_{\boldsymbol{p}}$. As a result, $\ell_{\boldsymbol{q}} = (\ell_{\boldsymbol{q}}^*)^* = (\ell_{\boldsymbol{p}})^*$. □

The next lemma follows immediately from Lemmas 12.8 and 12.10.

Lemma 12.11 *Suppose that $\boldsymbol{p} = (p_i)_{i\in\mathbb{Z}}$ is such that $p^- \ge 2$ or $1 < p^- \le p_i \le p^+ \le 2$. Then $\ell_{\boldsymbol{p}}$ is reflexive.*

Proof If $p^- \ge 2$, then we know from Lemma 12.8 that $\ell_{\boldsymbol{p}}$ is reflexive.

If $1 < p^- \le p_i \le p^+ \le 2$, then by Lemma 12.10, $\ell_{\boldsymbol{p}}$ is the dual space of the reflexive space $\ell_{\boldsymbol{q}}$ with $q^- \ge 2$, and consequently also reflexive. Indeed, $\ell_{\boldsymbol{p}}^{**} = (\ell_{\boldsymbol{q}}^*)^{**} = (\ell_{\boldsymbol{q}}^{**})^* = (\ell_{\boldsymbol{q}})^* = \ell_{\boldsymbol{p}}$. □

Finally we reach a general statement on the reflexivity of $\ell_{\boldsymbol{p}}$.

Theorem 12.7 *Assume that $p^- > 1$. Then the space $\ell_{\boldsymbol{p}}$ is reflexive and $\mathscr{Q}$ is the dual space of $\mathscr{P}$.*

Proof For $\boldsymbol{p} = (p_i)_{i\in\mathbb{Z}}$ define

$$\mathscr{J}_{\boldsymbol{p}}^1 := \{i \in \mathbb{Z} : p_i > 2\} \quad \text{and} \quad \mathscr{J}_{\boldsymbol{p}}^2 := \{i \in \mathbb{Z} : 1 < p_i \le 2\}.$$

Then for each $\boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{p}}$, write

$$\boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} = (v_i)_{i\in\mathbb{Z}} + (w_i)_{i\in\mathbb{Z}} = \boldsymbol{v} + \boldsymbol{w},$$

where

$$v_i = \begin{cases} u_i, & i \in \mathscr{J}_{\boldsymbol{p}}^1 \\ 0, & i \in \mathscr{J}_{\boldsymbol{p}}^2 \end{cases}, \quad \text{and} \quad w_i = \begin{cases} 0, & i \in \mathscr{J}_{\boldsymbol{p}}^1 \\ u_i, & i \in \mathscr{J}_{\boldsymbol{p}}^2 \end{cases}. \tag{12.16}$$

Moreover, define the spaces $\ell_{\boldsymbol{p}}^1 := \{\boldsymbol{v} : \boldsymbol{u} \in \ell_{\boldsymbol{p}}\}$ and $\ell_{\boldsymbol{p}}^2 := \{\boldsymbol{w} : \boldsymbol{u} \in \ell_{\boldsymbol{p}}\}$, where $\boldsymbol{v}$ and $\boldsymbol{w}$ are defined as in (12.16). By Lemma 12.11 we have $\ell_{\boldsymbol{p}}^1$ and $\ell_{\boldsymbol{p}}^2$ are reflexive spaces with $(\ell_{\boldsymbol{p}}^1)^* = \ell_{\boldsymbol{q}}^2$ and $(\ell_{\boldsymbol{p}}^2)^* = \ell_{\boldsymbol{q}}^1$. Since $\ell_{\boldsymbol{p}} = \ell_{\boldsymbol{p}}^1 \oplus \ell_{\boldsymbol{p}}^2$, $\ell_{\boldsymbol{p}}$ is also a reflexive space and $\ell_{\boldsymbol{p}}^* \equiv \ell_{\boldsymbol{q}}^2 \oplus \ell_{\boldsymbol{q}}^1 \equiv \ell_{\boldsymbol{q}}$. □

The rest of this section concerns the relation between two sequence spaces with variable exponents.

Theorem 12.8 *Suppose $\boldsymbol{p} = (p_i)_{i\in\mathbb{Z}}$ and $\boldsymbol{r} = (r_i)_{i\in\mathbb{Z}}$ are such that $p_i \ge r_i \ge 1$, for all $i \in \mathbb{Z}$. Then $\ell_{\boldsymbol{r}}$ is densely and continuously embedded in the space $\ell_{\boldsymbol{p}}$.*

Proof We first show the inclusion $\ell_{\boldsymbol{r}} \subset \ell_{\boldsymbol{p}}$. To this end, consider $\boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{r}}$ with $\sum_{i\in\mathbb{Z}} |u_i|^{r_i} < \infty$. Then $\lim_{|i|\to+\infty} |u_i|^{r_i} = 0$ and there exists $I \in \mathbb{N}$ such that $|u_i| < 1$ for all $|i| \ge I$. This gives $|u_i|^{p_i} \le |u_i|^{r_i}$ for all $|i| \ge I$ and as a result

$$\sum_{i\in\mathbb{Z}} |u_i|^{p_i} \le \sum_{|i|<I} |u_i|^{p_i} + \sum_{|i|\ge I} |u_i|^{r_i} < \infty,$$

i.e, $\boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{p}}$.

We next show the inclusion $\ell_{\boldsymbol{r}} \subset \ell_{\boldsymbol{p}}$ is dense. For $\boldsymbol{u} = (u_i)_{i\in\mathbb{Z}} \in \ell_{\boldsymbol{p}}$, consider the truncation sequence of elements in $\ell_{\boldsymbol{r}}$ given by

$$\boldsymbol{u}^n := (\cdots, 0, 0, \cdots, 0, u_{-n}, \cdots, u_{-1}, u_0, u_1, \cdots, u_n, 0, \cdots, 0, 0, \cdots).$$

Clearly $\rho(\boldsymbol{u}^n - \boldsymbol{u}) \to 0$ as $n \to +\infty$ and hence by Theorem 12.3

$$\lim_{n\to+\infty} \|\boldsymbol{u}^n - \boldsymbol{u}\|_{\boldsymbol{p}} = 0.$$

It then remains to show that the inclusion $\ell_{\boldsymbol{r}} \subset \ell_{\boldsymbol{p}}$ is continuous. For any $\boldsymbol{u} \in \ell_{\boldsymbol{r}}$, define $\gamma(\boldsymbol{u}) := \sum_{i\in\mathbb{Z}} |u_i|^{r_i}$. Then γ is a semi-modular for the space $\ell_{\boldsymbol{r}}$ and satisfies all properties developed in Sect. 12.3. Consider the cases (i) $\gamma(\boldsymbol{u}) \le 1$, and (ii) $\gamma(\boldsymbol{u}) > 1$.

(i) When $\gamma(\boldsymbol{u}) \le 1$, $|u_i| \le 1$ for all $i \in \mathbb{Z}$ since $r_i \ge 1$. Thus $|u_i|^{r_i} \ge |u_i|^{p_i}$ for all $i \in \mathbb{Z}$, which implies that $\rho(\boldsymbol{u}) \le \gamma(\boldsymbol{u})$.

(ii) When $\gamma(\boldsymbol{u}) > 1$, for any $k \in \mathbb{Z}$ we have $0 \le \dfrac{|u_k|}{\gamma(\boldsymbol{u})^{1/r_k}} \le 1$, and thus

$$\frac{|u_k|^{p_k}}{\gamma(\boldsymbol{u})^{p^+/r^-}} \le \left[\frac{|u_k|}{\gamma(\boldsymbol{u})^{1/r_k}}\right]^{p_k} \le \left[\frac{|u_k|}{\gamma(\boldsymbol{u})^{1/r_k}}\right]^{r_k}, \tag{12.17}$$

where $r^- := \inf_{i\in\mathbb{Z}} r_i$.

Summing inequalities (12.17) over $k \in \mathbb{Z}$ gives

$$\frac{\rho(\boldsymbol{u})}{[\gamma(\boldsymbol{u})]^{p^+/r^-}} \le \frac{\gamma(\boldsymbol{u})}{\gamma(\boldsymbol{u})} = 1,$$

i.e., $\rho(\boldsymbol{u}) \le [\gamma(\boldsymbol{u})]^{p^+/r^-}$. This implies that

$$\rho(\boldsymbol{u}) \le \max\{\gamma(\boldsymbol{u}), [\gamma(\boldsymbol{u})]^{p^+/r^-}\} \quad \forall \boldsymbol{u} \in \ell_{\boldsymbol{r}}. \tag{12.18}$$

It follows from Theorem 12.3 and (12.18) that $\boldsymbol{u}^{(n)} \to \hat{\boldsymbol{u}}$ in $\ell_{\boldsymbol{r}}$ as $n \to \infty$ implies $\boldsymbol{u}^{(n)} \to \hat{\boldsymbol{u}}$ in $\ell_{\boldsymbol{p}}$ as $n \to \infty$. Therefore, the inclusion $\ell_{\boldsymbol{r}} \hookrightarrow \ell_{\boldsymbol{p}}$ is continuous. □

Remark 12.1 As an important particular case, the Hilbert space ℓ^2 is densely and continuously embedded in the space $\ell_{\boldsymbol{p}}$ with $p^- \ge 2$.

Note that the inclusion $\ell_{\boldsymbol{r}} \hookrightarrow \ell_{\boldsymbol{p}}$ is continuous and dense but not compact. In fact, consider the sequence $\{\boldsymbol{e}^n = (e_i^n)_{i\in\mathbb{Z}}\}_{n\in\mathbb{N}}$ with $e_i^n = \delta_{ni}$ where δ is the Kronecker delta. Then $\rho(\boldsymbol{e}^n) = \gamma(\boldsymbol{e}^n) = 1$ for all $n \in \mathbb{N}$. But there is no convergent subsequence of $\{\boldsymbol{e}^n\}_{n\in\mathbb{N}}$ in $\ell_{\boldsymbol{p}}$, because for all subsequences $\{\boldsymbol{e}^{n_j}\}$ of $\{\boldsymbol{e}^n\}$ we have

$$\rho\left(\boldsymbol{e}^{n_j} - \boldsymbol{e}^{n_l}\right) = 2 \quad \forall j \ne l.$$

However, with Theorem 12.7 we can show that the inclusion is weakly compact, presented in the following theorem.

Theorem 12.9 *Suppose* $\boldsymbol{p} = (p_i)_{i\in\mathbb{Z}}$ *and* $\boldsymbol{r} = (r_i)_{i\in\mathbb{Z}}$ *are such that* $p_i \geq r_i \geq 1$, *for all* $i \in \mathbb{Z}$. *Then the inclusion* $\ell_{\boldsymbol{r}} \subset \ell_{\boldsymbol{p}}$ *is weakly compact.*

Proof Let $\{\boldsymbol{u}^{(n)}\}_{n\in\mathbb{N}}$ be a bounded sequence in $\ell_{\boldsymbol{r}}$. Then there exists $c_1 > 0$ such that

$$\gamma(\boldsymbol{u}^{(n)}) = \sum_{i\in\mathbb{Z}} \left|u_i^{(n)}\right|^{r_i} \leq c_1 \quad \forall n \in \mathbb{N}.$$

It follows immediate from (12.18) that

$$\rho(\boldsymbol{u}^{(n)}) \leq \max\{\gamma(\boldsymbol{u}^{(n)}), [\gamma(\boldsymbol{u}^{(n)})]^{p^+/r^-}\} \leq c_2 \quad \forall n \in \mathbb{N}$$

for some $c_2 > 0$. Hence $\{\boldsymbol{u}^{(n)}\}_{n\in\mathbb{N}}$ is a bounded sequence in the space $\ell_{\boldsymbol{p}}$.

Noting that $\ell_{\boldsymbol{p}}$ is reflexive, there exists $\hat{\boldsymbol{u}} \in \ell_{\boldsymbol{p}}$ and a subsequence $\{\boldsymbol{u}^{(n_j)}\}$ of $\{\boldsymbol{u}^{(n)}\}_{n\in\mathbb{N}}$ such that

$$\boldsymbol{u}^{(n_j)} \to \hat{\boldsymbol{u}} \text{ (weakly) as } j \to \infty.$$

The proof is complete. □

12.5 Closing Remarks

The motivation of this work is to investigated the long term dynamics of lattice dynamical systems with state dependent and nonlinear diffusion. In particular, we are interested in lattice dynamical systems with a leading operator defined by the discretization of the $p(x)$-Laplacian $\operatorname{div}\left(|\nabla u|^{p(x)-2}\nabla u\right)$ considered in Kloeden & Simsen [9, 10]. Applying the chain rule and the finite difference quotient to the operator $\operatorname{div}\left(|\nabla u|^{p(x)-2}\nabla u\right)$, we obtain a generalized version of the operator Γ defined in (12.3) for variable exponents:

$$(\Gamma_{\boldsymbol{p}} u)_i := |D^- u_i|^{p_i-2}\left[(D^+ p_i)(D^- u_i)\ln|D^- u_i| + (p_i - 1)D^+ D^- u_i\right]. \tag{12.19}$$

The ultimate goal is to investigate long term dynamics of lattice dynamical systems with the leading operator $\Gamma_{\boldsymbol{p}}$, which has never been done in the past, as the coercive properties of $\Gamma_{\boldsymbol{p}}$ may not hold in classical sequence spaces such as ℓ^2 and ℓ^p. However, this is not the main focus of this work. This work is to construct the proper sequence space to obtain desired properties of the leading operator $\Gamma_{\boldsymbol{p}}$ and hence can be viewed as one necessary and crucial step for further studies on lattice systems with nonlinear and state dependent diffusion.

There is very little in the literature dealing directly with the $\ell_{\boldsymbol{p}}$ space, see [12, 13]. A majority of results in this work are established by adaptations from the constant exponent case ℓ^p with $p \geq 1$ constant, to the variable exponent case $\ell_{\boldsymbol{p}}$ and adaptations from the continuous case $L^{p(x)}(\Omega)$ [6, 11] to the discrete case $\ell_{\boldsymbol{p}}$. The adaptations, however, are nontrivial and require skillful calculations.

All results in this work can be considered new in the context of $\ell_{\boldsymbol{p}}$. In particular, the proofs of Theorems 12.2, 12.5, 12.7, 12.8 and 12.9 are new and nontrivial to derive from earlier results. Some of results can be deduced from abstract results for the Orlicz spaces $L^{p(x)}(A, \mu)$ in Diening et al. [5], by choosing $A = \mathbb{Z}$ and μ to be the counting measure. But here we provide the reader with proofs more direct than those in [5].

Acknowledgements This work has been partially supported by the National Science Foundation of China grant number 11571125 (PEK) and FAPEMIG process CEX-PPM-00329-16 (JS).

References

1. Abdallah, A.Y.: Uniform global attractors for first order non-autonomous lattice dynamical systems. Proc. Am. Math. Soc. **138**, 3219–3228 (2010)
2. Bates, P.W., Lisei, H., Lu, K.: Attractors for stochastic lattice dynamical systems. Stochastics Dyn. **6**, 1–21 (2006)
3. Brezis, H.: Analyse Fonctionnelle. Dunod, France (2005)
4. Chaplin, M.: Do we underestimate the importance of water in cell biology? Nat. Rev. Mol. Cell Biol. **7**, 861–866 (2006)
5. Diening, L., Harjulehto, P., Hästö, P., Ruzicka, M.: Lebesgue and Sobolev Spaces with Variable Exponents. Springer Lecture Notes in Mathematics, vol. 2017. Springer, Heidelberg (2011)
6. Fan, X.L., Zhang, Q.H.: Existence of solutions for $p(x)-$Laplacian Dirichlet problems. Nonlinear Anal. **52**, 1843–1852 (2003)
7. Gu, A., Kloeden, P.E.: Asymptotic behavior of a nonautonomous p-Laplacian lattice system. Int. J. Bifurcation Chaos **26**(10), 1650174 (2016)
8. Han, X.: Asymptotic dynamics of stochastic lattice differential equations: a review. Continuous Distrib. Syst. II Stud. Syst. Decis. Control **30**, 121–136 (2015)
9. Kloeden, P.E., Simsen, J.: Pullback attractors for non-autonomous evolution equations with spatially variable exponents. Commun. Pure Appl. Anal. **13**(6), 2543–2557 (2014)
10. Kloeden, P.E., Simsen, J.: Attractors of asymptotically autonomous quasi-linear parabolic equation with spatially variable exponents. J. Math. Anal. Appl. **425**, 911–918 (2015)
11. Kovácik, O., Rálosník, J.: On spaces $L^{p(x)}(\Omega)$ and $W_0^{1,p(x)}(\Omega)$. Czechoslovak Math. J. **41**(116), 592–618 (1991)
12. Nekvinda, A.: Equivalence of ℓ^{p_n} norms and shife operators. Math. Ineq. Appl. **5**, 711–723 (2002)
13. Nekvinda, A.: Embeddings between discrete weighted Lebesgue spaces with variable exponents. Math. Ineq. Appl. **10**, 165–172 (2007)
14. Orlicz, W.: Über konjugierte Exponentenfolgen. Studia Math. **3**, 200–211 (1931)
15. Persson, E., Halle, B.: Cell water dynamics on multiple time scales. Proc. Natl. Acad. Sci. **105** (17), 6266–6271 (2008)
16. Wang, B.: Dynamics of systems on infinite lattices. J. Differ. Equ. **221**, 224–245 (2006)
17. Zhou, S.: Attractors and approximations for lattice dynamical systems. J. Differ. Equ. **200**, 342–368 (2004)

Chapter 13
Attractors for a Random Evolution Equation with Infinite Memory: An Application

María J. Garrido-Atienza, Björn Schmalfuß, and José Valero

Abstract In this paper we study the existence of random pullback attractors for an integro-differential parabolic equation of reaction-diffusion type with both finite and infinite delays and also some kind of randomness.

13.1 Introduction

In our previous paper [4] we studied the existence of a random attractor for a rather general random integro-differential equation of reaction-diffusion type with some memory terms given by a convolution integral having infinite delays and a nonlinear term with a finite delay, and where all the nonlinear forcing functions of the equation depend on a random parameter.

Since uniqueness of the Cauchy problem was not guaranteed, we defined a multivalued random dynamical system and proved that a random global pullback attractor exists under some assumptions on the nonlinear terms of the equations. It is important to mention that the multivalued character of the system makes it difficult to study the measurability properties of both the dynamical system and the pullback attractor. This is why several technical conditions (involving the dependence of the nonlinear functions with respect to the random parameter) were needed to be assumed.

M. J. Garrido-Atienza
Dpto. Ecuaciones Diferenciales y Análisis Numérico, Facultad de Matemáticas, Universidad de Sevilla, Sevilla, Spain
e-mail: mgarrido@us.es

B. Schmalfuß
Institut für Mathematik, Institut für Stochastik, Jena, Germany
e-mail: bjoern.schmalfuss@uni-jena.de

J. Valero (✉)
Centro de Investigación Operativa, Universidad Miguel Hernández de Elche, Elche (Alicante), Spain
e-mail: jvalero@umh.es

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_13

Our aim now consists in applying this abstract result to a more concrete reaction-diffusion equation with delay, in which all the non-linear functions are given explicitly and the random parameter is defined. As the assumptions given in [4] are very hard to verify when the random equation comes from the transformation of a stochastic one, we consider in this paper a particular simple random perturbation for which we are able to check all the conditions.

The paper is organized as follows. In the second section we recall the main theorem from [4]. In the third section we develop a particular application and prove that, by using the abstract results from [4], the solutions of the considered system generate a multivalued random dynamical system possessing a random global pullback attractor. We will end the paper giving a few examples of random fields that can be considered in our application.

13.2 Preliminaries

Let Ω be a Polish space with the metric d_Ω and let $\mathscr{F}$ be the Borel σ-algebra of Ω. A pair (Ω, θ) where $\theta = (\theta_t)_{t\in\mathbb{R}}$ is a flow on Ω, that is,

$$\theta : \mathbb{R} \times \Omega \to \Omega,$$
$$\theta_0 = \mathrm{id}_\Omega, \qquad \theta_{t+\tau} = \theta_t \circ \theta_\tau =: \theta_t\theta_\tau \quad \text{for } t,\ \tau \in \mathbb{R},$$

is called a non-autonomous perturbation.

Let $\mathscr{P} := (\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. On $\mathscr{P}$ we consider a measurable non–autonomous flow θ:

$$\theta : (\mathbb{R} \times \Omega, \mathscr{B}(\mathbb{R}) \otimes \mathscr{F}) \to (\Omega, \mathscr{F}).$$

In addition, $\mathbb{P}$ is supposed to be ergodic with respect to θ, which means that every θ_t-invariant set has measure zero or one for $t \in \mathbb{R}$. Hence $\mathbb{P}$ is invariant with respect to θ_t. The quadruple $(\Omega, \mathscr{F}, \mathbb{P}, \theta)$ is called a metric dynamical system. We denote by $\mathscr{P}^c$ its completion: $\mathscr{P}^c := (\Omega, \bar{\mathscr{F}}^{\mathbb{P}}, \bar{\mathbb{P}})$.

Let $\mathscr{O} \subset \mathbb{R}^N$ be an open bounded set with C^∞-smooth boundary. Let $H = L^2(\mathscr{O})$, $V = H_0^1(\mathscr{O})$, with their norms and scalar products denoted by $\|\cdot\|$, $\|\cdot\|_V$ and by $(\cdot,\cdot)$, $((\cdot,\cdot))$, respectively. We shall use $\langle\cdot,\cdot\rangle$ for the pairing between the spaces V' (the dual space of V) and V, and $\langle\cdot,\cdot\rangle_{q,p}$ for the pairing between $L^p(\mathscr{O})$ and $L^q(\mathscr{O})$, where $\frac{1}{p} + \frac{1}{q} = 1$ with $p \geq 2$.

Let A be a positive symmetric operator on H with compact inverse. The eigenvalues of A, denoted by $0 < \lambda_1 \leq \lambda_2 \leq \cdots \to \infty$, have finite multiplicity. In the sequel, A will be $-\Delta$, where Δ is the Laplacian operator endowed with homogeneous Dirichlet boundary conditions.

We will consider the space $L^2(-\infty, r; V)$, $r \in \mathbb{R}$, of square integrable functions with values in V and with the measure $e^{\lambda_1 s} Leb$, where Leb is the standard Lebesgue

measure. The space $L^2(-\infty, r; V)$ is equipped with the following norm

$$\|\psi\|^2_{L^2(-\infty,r;V)} = \int_{-\infty}^{r} e^{\lambda_1 s} \|\psi(s)\|^2_V \, ds.$$

This Banach space is separable, see e.g. [7]. In what follows, we will denote $L^2_V := L^2(-\infty, 0; V)$. Also, for $h > 0$ we consider the space $C_h := C([-h, 0], H)$ of continuous functions on $[-h, 0]$ with values in H and with the supremum norm. Let us finally consider the space

$$\mathscr{H} = \{\psi \in L^2_V \text{ such that } P_h\psi \in C_h\},$$

where P_h is the restriction operator to the interval $[-h, 0]$. $\mathscr{H}$ is a Banach space endowed with the norm $\|\psi\|_{\mathscr{H}} = \|\psi\|_{L^2_V} + \|\psi\|_{C_h}$. It can be proved easily that $\mathscr{H}$ is separable.

In this section, we recall an abstract result proved in [4] concerning existence of random attractors for the following random delay system

$$\begin{cases} \dfrac{du}{dt} + Au(t) = f(\theta_t\omega, u_t) + h(\theta_t\omega, u_t) - g(\theta_t\omega, u(t)), \text{ for } t \in [0, T], \\ \\ u(t) = \psi(t), \text{ for } t \le 0, \end{cases} \tag{13.1}$$

where $T > 0$ and $u_t(\cdot)$ denotes the element of $\mathscr{H}$ given by $u_t(s) = u(t + s)$, $s \le 0$. The initial condition ψ belongs to $\mathscr{H}$, which means that

$$u(t) = \psi(t), \text{ for } t \in [-h, 0] \text{ but almost everywhere when } s < -h. \tag{13.2}$$

We impose the following assumptions on the operators f, g, h.

First, $g : \Omega \times L^p(\mathscr{O}) \to L^q(\mathscr{O})$, $q = p/(p-1)$, $h : \Omega \times C_h \to H$ and $f : \Omega \times L^2_V \to V'$ are such that the mappings

$$v \mapsto g(\omega, v), (\omega, \xi) \mapsto h(\omega, \xi), \ \zeta \mapsto f(\omega, \zeta) \tag{13.3}$$

are continuous in their respective spaces (ω is fixed for g and f), and for arbitrary fixed $v \in L^p(\mathscr{O})$, $\xi \in C_h$, $\zeta \in L^2_V$ we have that

$$\omega \mapsto g(\omega, v), \omega \mapsto h(\omega, \xi), \omega \mapsto f(\omega, \zeta) \tag{13.4}$$

are measurable. Since L^2_V, C_h, $L^p(\mathscr{O})$, V' and H are separable, the functions

$$(\omega, v) \mapsto g(\omega, v), (\omega, \xi) \mapsto h(\omega, \xi), \ (\omega, \zeta) \mapsto f(\omega, \zeta)$$

are jointly measurable with respect to their arguments, see Castaing and Valadier [5, Chapter 3]. We observe also that since Ω is separable, for the function h condition (13.3) implies (13.4).

Moreover, assume the following inequalities

$$\langle g(\omega, v), v\rangle_{q,p} \geq \eta \|v\|^p_{L^p(\mathscr{O})} - c_1(\omega), \text{ for } v \in L^p(\mathscr{O}), \tag{13.5}$$

$$\|g(\omega, v)\|^q_{L^q(\mathscr{O})} \leq \nu \|v\|^p_{L^p(\mathscr{O})} + c_2(\omega), \text{ for } v \in L^p(\mathscr{O}), \tag{13.6}$$

$$\|h(\omega, \xi)\| \leq c_3(\omega) + c_4(\omega) \|\xi\|_{C_h}, \text{ for } \xi \in C_h, \tag{13.7}$$

$$2\|f(\omega, \zeta)\|^2_{V'} \leq c_5(\omega) + K\|\zeta\|^2_{L^2_V}, \text{ for } \zeta \in L^2_V, \tag{13.8}$$

where $\eta, \nu, K > 0$ and $c_i : \Omega \to \mathbb{R}^+$ are measurable with respect to $\mathscr{F}$. Also, the functions $t \mapsto c_1(\theta_t\omega)$, $t \mapsto c_3^2(\theta_t\omega)$ are assumed to be integrable on any finite interval and subexponentially growing (that is, tempered), whereas $t \mapsto c_2(\theta_t\omega)$, $t \mapsto c_5(\theta_t\omega)$ are integrable on any finite interval. For c_4 we suppose that $\mathbb{E}(c_4^2) < \infty$ (and then $t \mapsto c_4^2(\theta_t\omega)$ is integrable on any finite interval by the ergodic theorem), so that

$$\lim_{t\to\pm\infty} \frac{1}{t} \int_0^t c_4^2(\theta_s\omega)\, ds = \mathbb{E}\left(c_4^2\right)$$

on a $(\theta_t)_{t\in\mathbb{R}}$-invariant set of full measure.

On the other hand, and as a consequence of the Young inequality, we have

$$\|u\|^p_{L^p(\mathscr{O})} \geq \mu \|u\|^2 - C_\mu, \tag{13.9}$$

where $\mu > 0$ can be chosen arbitrarily and C_μ denotes a positive constant. In particular, we take μ such that

$$\mathbb{E}(c_4^2) < \frac{\eta\mu\lambda_1}{4e^{\lambda_1 h}}. \tag{13.10}$$

Note that from (13.8) it is straightforward to derive that

$$2\int_r^t e^{\lambda_1 s}\|f(\theta_s\omega, u_s)\|^2_{V'} ds \leq \int_r^t e^{\lambda_1 s} c_5(\theta_s\omega) ds + K(t-r)\int_{-\infty}^t e^{\lambda_1 s}\|u(s)\|^2_V ds. \tag{13.11}$$

For f and $t > 0$ we also assume the following inequalities

$$2\int_0^t e^{\lambda_1 s}\|f(\theta_s\omega, u_s)\|^2_{V'} ds \leq \int_0^t e^{\lambda_1 s} c_6(\theta_s\omega) ds + \frac{d}{2}\int_{-\infty}^t e^{\lambda_1 s}\|u(s)\|^2_V ds, \tag{13.12}$$

$$2\int_0^t e^{\lambda_1 s}\|f(\theta_s\overline{\omega},u_s)-f(\theta_s\omega,v_s)\|_{V'}^2 ds \le \int_0^t e^{\lambda_1 s}c_7(\theta_s\overline{\omega},\theta_s\omega)ds + \frac{b}{2}\int_{-\infty}^t e^{\lambda_1 s}\|u(s)-v(s)\|_V^2 ds, \tag{13.13}$$

for $u, v \in L^2(-\infty, t; V)$, $\overline{\omega}, \omega \in \Omega$, where $d, b < 1$, and $c_6 : \Omega \to \mathbb{R}^+$, $c_7 : \Omega \times \Omega \to \mathbb{R}^+$ are measurable with respect to $\mathscr{F}$ and $\mathscr{F} \otimes \mathscr{F}$, respectively, and the functions $t \mapsto c_6(\theta_t\omega)$, $t \mapsto c_7(\theta_t\overline{\omega}, \theta_t\omega)$ are integrable on any finite interval. Also, $t \mapsto c_6(\theta_t\omega)$ is subexponentially growing. Moreover, for any $t \in \mathbb{R}$, $\varepsilon > 0$ there exists $\delta > 0$ such that if $d_\Omega(\overline{\omega}, \omega) < \delta$, then

$$\int_0^t e^{\lambda_1 s}c_7(\theta_s\overline{\omega},\theta_s\omega)ds < \varepsilon. \tag{13.14}$$

Also, we assume that the maps

$$\omega \mapsto \int_0^t c_i(\theta_s\omega)ds,\ i = 1, 3, 4, 5, \tag{13.15}$$

are continuous for any fixed $t \in \mathbb{R}$, and that for every $\omega_0 \in \Omega$, $t \in \mathbb{R}$ there exists a neighborhood $\mathscr{U}$ of ω_0 and a constant $C(t, \omega_0)$ such that

$$\int_0^t c_3^2(\theta_s\omega)ds \le C,\ \int_0^t c_i(\theta_s\omega)ds \le C \text{ for any } \omega \in \mathscr{U}, \tag{13.16}$$

where $i = 1, 2, 5, 6$. It is obvious that for $i = 1, 5$ the property (13.15) implies condition (13.16). Also, suppose that for every $\omega_0 \in \Omega$, $t_0 \in \mathbb{R}$ there exists a neighbourhood $\mathscr{U}$ of ω_0 and a constant $D(t_0, \omega_0)$ such that

$$\int_t^{t+t_0} c_4^2(\theta_s\omega)ds \le D, \quad \text{for any } \omega \in \mathscr{U}, t \in \mathbb{R}, \tag{13.17}$$

and, moreover, assume that for any $\omega_0 \in \Omega$ and $k > 0$ there exist a neighborhood $\mathscr{U}$ of ω_0 and $C(\omega_0, k)$, $\overline{t}(\omega_0, k) > 0$ such that

$$c_i(\theta_t\omega) \le C(\omega_0, k)e^{k|t|},\ i = 1, 6, \tag{13.18}$$

$$c_3^2(\theta_t\omega) \le C(\omega_0, k)e^{k|t|}, \tag{13.19}$$

for all $\omega \in \mathscr{U}$, $|t| \geq \bar{t}$. Additionally, assume that for any $\omega_0 \in \Omega$ there exists a neighborhood $\mathscr{U}$ such that for any $\gamma > 0$ there exists $r_0(\gamma, \omega_0) < 0$ such that

$$(-\mathbb{E}(c_4^2) + \gamma)r \leq \int_r^0 c_4^2(\theta_s\omega)ds \leq (-\mathbb{E}(c_4^2) - \gamma)r, \tag{13.20}$$

for all $\omega \in \mathscr{U}$ and $r \leq r_0$.

Finally, let us assume that if $\omega^n \to \omega$, $u^n \to u$ in $L^2(0, T; H)$, $u^n \to u$ weakly in $L^p(0, T; L^p(\mathscr{O}))$ and $u^n \to u$ weakly in $L^2(-\infty, T; V)$ then

$$f(\theta.\omega^n, u^n_\cdot) \to f(\theta.\omega, u.) \text{ weakly in } L^2\left(0, T; V'\right), \tag{13.21}$$

$$g\left(\theta.\omega^n, u^n(\cdot)\right) \to g\left(\theta.\omega, u(\cdot)\right) \text{ weakly in } L^q\left(0, T; L^q\left(\mathscr{O}\right)\right), \tag{13.22}$$

and

$$\begin{aligned} &\liminf_{n\to\infty} \int_0^T e^{-\lambda_1(T-s)} \left\langle g\left(\theta_s\omega^n, u^n(s)\right), u^n(s)\right\rangle_{q,p} ds \\ &\geq \int_0^T e^{-\lambda_1(T-s)} \left\langle g\left(\theta_s\omega, u(s)\right), u(s)\right\rangle_{q,p} ds. \end{aligned} \tag{13.23}$$

Given $T > 0$, the function $u(t) = u(t, \omega, \psi) \in L^2(-\infty, T; V) \cap C([-h, T], H) \cap L^p(0, T; L^p(\mathscr{O}))$ is called a weak solution of (13.1) on $(0, T)$ with initial data $\psi \in \mathscr{H}$, if for arbitrary $v \in V \cap L^p(\mathscr{O})$,

$$\frac{d}{dt}(u, v) + < Au, v > = \langle f(\theta_t\omega, u_t), v\rangle + (h(\theta_t\omega, u_t), v) - \left\langle g\left(\theta_t\omega, u(t)\right), v\right\rangle_{q,p}, \tag{13.24}$$

and u agrees with the initial condition ψ according to (13.2).

It is known [4, Theorem 4.5] that for any $\psi \in \mathscr{H}$ there exists at least one weak solution to problem (13.1), although in general it could be non-unique. Every weak solution of (13.1) can be extended to a globally defined one (i.e. for all $t \in \mathbb{R}^+$) simply by concatenating solutions. Let then $\mathscr{S}(\psi, \omega)$ be the set of all globally defined solutions to (13.1) corresponding to $\psi \in \mathscr{H}$ and $\omega \in \Omega$. Let $P(\mathscr{H})$ be the set of all non-empty subsets of $\mathscr{H}$. We define the multivalued map $\Phi : \mathbb{R}^+ \times \Omega \times \mathscr{H} \to P(\mathscr{H})$ as follows

$$\Phi(t, \omega, \psi) = \{u_t : u(\cdot, \omega, \psi) \in \mathscr{S}(\psi, \omega)\}.$$

It is proved in [4, Lemma 4.6 and Corollary 1] that Φ satisfies the strict cocycle property:

$$\Phi(t+\tau,\omega,\psi)=\Phi(t,\theta_\tau\omega,\Phi(\tau,\omega,\psi)),\quad \text{for any } t,\tau\in\mathbb{R}^+,\psi\in\mathscr{H},\omega\in\Omega,$$

and also that Φ has closed (in fact, compact) values. Therefore, Φ is a strict multivalued non-autonomous dynamical system.

Moreover, the map $(t,\omega,\psi)\mapsto\Phi(t,\omega,\psi)$ is $\mathscr{B}(\mathbb{R}^+)\otimes\mathscr{F}\otimes\mathscr{B}(\mathscr{H})$ measurable, see [4, Lemma 6.1], that is, the set

$$\{(t,\omega,x):\Phi(t,\omega,x)\cap O\neq\emptyset\}\in\mathscr{B}(\mathbb{R}^+)\otimes\mathscr{F}\otimes\mathscr{B}(\mathscr{H})$$

for any open set O of $\mathscr{H}$. Hence, Φ is a multivalued random dynamical system (MRDS).

Denote by $P_f(\mathscr{H})$ the set of all non-empty closed subsets of $\mathscr{H}$.

Let us consider the system $\mathscr{D}$ given by the multivalued mappings $D:\omega\to D(\omega)\in P_f(\mathscr{H})$ with $D(\omega)\subset B_{\mathscr{H}}(0,\varrho(\omega))$, the closed ball with center zero and radius $\varrho(\omega)$, which is supposed to have a subexponential growth (ρ is tempered), i.e.

$$\lim_{t\to\pm\infty}\frac{\log^+\varrho(\theta_t\omega)}{t}=0\quad\text{for }\omega\in\Omega.$$

$\mathscr{D}$ is called the family of subexponentially growing (or tempered) multi-functions.

Definition 13.1 The family $\mathscr{A}\in\mathscr{D}$ is said to be a global pullback $\mathscr{D}$-attractor for the MRDS Φ if it satisfies:

i) $\mathscr{A}(\omega)$ is compact for any $\omega\in\Omega$.
ii) $\mathscr{A}$ is pullback $\mathscr{D}$-attracting, i.e., for every $D\in\mathscr{D}$,

$$\lim_{t\to+\infty}\operatorname{dist}(\Phi(t,\theta_{-t}\omega,D(\theta_{-t}\omega)),\mathscr{A}(\omega))=0,\ \text{for all }\omega\in\Omega,$$

where $dist(A,B)=\sup_{x\in A}\inf_{y\in B}\|x-y\|_{\mathscr{H}}$ denotes the Hausdorff semi-distance of two non-empty sets A,B.
iii) $\mathscr{A}$ is negatively invariant, that is, $\mathscr{A}(\theta_t\omega)\subset\Phi(t,\omega,\mathscr{A}(\omega))$ for all $\omega\in\Omega$, $t\geq 0$.

$\mathscr{A}$ is said to be a strict global pullback $\mathscr{D}$-attractor if, additionally, $\mathscr{A}(\theta_t\omega)=\Phi(t,\omega,\mathscr{A}(\omega))$ for all $\omega\in\Omega$, $t\geq 0$. If, moreover, $\mathscr{A}$ is a random set with respect to $\mathscr{P}^c$, then $\mathscr{A}$ is called a random global pullback $\mathscr{D}$-attractor.

Theorem 13.1 ([4, Theorem 6.5]) *The MRDS Φ possesses a random global pullback $\mathscr{D}$-attractor $\mathscr{A}$ in $\mathscr{H}$, which is strictly invariant.*

13.3 Application

We consider the following random heat equation in materials with memory:

$$\begin{cases} \dfrac{\partial u}{\partial t} - \Delta u - \int_{-\infty}^{t} \gamma(t-s)(\Delta u(s,x) + \Delta z(\theta_s\omega, x))ds + G(u(t,x) + z(\theta_t\omega, x)) \\ \qquad = L(u(t-h,x) + z(\theta_{t-h}\omega, x)), \\ u(t,x) = 0 \text{ on } \partial\mathscr{O}, t > 0, \\ u(s,x) = u_0(s,x) = \psi(s,x), \text{ for } s \le 0, x \in \mathscr{O}, \end{cases} \tag{13.25}$$

in the bounded open subset $\mathscr{O} \subset \mathbb{R}^N$ with C^∞–smooth boundary $\partial\mathscr{O}$, where $h > 0$, and $z : \Omega \times \mathscr{O} \to \mathbb{R}$, where the probability space $\mathbb{P}$ and the map $\theta : (\mathbb{R} \times \Omega, \mathscr{B}(\mathbb{R}) \otimes \mathscr{F}) \to (\Omega, \mathscr{F})$ satisfy the assumptions of Sect. 13.2. In particular, Ω is a Polish space with the metric d_Ω and $\mathscr{F}$ is the Borel σ-algebra of Ω. As before, the operator $A : H^2(\mathscr{O}) \cap H_0^1(\mathscr{O}) \to L^2(\mathscr{O})$ will be $-\Delta$, where Δ is the Laplacian operator endowed with homogeneous Dirichlet boundary conditions.

The particular form of the random perturbation in Eq. (13.25) appears naturally when we make a suitable change of variable in the following stochastic heat equation in materials with memory [2]:

$$\begin{cases} \dfrac{\partial v}{\partial t} - \Delta v + \int_{-\infty}^{t} \gamma(t-s)\, \Delta v(s,x)\, ds + G(v(t,x)) = L(v(t-h,x)) + \dfrac{dW}{dt}, \\ v(t,x) = 0 \text{ on } \partial\mathscr{O}, t > 0, \\ v(s,x) = v_0(s,x), \text{ for } s \le 0, x \in \mathscr{O}, \end{cases} \tag{13.26}$$

where $W(t)$, $t \in \mathbb{R}$, is a two-sided Wiener process in $L^2(\mathscr{O})$. Indeed, we consider the metric dynamical system $(\Omega, \mathscr{F}, \mathbb{P}, \theta)$ generated by W (see [3] for more details) and the following linear stochastic differential equation

$$dz^* = Az^*\, dt + dW. \tag{13.27}$$

It is known that this equation has a unique stationary solution which we denote by $z : \Omega \to L^2(\mathscr{O})$. From z we can define the well-known stationary Ornstein-Uhlenbeck process $\bar{z} : \mathbb{R} \times \Omega \to L^2(\mathscr{O})$ given by $\bar{z}(t,\omega) := z(\theta_t\omega)$. Let Q be the covariance operator of $W(t)$. Assuming that $tr_{L^2(\mathscr{O})}(QA^\epsilon) < \infty$, for some $\epsilon > 0$, it is known that $\bar{z}(t,\omega) \in H_0^1(\mathscr{O})$ [6, Proposition 3.1]. Then by means of the change

of variable $u = v - \overline{z}$ (13.26) is transformed into

$$\begin{cases} \dfrac{\partial u}{\partial t} - \nu\Delta u + \int_{-\infty}^{t} \gamma(t-s)\left(\Delta u(s,x) + \Delta\overline{z}(s,\omega)(x)\right) ds + G\left(u(t,x) + \overline{z}(t,\omega)(x)\right) \\ = L\left(u(t-h,x) + \overline{z}(t-h,\omega)(x)\right), \\ u(t,x) = 0 \text{ on } \partial\mathscr{O},\ t > 0, \\ u(s,x) = u_0(s,x) = v_0(s,x) - \overline{z}(s,x), \text{ for } s \le 0,\ x \in \mathscr{O}, \end{cases} \tag{13.28}$$

which has the form given in (13.25).

We assume the following conditions:

$(H1)$ $G \in C(\mathbb{R})$ and

$$G(v)v \ge \alpha_0 |v|^p - \alpha_1, \tag{13.29}$$

$$|G(v)| \le \gamma_1 |v|^{p-1} + \gamma_2, \tag{13.30}$$

where $\alpha_i, \gamma_i > 0$ and $p \ge 2$.

$(H2)$ $L \in C(\mathbb{R})$ and

$$|L(v)| \le \gamma_3 |v| + \gamma_4, \tag{13.31}$$

where $\gamma_i > 0$.

$(H3)$ $\gamma \in C(\mathbb{R}^+)$ and

$$0 \le \gamma(s) \le \gamma_5 e^{-\delta s}, \forall s \ge 0, \tag{13.32}$$

where $\gamma_5 > 0$ and $\delta > 0$ satisfy

$$\delta > \lambda_1, \ \frac{8\gamma_5^2}{\delta(\delta - \lambda_1)} < 1. \tag{13.33}$$

$(H4)$ The map z satisfies the following assumptions:

1. The maps

$$\omega \mapsto \int_{\mathscr{O}} |z(\omega, x)|^r dx = \|z(\omega)\|_{L^r}^r,$$

$$\omega \mapsto \int_{-\infty}^{0} \gamma(-r) \|z(\theta_r\omega)\|_V^2 dr,$$

$$(\overline{\omega}, \omega) \mapsto \int_{-\infty}^{0} \gamma(-r) \|z(\theta_r\overline{\omega}) - z(\theta_r\omega)\|_V^2 dr$$

are measurable, where $r = 1, 2, p$.

2. The maps

$$t \mapsto \int_{\mathscr{O}} |z(\theta_t\omega, x)|^r \, dx = \|z(\theta_t\omega)\|_{L^r}^r,$$
$$t \mapsto \int_{-\infty}^{0} \gamma(-r) \, \|z(\theta_{r+t}\omega)\|_V^2 \, dr$$

are integrable on any finite interval of $\mathbb{R}$ for any $\omega \in \Omega$, where $r = 1, 2, p$.

3. The maps

$$\omega \mapsto \int_0^t \int_{\mathscr{O}} |z(\theta_s\omega, x)|^r \, dx ds = \int_0^t \|z(\theta_s\omega)\|_{L^r}^r ds,$$
$$\omega \mapsto \int_0^t \left(\int_{\mathscr{O}} |z(\theta_s\omega, x)|^2 \, dx \right)^{\frac{1}{2}} ds = \int_0^t \|z(\theta_s\omega)\| ds,$$
$$\omega \mapsto \int_0^t \int_{-\infty}^{0} \gamma(-r) \, \|z(\theta_{r+s}\omega)\|_V^2 \, dr ds$$

are continuous for any $t \in \mathbb{R}$, where $r = 1, 2, p$.

4. For any $\omega_0 \in \Omega$ and $k > 0$ there exist a neighborhood $\mathscr{U}$ of ω_0 and $C(\omega_0, k)$, $\overline{t}(\omega_0, k) > 0$ such that

$$\|z(\theta_t\omega)\|_{L^r}^r \leq C(\omega_0, k) e^{k|t|},$$
$$\int_{-\infty}^{0} \gamma(-r) \, \|z(\theta_{r+t}\omega)\|_V^2 \, dr \leq C(\omega_0, k) e^{k|t|},$$

for all $\omega \in \mathscr{U}$, $|t| \geq \overline{t}$, where $r = 1, 2, p$.

5. The maps $f_z : \Omega \to L^2(0, T, L^2_\gamma(-\infty, 0; V))$, $g_z : \Omega \to L^p((0, T), L^p(\mathscr{O}))$, $h_z : \Omega \to L^p(\mathscr{O})$, $w_z : \Omega \to L^2_\gamma(-\infty, 0; V)$ are continuous, where

$$\begin{aligned}
f_z(\omega) &= \left\{ \begin{array}{c} y \in L^2(0, T, L^2_\gamma(-\infty, 0; V)) : y(t, r, x) = z(\theta_t\theta_r\omega, x), \\ \text{for a.a. } t \in (0, T), \ r \in (-\infty, 0), \ x \in \mathscr{O}, \end{array} \right. \\
g_z(\omega) &= \left\{ \begin{array}{c} y \in L^p((0, T) \times \mathscr{O}) : y(t, x) = z(\theta_t\omega, x), \\ \text{for a.a. } t \in (0, T) \ x \in \mathscr{O}, \end{array} \right. \\
h_z(\omega) &= \left\{ y \in L^p(\mathscr{O}) : y(t, x) = z(\omega, x) \text{ for a.a. } x \in \mathscr{O}, \right. \\
w_z(\omega) &= \left\{ \begin{array}{c} y \in L^2_\gamma(-\infty, 0; V) : y(r, x) = z(\theta_r\omega, x), \\ \text{for a.a. } r \in (-\infty, 0), \ x \in \mathscr{O}. \end{array} \right.
\end{aligned} \tag{13.34}$$

Here, $L^2_\gamma(-\infty, 0; V)$ is the space of square integrable functions $u(\cdot)$ with values in the space V and with the measure $\gamma(r)Leb$, equipped with the norm

$$\|u\|_{L^2_\gamma(-\infty,0;V)} := \int_{-\infty}^{0} \gamma(-r)\,\|u(r)\|_V^2\,dr.$$

Let us consider some sufficient conditions on z ensuring that $(H4)$ holds.

Let us assume that the mappings $t \mapsto \theta_t\omega$ and $\omega \mapsto \theta_t\omega$ are continuous. This is the situation when choosing for Ω the Fréchet space $C_0(\mathbb{R}, H)$ (the space of continuous functions from $\mathbb{R}$ into H that are zero at zero) and θ is the Wiener shift

$$\theta_t\omega(\cdot) = \omega(\cdot + t) - \omega(t). \tag{13.35}$$

In this case

$$(t, \omega) \mapsto \theta_t\omega$$

is continuous, see Arnold [1, Appendix A]. We assume that the mapping $z(\omega, x) \in \mathbb{R}$ is jointly measurable with respect to both arguments. In particular, we assume that

$$\Omega \ni \omega \mapsto z(\omega) \in V \cap L^p(\mathcal{O})$$

is continuous and bounded. Note that if we assume $1/2 - 1/N \le 1/p$, then the continuous embedding $V \subset L^p(\mathcal{O})$ is ensured, therefore the previous assumption could be reduced to assume the continuity and the boundedness of the mapping

$$\Omega \ni \omega \mapsto z(\omega) \in V.$$

At the end of this section, we will give three examples of random fields fulfilling the above assumptions.

Let us check now that the assumptions (H4.1)–(H4.5) hold true. Since $p \ge 2$ we have the measurability of the first mapping in (H4.1) by the continuous embedding $L^p(\mathcal{O}) \subset L^2(\mathcal{O}) \subset L^1(\mathcal{O})$. By the measurability of $(t, \omega) \mapsto \theta_t\omega$, the measurability of the second and third expressions of (H4.1) follows by Tonelli's theorem.

By the continuity of

$$\begin{aligned} &t \mapsto \theta_t\omega, \\ &\omega \mapsto \|z(\omega)\|_{L^q(\mathcal{O})} \quad \text{for } q = 1,\ 2,\ p, \\ &\omega \mapsto \|z(\omega)\|_V, \end{aligned}$$

and the boundedness of $\|z(\omega)\|^p_{L^p(\mathscr{O})}$ and $\|z(\theta_{r+t}\omega)\|^2_V$, the mappings in (H4.2) are integrable on any bounded interval. In particular, since $\|z(\theta_{r+t}\omega)\|_V$ is bounded in r and t, the last expression of (H4.2) is bounded in t.

The majorant theorem applies to obtain the continuity of the expressions in (H4.3). In addition, (H4.4) follows trivially with $k = 0$, $C(\omega_0, k) = C$ and $\mathscr{U} = \Omega$.

To deal with first expression f_z of (H4.5) we take a sequence $(\omega_n)_{n\in\mathbb{N}}$ converging in the metric of Ω to ω_0. The convergence

$$\int_0^T \int_{-\infty}^0 \gamma(-r)\|z(\theta_{r+t}\omega_n) - z(\theta_{r+t}\omega_0)\|^2_V dr dt \to 0$$

follows by the convergence of the inner integral by Lebesgue's theorem for any $t \in [0, T]$. However, the inner integrals are uniformly bounded with respect to $t \in [0, T]$. Applying the majorant theorem again gives the desired convergence. The continuity of the expressions g_z, h_z and w_z can be studied in a similar manner. By the considerations above we obtain the desired continuity.

We define the maps $g : \Omega \times L^p(\mathscr{O}) \to L^q(\mathscr{O}), q = p/(p-1)$, $h : \Omega \times C_h \to H$ and $f : \Omega \times L^2_V \to V'$ in the following way:

$$\begin{aligned} g(\omega, v) &= G(v(\cdot) + z(\omega, \cdot)), \\ h(\omega, \xi) &= L(\xi(-h)(\cdot) + z(\theta_{-h}\omega, \cdot)), \\ f(\omega, \psi) &= -\int_{-\infty}^0 \gamma(-r)(A\psi(r) + Az(\theta_r\omega))dr. \end{aligned}$$

We check now the properties of the maps g, h, f.

Lemma 13.1 *Conditions (13.3)–(13.23) are satisfied.*

Proof First, we check conditions (13.5)–(13.8). For any $v \in L^p(\mathscr{O})$ from (13.29)–(13.30) in (H1) and Hölder's inequality we get

$$\begin{aligned} \langle g(\omega, v), v\rangle_{q,p} &= \int_{\mathscr{O}} G(v(x) + z(\omega, x))v(x)dx \\ &= \int_{\mathscr{O}} G(v(x) + z(\omega, x))(v(x) + z(\omega, x))dx \\ &\quad - \int_{\mathscr{O}} G(v(x) + z(\omega, x))z(\omega, x)dx \\ &\geq \alpha_0 \int_{\mathscr{O}} |v(x) + z(\omega, x)|^p dx - \alpha_1 |\mathscr{O}| - \gamma_2 \int_{\mathscr{O}} |z(\omega, x)| dx \end{aligned}$$

$$
\begin{aligned}
&- \gamma_1 \int_{\mathscr{O}} |z(\omega, x)| \ |v(x) + z(\omega, x)|^{p-1} \, dx \\
&\geq \frac{\alpha_0}{2} \int_{\mathscr{O}} |v(x) + z(\omega, x)|^p \, dx - \alpha_1 |\mathscr{O}| - \gamma_2 \int_{\mathscr{O}} |z(\omega, x)| \, dx \\
&- K_1 \int_{\mathscr{O}} |z(\omega, x)|^p \, dx.
\end{aligned}
$$

Using

$$
|v|^p = |v + z - z|^p \leq 2^{p-1} \left(|v + z|^p + |z|^p \right)
$$

we obtain

$$
\begin{aligned}
\langle g(\omega, v), v \rangle_{q,p} &\geq 2^{-p} \alpha_0 \int_{\mathscr{O}} |v(x)|^p \, dx - \alpha_1 |\mathscr{O}| \\
&\quad - \gamma_2 \int_{\mathscr{O}} |z(\omega, x)| \, dx - K_2 \int_{\mathscr{O}} |z(\omega, x)|^p \, dx \\
&= \eta \, \|v\|^p_{L^p(\mathscr{O})} - c_1(\omega),
\end{aligned}
$$

where $c_1(\omega) = \alpha_1 |\mathscr{O}| + \gamma_2 \int_{\mathscr{O}} |z(\omega, x)| \, dx + K_2 \int_{\mathscr{O}} |z(\omega, x)|^p \, dx$, $\eta = 2^{-p} \alpha_0$.

Furthermore,

$$
\begin{aligned}
\|g(\omega, v)\|^q_{L^q(\mathscr{O})} &= \int_{\mathscr{O}} |G(v(x) + z(\omega, x))|^q \, dx \\
&\leq \int_{\mathscr{O}} \left(\gamma_2 + \gamma_1 |v(x) + z(\omega, x)|^{p-1} \right)^q dx \\
&\leq 2^{q-1} \left(\gamma_2^q |\mathscr{O}| + 2^{p-1} \gamma_1^q \int_{\mathscr{O}} \left(|v(x)|^p + |z(\omega, x)|^p \right) dx \right) \\
&= \nu \, \|v\|^p_{L^p(\mathscr{O})} + c_2(\omega),
\end{aligned}
$$

where $c_2(\omega) = 2^{q-1} (\gamma_2^q |\mathscr{O}| + 2^{p-1} \gamma_1^q \int_{\mathscr{O}} |z(\omega, x)|^p \, dx)$, $\nu = 2^q 2^{p-1} \gamma_1^q$.

Also, on account of (13.31), for $\xi \in C_h$,

$$
\begin{aligned}
&\|h(\omega, \xi)\|^2 \\
&= \int_{\mathscr{O}} (L(\xi(-h)(x) + z(\theta_{-h}\omega, x)))^2 dx \\
&\leq 2 \int_{\mathscr{O}} (\gamma_3^2 |\xi(-h)(x) + z(\omega, x)|^2 + \gamma_4^2) dx \\
&\leq 4\gamma_3^2 \int_{\mathscr{O}} |\xi(-h)(x)|^2 \, dx + 2\gamma_4^2 |\mathscr{O}| + 4\gamma_3^2 \int_{\mathscr{O}} |z(\omega, x)|^2 \, dx.
\end{aligned}
$$

Hence

$$\|h(\omega,\xi)\| \le c_3(\omega) + c_4(\omega)\,\|\xi\|_{C_h},$$

where $c_3(\omega) = 2^{\frac{1}{2}}\gamma_4\,|\mathscr{O}|^{\frac{1}{2}} + 2\gamma_3\big(\int_{\mathscr{O}} |z(\omega,x)|^2\,dx\big)^{\frac{1}{2}}$, $c_4(\omega) = 2\gamma_3$.

For $\zeta \in L^2_V$ and $v \in V$ we have

$$|\langle f(\omega,\zeta), v\rangle| \le \int_{-\infty}^{0} \gamma(-r)(\|\zeta(r)\|_V + \|z(\theta_r\omega)\|_V)\,\|v\|_V\,dr,$$

thus

$$\begin{aligned}
2\|f(\omega,\zeta)\|^2_{V'} &\le 2\left(\int_{-\infty}^{0} \gamma(-r)(\|\zeta(r)\|_V + \|z(\theta_r\omega)\|_V)dr\right)^2 \\
&\le 4\left(\int_{-\infty}^{0} \gamma_5 e^{\delta r}\,\|\zeta(r)\|_V\,dr\right)^2 + 4\left(\int_{-\infty}^{0} \gamma(-r)\,\|z(\theta_r\omega)\|_V\,dr\right)^2 \\
&\le 4\|\zeta\|^2_{L^2_V}\int_{-\infty}^{0} \gamma_5^2 e^{(2\delta-\lambda_1)r}dr + \frac{4\gamma_5}{\delta}\int_{-\infty}^{0} \gamma(-r)\,\|z(\theta_r\omega)\|^2_V\,dr \\
&= K\|\zeta\|_{L^2_V} + c_5(\omega),
\end{aligned}$$

where $K = \frac{4\gamma_5^2}{2\delta-\lambda_1}$ and $c_5(\omega) = \frac{4\gamma_5}{\delta}\int_{-\infty}^{0} \gamma(-r)\,\|z(\theta_r\omega)\|^2_V\,dr$. Note that above we have used that $\delta > \frac{\lambda_1}{2}$ and (13.32).

It follows from $(H4)$ that the maps $c_i : \Omega \to \mathbb{R}^+$, $i = 1,2,3,5$ are measurable and that $t \mapsto c_i(\theta_t\omega)$, $i = 1,2$, $t \mapsto c_3^2(\theta_t\omega)$ are integrable on any finite interval and tempered. Also, $t \mapsto c_5(\theta_t\omega)$ is integrable on any finite interval. Finally, as $c_4(\omega)$ is constant, we have

$$c_4^2(\omega) = \mathbb{E}(c_4^2) = 4\gamma_3^2 < +\infty.$$

Therefore, conditions (13.5)–(13.8) hold.

Next, we check conditions (13.12)–(13.16).

First, for $u \in L^2(-\infty,t;V)$, $\omega \in \Omega$, $t > 0$, using (13.32) we get

$$\begin{aligned}
&2\int_0^t e^{\lambda_1 s}\|f(\theta_s\omega,u_s)\|^2_{V'}ds \\
&\le 2\int_0^t e^{\lambda_1 s}\left(\int_{-\infty}^{0} \gamma(-r)(\|u_s(r)\|_V + \|z(\theta_{r+s}\omega)\|_V)dr\right)^2 ds \\
&\le 4\int_0^t e^{\lambda_1 s}\left(\int_{-\infty}^{s} \gamma(s-r)\,\|u(r)\|_V\,dr\right)^2 ds
\end{aligned}$$

$$
\begin{aligned}
&+4\int_0^t e^{\lambda_1 s}\left(\int_{-\infty}^0 \gamma(-r)\,\|z(\theta_{r+s}\omega)\|_V\,dr\right)^2 ds\\
&\le 4\int_0^t e^{\lambda_1 s}\left(\int_{-\infty}^s \gamma_5 e^{-\delta(s-r)}\,\|u(r)\|_V\,dr\right)^2 ds\\
&+\frac{4\gamma_5}{\delta}\int_0^t e^{\lambda_1 s}\left(\int_{-\infty}^0 \gamma(-r)\,\|z\,(\theta_{r+s}\omega)\|_V^2\,dr\right) ds.
\end{aligned}
$$

The first term is estimated as follows:

$$
\begin{aligned}
&4\int_0^t e^{\lambda_1 s}\left(\int_{-\infty}^s \gamma_5 e^{-\delta(s-r)}\,\|u(r)\|_V\,dr\right)^2 ds\\
&\le 4\gamma_5^2\int_0^t e^{\lambda_1 s}\int_{-\infty}^s e^{-\delta(s-r)}dr\int_{-\infty}^s e^{-\delta(s-r)}\,\|u(r)\|_V^2\,drds\\
&\le \frac{4\gamma_5^2}{\delta}\int_0^t e^{\lambda_1 s}\int_{-\infty}^s e^{-\delta(s-r)}\,\|u(r)\|_V^2\,drds\\
&= \frac{4\gamma_5^2}{\delta}\int_{-\infty}^0\int_0^t e^{\lambda_1 s}e^{-\delta(s-r)}\,\|u(r)\|_V^2\,dsdr\\
&+\frac{4\gamma_5^2}{\delta}\int_0^t\int_r^t e^{\lambda_1 s}e^{-\delta(s-r)}\,\|u(r)\|_V^2\,dsdr\\
&\le \frac{4\gamma_5^2}{\delta(\delta-\lambda_1)}\left(\int_{-\infty}^0 e^{\delta r}\,\|u(r)\|_V^2\,dr+\int_0^t e^{\delta r}e^{(\lambda_1-\delta)r}\,\|u(r)\|_V^2\,dr\right)\\
&\le \frac{4\gamma_5^2}{\delta(\delta-\lambda_1)}\int_{-\infty}^t e^{\lambda_1 r}\,\|u(r)\|_V^2\,dr.
\end{aligned}
$$

Hence,

$$
2\int_0^t e^{\lambda_1 s}\|f(\theta_s\omega,u_s)\|_{V'}^2 ds\le\frac{d}{2}\int_{-\infty}^t e^{\lambda_1 s}\,\|u(s)\|_V^2\,ds+\int_0^t e^{\lambda_1 s}c_6(\theta_s\omega)ds,
$$

where $d=\frac{8\gamma_5^2}{\delta(\delta-\lambda_1)}$, $c_6(\omega)=\frac{4\gamma_5^2}{\delta}\int_{-\infty}^0\gamma(-r)\,\|z\,(\theta_r\omega)\|_V^2\,dr$. In view of (13.33) we have $d<1$. Also, $(H4)$ implies that $c_6:\Omega\to\mathbb{R}^+$ is measurable and that $t\mapsto c_6(\theta_t\omega)$ is integrable on any finite interval and tempered. Thus, (13.12) holds.

For $u, v \in L^2(-\infty, t; V)$, $\omega, \overline{\omega} \in \Omega$ and $t > 0$, in the same way as before we obtain

$$
\begin{aligned}
&2\int_0^t e^{\lambda_1 s}\|f(\theta_s\overline{\omega}, u_s) - f(\theta_s\omega, v_s)\|_{V'}^2 ds \\
&\leq 4\int_0^t e^{\lambda_1 s}\left(\int_{-\infty}^s \gamma_5 e^{-\delta(s-r)}\|u(r) - v(r)\|_V\, dr\right)^2 ds \\
&+ 4\int_0^t e^{\lambda_1 s}\left(\int_{-\infty}^0 \gamma(-r)\,\|z(\theta_{r+s}\overline{\omega}) - z(\theta_{r+s}\omega)\|_V\, dr\right)^2 ds \\
&\leq \frac{4\gamma_5^2}{\delta(\delta-\lambda_1)}\int_{-\infty}^t e^{\lambda_1 r}\|u(r) - v(r)\|_V^2\, dr \\
&+ \frac{4\gamma_5}{\delta}\int_0^t e^{\lambda_1 s}\int_{-\infty}^0 \gamma(-r)\,\|z(\theta_{r+s}\overline{\omega}) - z(\theta_{r+s}\omega)\|_V^2\, dr ds \\
&= \frac{b}{2}\int_{-\infty}^t e^{\lambda_1 r}\|u(s) - v(s)\|_V^2\, ds + \int_0^t e^{\lambda_1 s} c_7(\theta_s\overline{\omega}, \theta_s\omega) ds,
\end{aligned}
$$

where $b = d < 1$ and

$$
c_7(\overline{\omega}, \omega) = \frac{4\gamma_5}{\delta}\int_{-\infty}^0 \gamma(-r)\,\|z(\theta_{r+s}\overline{\omega}) - z(\theta_{r+s}\omega)\|_V^2\, dr.
$$

From $(H4)$ we can see that $c_7 : \Omega \to \mathbb{R}^+$ is measurable and that $t \mapsto c_7(\theta_t\omega)$ is integrable on any finite interval. Therefore, we have checked (13.13). Assumption (13.14) follows from $(H4)$ as well.

From the third assumption in $(H4)$ we obtain that (13.15)–(13.16) are true.

From the fourth condition in $(H4)$ it is obvious that (13.18)–(13.19) hold. Moreover, since $c_4(\omega)$ is a constant, it is clear that assumptions (13.17) and (13.20) are true.

Let us consider now condition (13.22). Since $L^2((0,T)\times\mathcal{O})) \cong L^2(0,T;H)$ (see [9, Chapter 7]), $u^n \to u$ in $L^2(0,T;H)$ and the fifth condition in $(H4)$ imply that $u^n(t,x) \to u(t,x)$, $z(\theta_t\omega^n, x) \to z(\theta_t\omega, x)$ for a.a. $(t,x) \in (0,T)\times\mathcal{O}$. Hence, the continuity of G implies that $G(u^n(t,x) + z(\theta_t\omega^n, x)) \to G(u(t,x) + z(\theta_t\omega, x))$ for a.a. (t,x). Condition (13.30), $u^n \to u$ weakly in $L^p(0,T;L^p(\mathcal{O}))$ and the third condition in $(H4)$ imply that

$$
\left\|g\left(\theta.\omega^n, u^n(\cdot)\right)\right\|_{L^q(0,T;L^q(\mathcal{O}))} \leq C_1\int_0^T\int_{\mathcal{O}}\left(1 + \left|u^n(t,x) + z\left(\theta_t\omega^n, x\right)\right|^p\right) dxdt \leq C_2,
$$

and also that $g(\theta.\omega, u(\cdot)) \in L^q(0,T;L^q(\mathcal{O}))$. Hence, a standard result (see e.g. [8]) implies that $g(\theta.\omega, u^n(\cdot)) \to g(\theta.\omega, u(\cdot))$ weakly in $L^q(0,T;L^q(\mathcal{O}))$. Therefore, (13.22) is satisfied.

For condition (13.21) we note that for any $\psi \in L^2(0,T;V)$,

$$\begin{aligned}
&\int_0^T \left\langle f(\theta_t\omega^n, u_t^n), \psi(t)\right\rangle dt \\
&= \int_0^T \left\langle -\int_{-\infty}^0 \gamma(-r)\left(Au^n(t+r) + Az(\theta_{t+r}\omega^n)\right) dr, \psi(t)\right\rangle dt \\
&= -\int_0^T \int_{-\infty}^0 \gamma(-r)\left(\left\langle Au^n(t+r), \psi(t)\right\rangle + \left\langle Az\left(\theta_{t+r}\omega^n\right), \psi(t)\right\rangle\right) drdt.
\end{aligned}$$

For a.a. $t \in (0,T)$ we have $\gamma(-\cdot)\psi(t) \in L_V^2$, so that $u^n(t+\cdot) \to u(t+\cdot)$ weakly in L_V^2 gives

$$\int_{-\infty}^0 \gamma(-r)\left\langle Au^n(t+r), \psi(t)\right\rangle dr \to \int_{-\infty}^0 \gamma(-r)\left\langle Au(t+r), \psi(t)\right\rangle dr.$$

Since u^n is bounded in $L^2(-\infty,T;V)$ and $\delta > \lambda_1$, using (13.32) we have

$$\begin{aligned}
&\int_{-\infty}^0 \gamma(-r)\left\langle Au^n(t+r), \psi(t)\right\rangle dr \\
&\le \int_{-\infty}^0 \gamma(-r)\left\|u^n(t+r)\right\|_V dr \left\|\psi(t)\right\|_V \\
&\le \left(\int_{-\infty}^0 e^{\lambda_1 r}\left\|u^n(t+r)\right\|_V^2 dr\right)^{\frac{1}{2}} \left(\int_{-\infty}^0 e^{-\lambda_1 r}\gamma^2(-r)dr\right)^{\frac{1}{2}} \left\|\psi(t)\right\|_V \\
&\le C\gamma_5 \left\|\psi(t)\right\|_V \left(\int_{-\infty}^0 e^{(2\delta-\lambda_1)r} dr\right)^{\frac{1}{2}} \\
&= C\frac{\gamma_5}{(2\delta-\lambda_1)^{\frac{1}{2}}} \left\|\psi(t)\right\|_V, \text{ for a.a. } t \in (0,T).
\end{aligned}$$

By Lebesgue's theorem we obtain that

$$\int_0^T \int_{-\infty}^0 \gamma(-r)\left\langle Au^n(t+r), \psi(t)\right\rangle drdt \to \int_0^T \int_{-\infty}^0 \gamma(-r)\left\langle Au(t+r), \psi(t)\right\rangle drdt.$$

Finally, by the fifth condition in $(H4)$ the map $\omega \mapsto f_z(\omega)$ defined in (13.34) is continuous, which implies that

$$\int_0^T \int_{-\infty}^0 \gamma(-r)\left\langle Az(\theta_{t+r}\omega^n), \psi(t)\right\rangle drdt \to \int_0^T \int_{-\infty}^0 \gamma(-r)\left\langle Az(\theta_{t+r}\omega), \psi(t)\right\rangle drdt.$$

Thus,

$$\left\langle f(\theta_{\cdot}\omega^n, u^n_{\cdot}), \psi \right\rangle \to \langle f(\theta_{\cdot}\omega, u_{\cdot}), \psi \rangle ,$$

that is, (13.21) holds.

Further, we will check condition (13.23). Since $G(u)$ is continuous and $u^n(t,x) \to u(t,x)$, $z(\theta_t\omega^n, x) \to z(\theta_t\omega, x)$ for a.a. $(t,x) \in (0,T) \times \mathscr{O}$, we have the convergence $G(u^n(t,x) + z(\theta_t\omega^n, x)) \to G(u(t,x) + z(\theta_t\omega, x))$ for a.a. (t,x). It follows from (13.29)–(13.30) and the fifth condition in $(H4)$ that there exists $r(\cdot,\cdot) \in L^1((0,T) \times \mathscr{O})$ such that

$$\begin{aligned}
&G(u^n(t,x) + z(\theta_t\omega^n, x))u^n(t,x) \\
&\geq -\alpha_1 + \alpha_0 \left|u^n(t,x) + z(\theta_t\omega^n, x)\right|^p - G(u^n(t,x) + z(\theta_t\omega, x))z(\theta_t\omega^n, x) \\
&\geq -\alpha_1 + \alpha_0 \left|u^n(t,x) + z(\theta_t\omega^n, x)\right|^p - \gamma_2 \left|z(\theta_t\omega^n, x)\right| \\
&- \gamma_1 \left|u^n(t,x) + z(\theta_t\omega^n, x)\right|^{p-1} \left|z(\theta_t\omega^n, x)\right| \\
&\geq \frac{\alpha_0}{2} \left|u^n(t,x) + z(\theta_t\omega^n, x)\right|^p - R_1 \left|z(\theta_t\omega^n, x)\right|^p - R_2 \\
&\geq -R_1 \left|z(\theta_t\omega^n, x)\right|^p - R_2 \geq r(t,x), \text{ for a.a. } (t,x).
\end{aligned}$$

Then Lebesgue-Fatou's lemma (see [10]) implies

$$\begin{aligned}
&\liminf_{n\to\infty} \int_0^T e^{-\lambda_1(T-s)} (g(\theta_s\omega^n, u^n(s)), u^n(s)) ds \\
&= \liminf_{n\to\infty} \Big(\int_0^T \int_{\mathscr{O}} e^{-\lambda_1(T-s)} G(u^n(t,x) + z(\theta_t\omega^n, x)) u^n(t,x) dx ds \Big) \\
&\geq \int_0^T \int_{\mathscr{O}} e^{-\lambda_1(T-s)} \liminf_{n\to\infty} G(u^n(t,x) + z(\theta_t\omega^n, x)) u^n(t,x) dx ds \\
&= \int_0^T \int_{\mathscr{O}} e^{-\lambda_1(T-s)} G(u(t,x) + z(\theta_t\omega, x)) u(t,x) dx ds \\
&= \int_0^T e^{-\lambda_1(T-s)} (g(\theta_s\omega, u(s)), u(s)) ds,
\end{aligned}$$

so that (13.23) is satisfied.

Finally, we need to verify (13.3)–(13.4). Since all the spaces are separable and metrizable, for this it suffices to prove that the three maps g, h, f are jointly continuous with respect to both variables, see Castaing and Valadier [5, Chapter 3].

Let us consider first the continuity of the map $(\omega, u) \mapsto g(\omega, u)$. Let $\omega^n \to \omega$, $u^n \to u$ in $L^p(\mathscr{O})$. In view of the fifth condition in $(H4)$ the map $z(\omega^n, \cdot)$ converges to $z(\omega, \cdot)$ in $L^p(\mathscr{O})$. Thus, up to a subsequence, the continuity of G implies that

$G(u^n(x) + z(\omega^n, x)) \to G(u(x) + z(\omega, x))$ for a.a. x. Thus, by using (13.30) and Lebesgue's theorem the result follows.

The continuity of $(\omega, u) \mapsto h(\omega, u)$ is proved in a rather similar way by using (13.31) and the continuity of L.

The continuity of $(\omega, \psi) \to f(\omega, \psi)$ follows from the fifth condition in $(H4)$ and

$$\begin{aligned}
&\|f(\omega_1, \psi_1) - f(\omega_2, \psi_2)\|_{V'} \\
&\leq \int_{-\infty}^{0} \gamma(-r)\, \|\psi_1(r) - \psi_2(r)\|_V \, dr + \int_{-\infty}^{0} \gamma(-r)\, \|z(\theta_r\omega_1) - z(\theta_r\omega_2)\|_V \, d\tau \\
&\leq \left(\int_{-\infty}^{0} \gamma_5^2 e^{(2\delta - \lambda_1)r} dr \right)^{\frac{1}{2}} \|\psi_1 - \psi_2\|_{L_V^2} + \frac{\gamma_5}{\delta} \int_{-\infty}^{0} \gamma(-r)\, \|z(\theta_r\omega_1) - z(\theta_r\omega_2)\|_V^2 \, dr \\
&\leq \frac{\gamma_5}{(2\delta - \lambda_1)^{\frac{1}{2}}} \|\psi_1 - \psi_2\|_{L_V^2} + \frac{\gamma_5}{\delta} \int_{-\infty}^{0} \gamma(-r)\, \|z(\theta_r\omega_1) - z(\theta_r\omega_2)\|_V^2 \, dr,
\end{aligned}$$

where we have used (13.32) and $\delta > \frac{\lambda_1}{2}$.

Remark 13.1 In Sect. 13.2, in (13.9) we chose the constant $\mu > 0$ such that inequality (13.10) is satisfied. In this case, since $c_4(\omega) = 2\gamma_3$, μ has to be chosen such that

$$\frac{8\gamma_3 e^{\lambda_1 h}}{\eta\lambda_1} < \mu.$$

Applying Lemma 13.1 and Theorem 13.1 we obtain the existence of the random attractor.

Theorem 13.2 *The solutions of system (13.25) generate a MRDS having a random global pullback $\mathscr{D}$-attractor $\mathscr{A}$, which is strictly invariant.*

We would like to give some examples of the random fields z introduced above.

For simplicity, we assume that $1/2 - 1/N \leq 1/p$, hence V is continuously embedded into $L^p(\mathscr{O})$.

Let Ω be the Fréchet space $C_0(\mathbb{R}, H)$ and θ is the Wiener shift defined by (13.35). We consider a bounded mapping

$$z : \Omega \times \mathscr{O} \to \mathbb{R}$$

such that

$$x \mapsto z(\omega, x) \in C^1(\overline{\mathscr{O}}),$$

and the derivative with respect to its second variable

$$D_2 z : \Omega \times \mathscr{O} \mapsto \mathbb{R}$$

is bounded too. We assume that for any $x \in \mathscr{O}$,

$$\Omega \ni \omega \mapsto z(\omega, x), \ \Omega \ni \omega \mapsto D_2 z(\omega, x)$$

are continuous. By the separability of Ω the mapping

$$(\omega, x) \mapsto z(\omega, x)$$

is measurable with respect to $\mathscr{F} \otimes \mathscr{B}(\mathscr{O})$.

Under these conditions, the map

$$\omega \mapsto z(\omega, \cdot) \in V$$

is continuous and bounded. Indeed, if $\omega_n \to \omega$ for any $x \in \mathscr{O}$ we have that $z(\omega_n, x) \to z(\omega, x)$, $D_2 z(\omega_n, x) \to D_2 z(\omega, x)$. Thus, the boundedness of z, $D_2 z$ and the majorant theorem imply that

$$\int_{\mathscr{O}} \Big((z(\omega_n, x) - z(\omega, x))^2 + (D_2 z(\omega_n, x) - D_2 z(\omega, x))^2 \Big) dx \to 0,$$

which proves the continuity. Also, boundedness in V is straightforward.

Therefore, condition $(H4)$ is satisfied for such z.

Let us give some precise examples in which the above conditions are true.

Example 13.1 Let

$$Z : H \times \mathscr{O} \mapsto \mathbb{R}$$

that is supposed to be bounded and continuous with respect to each argument, and such that for a fixed argument in H the mapping is in $C^1(\overline{\mathscr{O}})$ with respect to x. Assume also that $D_2 Z$ is bounded and continuous with respect to the first argument. In addition, for some $T \in \mathbb{R}$ let $\delta_T \omega = \omega(T)$, which is a continuous mapping from Ω into H. Define

$$z(\omega, x) = Z(\delta_T \omega, x).$$

It is obvious that z is bounded, continuous with respect to each argument and belongs to $C^1(\overline{\mathscr{O}})$ for any fixed ω. On top of that, it is straightforward to see that $D_2 z$ is bounded and continuous with respect to the first argument. Thus, z satisfies all the above conditions.

Example 13.2 Let now ζ be a bounded mapping on $H \times \mathscr{O}$, which is continuous in the first argument if the second argument is fixed, and for fixed first argument $h \in H$ the mapping $\zeta(h, \cdot) \in C^1(\overline{\mathscr{O}})$. We assume for $D_2\zeta$ boundedness and continuity with respect to the first argument. Then for any $T_1 < T_2 \in \mathbb{R}$ we define the mapping

$$z(\omega, x) = \int_{T_1}^{T_2} \zeta(\omega(t), x)dt.$$

The convergence in Ω of a sequence $(\omega_n)_{n\in\mathbb{N}}$ induces the uniform convergence of this sequence on $[T_1, T_2]$. Therefore, z is bounded as ζ is bounded. The boundedness of $D_2\zeta$ ensures that we can exchange D_2 and the integral, that is,

$$D_2 z\,(\omega, x) = \int_{T_1}^{T_2} D_2\zeta(\omega(t), x)dt,$$

so that by the majorant theorem we obtain easily that $x \mapsto z(\omega, x) \in C^1(\overline{\mathscr{O}})$ and also that $D_2 z$ is bounded and $z,\ D_2 z$ are continuous with respect to ω. Thus, z satisfies all the above assumptions.

Example 13.3 Consider a sequence of functions $(\zeta_i)_{i\in\mathbb{Z}}$ on $H \times \mathscr{O}$, continuous in the first variable and contained in $C^1(\overline{\mathscr{O}})$ with respect to the second variable. We suppose that there exists a $C > 0$ such that

$$\sup_{H\times\mathscr{O}} |\zeta_i(h, x)| \le C2^{-|i|}, \qquad \sup_{H\times\mathscr{O}} |D_2\zeta_i(h, x)| \le C2^{-|i|}$$

and that $D_2\zeta_i$ are continuous with respect to the first argument. We define

$$z(\omega, x) = \sum_{i\in\mathbb{Z}} \int_i^{i+1} \zeta_i(\omega(t), x)dt.$$

It is clear that z and $D_2 z$ are bounded in the whole space $\Omega \times \mathscr{O}$. Indeed,

$$|D_2 z(\omega, x)| = \sum_{i\in\mathbb{Z}} \int_i^{i+1} |D_2\zeta_i(\omega(t), x)|\,dt \le C \sum_{i\in\mathbb{Z}} 2^{-|i|} = 3C,$$

and the same estimate is true for z. Also, using the majorant theorem and estimating the tails we obtain that $x \mapsto z(\omega, x) \in C^1(\overline{\mathscr{O}})$ and also that $z,\ D_2 z$ are continuous with respect to ω. Thus, z satisfies all the above assumptions.

Acknowledgements This work has been partially supported by Spanish Ministry of Economy and Competitiveness and FEDER, projects MTM2015-63723-P and MTM2016-74921-P, and by Junta de Andalucía (Spain), project P12-FQM-1492.

We would like to thank the referees for their valuable remarks and suggestions.

References

1. Arnold, L.: Random Dynamical Systems. Springer Monographs in Mathematics. Springer, Berlin (1998)
2. Caraballo, T., Chueshov, I.D., Real, J.: Pullback attractors for stochastic heat equations in materials with memory. Discrete Contin. Dyn. Syst. Ser. B **9**, 525–539 (2008)
3. Caraballo, T., Garrido-Atienza, M.J., Schmalfuss, B., Valero, J.: Asymptotic behaviour of a stochastic semilineardissipative functional equation without uniqueness of solutions. Discrete Contin. Dyn. Syst. Ser. B **14**, 439–455 (2010)
4. Caraballo, T., Garrido-Atienza, M.J., Schmalfuss, B., Valero, J.: Attractors for a random evolution equation with infinite memory: theoretical results. Discrete Contin. Dyn. Syst. Ser. B **22**, 1779–1800 (2017)
5. Castaing, C., Valadier, M.: Convex Analysis and Measurable Multifunctions. Springer, Berlin (1977)
6. Chueshov, I.D., Scheutzow, M.: Inertial manifolds and forms for stochastically perturbed retarded semilinear parabolic equations. J. Dyn. Differ. Equ. **13**, 355–380 (2001)
7. Hino, Y., Murakami, S., Naito, T.: Functional Differential Equations with Infinite Delay. Lecture Notes in Mathematics, vol. 1473. Springer, Berlin (1991)
8. Lions, J.L.: Quelques méthodes de résolution des problèmes aux limites non linéaires. Gauthier-Villar, Paris (1969)
9. Robinson, J.: Infinite-Dimensional Dynamical Systems. Cambridge University Press, Cambridge (2001)
10. Yosida, K.: Functional Analysis. Springer, Berlin (1965).

Chapter 14
Non-Lipschitz Homogeneous Volterra Integral Equations

Theoretical Aspects and Numerical Treatment

M. R. Arias, R. Benítez, and V. J. Bolós

Abstract In this chapter we introduce a class of nonlinear Volterra integral equations (VIEs) which have certain properties that deviate from the standard results in the field of integral equations. Such equations arise from various problems in shock wave propagation with nonlinear flux conditions. The basic equation we will consider is the nonlinear homogeneous Hammerstein–Volterra integral equation of convolution type

$$u(t) = \int_0^t k(t-s)g(u(s))\,\mathrm{d}s.$$

When $g(0) = 0$, this equation has function $u \equiv 0$ as a solution (trivial solution). It is interesting to determine whether there exists a nontrivial solution or not. Classical results on integral equations are not to be applied here since most of them fail to assure the existence of other solution than the trivial one. Several characterizations of the existence of nontrivial solutions under different hypothesis on the kernel k and the nonlinearity g will be presented. We will also focus on the uniqueness of nontrivial solutions for such equations. In this regard, it is important to note that this equation can be written as a fixed point equation, so we shall also discuss the attracting character of the solutions with respect to the Picard iterations of the nonlinear integral operator defined by the RHS of the equation. Indeed we will give some examples for which those iterations do not converge to the nontrivial solutions for some initial conditions and we will study the attraction basins for such repelling solutions. Numerical estimation of the solutions is also discussed. Collocation methods have proven to be a suitable technique for such equations.

M. R. Arias
Department of Mathematics, University of Extremadura, Badajoz, Spain
e-mail: arias@unex.es

R. Benítez (✉) · V. J. Bolós
Department of Business Mathematics, University of Valencia, Valencia, Spain
e-mail: rabesua@uv.es; vicente.bolos@uv.es

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*,
Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_14

However, classical results on numerical analysis of existence and convergence of collocation solutions cannot be considered here either since the non-Lipschitz character of the nonlinear operator prevents these results from being applied. New concepts on collocation solutions are introduced along with their corresponding results on existence and uniqueness of collocation solutions.

14.1 Introduction

Integral equations in general, and Volterra integral equations in particular, have been a source of very interesting problems within the realm of functional analysis since the early works of Volterra more than a century ago. Moreover, over the years, Volterra integral equations have been successfully applied to numerous problems of physics, engineering and ecology, among others.

The present review is devoted to establish the strong relation between the existence and uniqueness of solutions of nonlinear Volterra integral equations of convolution type with kernel k and nonlinearity g

$$u(x) = \int_0^x k(x-s)g(u(s))\,\mathrm{d}s, \quad x \geq 0, \tag{14.1}$$

and their attractive character. The main objective of this work is to present in a self-contained way the problems related to the existence and uniqueness of the above mentioned equation so the reader can get a general idea of the current status of the proposed problems.

We will say that a function f is attracted by u if

$$\lim_{n\to\infty} T_{kg}^n f(x) = u(x), \quad x \geq 0,$$

where T_{kg}^n denotes the composition of T_{kg} with itself n-times; being T_{kg} the operator

$$T_{kg} f(x) = \int_0^x k(x-s)g(f(s))\,\mathrm{d}s, \quad x \geq 0.$$

This operator will be referred to as *associated operator* to Eq. (14.1).

This equation is the homogeneous case of the nonlinear Volterra-Hammerstein integral equation

$$u(x) = f(x) + \int_0^x k(x-s)g(u(s))\,\mathrm{d}s, \quad x \geq 0; \tag{14.2}$$

which is also a particular case of the more general Volterra integral equation

$$u(x) = f(x) + \int_0^x h(x,s,u(s))\,\mathrm{d}s, \quad x \geq 0. \tag{14.3}$$

The following two theorems, whose proofs were given in 1953 by Sato in [18], are an example of classical results about the existence and uniqueness of solutions for Eq. (14.3), where the relationship between existence and uniqueness of solutions and their attractive character become clear.

Theorem 14.1 *Let f be a continuous function on $0 \leq x \leq \delta$, and let*

$$D := \left\{(x, s, u) \in \mathbb{R}^3 : 0 \leqslant s \leqslant x \leqslant \delta, \qquad |u - f(x)| \leqslant p\right\}.$$

Let $h \in C(D)$ have the bound $|h(x, s, u)| \leq M$ in D. Then, if $r := \min\{\delta, p/M\}$, the integral equation (14.3) has a solution in $C([0, r])$.

Theorem 14.2 *Under the hypotheses of the previous theorem, let h in addition satisfy a Lipschitz condition*

$$|h(x, s, z) - h(x, s, v)| \leq L|z - v|,$$

for all (x, s, z) and (x, s, v) in D. Then there is a unique continuous solution u of the integral equation (14.3) on the interval $[0, r]$, which satisfies $|u(x) - f(x)| \leqslant p$ on this interval. Moreover, this solution is the uniform limit of the successive approximations $u_n(x)$, defined by

$$u_0(x) = f(x)$$
$$u_{n+1}(x) = f(x) + \int_0^x h(x, s, u_n(s))\,\mathrm{d}s.$$

Similar results about existence, uniqueness and convergence of the successive approximations for systems of Volterra integral equations of the form

$$u(x) = f(x) + \int_0^x k(x - s)g(s, u(s))\,\mathrm{d}s, \quad x \geq 0,$$

where f and g are vectors of n components and k is a $n \times n$ matrix, defined on $[0, x_0)$, were presented by J. A. Nohel in 1962 (see [16]).

Let us go back to the convolution equations, (14.1) and (14.2), and let us consider monotone increasing associated operators T_{kg} (i.e. let g be an increasing function). In the homogeneous case, assuming without loss of generality[1] that $g(0) = 0$, we have that $u \equiv 0$ is a continuous solution and that if u is a solution, then for any $c > 0$ the function

$$u_c(x) = \begin{cases} 0 & \text{if } x \in [0, c) \\ u(x - c) & \text{if } x \geq c, \end{cases}$$

[1] If $g(0) = g_0 > 0$, then Eq. (14.1) could be writen as $u(x) = f(x) + \int_0^x k(x - s)\tilde{g}(u(s))ds$, being $f(x) = g_0 \int_0^x k(s)ds$ and $\tilde{g}(u) = g(u) - g_0$ and therefore the equation would be non-homogeneous.

is also a solution of the same equation. This kind of solutions will be referred to as c-translations of u. Let M be any positive constant. It is also known that the sequence of successive approximations

$$\begin{aligned} u_0 &= M \\ u_n &= T_{kg} u_{n-1}, \qquad n \geq 1, \end{aligned}$$

converges to 0, on an interval $[0, \delta)$, if and only if there is no other continuous solution than the trivial one (see [6]).

It is very important to note that in this homogeneous case, if the nonlinearity g satisfies a Lipschitz condition like the one in Theorem 14.2, then it can be easily proved that $g(u) \leq Lu$ for any positive u. This sublinearity condition implies that any solution for Eq. (14.1) will satisfy

$$u(x) = \int_0^x k(x-s)g(u(s))\,\mathrm{d}s \leq L \int_0^x k(x-s)u(s)\,\mathrm{d}s.$$

Thus, by means of a generalization of Gronwall's Inequality (see [11], page 69), it follows that $u(x) \leq 0$, which implies that Eq. (14.1) would only have the trivial solution, since the kernel k and the nonlinearity g are assumed to vanish in $(-\infty, 0]$.

In the non-homogeneous case (14.2), assuming that f is continuous, it is obvious that function zero is a subsolution, i.e., $0 \leq f + T_{kg}0$, so the sequence of successive approximations, starting with function 0, is a monotonic increasing sequence upper bounded by any positive constant, M, on an interval $[0, \delta)$, with δ depending on M. This implies the existence of a continuous solution for Eq. (14.2), at least near zero (see [11, Theorem 2.1.10]).

In both cases, the existence of continuous solutions is obtained from a simple argument in which the attractive character of the solutions has an important role. Therefore Theorem 14.1 gives us no more information about the existence of continuous solutions than that obtained in both previous paragraphs. This is an example of the relationship between the existence, the uniqueness and the attractive behaviour of the solutions.

14.2 Increasing Nonlinear Volterra Operators with Locally Bounded Kernels

In our first step in this analysis we will deal with Eq. (14.1), that is, the nonlinear homogeneous Volterra equation

$$u(x) = \int_0^x k(x-s)g(u(s))\,\mathrm{d}s, \quad x \geq 0,$$

with kernel k and nonlinearity $g$satisfying, at least, the following conditions:

- $k : [0, +\infty) \to [0, +\infty)$ is a locally bounded, measurable function such that

$$K(x) = \int_0^x k(s)\,\mathrm{d}s$$

  is a strictly increasing function.
- $g : [0, +\infty) \to [0, +\infty)$ is a strictly increasing function vanishing at 0.

Other extra assumptions on the kernel and nonlinearity will be indicated when necessary.

There is a broad bibliography about Volterra integral equation, where it is immediate to appreciate the closeness between existence and uniqueness of solutions and their attractive behaviour. One nice example is the paper by Bushell [12], where some results about homogeneous operators of degree p are proved. Recall that T is an *homogeneous operator of degree p* if $T(mv) = m^p T(v)$, for any positive m. Also, let us denote

$$\begin{aligned}
\mathscr{K} &= \{v \in C([0,1]) : v \geq 0 \text{ in } [0,1]\},\\
\mathscr{K}^0 &= \{v \in K : v(0) = 0 \text{ and } v > 0 \text{ in } (0,1]\} \text{ and}\\
\mathscr{K}_h^0 &= \{v \in C_h : \inf_{(0,1]} \frac{v(t)}{h(t)} > 0\},
\end{aligned}$$

being

$$C_h = \{v \in C([0,1]) : \|v\|_h = \sup_{(0,1]} \frac{|v(t)|}{h(t)} < \infty\}$$

and $\|\cdot\|$ denotes the supremum norm in $C([0,1])$. Under this notation, in [12, Theorem 1.1 and Theorem 1.2] they are proved the following results:

Theorem 14.3 *Suppose that $T : \mathscr{K} \to \mathscr{K}$ is a monotone increasing mapping which is homogeneous of degree p with $0 < p < 1$. If there exists a function $h \in C([0,1])$ such that $h > 0$ in $(0,1]$ and $Th \in \mathscr{K}_h^0$, then T has a unique fixed point in $\mathscr{K}_h^0$.*

Theorem 14.4 *Suppose that $T : \mathscr{K} \to \mathscr{K}$ is a monotone increasing mapping which is homogeneous of degree p with $0 < p < 1$. Take $f \in \mathscr{K}^0$ and define $S : \mathscr{K}^0 \to \mathscr{K}^0$ by $Sv = f + Tv$.*

If, in addition, T is continuous and uniformly bounded on $\mathscr{K}^0 \cap \{v : \|v\| = 1\}$ then exists unique solutions in $\mathscr{K}^0$ to the equations $u = Su$ and $u = S(1/u)$.

To prove the above theorems, first, the author finds the complete metric space $\{E_h, d_h\}$, being d_h the Hilbert's projective metric on $\mathscr{K}_h^0$ and $E_h = \mathscr{K}_h^0 \cap \{v : v \in C_h, \|v\|_h = 1\}$. The aim is to prove that operator $F : E_h \to E_h$ defined by

$F(v) = T(v)/\|T(v)\|_h$ is a strict contraction. Next, the author is ready to prove the existence and uniqueness of a fixed point of the operator T, by using Banach's contraction mapping theorem.

Using these theorems, the author proved the existence and uniqueness of solutions for an homogeneous Volterra integral equation with a more general non-convolution smooth kernel and a power nonlinearity of the form

$$u(x) = \int_0^x k(x,s)(u(s))^p \, \mathrm{d}s, \qquad 0 < p < 1.$$

Note that the nonlinearity of such equations is a non-Lipschitz function.

14.2.1 Continuous and Increasing Kernels

In [17], Szwarc considers a generalization of continuous increasing convolution kernels. He studies the non-convolution equation

$$u(x) = \int_0^x k(x,s)g(u(s)) \, \mathrm{d}s, \quad x \geq 0, \tag{14.4}$$

where $k : \mathbb{R}^2 \to \mathbb{R}$ is continuous in $0 \leq s \leq x$ with $k(x,s) = 0$ whenever $0 \leq x \leq s$ and such that,

$$k(s,t) \leq k(x,y), \text{ if } x \geqslant s, \ y \geqslant t, \ x - y > s - t. \tag{14.5}$$

This condition is a generalization of increasing convolution kernels, since if μ is an increasing function, the kernel $k(x,s) = \mu(x-s)$ satisfies condition (14.5). For such equations, the associated operator T_{kg} has an important property: if $x < y$, then, for all positive and measurable function f, we have that

$$\begin{aligned} T_{kg}f(x) &= \int_0^x k(x,s)g(f(s)) \, \mathrm{d}s = \int_0^y k(x,s)g(f(s)) \, \mathrm{d}s \\ &\leq \int_0^y k(y,s)g(f(s)) \, \mathrm{d}s = T_{kg}f(y), \end{aligned}$$

so the associated operator, T_{kg}, transforms measurable positive functions, defined on $\mathbb{R}^+$, into increasing functions.

R. Szwarc proves first the uniqueness of positive solutions and secondly he proves that when it exists, the solution is a global attractor of all positive and measurable functions. In order to prove the uniqueness of positive solution he first proves, in [17, Lemma 2], the following result:

Lemma 14.1 *Let u be a subsolution of Eq. (14.4), then for all positive c, there exists $\varepsilon > 0$ such that, the function v, defined as*

$$v(x) = \begin{cases} u(x) & \text{if } x \in [0, c] \\ u(c) & \text{if } x > c, \end{cases}$$

satisfies

$$\liminf_{n\to\infty} T_{kg}^n v(x) \geq u(x),$$

for all $x \in (c, c + \varepsilon)$.

Using this local attraction result and the fact that T_{kg} transforms horizontal translations of subsolutions into subsolutions, the uniqueness of positive solutions for Eq. (14.4) is proved in [17, Proposition 1].

Once the uniqueness is granted, it is proved that if it exists, the unique positive solution is a global attractor of all positive and measurable functions. This result is proved in two steps: first it is shown that every subsolution and every supersolution is globally attracted by the solution; and secondly it is proved that every positive function, that without losing generality can be considered increasing, can be upper-bounded and lower-bounded by a supersolution and a subsolution respectively.

A particular case of Eq. (14.4) is the convolution equation (14.1) with a continuous and increasing kernel k. In this case we have the following result:

Theorem 14.5 *Let k be a continuous and increasing function. If Eq. (14.1) admits a positive solution, then it is unique and attracts all positive functions.*

14.2.2 Continuous Like *Increasing Kernels*

With the aim of extending the results of Szwarc to a wider class of kernels, Arias and Castillo study, in [6], a Volterra integral equation (k, g), with a continuous kernel k being *"like"* a continuous increasing function. That is, a kernel k such that there exists a continuous increasing function φ, and two positive constants, m and M, satisfying $m\varphi \leq k \leq M\varphi$.

Let T_{mg} and T_{Mg} be the associated operators to the equations $(m\varphi, g)$ and $(M\varphi, g)$, respectively. Let us assume that equation (k, g) admits a solution u. On one hand, since $u = T_{kg}u \leq T_{Mg}u$, u is a subsolution of equation $(M\varphi, g)$. This implies the existence of a solution, u_M, for equation $(M\varphi, g)$. On the other hand, next theorem assures that then, equation $(m\varphi, g)$ also admits a solution u_m.

Theorem 14.6 *Let (φ, g) be an equation having a positive increasing kernel. The equation (φ, g) admits a solution if and only if for every $\lambda > 0$ the equation $(\lambda\varphi, g)$ admits a solution.*

Since φ is an increasing function, equations $(m\varphi, g)$ and $(M\varphi, g)$ are one of those considered by Szwarc in [17], thus, both u_m and u_M, are unique and global attractor of all positive functions. Taking this into account, they prove the following preliminary result:

Proposition 14.1 *Let (k, g) be an equation with a continuous kernel being like an increasing continuous function φ. If the equation (k, g) admits a solution then it admits a maximum solution and a minimum solution, which are increasing functions.*

Proof (Sketch of the Proof) If equation (k, g) has a solution u, since $m\varphi \leq k \leq M\varphi$, last theorem guarantees the existence of solutions u_m and u_M of equations $(m\varphi, g)$ and $(M\varphi, g)$, respectively. First it is proved that $u_m \leq u \leq u_M$. Thus, by the monotony of the operator T_{kg}, we get that $(T_{kg}^n u_M)_{n\in\mathbb{N}}$ is a decreasing sequence bounded by below by u. Defining

$$u_1 = \lim_{n\to\infty} T_{kg}^n u_M,$$

the Monotone Convergence Theorem yields that u_1 is a solution of equation (k, g). Similarly, the sequence $(T_{kg}^n u_m)_{n\in\mathbb{N}}$ is increasing and upperbounded by u, and its pointwise limit, u_2, is also a solution for (k, g).

In order to prove that u_1 and u_2 are a maximum and a minimum solution, respectively, it suffices to show that u_1 attracts all functions $f \geq u_1$, and that u_2 attracts all functions $f \leq u_2$.

Using the same method of demonstration as in [17, Proposition 1], the uniqueness of increasing solutions is obtained.

Proposition 14.2 *The equation (k, g) has at most one positive increasing solution.*

Note that, since u_m and u_M are increasing functions, u_1 and u_2 are also increasing functions. Thus, from last proposition, it follows that if equation (k, g) has a positive solution u, then we have that $u_1 = u_2 = u$. And from the attracting properties shown in the proof of that proposition, we get the following theorem:

Theorem 14.7 *Let k be a kernel like an increasing kernel. If the equation (k, g) admits a solution, then it is unique and attracts all positive functions.*

Thus, in [6], Arias and Castillo obtain for continuous and like increasing kernels, the same results Szwarc obtained in [17] for continuous increasing kernels.

14.2.3 Continuous Kernels

Another extension of the results obtained by Szwarc was carried out in [2]. In that paper, an equation (k, g) with a continuous kernel was considered. First, it is presented a characterization of the existence of continuous strictly increasing

functions, by means of an operator, denoted by H, and defined as

$$Hf(x) = \int_0^x K(F(x) - F(s))\,\mathrm{d}s, \tag{14.6}$$

being

$$K(x) = \int_0^x k(s)\,\mathrm{d}s, \qquad F(x) = \int_0^x f(s)\,\mathrm{d}s.$$

Using classical results about the properties of the convolution operator (see [13, p. 99]), it is immediate that continuous strictly increasing solutions of equation (k, g) are locally absolutely continuous functions with positive derivatives a.e. and their inverses are also locally absolutely continuous functions on $\mathbb{R}^+$. With this properties of continuous strictly increasing solutions, first it is proved the following technical lemma:

Lemma 14.2 *Let u be a continuous strictly increasing function. Then u is a solution of (k, g) if and only if*

$$x = \int_0^{g(x)} K(u^{-1}(x) - u^{-1} \circ g^{-1}(s))\,\mathrm{d}s.$$

This lemma leads to the characterization of the existence of continuous strictly increasing solutions of equation (k, g).

Theorem 14.8 *Equation (k, g) has a continuous strictly increasing solution if and only if there exists a positive integrable function f such that $Hf \geq g^{-1}$.*

Using the remarkable fact that there exits a bijection between the set of all continuous strictly increasing solutions of (k, g), and the set $L^1_{loc}(\mathbb{R}^+, \mathbb{R}^+) \cap H^{-1}(g^{-1})$, it is proved the following uniqueness theorem, by showing that the operator H is injective on $L^1_{loc}(\mathbb{R}^+, \mathbb{R}^+)$.

Theorem 14.9 *The equation (k, g) has at most one continuous strictly increasing solution.*

Once the uniqueness of continuous strictly increasing solutions is proved, the question is: do there exist another kind of positive solutions? As it was shown in [3], if k is a continuous function, we can define the auxiliary kernel

$$\overline{k}(x) = \max\{k(s) : s \in [0, x]\}.$$

Obviously $\overline{k}$ is increasing and continuous, so the equation $(\overline{k}, g)$ is one of those considered by Szwarc in [17]. Thus, the associated operator to equation $(\overline{k}, g)$, $T_{\overline{k}g}$, transforms positive measurable functions into increasing functions and satisfies the

inequality $T_{kg} \leq T_{\overline{k}g}$. So if u is a positive solution of equation (k, g), then

$$u = T_{kg}u \leq T_{\overline{k}g}u.$$

Since $T_{\overline{k}g}u$ is an increasing function, u is a locally bounded function. Also, since the kernel k is continuous, from the properties of the convolution, the continuity of locally bounded solutions is obtained. To sum up, we know, from Theorem 14.9, that there is at most one unique continuous strictly increasing solution of equation (k, g), and we also have obtained that any positive solution of such equation is a continuous function. Then, what we have to deal now is with the problem of the existence of continuous non-increasing solutions.

In order to prove the uniqueness of positive solutions, an equation (k, g) with a positive solution v is considered. First it is proved that the existence of a positive solution implies the existence of a continuous strictly increasing solution u, which is the maximum solution of (k, g). Then, some attracting properties of u (Properties of Inertia) are obtained. For the first one, given a positive real function, f, defined on $\mathbb{R}^+$, and a positive constant α, we shall define the *α-shift* of f by

$$f_\alpha(x) = \begin{cases} 0 & \text{if } x \in [0, \alpha] \\ f(x - \alpha) & \text{if } x > \alpha, \end{cases}$$

and the *α-cut* of f by

$$f^\alpha(x) = \begin{cases} f(x) & \text{if } x \in [0, \alpha] \\ f(\alpha) & \text{if } x > \alpha. \end{cases}$$

Lemma 14.3 (First Property of Inertia) *Let α and β be two positive constants with $\beta > \alpha$. Then $\lim_{n\to\infty} T^n_{kg} u^\beta_\alpha = u_\alpha$.*

The second property asserts that there cannot exist two different solutions that coincide on a neighborhood of 0.

Lemma 14.4 (Second Property of Inertia) *Let v be a solution of (k, g). If $v = u$ in an interval $[0, \delta]$, then $v \equiv u$.*

Using these two Properties of Inertia, one can show the following uniqueness result:

Theorem 14.10 *The unique solution of (k, g) is u.*

With a similar proof an attraction property is obtained.

Corollary 14.1 *Le f be a positive and measurable function such that $f \leq u$. Then f is attracted by u.*

To obtain a similar attraction result, for functions $f \geq u$, in [3] the auxiliary increasing kernel $\overline{k}$ was considered again. Note that if u is the positive solution of equation (k, g), then $u \leq T_{\overline{k}g}u$, that is, u is a subsolution of equation $(\overline{k}, g)$. Thus,

equation $(\overline{k}, g)$ has a positive solution $\overline{u}$, which is unique and a global attractor of all positive functions (see Theorem 14.5). From the attracting behaviour of $\overline{u}$, and the inequality $u \leq T_{\overline{k}g}u$, it is also obtained that $u \leq \overline{u}$. Since T_{kg} is a monotone operator, using the Monotone Convergence Theorem, it can be shown that the sequence $(T_{kg}\overline{u})_{n\in\mathbb{N}}$ converges to a solution of equation (k, g). From the uniqueness of solutions of equation (k, g) we get that $\lim_{n\to\infty} T^n_{kg}\overline{u} = u$.

With a proof similar to the proof of Proposition 14.1, we find another attracting property of u.

Lemma 14.5 *Let f be a positive measurable function such that $f \geq u$. Then*

$$\lim_{n\to\infty} T^n_{kg} f = u.$$

As a consequence of Corollary 14.1 and Lemma 14.5, we finally get the following theorem:

Theorem 14.11 *Let k be a continuous kernel. If equation (k, g) admits a positive solution, then it is unique and attracts (globally) all positive functions.*

14.2.4 Locally Bounded Kernels

A brief analysis of the results above mentioned reveals that one of the keys to prove the uniqueness and the attractive behaviour of the solutions of equation (k, g), is the construction of the auxiliary increasing kernel $\overline{k}$. But in order to define the kernel $\overline{k}$ we do not need k to be a continuous function; it is only necessary that k is locally bounded. Although Szwarc considers in [17] continuous increasing kernels, after a close inspection of that paper, one can conclude that the continuity of k is only used in Lemma 2, in order to find a bound of the kernel in a closed and bounded interval. That can be done with a locally bounded kernel.

In [5, Sec. 2] it is analyzed where, in [2] and [3], is needed the continuity of the kernel. It is shown that there is no problem to replace continuous kernels by locally bounded ones. Thus we can get the following extension of Theorem 14.11:

Theorem 14.12 *Let k be a locally bounded kernel. If equation (k, g) admits a positive solution, then it is unique and attracts (globally) all positive functions.*

Then we have the following progression

$$\begin{array}{ccc} \text{Continuous} & & \text{Continuous like} \\ \text{increasing kernels} & \Rightarrow & \text{increasing kernels} \\ & & \Downarrow \\ \text{Locally bounded} & & \text{Continuous} \\ \text{kernels} & \Leftarrow & \text{kernels} \end{array}$$

The question that arises now in a natural way is: can Theorem 14.12 be extended to non-locally bounded kernels? As we shall see in next section, the answer is no.

14.3 Increasing Nonlinear Volterra Operators with Locally Integrable Kernels

In order to study the equation (k, g) when the kernel is a locally integrable but non-locally bounded function, it was considered in [5], the Abel integral equation

$$u(x) = \int_0^x (x - s)^\alpha g(u(s))\, \mathrm{d}s, \tag{14.7}$$

where $-1 < \alpha < 0$. Such equation has been deeply studied in the literature. W. Mydlardzyc considered, in [15], the more general integral equation

$$u(x) = \int_0^x k(x - s)r(s)g(u(s) + h(s))\, \mathrm{d}s, \tag{14.8}$$

where

1. $k > 0$, $\int_0^x k(s)\, \mathrm{d}s < \infty$ for $x > 0$.
2. r and g are nondecreasing and continuous functions, vanishing in $(-\infty, 0]$, and positive in $(0, +\infty)$.
3. h is a nondecreasing and continuous function, vanishing in $(-\infty, 0]$ and $h(x) > 0$ for $x > 0$, or $h \equiv 0$.

For such equations, W. Mydlardzyc proves that Eq. (14.8) has at most one continuous solution u. Note that Eq. (14.8) reduces to equation (k, g), when $r \equiv 1$ on $(0, +\infty)$, and $h \equiv 0$. Thus, the uniqueness result introduced in [15] is an alternative proof of Theorem 14.9. So, denoting Eq. (14.7) by (x^α, g), we have the following uniqueness result:

Proposition 14.3 *Equation (x^α, g) has at most one continuous solution.*

Taking into account that the convolution of an integrable function with a bounded function is a uniformly continuous function, it can be proved that every locally bounded solution of equation (x^α, g) is a continuous function. Thus last proposition can be rewritten as a more general uniqueness theorem:

Theorem 14.13 *Equation (x^α, g) has at most one locally bounded solution.*

Recall that the aim of this section is to study whether Theorem 14.12 can be extended to Abel kernels or not. To study the attracting character of the locally bounded solutions of equation (x^α, g), it suffices that the kernel is a locally integrable function. Thus, we have that

Proposition 14.4 *Let u be a locally bounded solution of equation (x^α, g), and f a positive and measurable function such that $f \leq u$. Then f is globally attracted by u.*

The next step is to prove the same attraction property for functions lower bounded by u. In Sect. 14.2 that was proved by defining an auxiliary increasing kernel, $\overline{k}$ such that $k \leq \overline{k}$. Since Abel kernels are non-locally bounded and strictly decreasing, we cannot find an increasing function bounding the Abel kernel. Thus we cannot use the same arguments as those considered in last section with locally bounded kernels. Nevertheless, constant functions are supersolutions on an interval $[0, \delta)$, with δ depending on the constant. So constant functions are locally attracted by the locally bounded solution. Using a comparison argument, any locally bounded function will be locally attracted by such solution. Adding some conditions on the nonlinearity, this local attraction property can be extended (see [4]), to a global attraction one.

Theorem 14.14 *Let f be a positive, measurable and locally bounded function, and g a nonlinearity satisfying*

$$\lim_{x \to +\infty} \frac{x}{g(x)} = +\infty. \tag{14.9}$$

Then f is globally attracted by the locally bounded solution of equation (x^α, g).

In conclusion, for Abel equations we have the following result:

Theorem 14.15 *Equation (x^α, g) has at most one locally bounded solution. Moreover, if g satisfies (14.9) then such solution is a global attractor of any positive, measurable and locally bounded function.*

Note that results in Theorem 14.15 are not as strong as in Theorem 14.12. While in Theorem 14.12 we have uniqueness of positive solutions and global attraction of all positive functions, in Theorem 14.15 we can only guarantee the uniqueness of locally bounded solutions and the attraction of locally bounded functions.

The first question that arises in view of Theorem 14.15 is: is there any positive and non-locally bounded solution of equation (x^α, g)? As we shall see next, the answer is positive.

14.3.1 Non-locally Bounded and Multiple Solutions

Simple examples of Abel integral equations with non-locally bounded solutions can be obtained considering nonlinearities $g(x) = x^\beta$. Indeed, the equation (x^α, x^β), with $-1 < \alpha < 0$ and $\beta > -\alpha^{-1}$ has as solution, the function $u(x) = dx^\gamma$, being

$$\gamma = \frac{\alpha + 1}{1 - \beta}, \qquad d = B(\alpha + 1, \gamma\beta + 1)^{1/(1-\beta)},$$

where B denotes the Beta function. Since $\beta > -\alpha^{-1} > -1$, then $\gamma < 0$, and therefore u is a positive non-locally bounded solution of equation (x^α, x^β).

To the question of whether this non-locally bounded solution is an attractor or not, a partial answer was given in [5]. There it was proved that u does not attract the functions of the family

$$\mathscr{U} = \{cx^\gamma : c > 0,\ c \neq d\}.$$

In the next theorem, whose proof can be found in the aforementioned reference [5], for the sake of simplicity, $T_{\alpha\beta}$ denotes the associated operator to equation (x^α, x^β).

Theorem 14.16 *Let $f \in \mathscr{U}$. Then we have that:*

(a) If $c > d$ then $\lim_{n\to\infty} T^n_{\alpha\beta} f(x) = +\infty$, for all $x \in (0, +\infty)$
(b) If $c < d$ then $\lim_{n\to\infty} T^n_{\alpha\beta} f(x) = 0$, for all $x \in (0, +\infty)$

The existence of Abel equations with non-locally bounded solutions leads to the existence of Abel equations with two positive solutions. Such equations were obtained in [4], combining two Abel equations: one with a locally bounded solution, and one with a non-locally bounded solution. Consider an Abel equation $(x^\alpha, C\tilde{g})$ with a locally bounded solution $\tilde{u}$, and an Abel equation (x^α, x^β), as above, with a nonlocally bounded solution dx^γ. Then, given a positive δ, we define the nonlinearity

$$g(x) = \begin{cases} C\tilde{g}(x) & \text{if } x \in [0, \delta) \\ x^\beta & \text{if } x \geq \delta, \end{cases}$$

where $C = \delta^\beta \tilde{g}(\delta)^{-1}$ is such that g is continuous. For such nonlinearity we have that equation (x^α, g) has two positive solutions: one, u_1, locally bounded and the other, u_2, non-locally bounded. Moreover, we have that, in a neighbourhood of zero, u_1 coincides with $\tilde{u}$ and u_2 coincides with dx^γ.

The attracting properties of the solutions of equation (x^α, g), when two positive solutions are present, was studied in [5]. First it was proved the following lemma:

Lemma 14.6 *Let u_1 and u_2 denote the locally bounded and the non-locally bounded solution of equation (x^α, g), respectively. Then $u_1 \leq u_2$.*

Theorem 14.17 *Let f be a function such that $f(x) = cx^\gamma$ in a neighbourhood of zero. Then*

(a) If $c > d$, $\lim_{n\to\infty} T^n_{\alpha g} f(x) = +\infty$ for all $x \in (0, +\infty)$.
(b) If $c < d$, $\lim_{n\to\infty} T^n_{\alpha g} f(x) = u_1(x)$ on the domain of u_1.

However, the structure of the attraction basins of the solutions u_1 and u_2 is far more complex that what may appear from Theorem 14.17. In fact, the structure of the attraction basin of the non-bounded solution u_2 was studied in [7]. There, it was found that the basins of attraction could not be separated with the $L^1(0, \delta)$ topology,

in the sense that for every ball, centered on the unbounded fixed point u_2, $\mathscr{B}(u_2)$, we could find functions $u_0 \in \mathscr{B}(u_2)$ such that the iteration sequence $(T^n u_0)_{n\in\mathbb{N}}$ converges to u_2, converges to u_1 or diverges to ∞.

14.3.2 Abel Equations as Limit of Volterra Equations

Let (x^α, g) be an Abel integral equation with a locally bounded solution u and, for any natural n, we define

$$k_n(x) = \begin{cases} \frac{1}{n^\alpha} & \text{if } x \in [0, \frac{1}{n}] \\ x^\alpha & \text{if } x > \frac{1}{n}. \end{cases}$$

The sequence $(k_n)_{n\in\mathbb{N}}$ converges pointwise to the Abel kernel x^α on $(0, +\infty)$. Next we shall see that under some assumptions, any equation (k_n, g) has a solution u_n.

Proposition 14.5 *If there exists a positive δ such that*

$$\int_0^\delta \frac{\mathrm{d}s}{g(s)} < +\infty, \tag{14.10}$$

then for any natural n the equation (k_n, g) admits a solution u_n.

In the next result, which was proved in [4], T_{ng} and $T_{\alpha g}$ denote the associated operators to equation (k_n, g) and (x^α, g), respectively.

Proposition 14.6 *If $T_{\alpha g} f$ is continuous and $m \in \mathbb{N}$, then the sequence $(T_{ng}^m f)_{n\in\mathbb{N}}$ converges to $T_{\alpha g}^m f$ uniformly on any compact in $\mathbb{R}^+$.*

With this convergence result we obtain the following theorem:

Theorem 14.18 *The sequence $(u_n)_{n\in\mathbb{N}}$ converges to u uniformly on any compact in $\mathbb{R}^+$.*

Note that, since the existence of solutions for the equation (k_n, g) implies the existence of a locally bounded solution of (x^α, g), this theorem shows a way to construct such solution.

We also have to note that hypothesis (14.10) cannot be dismissed. Since the kernel k_n is constant in the interval $[0, 1/n]$, equation (k_n, g) takes the form

$$u_n(x) = \frac{1}{n^\alpha} \int_0^x g(u_n(s))\, \mathrm{d}s.$$

And this equation is equivalent to the initial value problem

$$u_n'(x) = \frac{1}{n^\alpha} g(u_n(x)), \qquad x \in [0, \frac{1}{n}]; \qquad u_n(0) = 0. \tag{14.11}$$

But it follows from the Osgood Uniqueness Theorem (see [13]), that if

$$\int_0^{\delta} \frac{\mathrm{d}s}{g(s)} = +\infty,$$

then $u_n \equiv 0$ is the only continuous solution of the initial value problem (14.11). So if (14.10) does not hold, equation (k_n, g) has only the trivial solution. Moreover, in such case the Abel equation (x^{α}, g) cannot be written as the (pointwise) limit of a Volterra integral equation sequence $(k_n, g)_{n\in\mathbb{N}}$ with locally bounded kernels.

Finally, recall that, since k_n is a locally bounded function, the locally bounded solution u_n is unique. So if we consider an Abel integral equation (x^{α}, g) with two positive solutions, then equations (k_n, g) may have only the trivial solution or the locally bounded solution u_n, depending on whether (14.10) holds or not. Therefore, neither the existence of nontrivial solutions, nor the uniqueness of such solutions are properties preserved under the pass to the (pointwise) limit.

14.4 Numerical Study

Over this section we are going to deal with numerical methods to estimate non trivial solutions for homogeneous Volterra integral equations. In previous sections of this chapter we have seen results about existence and uniqueness of nontrivial solutions and also, certain results describing some attractive and repelling behaviours of such solutions considered as fixed points of some discrete dynamical systems. Among these results, some of them bound their basins of attraction or, at least, determine subsets enclosed in them. From this information, to raise some numerical methods for estimating nontrivial solutions is relatively straightforward. A couple of these methods can be seen in [1]. Next, we will give a brief description of two direct methods. In both of them, the equation considered will be Eq. (14.1)

$$u(x) = \int_0^x k(x-s)g(u(s))ds, \qquad x \in [0, \delta].$$

We assume that last equation has a unique bounded nontrivial solution. In this section, for the sake of readability, operator T_{kg} will be denoted by T, since there is no ambiguity in either the kernel or the nonlinearity.

From the previous sections it becomes clear that u is an absolutely continuous function on $[0, \delta]$ and therefore the existence of positive constants M and $\delta_0 \leq \delta$ such that $u \leq M$ on $[0, \delta]$ and $M \geq TM \geq u$ on $[0, \delta_0]$ is an immediate fact. The last inequalities are the basis for the first method presented in [1, Chapter 5]. Since T is an increasing operator and u is the unique continuous fixed point; the immediate consequences are the proper definition of the sequence $(T^n M)_{n\in\mathbb{N}}$ and their monotonous decreasing character. Obviously, $(T^n M)_{n\in\mathbb{N}}$, the orbit of M, must be a convergent sequence lower bounded by u. Since the limit of this orbit must be

a fixed point of T we can assert that

$$(T^n M)_{n\in\mathbb{N}} \to u,$$

on $[0, \delta_0]$. It is not necessary a deep analysis to recognize the drawback of this method to estimate u. The main problem is the required computations in order to determine the different terms of $(T^n M)_{n\in\mathbb{N}}$; namely, to calculate the n-fold convolution. To overcome this problem we consider the *downward-upward method* [1]. We know that

$$M \geq TM \geq u, \qquad \text{on } [0, \delta_0].$$

Since TM is an absolute continuous function, for $\epsilon/2 > 0$, there exists $\delta_1 > 0$ and P_1, a partition of $[0, \delta_0]$ with size δ_1, and it is possible to define the step function

$$M_1(x) = \sum_{i=1}^{n_1} TM(x_i)\chi_i(x), \qquad x \in [0, \delta_0]$$

where $x_i \in P_1$ with $i = 1, \ldots, n_1$ and χ_i, the characteristic function of $[x_{i-1}, x_i)$; in order to ensure that $0 \leq M_1 - TM \leq \epsilon/2$ and $M \geq M_1 \geq TM$ on $[0, \delta_0]$. In this sense, TM would be the drop and M_1, the rise. Considering the monotone increasing character of T and the last inequalities we obtain that $M_1 \geq TM \geq TM_1$ on $[0, \delta_0]$.

From TM_1, analogously to the construction of M_1 from TM, it is possible to define the step function

$$M_2(x) = \sum_{i=1}^{n_2} TM_1(x_i)\chi_i(x), \qquad x \in [0, \delta_0],$$

where $x_i \in P_2$ with $i = 1, \ldots, n_2$; being P_2, a partition of $[0, \delta_0]$ with size δ_2. The partition P_2 and δ_2 are selected in order to ensure that $0 \leq M_2 - TM_1 \leq \epsilon/2^2$ and $M_1 \geq M_2 \geq TM_1$ on $[0, \delta_0]$. Therefore $M_2 \geq TM_1 \geq TM_2$ holds on $[0, \delta_0]$. In this case TM_1 would be the drop and M_2, the rise.

And so on for any $m \in \mathbb{N}$, once TM_m is obtained, it is possible to define the step function

$$M_{m+1}(x) = \sum_{i=1}^{n_{m+1}} TM_m(x_i)\chi_i(x), \qquad x \in [0, \delta_0],$$

where $x_i \in P_{m+1}$ with $i = 1, \ldots, n_{m+1}$; being P_{m+1}, in this case, a partition of $[0, \delta_0]$ with size δ_{m+1}. The partition P_{m+1} and δ_{m+1} are selected in order to ensure that $0 \leq M_{m+1} - TM_m \leq \epsilon/2^{m+1}$ and $M_m \geq M_{m+1} \geq TM_m$ on $[0, \delta_0]$. Therefore

$M_{m+1} \geq TM_m \geq TM_{m+1}$ holds on $[0, \delta_0]$. In this case TM_m would be the drop and M_{m+1}, the rise.

It can be seen that, for any $m \in \mathbb{N}$, the sequence $(T^n M_m)_{n\in\mathbb{N}} \to u$ pointwise on $[0, \delta_0]$ and the same will happen with the sequences $(T^n M_m)_{m\in\mathbb{N}}$, for any $n \in \mathbb{N}$, see [1]. Therefore, an easy way to estimate u locally, near to zero, is to consider the sequence $(TM_m)_{m\in\mathbb{N}}$. In our opinion the major problem with *downward-upward method* is the slow convergence rate. To solve this and other problems, other methods can be considered, as we shall see below.

14.4.1 Collocation Methods

Let us consider the nonlinear homogeneous Volterra-Hammerstein integral equation (HVHIE) with non-convolution kernel given by

$$u(x) = \int_0^x k(x, s)\, g(u(s))\, \mathrm{d}s, \qquad x \in I := [0, \delta]. \tag{14.12}$$

We assume that the following *general conditions* are always held, even if they are not explicitly mentioned.

- **Over** k. The kernel $k : \mathbb{R}^2 \to [0, +\infty)$ is a locally bounded function and its support is in $\left\{(x, s) \in \mathbb{R}^2 : 0 \leq s \leq x\right\}$.
  For every $x > 0$, the map $s \mapsto k(x, s)$ is locally integrable, and $\int_0^x k(x, s)\, \mathrm{d}s$ is a strictly increasing function.
- **Over** g. The nonlinearity $g : [0, +\infty) \to [0, +\infty)$ is a continuous, strictly increasing function, and $g(0) = 0$.

Note that since $g(0) = 0$, the zero function is a solution of Eq. (14.12) and so, uniqueness of solutions is no longer a desired property for Eq. (14.12) because we are obviously interested in nontrivial solutions. Therefore, it will be necessary that the nonlinearity g does not satisfy a Lipschitz condition. This is an important issue because the main theorems about the existence, uniqueness and convergence of collocation solutions for Volterra integral equations rely on the Neumann Lemma, which states that under certain condition, some matrices must have a spectral radius lower than 1, and use this fact in order to prove the convergence. Such conditions are, namely, the Lipschitz continuity of the nonlinearity (see [11]).

Taking $z := g \circ u$, Eq. (14.12) can be written as an *implicitly linear* homogeneous Volterra integral equation (HVIE) for z:

$$z(x) = g\left(\int_0^x k(x, s)\, z(s)\, \mathrm{d}s\right), \qquad x \in I. \tag{14.13}$$

So, if z is a solution of (14.13), then $u := g^{-1} \circ z$ is a solution of (14.12).

14.4.1.1 Collocation Problems for Implicitly Linear HVIEs

Let $I_h := \{x_n : 0 = x_0 < x_1 < \ldots < x_N = \delta\}$ be a mesh (not necessarily uniform) on the interval $I = [0, \delta]$, and set $\sigma_n := (x_n, x_{n+1}]$ with lengths $h_n := x_{n+1} - x_n$ $(n = 0, \ldots, N-1)$. The quantity $h := \max\{h_n : 0 \leq n \leq N-1\}$ is called the *stepsize*. Given a set of m *collocation parameters* $\{c_i : 0 \leq c_1 < \ldots < c_m \leq 1\}$, the *collocation points* are given by $x_{n,i} := x_n + c_i h_n$ $(n = 0, \ldots, N-1)$ $(i = 1, \ldots, m)$, and the set of collocation points is denoted by X_h. All this defines a *collocation problem* for Eq. (14.13) (see [10], [11, p. 117]), and a *collocation solution* z_h is given by the *collocation equation*

$$z_h(x) = g\left(\int_0^x k(x,s)\, z_h(s)\, \mathrm{d}s\right), \qquad x \in X_h, \tag{14.14}$$

where z_h is in the space of piecewise polynomials of degree less than m (see [11, p. 85]). Note that the identically zero function is always a collocation solution, since $g(0) = 0$.

As it is stated in [11], a collocation solution z_h is completely determined by the coefficients $Z_{n,i} := z_h\left(x_{n,i}\right)$ $(n = 0, \ldots, N-1)$ $(i = 1, \ldots, m)$, since $z_h(x_n + vh_n) = \sum_{j=1}^m L_j(v)\, Z_{n,j}$ for all $v \in (0, 1]$, where $L_j(v) := \prod_{k \neq j}^m \frac{v - c_k}{c_j - c_k}$ $(j = 1, \ldots, m)$ are the Lagrange fundamental polynomials with respect to the collocation parameters. The values of $Z_{n,i}$ are given by the systems

$$Z_{n,i} = g\left(F_n\left(x_{n,i}\right) + h_n \sum_{j=1}^m B_n(i,j)\, Z_{n,j}\right), \tag{14.15}$$

where

$$B_n(i,j) := \int_0^{c_i} k\left(x_{n,i}, x_n + sh_n\right) L_j(s)\, \mathrm{d}s. \tag{14.16}$$

and

$$F_n(x) := \int_0^{x_n} k(x,s)\, z_h(s)\, \mathrm{d}s. \tag{14.17}$$

14.4.1.2 Existence and Uniqueness of Nontrivial Collocation Solutions

Given a kernel k, a nonlinearity g and some collocation parameters $\{c_1, \ldots, c_m\}$, our aim is to study the existence of collocation solutions in an interval $I = [0, \delta]$ using a mesh I_h. Specifically, we are interested in "nontrivial" collocation solutions that are not identically zero in σ_0. Next, we are going to define three different kinds of existence of nontrivial collocation solutions.

- We say that there is **existence near zero** if there exists $H_0 > 0$ such that if $0 < h_0 \leq H_0$ then there are nontrivial collocation solutions in $[0, x_1]$; moreover, there exists $H_n > 0$ such that if $0 < h_n \leq H_n$ then there are nontrivial collocation solutions in $[0, x_{n+1}]$ (for $n = 1, \ldots, N-1$ and given $h_0, \ldots, h_{n-1} > 0$ such that there are nontrivial collocation solutions in $[0, x_n]$). That is, collocation solutions can always be extended a bit more.
- We say that there is **existence for fine meshes** if there exists $H > 0$ such that if $0 < h \leq H$ then the corresponding collocation problem has nontrivial collocation solutions in any interval I.
- We say that there is **unconditional existence** if there exist nontrivial collocation solutions in any interval I and for any mesh I_h.

Collocation solutions should converge to solutions of (14.13) (if they exist) when $h \to 0^+$ (and $N \to +\infty$), but these convergence problems are, in general, very complex. Taking this into account, we are interested in the study of existence of nontrivial collocation solutions using meshes I_h with arbitrarily small $h > 0$ and moreover, we are not interested in collocation problems whose collocation solutions "escape" to $+\infty$ in a certain σ_n when $h_n \to 0^+$.

Let S be an index set of all the nontrivial collocation solutions of a collocation problem with mesh I_h. For any $s \in S$ we denote by $Z_{s;n,i}$ the coefficients satisfying Eq. (14.15) (with $n = 0, \ldots, N-1$ and $i = 1, \ldots, m$) and such that, at least, one of the coefficients $Z_{s;0,i}$ is different from zero (for some $i \in \{1, \ldots, m\}$). Given $I_h = \{0 = x_0 < \ldots < x_{N-1}\}$ such that there exist nontrivial collocation solutions using this mesh, we say that there is *nondivergent existence* if, for $n = 0, \ldots, N-1$,

$$\mathscr{Z}_{h_n} := \inf_{s \in S_{h_n}} \left\{ \max_{i=1,\ldots,m} \left\{ Z_{s;n,i} \right\} \right\}$$

exists for small enough $h_n > 0$ and it does not diverge to $+\infty$ when $h_n \to 0^+$. Moreover, we say that there is *nondivergent uniqueness* if

$$\min_{s \in S_{h_n}} \left\{ \max_{i=1,\ldots,m} \left\{ Z_{s;n,i} \right\} \right\} = \mathscr{Z}_{h_n}$$

exists for small enough $h_n > 0$ and it does not diverge to $+\infty$ when $h_n \to 0^+$, but

$$\inf_{s \in S_{h_n}} \left(\left\{ \max_{i=1,\ldots,m} \left\{ Z_{s;n,i} \right\} \right\} - \left\{ \mathscr{Z}_{h_n} \right\} \right)$$

diverges. In case of nondivergent uniqueness, the *nondivergent collocation solution* that makes sense is the one whose coefficients $Z_{n,i}$ satisfy $\max_{i=1,\ldots,m} \left\{ Z_{n,i} \right\} = \mathscr{Z}_{h_n}$ for $n = 0, \ldots, N-1$.

Next, we are going to present some results about nondivergent existence and uniqueness for cases 1 ($m = 1$ with $c_1 > 0$) and 2 ($m = 2$ with $c_1 = 0$) given in [8]. In these results, we say that a property $\mathscr{P}$ holds *near zero* if there exists $\epsilon > 0$

such that $\mathscr{P}$ holds on $(0, \delta)$ for all $0 < \delta < \epsilon$. On the other hand, we say that $\mathscr{P}$ holds *away from zero* if $\mathscr{P}$ holds on $(x, +\infty)$ for all $x > 0$. Finally, we say that g is "well-behaved" if $\frac{g(\alpha+u)}{u}$ is strictly decreasing near zero for all $\alpha > 0$. This condition is in fact very weak, since $\lim_{u\to 0^+} \frac{g(\alpha+u)}{u} = +\infty$. Now, we are able to present these results:

- $k(x, s) \leq k\left(x', s\right)$ for all $0 \leq s \leq x < x'$;
  $\frac{g(u)}{u}$ is unbounded near zero if and only if there is nondivergent existence near zero.
  Moreover, if g is "well-behaved" then there is nondivergent uniqueness near zero.
- (Hypothesis only for case 2: $c_2 = 1$, or $k(x, s) \leq k\left(x', s\right)$ for all $0 \leq s \leq x < x'$);
  If $\frac{g(u)}{u}$ is unbounded near zero but bounded away from zero then there is nondivergent existence near zero.
  Moreover, if g is "well-behaved" then there is nondivergent uniqueness near zero.
  With these hypotheses, for convolution kernels $k(x - s)$ there is nondivergent existence or uniqueness (resp.) for fine meshes instead of near zero.
- (Hypothesis only for case 2: $c_2 = 1$, or $k(x, s) \leq k\left(x', s\right)$ for all $0 \leq s \leq x < x'$);
  If $\frac{g(u)}{u}$ is unbounded near zero and there exists a sequence $\{u_n\}_{n=1}^{+\infty}$ of positive real numbers and divergent to $+\infty$ such that $\lim_{n\to+\infty} \frac{g(u_n)}{u_n} = 0$ then there is unconditional nondivergent existence.
  Moreover, if g is "well-behaved" then there is unconditional nondivergent uniqueness.

It should be noted that if the collocation problem is not in the scope of cases 1 and 2, the existence of nontrivial collocation solutions is not assured. Some examples can be found in [8].

14.4.1.3 Blow-Up Collocation Solutions

Some problems in engineering and physics exhibiting explosive behaviour are described by nonlinear integral equations whose solutions present a blow-up at finite time. It has been an important issue regarding these equations, not only to determine whether a nonlinear Volterra integral equation has blow-up solutions, but also, to give estimates of the location of such blow-up time (see [14, 19]).

In this section, we are going to extend the concept of collocation problem and collocation solution in order to consider the case of "blow-up collocation solutions". But first, since non locally bounded kernels appear in many cases of blow-up solutions, we have to make some changes in the general conditions over the kernel k in order to not exclude this kind of kernels. Specifically, we are going to replace the condition "locally bounded" on the kernel by "$\lim_{x\to a^+} \int_a^x k(x, s)\, \mathrm{d}s = 0$ for all $a \geq 0$". In [9] can be found a list of results about nondivergent existence and uniqueness taking into account these new general conditions.

We say that a collocation problem is a *blow-up collocation problem* (or has a blow-up) if the following conditions are held:

1. There exists $t^* > 0$ such that there is no collocation solution in $I = \left[0, t^*\right]$ for any mesh I_h.
2. Given $M > 0$ there exists $0 < \tau < t^*$, and a collocation solution z_h defined on $[0, \tau]$ such that $|z_h(x)| > M$ for some $x \in [0, \tau]$.

We can not speak about "blow-up collocation solutions" in the classic sense, since "collocation solutions" are defined in compact intervals and obviously they are bounded; so, we have to extend first the concept of "collocation solution" to half-closed intervals $I = [0, t^*)$ before we are in position to define the notion of "blow-up collocation solution".

Let $I := [0, t^*)$ and I_h be an infinite mesh given by a strictly increasing sequence $\{x_n\}_{n=0}^{+\infty}$ with $x_0 = 0$ and convergent to t^*.

- A *collocation solution on I* using the mesh I_h is a function defined on I such that it is a collocation solution (in the classic sense) for any finite submesh $\{x_n\}_{n=0}^{N}$ with $N \in \mathbb{N}$.
- A collocation solution on I is a *blow-up collocation solution* (or has a blow-up) with *blow-up time* t^* if it is unbounded.

Given a collocation problem with nondivergent uniqueness near zero, a necessary condition for the nondivergent collocation solution to blow-up is that there is neither existence for fine meshes nor unconditional existence. So, for example, given a convolution kernel $k(x-s)$, in cases 1 and 2 we must require that $\frac{g(u)}{u}$ is unbounded away from zero; moreover, in case 2 with $c_2 \neq 1$, we must demand that there exists $\epsilon > 0$ such that k is continuous in $(0, \epsilon)$. More examples and a numerical algorithm for detecting blow-ups can be found in [9].

References

1. Arias, M.: Un nuevo tipo de ecuaciones integrales de abel no lineales. Ph.D. thesis, Dpto. Matemticas. Universidad de Extremadura, Av. Elvas s/n. 06011 Badajoz (1999)
2. Arias, M.R.: Existence and uniqueness of solutions for nonlinear volterra equations. Math. Proc. Camb. Philos. Soc. **129**(2), 361370 (2000)
3. Arias, M., Benítez, R.: A note of the uniqueness and the attractive behaviour of solutions for nonlinear volterra equations. J. Integral Equ. Appl. **13**(4), 305–310 (2001)
4. Arias, M., Benítez, R.: Aspects of the behaviour of solutions of nonlinear abel equations. Nonlinear Anal. Theory Methods Appl. **54**(7), 1241–1249 (2003)
5. Arias, M., Benítez, R.: Properties of solutions for nonlinear volterra integral equations. In: Discrete and Continuous Dynamical Systems, Proceedings of the 4th International Conference on Dynamical Systems and Differential Equations, pp. 42–47 (2003)
6. Arias, M., Castillo, J.: Attracting solutions of nonlinear volterra integral equations. J. Integral Equ. Appl. **11**(3), 299–308 (1999). https://doi.org/10.1216/jiea/1181074279
7. Arias, M., Bentez, R., Bols, V.: Attraction properties of unbounded solutions for a nonlinear abel integral equation. J. Integral Equ. Appl. **19**(4), 439–452 (2007)

8. Benítez, R., Bolós, V.: Existence and uniqueness of nontrivial collocation solutions of implicitly linear homogeneous volterra integral equations. J. Comput. Appl. Math. **235**(12), 3661–3672 (2011)
9. Benítez, R., Bolós, V.: Blow-up collocation solutions of nonlinear homogeneous volterra integral equations. Appl. Math. Comput. **256**, 754–768 (2015)
10. Brunner, H.: Implicitly linear collocation methods for nonlinear volterra equations. Appl. Numer. Math. **9**(3), 235–247 (1992). https://doi.org/10.1016/0168-9274(92)90018-9
11. Brunner, H.: Collocation Methods for Volterra Integral and Related Functional Differential Equations, vol. 15. Cambridge University Press, Cambridge (2004)
12. Bushell, P.J.: On a class of volterra and fredholm non-linear integral equations. Math. Proc. Cambridge Philos. Soc. **79**(2), 329335 (1976). https://doi.org/10.1017/S0305004100052324
13. Gripenberg, G., Londen, S., Staffans, O.: Volterra Integral and Functional Equations. Cambridge Ocean Technology Series. Cambridge University Press, Cambridge (1990)
14. Maolepszy, T., Okrasiski, W.: Blow-up time for solutions to some nonlinear volterra integral equations. J. Math. Anal. Appl. **366**(1), 372–384 (2010). https://doi.org/10.1016/j.jmaa.2010.01.030
15. Mydlarczyk, W.: The blow-up solutions of integral equations. Colloq. Math. **79**(1), 147–156 (1999)
16. Nohel, J.: Some problems in nonlinear volterra integral equations. Bull. Am. Math. Soc. **68**(4), 323–329 (1962)
17. Szwarc, R.: Attraction principle for nonlinear integral operators of the volterra type. J. Math. Anal. Appl. **170**(2), 449–456 (1992). http://dx.doi.org/10.1016/0022-247X(92)90029-D
18. Sato, T.: Sur l'quation intgrale non linaire de volterra. Compos. Math. **11**, 271–290 (1953)
19. Yang, Z.W., Brunner, H.: Blow-up behavior of collocation solutions to hammerstein-type volterra integral equations. SIAM J. Numer. Anal. **51**(4), 2260–2282 (2013). https://doi.org/10.1137/12088238X

Chapter 15
Solving Random Ordinary and Partial Differential Equations Through the Probability Density Function: Theory and Computing with Applications

J. Calatayud, J.-C. Cortés, M. Jornet, and A. Navarro-Quiles

Abstract This contribution provides a practical view to the computation of the first probability density function of the solution stochastic process to ordinary and partial differential equations with randomness using the Random Variable Transformation technique. The analysis is performed via a set of simple examples, belonging to different areas like Physics, Biology and Engineering, with the aim of illustrating key ideas from a practical standpoint.

15.1 Introduction and Motivation

Differential equations play a prominent role in applications of Mathematics to other scientific areas such as Physics, Chemistry, Engineering, Biology, Epidemiology, Economy, etc. However, differential equations governing physical phenomena (in a wide sense) have inputs (initial and/or boundary conditions, forcing term and/or coefficients) that in practice need to be set from experimental data. Due to errors in the measurements and inherent uncertainty often involved in real phenomena, it is more realistic to consider those inputs as random variables or stochastic processes rather than numbers or deterministic functions, respectively. This approach leads to the area of random or stochastic differential equations. Accordingly, the solutions of such differential equations are stochastic processes. As a major difference with respect to what is done in the deterministic scenario, where the main goal is to calculate (exact or approximate) solutions of differential equations, in the random framework it is also important to find the relevant probabilistic information of

J. Calatayud · J.-C. Cortés (✉) · M. Jornet
Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Valencia, Spain
e-mail: jucagre@alumni.uv.es; jccortes@imm.upv.es; marjorsa@doctor.upv.es

A. Navarro-Quiles
DeustoTech, Universidadty of Deusto, Basque Country, Spain
e-mail: ananavarro@deusto.es

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_15

the solution stochastic process, say $X(t)$. This information is usually reported via the statistical moments of $X(t)$ such as mean, variance, kurtosis, asymmetry, etc. However, to complete this statistical information it is also desirable to compute the so-called *fidis* (finite distributions) of $X(t)$ [1, Ch. 3]. In particular, the first probability density function (1-PDF), $f_1(x,t)$, allows us to compute all one-dimensional statistical moments of $X(t)$,

$$\mathbb{E}[(X(t))^k] = \int_{-\infty}^{\infty} x^k f_1(x,t)\,\mathrm{d}x, \qquad k = 1, 2, \ldots. \tag{15.1}$$

As a consequence, from $f_1(x,t)$ we can obtain the mean, $\mathbb{E}[X(t)]$, and the variance, $\mathbb{V}[X(t)] = \mathbb{E}[X^2(t)] - (\mathbb{E}[X(t)])^2$.

In dealing with differential equations with uncertainty one usually distinguishes between two different kind of such equations, namely, stochastic differential equations (SDEs) and random differential equations (RDEs). Both SDEs and RDEs are distinctly different because of the nature of randomness acting on them. As a consequence, analysis and approximation of RDES and SDEs require completely different methods [2, pp. 96–97]. SDEs are forced by an irregular noise such as a Wiener process $W(t)$. They are usually written in terms of Riemann and Itô stochastic integrals. The calculation of the exact solution of SDEs is based on the application of the so-called Itô's lemma [3]. However, since most of the SDEs cannot be solved in an exact manner, a number of numerical schemes have been proposed [4]. In general, the order of numerical schemes designed for SDEs where uncertainty is considered via the Wiener process have lower order than their deterministic counterpart. RDEs are those in which random effects are directly manifested with inputs that are assumed to possess sample regular behaviour (e.g., sample continuous or differentiable) with respect to time. The computation of the 1-PDF of the solution stochastic process for both types of differential equations is a major goal. On the one hand, in the context of SDEs, it can be proved that this function solves the so-called Fokker-Planck-Kolmogorov partial differential equation [5]. In general, the exact solution of this equation is extremely difficult and one must rely on numerical techniques. On the other hand, a method that has demonstrated its effectiveness to determine the 1-PDF of RDEs is the Random Variable Transformation (RVT) technique.

The objective of this chapter is twofold. Firstly, we will show how the RVT technique can be applied to compute the 1-PDF of RDEs. With the aim of providing a broader view of the application of this technique, we shall obtain the 1-PDF of the solution stochastic process of ordinary and partial RDEs. Secondly, we will illustrate the usefulness of computing the 1-PDF to fit a mathematical model using real data. In the context of this application, it is also a primary goal of this contribution to address the inverse problem which consists of determining appropriate probability distributions for model inputs using a Bayesian approach.

The chapter is organized as follows. Section 15.2 is devoted to introduce RVT technique and to illustrate its application to determine the 1-PDF of an absolutely continuous stochastic process. Section 15.3 is divided into two parts, the first one

addresses the computation of the 1-PDF of the solution stochastic process of a random ordinary differential equation that appears in a physical problem, while in the second part we show a Bayesian technique for determining the probability distributions of input data of a randomized differential equation modelling the fish weight using real data. In Sect. 15.4 we illustrate, by means of a randomized nonlinear partial differential equation, how the RVT technique can also be applied to obtain the 1-PDF of the solution stochastic process in the context of partial differential equations. Conclusions are drawn in Sect. 15.5.

15.2 A Glance to the RVT Technique

The RVT technique is a powerful method that permits the computation of the PDF of a random vector which is obtained after mapping another random vector whose PDF is given. There exist several formulations of this result [6–8], below we state the one that will be used throughout this chapter.

Lemma 15.1 ([9] Random Variable Transformation Technique) *Let X be an absolutely continuous random vector with density f_X and with support D_X contained in an open set $D \subseteq \mathbb{R}^n$. Let $g : D \to \mathbb{R}^n$ be a $C^1(D)$ function, injective on D such that $J[g](x) \neq 0$ for all $x \in D$ (J stands for Jacobian). Let $h = g^{-1} : g(D) \to \mathbb{R}^n$. Let $Y = g(X)$ be a random vector. Then Y is absolutely continuous with density*

$$f_Y(y) = \begin{cases} f_X(h(y))|J[h](y)|, & y \in g(D), \\ 0, & y \notin g(D). \end{cases}$$

Although this result is formulated for random variables (vectors), we can still take advantage of it in the context of stochastic processes, say $X(t)$, as by definition for each $\hat{t}$ fixed, $X(\hat{t})$ is a random variable. Then, applying Lemma 15.1 one obtains the PDF of the random variable $X(\hat{t})$. If this expression is valid for every t, then one gets the so-called 1-PDF of $X(t)$, $f_1(x, t)$. Analogously, if two different time instants, say $\hat{t}_1$ and $\hat{t}_2$, are fixed, then applying Lemma 15.1 one can calculate the joint PDF of the random vector $(X(\hat{t}_1), X(\hat{t}_2))$, and letting (t_1, t_2) be arbitrary, one can deduce the so-called second probability density function (2-PDF) of $X(t)$, $f_2(x_1, t_1; x_2, t_2)$, and so on.

To illustrate this procedure, next we show an example where the 1-PDF is computed. Despite the simplicity of the stochastic process involved in this example (it is just a linear transformation of two independent uniform random variables), it is interesting to notice that the computation of the 1-PDF is not trivial. In the context of random ordinary and partial differential equations the computation of the 1-PDF becomes harder.

Example 15.1 Let us consider the stochastic process $Y(t) = At + B$, $t > 0$ where A and B are independent and identically random variables uniformly distributed on

the interval]0, 1[($A, B \sim$ Un(]0, 1[)). Let us fix $t > 0$ and consider the following transformation mapping $g : D =]0, 1[\times]0, 1[\longrightarrow \mathbb{R}^2$, where $y_1 = g_1(a, b) = at + b$ and $y_2 = g_2(a, b) = b$ which is $C^1(D)$ and injective. Moreover, its inverse transformation mapping is $h = g^{-1} :]0, t + 1[\times]0, 1[\longrightarrow \mathbb{R}^2$, $h_1(y_1, y_2) = (y_1 - y_2)/t$ and $h_2(y_1, y_2) = y_2$. It is straightforward to check that the Jacobian of h is $|J[h](y_1, y_2)| = 1/t \neq 0$. Then, according to Lemma 15.1, the joint PDF of the random vector (Y_1, Y_2) is given by

$$f_{Y_1,Y_2}(y_1, y_2) = f_{A,B}\left(\frac{y_1 - y_2}{t}, y_2\right)\frac{1}{t}.$$

Observe that as usual, we write random variables/vectors by capital letters while deterministic quantities are denoted in lower case letters. By marginalizing with respect to Y_2 and taking into account that t is arbitrary, one obtains the 1-PDF of $Y(t)$:

$$\begin{aligned} f_1(y, t) &= \int_0^1 f_{Y_1,Y_2}(y_1, y_2)\mathrm{d}y_2 \qquad = \int_0^1 f_{A,B}\left(\frac{y-b}{t}, b\right)\frac{1}{t}\mathrm{d}b \\ &= \frac{1}{t}\int_0^1 f_A\left(\frac{y-b}{t}\right) f_B(b)\mathrm{d}b = \frac{1}{t}\int_0^1 f_A\left(\frac{y-b}{t}\right)\mathrm{d}b \\ &= \frac{1}{t}, \qquad b \leq y \leq b + t, \end{aligned}$$

where independence between the random variables A and B and that the PDF of $B \sim$ Un(]0, 1[) is $f_B(b) = 1$, $0 < b < 1$, have been used. Observe that the interval where y belongs to guarantees $(y - b)/t \in]0, 1[$, as requested, so that the last integral makes sense. Notice that $b < y < b + t$ entails $y - t < b < y$, hence $f_1(y, t)$ can be expressed as follows:

$$f_1(y, t) = \frac{1}{t}\int_{\max(0,y-t)}^{\min(1,y)} 1\,\mathrm{d}b.$$

To give an explicit expression of $f_1(y, t)$ it is necessary to split the integral in four cases:

Case 1: $y \geq t$ and $y \geq 1$

$$f_1(y, t) = \frac{1}{t}\int_{y-t}^1 1\mathrm{d}b = 1 + \frac{1}{t}(1 - y).$$

Case 2: $y \geq t$ and $0 \leq y \leq 1$

$$f_1(y, t) = \frac{1}{t}\int_{y-t}^y 1\mathrm{d}b = 1.$$

Case 3: $y \leq t$ and $y \geq 1$

$$f_1(y,t) = \frac{1}{t}\int_0^1 1\mathrm{d}b = \frac{1}{t}.$$

Case 4: $y \leq t$ and $0 \leq y \leq 1$

$$f_1(y,t) = \frac{1}{t}\int_0^y 1\mathrm{d}b = \frac{y}{t}.$$

Now, we check that the integral of this function is the unit:

$$\int_{-\infty}^{\infty} f_1(y,t)\mathrm{d}y = \int_1^{1+t} 1 + \frac{1}{t}(1-y)\mathrm{d}y + \int_t^1 1\mathrm{d}y + \int_0^t \frac{y}{t}\mathrm{d}y = 1$$

where we have taken into account that $0 < y < 1 + t$, since $Y(t) = At + B$, $t > 0$, and A, B are uniformly distributed on $]0, 1[$.

15.3 Computing the 1-PDF in the Context of Random Ordinary Differential Equations

This section is addressed to show, in a practical way, how the 1-PDF of the solution stochastic process can be computed in the context of random ordinary differential equations. Moreover, we will illustrate the Bayesian technique to determine the probability distributions of inputs when a RDE is considered to model real data.

Before addressing our particular study, it is convenient to point out that the computation of the 1-PDF of the solution stochastic process of random ordinary differential equations has been studied in recent contributions. On the one hand, in [10] random linear first-order differential equations have been studied; in [11] one extends the corresponding analysis for random linear second-order differential equations; in [12], one addresses the study of an important class of random nonlinear differential equation, namely, the Bernoulli equation. On the other hand, some applications in Epidemiology and Biology are shown in [13, 14], respectively.

15.3.1 The Nonlinear Random Differential Equation for a Falling Body

Let us consider the differential equation governing the velocity $V(t)$ of a falling body of mass m in a medium where the resistance is proportional to the square of the velocity

$$V'(t) + \frac{p}{m}V^2(t) = g, \quad t > 0, \tag{15.2}$$

being $p > 0$ a physical constant related to the resistance and g the gravity constant. Let us assume that the initial velocity is unknown in a deterministic way because of measurement errors and hence it is assumed to be a positive random variable, say V_0. Hereinafter, $f_{V_0}(v_0)$ will denote its PDF. For the sake of simplicity in this first example, uncertainty is only considered through the initial condition. The solution of the Riccati differential equation (15.2) with random initial condition $V(0) = V_0$ is given by

$$V(t) = a\frac{a\tanh(bt) + V_0}{a + V_0\tanh(bt)}, \quad t \geq 0, \quad \text{where} \quad a = \sqrt{\frac{m\,g}{p}} > 0, \quad b = \sqrt{\frac{p\,g}{m}} > 0. \tag{15.3}$$

For each $t \geq 0$ fixed, $V(t) = V$ can be seen as a transformation mapping $g : D =]0, +\infty[\longrightarrow \mathbb{R}$ defined by $g(v_0) = (a^2\tanh(bt) + a v_0)/(a + v_0\tanh(bt))$, which satisfies

$$\frac{\mathrm{d}g}{\mathrm{d}v_0} = \frac{ab(a - v_0)(a + v_0)}{(a\cosh(bt) + v_0\sinh(bt))^2}.$$

Therefore g is injective (increasing if $0 < v_0 < a$ and decreasing if $v_0 > a$). Moreover, its inverse h and the Jacobian of its inverse are given by

$$v_0 = h(v) = \frac{a v - a^2\tanh(bt)}{a - v\tanh(bt)}, \qquad J[h](v) = \frac{\mathrm{d}h}{\mathrm{d}v} = \frac{a^2}{(a\cosh[bt] - v\sinh[bt])^2} > 0,$$

respectively. According to Lemma 15.1, the PDF f_V of random variable V can be computed in terms of the PDF f_{V_0}, and taking $t \geq 0$ arbitrary one also obtains the 1-PDF of the velocity stochastic process

$$f_1(v, t) = f_{V_0}\left(\frac{a v - a^2\tanh(bt)}{a - v\tanh(bt)}\right)\frac{a^2}{(a\cosh(bt) - v\sinh(bt))^2}, \tag{15.4}$$

where the positive constants a and b are defined in (15.3).

In order to illustrate numerically the above theoretical results, let us assume that $m = 100\,\mathrm{kg}$, $p = 10\,\mathrm{kg/m}$, $g = 9.8\,\mathrm{m/s^2}$ and that the initial velocity V_0 has a uniform distribution on the interval $[1.4, 1.6]$. In agreement with (15.4), the 1-PDF of the velocity of the falling body is given by

$$f_1(v, t) = \frac{490}{\left(\sqrt{98}\cosh\left(\frac{\sqrt{98}}{10}t\right) - v\sinh\left(\frac{\sqrt{98}}{10}t\right)\right)^2} \geq 0, \quad t \geq 0, \;\; v_{0,1} \leq v \leq v_{0,2}, \tag{15.5}$$

where

$$v_{0,1} = 70 - \frac{343\sqrt{2}}{\tanh\left(\frac{7t}{5\sqrt{2}}\right) + 5\sqrt{2}}, \qquad v_{0,2} = \frac{7}{4}\left(35 - \frac{1193\sqrt{2}}{8\tanh\left(\frac{7t}{5\sqrt{2}}\right) + 35\sqrt{2}}\right).$$

It can be checked that $\int_{v_{0,1}}^{v_{0,2}} f_1(v,t)\mathrm{d}v = 1$, thus $f_1(v,t)$ is really a PDF for every $t \geq 0$. Furthermore, the mean of the velocity is

$$\begin{aligned}\mathbb{E}[V(t)] &= \int_{v_{0,1}}^{v_{0,2}} v f_1(v,t)\mathrm{d}v \\ &= \tfrac{7}{2}\mathrm{csch}^2\left(\tfrac{7t}{5\sqrt{2}}\right)\left(\sqrt{2}\sinh\left(\tfrac{7\sqrt{2}t}{5}\right) - 140\log\left(\frac{1}{35\sqrt{2}\coth\left(\frac{7t}{5\sqrt{2}}\right)+7} + 1\right)\right),\end{aligned}$$

where $\mathrm{csch}(\cdot)$ and $\coth(\cdot)$ stand for the hyperbolic cosecant and cotangent, respectively. It is instructive to observe that

$$\lim_{t\to 0}\mathbb{E}[V(t)] = 1.5, \quad \lim_{t\to\infty}\mathbb{E}[V(t)] = 0,$$

as expected.

15.3.2 Bayesian Computation of the Parameters of a Fish Weight Growth Model

In this section we consider the random Bertalanffy model that has been applied to describe the evolution of the time of fish weights [15, p. 331]. In [12], authors randomized the deterministic Bertalanffy model

$$\begin{cases} W'(t) = -\lambda W(t) + \eta(W(t))^{2/3}, \ \ t \geq t_0, \\ W(t_0) = W_0, \end{cases} \tag{15.6}$$

where $W(t)$ denotes the fish weight at the time instant t and λ, η and W_0 are assumed to be absolutely continuous random variables on a common complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Notice that Bertalanffy model (15.6) is formulated via a Bernoulli differential equation. Assuming that $\lambda \neq 0$ almost surely, the solution stochastic process of this random Cauchy problem is given by

$$W(t) = \left(W_0^{\frac{1}{3}} e^{-\frac{1}{3}\lambda(t-t_0)} - \frac{\eta}{\lambda} e^{\frac{1}{3}\lambda(t-t_0)} + \frac{\eta}{\lambda}\right)^3. \tag{15.7}$$

Taking advantage of the RVT method stated in Lemma 15.1, in [12] one determines the 1-PDF of the solution stochastic process $W(t)$,

$$f_1(w,t) = \int_{\mathscr{D}(\eta)} \int_{\mathscr{D}(\lambda)} f_{W_0,\eta,\lambda}\left(\left(\frac{e^{(1/3)\lambda(t-t_0)}\lambda w^{1/3} + \eta - e^{(1/3)\lambda(t-t_0)}\eta}{\lambda}\right)^3, \eta, \lambda\right) \times \left(\frac{e^{(1/3)\lambda(t-t_0)}\lambda w^{1/3} + \eta - e^{(1/3)\lambda(t-t_0)}\eta}{\lambda}\right)^2 e^{(1/3)\lambda(t-t_0)}|w|^{-2/3} \mathrm{d}\lambda\, \mathrm{d}\eta, \tag{15.8}$$

where $\mathscr{D}(\eta)$ and $\mathscr{D}(\lambda)$ denote the domains of random inputs η and λ, respectively. In [12] it is presented an inverse frequentist technique to assign a reliable distribution to input model parameter (W_0, η, λ), [2, ch. 7]. Due to the own foundations of frequentist technique, a trivariate Gaussian distribution is attributed to random vector (W_0, η, λ). Now, we shall illustrate an alternative approach to deal with the key problem of assigning an adequate probability distribution to parametres of a random ordinary differential equation in the context of the Bertalanffy model. This method is based on the so-called Bayesian approach [16] and [2, ch. 8].

In Fig. 15.1, we show the fish weight in lbs (vertical axis) per year (horizontal axis). The fish weight datum at the i-th year will be denoted by w_i, for $1 \leq i \leq 33$.

We have data on fish weights $w_1, \ldots, w_{33}$ at years $t_1 = 1, \ldots, t_{33} = 33$ (that is, realizations $w_i = W_i(\omega)$, $\omega \in \Omega$ of $W_i := W(t_i)$ for $i = 1, \ldots, 33$). In this framework, we deal with the following Random Inverse Parameter Estimation Problem, namely, to find random variables W_0, η and λ that fit data shown in Fig. 15.1 to the random Bertalanffy model whose solution is given in (15.7). To

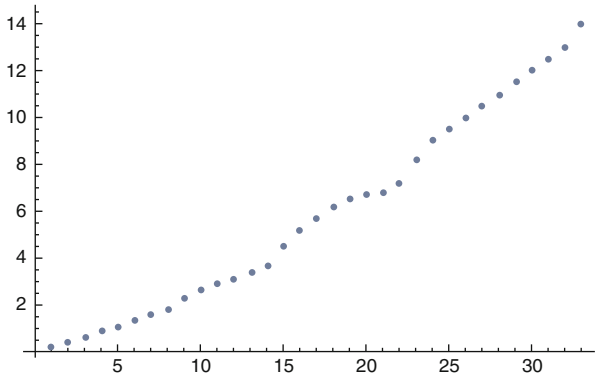


Fig. 15.1 Fish weights data [12]. The horizontal axis represents time measure in years and the vertical axis represents weight measure in lbs

this aim, we propose the following hierarchical Bayesian model:

$$(W_1, \ldots, W_N)|(W_0, \eta, \lambda, \tau) \sim \prod_{i=1}^{N} \mathrm{N}\left(\left\{ W_0^{\frac{1}{3}} e^{-\frac{1}{3}\lambda(t_i - t_0)} - \frac{\eta}{\lambda} e^{\frac{1}{3}\lambda(t_i - t_0)} + \frac{\eta}{\lambda} \right\}^3, \tau\right),$$

$$(W_0, \eta, \lambda)|(\mu, A) \sim \mathrm{N}_3(\mu, A),\ \mu \sim \mathrm{N}_3(0, 0.1\, I_3),\ A \sim \text{Wishart}(I_3, 3),\ ^{1}\ \tau \sim \mathrm{Ga}(0.1, 0.1),$$

where $N = 33$, τ is the precision (inverse of the variance σ^2) and A is the inverse of the covariance matrix (following the notation from the software `WinBUGS`, where the Bayesian model is implemented). The random variables/vectors τ, μ and A are assumed to be independent.

The joint posterior density of the vector parameter $(W_0, \eta, \lambda, \tau, \mu, A)$ is:

$$\begin{aligned} &p_{W_0,\eta,\lambda,\tau,\mu,A|W_1,\ldots,W_N}(w_0, \eta, \lambda, \tau, \mu, a|w_1, \ldots, w_N) \\ &= \frac{\prod_{i=1}^{N} p_{W_i|W_0,\eta,\lambda,\tau}(w_i|w_0, \eta, \lambda, \tau) p_\tau(\tau) p_{W_0,\eta,\lambda|\mu,A}(w_0, \eta, \lambda|\mu, a) p_\mu(\mu) p_A(a)}{\int_{\mathscr{D}} \prod_{i=1}^{N} p_{W_i|W_0,\eta,\lambda,\tau}(w_i|\tilde{w}_0, \tilde{\eta}, \tilde{\lambda}, \tilde{\tau}) p_\tau(\tilde{\tau}) p_{W_0,\eta,\lambda|\mu,A}(\tilde{w}_0, \tilde{\eta}, \tilde{\lambda}|\tilde{\mu}, \tilde{a}) p_\mu(\tilde{\mu}) p_A(\tilde{a})\, \mathrm{d}\tilde{a}\, \mathrm{d}\tilde{\mu}\, \mathrm{d}\tilde{\tau}\, \mathrm{d}\tilde{\lambda}\, \mathrm{d}\tilde{\eta}\, \mathrm{d}\tilde{w}_0}, \end{aligned} \tag{15.9}$$

where $\mathscr{D} = \mathscr{D}(A) \times \mathbb{R}^3 \times (0, \infty) \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ and $\mathscr{D}(A)$ is the set of vectors in $\mathbb{R}^{3\times 3}$ such that before vectorization they formed a symmetric positive definite matrix. The posterior density for each of the parameters, $p(\cdot|w_1, \ldots, w_N)$, is computed via the marginal distributions of (15.9). The posterior predictive distribution of W_i (in a more comprehensible language, its 1-PDF) is given by

$$\begin{aligned} &p_{W_i|W_1,\ldots,W_N}(\tilde{w_i}|w_1, \ldots, w_N) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} \int_0^\infty p_{W_i|W_0,\eta,\lambda,\tau}(\tilde{w_i}|w_0, \eta, \lambda, \tau) p_{W_0,\eta,\lambda,\tau|W_1,\ldots,W_N}(w_0, \eta, \lambda, \tau|w_1, \ldots, w_N)\, \mathrm{d}\tau\, \mathrm{d}\lambda\, \mathrm{d}\eta\, \mathrm{d}w_0. \end{aligned} \tag{15.10}$$

Since these formulas cannot be computed analytically, the computations are performed in `WinBUGS`. We simulate a sample from the posterior distribution of the parameters. We run 900,000 iterations with a burn-in period of 600,000. We assess convergence via two chains with different initial conditions. Table 15.1 shows the

[1] The Whishart distribution is a probability distribution for random matrices. A random matrix is a matrix whose entries are random variables. We say that a $k \times k$ random matrix A is absolutely continuous (respectively discrete) if its vectorization, $\mathrm{vec}(A)$, is an absolutely continuous (respectively discrete) $k^2 \times 1$ random vector. In this case, the density or mass function of A is defined as the density or mass function of $\mathrm{vec}(A)$. We say that a $k \times k$ random matrix A follows a Wishart distribution, $A \sim \text{Wishart}(H, n)$, H being $k \times k$ symmetric and positive definite matrix and $n \geq k$, if it is symmetric, positive definite and absolutely continuous, with density

$$P_A(a) = c[\det(a)]^{\frac{n}{2} - \frac{k+1}{n}} e^{-\frac{n}{2}\mathrm{tr}(H^{-1}a)},$$

being c a constant depending on H and n.

Table 15.1 Posterior mean and 0.95 credible interval for the model parameters

Parameter	Mean	Credible interval
σ	0.242	(0.188, 0.315)
μ_1	0.296	(−2.375, 2.844)
μ_2	0.267	(−2.391, 2.831)
μ_3	0.079	(−2.525, 2.680)
a_{11}	3.148	(0.289, 9.524)
a_{12}	0.003	(−3.625, 3.632)
a_{13}	0.001	(−3.626, 3.635)
a_{21}	0.003	(−3.625, 3.632)
a_{22}	3.160	(0.294, 9.555)
a_{23}	−0.006	(−3.630, 3.623)
a_{31}	0.001	(−3.626, 3.635)
a_{32}	−0.006	(−3.630, 3.623)
a_{33}	3.158	(0.292, 9.568)

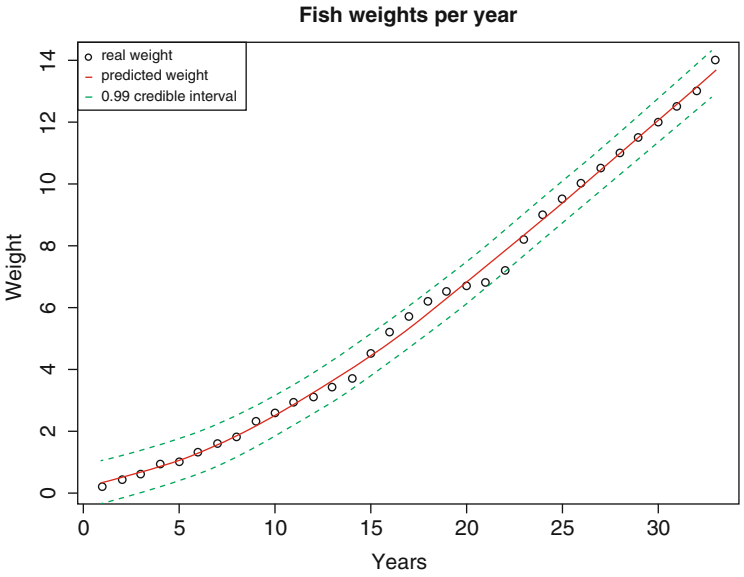


Fig. 15.2 Fish weights given by the model

mean and a 0.95 credible interval for $\sigma = 1/\sqrt{\tau}$, $\mu = (\mu_1, \mu_2, \mu_3)$ and A (with entries a_{ij}).

In Fig. 15.2 we show the fit of the model to the data. All points lie in the credible region. We conclude that the model explains data in a good way. These results are in agreement with the ones reported by some of the authors in [12], where data were fitted using an Inverse Frequentist technique for parameter estimation.

We plot the PDF of random variables $W_1, \ldots, W_{15}$ in Fig. 15.3. The density function of W_i is given by its posterior predictive distribution (15.10). We use

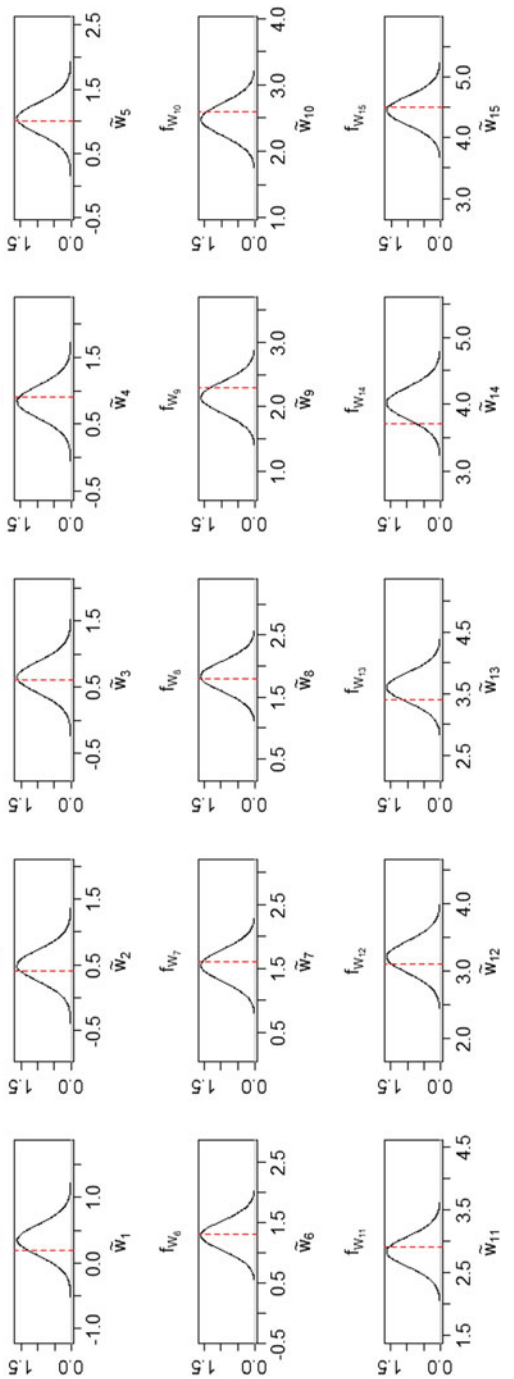


Fig. 15.3 Density function (posterior predictive distribution) of the weight random variables $W_1, \ldots, W_{15}$. The red lines represent the data weights $w_1, \ldots, w_{15}$

the samples for the parametres generated by `WinBUGS` to simulate the posterior predictive distribution. Notice that Fig. 15.3 is an alternative to the computation of the 1-PDF (15.8) obtained via an inverse frequentist technique and RVT technique in [12].

15.4 Probability Density Function of a Soliton Solution of the Random Nonlinear Dispersive Partial Differential Equation

As indicated in Sect. 15.1, it is also an objective of this chapter to illustrate the computation of the 1-PDF of the solution stochastic process in the context of random partial differential equations. To the best of our knowledge, there are few contributions dealing with these kind of problems. In [17, 18] one can find some contributions in this regard in the context of particular problems of great interest in Physics and Engineering.

Now, our goal is to obtain the 1-PDF of a soliton solution of the random nonlinear dispersive partial differential equation (PDE). The nonlinear dispersive PDE that we will deal with, in its deterministic version, is given by

$$u^n(u^n)_t + a(u^{3n})_x + d\,u^n(u^n)_{xxx} = 0, \tag{15.11}$$

where $n > 0$ and a and d are real constants.

Determinsitic nonlinear dispersive PDEs, including existence and stability of solitary and periodic travelling wave solutions, have been studied in many contributions [19]. To find a soliton solution to this deterministic PDE, we will use the Bernoulli method [20, 21]. Once we have the soliton solution, we will randomize this deterministic PDE. The soliton solution to the random nonlinear dispersive PDE will become a stochastic process. In order to compute its probability density function, we will make use of the Random Variable Transformation (RVT) technique.

15.4.1 Bernoulli Method

Consider a general deterministic PDE problem for the function $u(x,t)$:

$$f(u, u_t, u_x, u_{tx}, u_{xx}, u_{txx}, \ldots) = 0.$$

We look for a solution of the form $u(x,t) = U(x - ct)$, where $c \in \mathbb{R}$. This type of solutions are called solitons. Physically, they represent a solitary travelling wave, that is, a wave that propagates with no change of its properties (velocity, shape, etc.).

Using the chain rule, we derive that $U(\xi)$ verifies an ODE:

$$F(U, U', U'', \ldots) = 0. \tag{15.12}$$

We look for solutions U of the form

$$U(\xi) = a_0 + \sum_{i=1}^{M} a_i G^i(\xi), \tag{15.13}$$

where $G(\xi)$ satisfies a Bernoulli ODE

$$G' + \lambda G = \mu G^2, \tag{15.14}$$

where $\lambda \neq 0$ and $\mu \neq 0$. The solution to this Bernoulli ODE is given by

$$G(\xi) = \frac{1}{\frac{\mu}{\lambda} + be^{\lambda \xi}}, \tag{15.15}$$

where b is a constant. Substituting (15.13) in the ODE (15.12) and using the relation (15.14), we obtain a polynomial on $G(\xi)$ that is equal to 0. We equate the coefficients of this polynomial to 0 to obtain $a_0, a_1, \ldots, a_M$. The superscript M is chosen in such a way that there exist $a_0, a_1, \ldots, a_M$.

15.4.2 Application of the Bernoulli Method to Find a Soliton Solution for the Deterministic Nonlinear Dispersive PDE

The goal in this subsection is to find a solution for the deterministic nonlinear dispersive equation (15.11) using the Bernoulli method described in Sect. 15.4.1.

First of all, we simplify the deterministic PDE problem (15.11). Set $v = u^n$, so that

$$0 = vv_t + a(v^3)_x + dvv_{xxx} = vv_t + 3av^2v_x + dvv_{xxx}.$$

Divide by v:

$$0 = v_t + 3avv_x + dv_{xxx} = v_t + \frac{3}{2}a(v^2)_x + dv_{xxx}.$$

Now, we look for a solution of the form $v(x, t) = V(x - ct)$. This transforms the PDE problem in the following ODE problem for $V(\xi)$ with $\xi = x - ct$:

$$-cV' + \frac{3}{2}a(V^2)' + dV''' = 0.$$

Integrating with respect to ξ and setting the constant of integration to zero,

$$-cV + \frac{3}{2}aV^2 + dV'' = 0. \tag{15.16}$$

Bernoulli method proposes to find a solution of the form

$$V(\xi) = a_0 + \sum_{i=1}^{M} a_i G^i(\xi),$$

where G satisfies the Bernoulli ODE (15.14).

Suppose that $M = 1$, so that $V = a_0 + a_1 G$, $a_1 \neq 0$. Then $V'' = a_1 G''$, where

$$G'' = (-\lambda G + \mu G^2)' = -\lambda G' + 2\mu G G' = (-\lambda + 2\mu G)G' = (-\lambda + 2\mu G)(-\lambda G + \mu G^2).$$

Therefore, $V'' = a_1(-\lambda + 2\mu G)(-\lambda G + \mu G^2)$, which is a polynomial of degree three on $G(\xi)$ (because $a_1 \neq 0$ and $\mu \neq 0$). But $-cV + (3/2)aV^2$ is a polynomial of degree two on $G(\xi)$, so the equality in (15.16) is not possible.

Thus, we need to try with $M = 2$. In this case, as we shall see, we will obtain a solution for (15.16). Let $V = a_0 + a_1 G + a_2 G^2$. Differentiate twice:

$$V' = a_1 G' + 2a_2 G G',$$

$$\begin{aligned} V'' &= a_1 G'' + 2a_2 (G')^2 + 2a_2 G G'' = G''(a_1 + 2a_2 G) + 2a_2 (G')^2 \\ &= (-\lambda + 2\mu G)(-\lambda G + \mu G^2)(a_1 + 2a_2 G) + 2a_2 G^2 (-\lambda + \mu G)^2 \\ &= \lambda^2 a_1 G + (4\lambda^2 a_2 - 3\lambda\mu a_1) G^2 + (-10\lambda\mu a_2 + 2\mu^2 a_1) G^3 + 6\mu^2 a_2 G^4. \end{aligned}$$

From (15.16),

$$\begin{aligned} 0 = & -cV + \frac{3}{2}aV^2 + dV'' = -ca_0 - ca_1 G - ca_2 G^2 \\ & + \frac{3}{2}a[a_0^2 + 2a_0 a_1 G + (2a_0 a_2 + a_1^2) G^2 + 2a_1 a_2 G^3 + a_2^2 G^4] \\ & + d\lambda^2 a_1 G + (4d\lambda^2 a_2 - 3d\lambda\mu a_1) G^2 + (-10d\lambda\mu a_2 + 2d\mu^2 a_1) G^3 + 6d\mu^2 a_2 G^4 \\ = & \left(\frac{3}{2}aa_0^2 - ca_0\right) + (d\lambda^2 a_1 + 3aa_0 a_1 - ca_1) G \\ & + \left(-3d\lambda\mu a_1 + 3aa_0 a_2 + \frac{3}{2}aa_1^2 - ca_2 + 4d\lambda^2 a_2\right) G^2 \\ & + (-10\lambda\mu a_2 d + 2\mu^2 a_1 d + 3aa_1 a_2) G^3 + \left(6\mu^2 a_2 d + \frac{3}{2}aa_2^2\right) G^4. \end{aligned}$$

Matching the coefficients of this polynomial on $G(\xi)$ to 0, we obtain a system of nonlinear equations. The solutions of this system, obtained with Mathematica®, are:

$$a_0 = 0,\ a_1 = \frac{4d\lambda\mu}{a},\ a_2 = -\frac{4d\mu^2}{a},\ c = d\lambda^2,$$

$$a_0 = -\frac{2d\lambda^2}{3a},\ a_1 = \frac{4d\lambda\mu}{a},\ a2 = -\frac{4d\mu^2}{a},\ c = -d\lambda^2.$$

Put these values on $V(\xi) = a_0 + a_1 G(\xi) + a_2 G^2(\xi)$, where $G(\xi)$ is given by the solution (15.15) of the Bernoulli ODE (15.14). Finally, $u(x,t) = V(x - ct)^{1/n}$ is a solution of the PDE (15.11). This procedure gives two solutions:

$$u_1(x,t) = \left[\frac{4d\mu\lambda^3 b e^{\lambda(x-dt\lambda^2)}}{a\left(\lambda b e^{\lambda(x-dt\lambda^2)} + \mu\right)^2}\right]^{\frac{1}{n}}, \tag{15.17}$$

and

$$u_2(x,t) = \left(\frac{2}{3}\right)^{\frac{1}{n}} \left(-\frac{d\lambda^2\left(b^2\lambda^2 e^{2\lambda\left(d\lambda^2 t+x\right)} - 4b\lambda\mu e^{\lambda\left(d\lambda^2 t+x\right)} + \mu^2\right)}{a\left(b\lambda e^{\lambda\left(d\lambda^2 t+x\right)} + \mu\right)^2}\right)^{\frac{1}{n}}.$$

In Fig. 15.4, we have plotted the travelling solution $u_1(x,t)$ for particular values of parameters a, b, d, μ, λ and n.

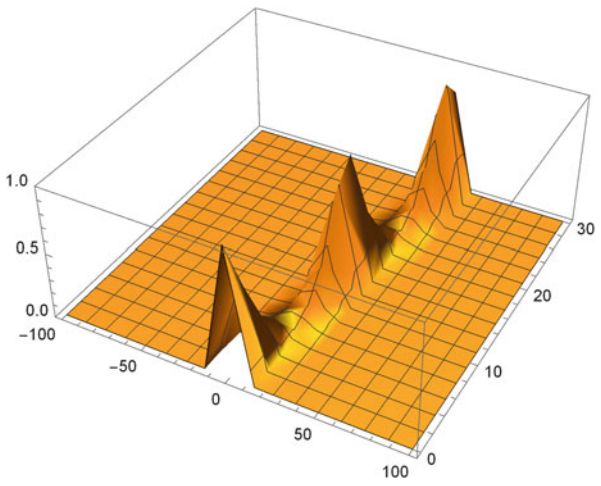


Fig. 15.4 Solution $u_1(x,t)$ for $a = 1$, $\mu = 1$, $\lambda = 1$, $d = 1$, $b = 1$ and $n = 2$. In this case, $u_1(x,t) = 2\sqrt{e^{t+x}/\left(e^t + e^x\right)^2}$

15.4.3 Obtaining the Probability Density Function of the Soliton Solution

Suppose that the coefficients a and d in the nonlinear dispersive equation (15.11) are random variables on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$. In this new setting, the random PDE (15.11) may have different interpretations. We could say that a stochastic process

$$u = \{u(x,t)(\omega) : (x,t) \in \mathscr{D},\ \omega \in \Omega\}$$

is a solution to the random PDE (15.11) if:

- it is an almost sure solution. That is, for almost every $\omega \in \Omega$, $u(\cdot,\cdot)(\omega) \in C^{3,1}(\mathscr{D})$ (it has three classical derivatives on x and one classical derivative on t) and

$$u^n(x,t)(\omega)(u^n(x,t)(\omega))_t + a(\omega)(u^{3n}(x,t)(\omega))_x + d(\omega)\, u^n(x,t)(\omega)(u^n(x,t)(\omega))_{xxx} = 0,$$

  for all $(x,t) \in \mathscr{D}$.
- it is a mean square solution. That is, u is three times mean square differentiable with respect to x, it is one time mean square differentiable with respect to t, and satisfies the random PDE (15.11) in the mean square sense. For a detailed exposition on the mean square differentiation theory, see Chapter 4 in [1].

Looking for solutions in an almost sure sense is exactly as in the deterministic case, for each fixed $\omega \in \Omega$. In the analysis performed in Sect. 15.4.2, we could consider that, for each $\omega \in \Omega$, we take a $c(\omega)$ for the change of variable and a $\mu(\omega)$ and $\lambda(\omega)$ to define a stochastic process $G(\xi)(\omega)$ that satisfies the random Bernoulli ODE

$$G'(\xi)(\omega) + \lambda(\omega)G(\xi)(\omega) = \mu(\omega)G(\xi)(\omega)^2$$

almost surely. In this way, a solution for the random PDE problem (15.11) is given by randomizing expression (15.17), i.e.,

$$u_1(x,t)(\omega) = \left[\frac{4d(\omega)\mu(\omega)\lambda(\omega)^3 b(\omega) e^{\lambda(\omega)(x-d(\omega)t\lambda(\omega)^2)}}{a(\omega)\left(\lambda(\omega)b(\omega)e^{\lambda(\omega)(x-d(\omega)t\lambda(\omega)^2)} + \mu(\omega)\right)^2}\right]^{\frac{1}{n}}. \tag{15.18}$$

The search of mean square solutions is slightly more complicated [1, Ch. 4], [22, 23]. Following the same reasoning as in Sect. 15.4.2, the key fact consists in proving that the stochastic process $G(\xi)(\omega)$ is a mean square solution of the random

Bernoulli ODE (15.14). To differentiate

$$G(\xi)(\omega) = \frac{1}{\frac{\mu(\omega)}{\lambda(\omega)} + b(\omega)e^{\lambda(\omega)\xi}}$$

in the mean square sense, we need to check, essentially, that the mean square differentiation of a quotient and of the exponential function follows the same rules as in the deterministic calculus.

Given a stochastic process $X(\xi)(\omega)$ that is mean fourth differentiable and such that $|X(\xi)(\omega)| \geq C$ for certain constant $C > 0$, the mean square derivative of $1/X$ exists and is given by $-X'/X^2$. We apply this result to $X(\xi)(\omega) = \mu(\omega)/\lambda(\omega) + b(\omega)e^{\lambda(\omega)\xi}$. Essentially, we need to prove that $b(\omega)e^{\lambda(\omega)\xi}$ is mean fourth differentiable, that is,

$$\lim_{h\to 0} \left\| b\frac{e^{\lambda(\xi+h)} - e^{\lambda\xi}}{h} - b\lambda e^{\lambda\xi} \right\|_{\mathrm{L}^4(\Omega)} = 0. \tag{15.19}$$

Here $(\mathrm{L}^4(\Omega), \|\cdot\|_4)$ stands for the Banach space of real random variables having fourth-order moment, i.e., $\mathbb{E}[X^4] < +\infty$ and $\|X\|_4 = (\mathbb{E}[X^4])^{1/4}$. Notice that (15.19) is a consequence of the Mean Value Theorem applied to the function e^x and of the Dominated Convergence Theorem, assuming that the dominating random variable $|b(\omega)|e^{\lambda(\omega)\xi}|\lambda(\omega)|(e^{|\lambda(\omega)|h_0} + 1)$ belongs to $\mathrm{L}^4(\Omega)$, for certain $h_0 > 0$ (for example, if b and λ are bounded random variables). Therefore, under this extra assumption, the stochastic process $u_1(x,t)(\omega)$ given by (15.18) is a mean square solution of the randomized PDE (15.11).

Hereinafter, the main goal is to obtain the PDF of the stochastic process $u_1(x,t)(\omega)$, for each (x,t). For the sake of generality, we assume that (a,d,λ,μ,b) is an absolutely continuous random vector with density function $f_{(a,d,\lambda,\mu,b)}$, so that $u_1(x,t)$ is absolutely continuous as well, with density function $f_{u_1(x,t)}$ to be computed. We use the RVT technique stated in the Lemma 15.1 with the following identification in the notation of this result: consider the transformation mapping

$$g(a,d,\lambda,\mu,b) = \left(\left[\frac{4d\mu\lambda^3 be^{\lambda(x-dt\lambda^2)}}{a\left(\lambda be^{\lambda(x-dt\lambda^2)} + \mu\right)^2} \right]^{\frac{1}{n}}, d, \lambda, \mu, b \right),$$

with domain

$$D = (\mathbb{R}\backslash\{0\}) \times \underbrace{\{(d,\lambda,\mu,b) \in (\mathbb{R}\backslash\{0\})^4 : be^{\lambda(x-dt\lambda^2)} + \frac{\mu}{\lambda} \neq 0\}}_{D'}. \tag{15.20}$$

Suppose that the random vector (a,d,λ,μ,b) has its support contained in D. In such a case, the solution stochastic process (15.18) is defined for all $(x,t) \in \mathbb{R}^2$

(because the denominator of the fraction does not vanish). In the domain D, the transformation mapping g is injective and has inverse

$$h(u, v, w, r, s) = \left(\frac{4vrw^3 se^{w(x-vtw^2)}}{u^n (wse^{w(x-vtw^2)} + r)^2}, v, w, r, s \right),$$

with domain $g(D) = D$. Its Jacobian is given by

$$|J[h](u, v, w, r, s)| = \frac{4n|v||r||w|^3|s|e^{w(x-vtw^2)}}{|u|^{n+1} \left(wse^{w(x-vtw^2)} + r\right)^2} \neq 0.$$

By RVT technique stated in Lemma 15.1,

$$f_{(u,v,w,r,s)}(u, v, w, r, s) = f_{(a,d,\lambda,\mu,b)} \left(\frac{4vrw^3 se^{w(x-vtw^2)}}{u^n (wse^{w(x-vtw^2)} + r)^2}, v, w, r, s \right) \cdot$$
$$\cdot \frac{4n|v||r||w|^3|s|e^{w(x-vtw^2)}}{|u|^{n+1} \left(wse^{w(x-vtw^2)} + r\right)^2},$$

for $(u, v, w, r, s) \in D$, and 0 otherwise. Computing the marginal PDF for U, we obtain the density of $u_1(x, t)$, for $x, t \in \mathbb{R}$:

$$f_{u_1(x,t)}(u) = \int_{D'} f_{(a,d,\lambda,\mu,b)} \left(\frac{4vrw^3 se^{w(x-vtw^2)}}{u^n (wse^{w(x-vtw^2)} + r)^2}, v, w, r, s \right)$$
$$\cdot \frac{4n|v||r||w|^3|s|e^{w(x-vtw^2)}}{|u|^{n+1} \left(wse^{w(x-vtw^2)} + r\right)^2} \, \mathrm{d}s \, \mathrm{d}r \, \mathrm{d}w \, \mathrm{d}v, \tag{15.21}$$

for $u \neq 0$ (the fact that this PDF is not defined for $u = 0$ does not suppose any problem, since density functions are defined only up to sets of Lebesgue measure zero).

Remark 15.1 In the case that any of the random variables $\lambda(\omega)$, $\mu(\omega)$ and $b(\omega)$ is not considered as a random variable, that is, any of them is a constant, the same reasoning applies. In the transformation mapping g, the new constant random variables do not appear in the evaluation of g nor in its domain D. In the final density function (15.21), the integration with respect to this constant parameter does not appear.

Another way of interpreting this fact is that the probability density function of a constant random variable $x_0 \in \mathbb{R}$ is a Dirac delta function δ_{x_0}, which satisfies

$$\int_{\mathscr{A}} H(\xi_1, \xi_2, \ldots, \xi_n) \delta_{x_0}(\xi_1) \, \mathrm{d}\xi_1 \, \mathrm{d}\xi_2 \cdots \mathrm{d}\xi_n = \int_{\mathscr{A}_{x_0}} H(x_0, \xi_2, \ldots, \xi_n) \, \mathrm{d}\xi_2 \cdots \mathrm{d}\xi_n, \tag{15.22}$$

where $\mathscr{A} \subseteq \mathbb{R}^n$, H is a function defined on $\mathscr{A}$ and

$$\mathscr{A}_{x_0} = \{(\xi_2, \ldots, \xi_n) \in \mathbb{R}^{n-1} : (x_0, \xi_2, \ldots, \xi_n) \in \mathscr{A}\}.$$

Using property (15.22) yields the same result as the first paragraph of this remark.

15.4.4 Example

Suppose that the data coefficients a and d in the PDE (15.11) are random variables, so that $a \sim \text{Gamma}(6, 4)$ and $d \sim \text{Beta}(5, 6)$. Suppose that the coefficients that appear in the Bernoulli random ODE (15.14) and in the solution stochastic process (15.15) are random variables as well, being (λ, μ) a multivariate Gaussian random vector, with mean vector and covariance matrix

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix},$$

respectively, truncated to the square $[2, 5] \times [1, 6]$, and $b \sim \text{Uniform}(1, 2)$. Under these distributions, the random vector (a, d, λ, μ, b) has its support contained in D, where D is the domain defined in (15.20). The random variables/vectors $a, d, (\lambda, \mu)$ and b are assumed to be independent. We take $n = 2$ in (15.11). Then, the density function (15.21) becomes

$$f_{u_1(x,t)}(u) = \int_0^1 \int_2^5 \int_1^6 \int_1^2 f_a\left(\frac{4vrw^3se^{w(x-vtw^2)}}{u^2(wse^{w(x-vtw^2)} + r)^2}\right) f_d(v) f_{(\lambda,\mu)}(w, r) f_b(s) \cdot \frac{8|v||r||w|^3|s|e^{w(x-vtw^2)}}{|u|^3\left(wse^{w(x-vtw^2)} + r\right)^2} \, \mathrm{d}s \, \mathrm{d}r \, \mathrm{d}w \, \mathrm{d}v, \tag{15.23}$$

for $u \neq 0$ and $x, t \in \mathbb{R}$. In Fig. 15.5, a three dimensional plot of the density function $f_{u_1(x,t=1)}(u)$ given in (15.23) is presented. In Fig. 15.6, the graph of the density function $f_{u_1(x=1,t=1)}(u)$ given in (15.23) is shown.

15.5 Conclusions

The aim of this chapter has been to provide an affordable view of the computation of the first probability density function of the solution stochastic process to both random ordinary and partial differential equations via the so-called Random Variable Transformation technique. Our presentation has been focused mainly on simple examples with the aim of illustrating key ideas. As setting probabilistic distribution of input parameters (initial/boundary condition, forcing term and/or coefficients)

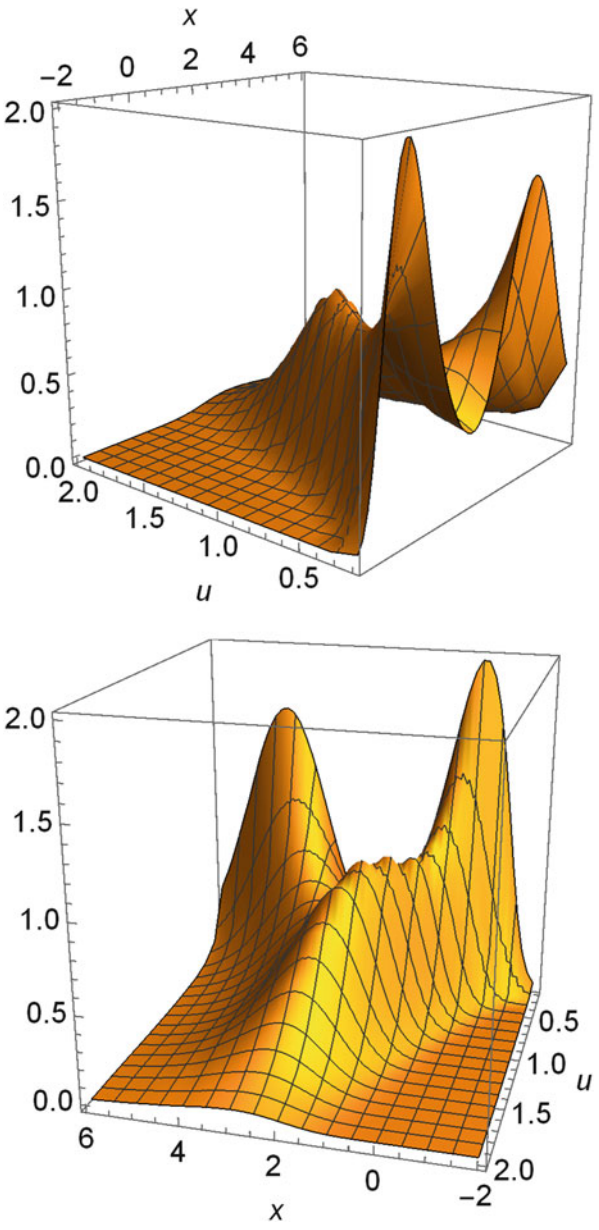


Fig. 15.5 Density function $f_{u_1(x,t=1)}(u)$ given in (15.23). Two different perspectives of the same graph

is a crucial point when Random Variable Transformation is set in practice, we have shown a Bayesian technique to assign the initial probabilistic distribution in the context of mathematical modelling using real data. It is important to point out that our approach has relied upon the availability of a closed expression for the solution of the corresponding differential equation. From a practical standpoint, this

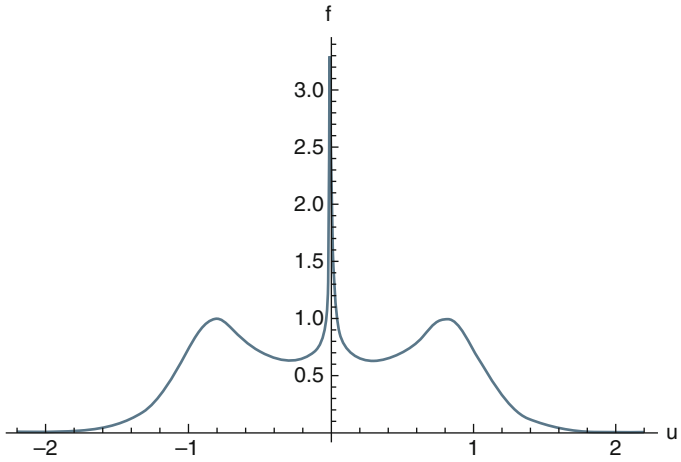


Fig. 15.6 Density function $f_{u_1(x=1,t=1)}(u)$ given in (15.23)

is a strong assumption since for the most part of differential equations numerical techniques are required. Therefore, it would be very interesting to extend the application of the Random Variable Transformation technique in that scenario in future contributions.

References

1. Soong, T.T.: Random Differential Equations in Science and Engineering. Academic, New York (1973)
2. Smith, R.C.: Uncertainty Quantification. Theory, Implementation and Applications. SIAM Computational Science & Engineering. SIAM, Philadelphia (2014)
3. Øksendal, B.: Stochastic Differential Equations. An Introduction with Applications. Stochastic Modelling and Applied Probability, vol. 23. Springer, Heidelberg (2003)
4. Kloeden, P., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (2011)
5. Allen, E.: Modeling with Itô Stochastic Differential Equations. Mathematical Modelling: Theory and Applications. Springer, Amsterdam (2007)
6. Papoulis, A., Pillai, U.: Probability, Random Variables and Stochastic Processes, 4th edn. McGraw-Hill, New York (2002)
7. Casella, G., Berger, R.: Statistical Inference. Cambridge Texts in Applied Mathematics. Brooks/Cole, New York (2002)
8. Kadry, S.: On the generalization of probabilistic transformation method. Appl. Math. Comput. **190**(15), 1284–1289 (2007). https://doi.org/10.1016/j.amc.2011.12.088
9. Lord, G.J., Powell, C.E., Shardlow, T.: An Introduction to Computational Stochastic PDEs. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York (2014)
10. Casabán, M.-C., Cortés, J.-C., Romero, J.-V., Roselló, M.-D.: Determining the first probability density function of linear random initial value problems by the random variable transformation (R.V.T.) technique: a comprehensive study. Abstr. Appl. Anal. **2014**, 1–25, 248512 (2014). https://doi.org/10.1155/2013/248512

11. Casabán, M.-C., Cortés, J.-C., Romero, J.-V., Roselló, M.-D.: Solving random homogeneous linear second-order differential equations: a full probabilistic description. Mediterr. J. Math. **13**(6), 3817–3836 (2016). https://doi.org/10.1007/s00009-016-0716-6
12. Casabán, M.-C., Cortés, J.-C., Navarro-Quiles, A., Romero, J.-V., Roselló, M.-D., Villanueva, R.-J.: Computing probabilistic solutions of the Bernoulli random differential equation. J. Comput. Appl. Math. **309**, 396–407 (2017). https://doi.org/10.1016/jcam.2016.02.034
13. Casabán, M.-C., Cortés, J.-C., Navarro-Quiles, A., Romero, J.-V., Roselló, M.-D., Villanueva, R.-J.: A comprehensive probabilistic solution of random SIS-type epidemiological models using the Random Variable Transformation technique. Commun. Nonlinear Sci. Numer. Simul. **32**, 199–210 (2016). https://doi.org/10.1016/j.cnsns.2015.08.009
14. Dorini, F.A., Cecconello, M.S., Dorini, L.B.: On the logistic equation subject to uncertainties in the environmental carrying capacity and initial population density. Commun. Nonlinear Sci. Numer. Simul. **33**, 160–173 (2016). https://doi.org/10.1016/j.cnsns.2015.09.009
15. Seber, G.A.F., Wild, C.J.: Nonlinear Regression. Wiley, Hoboken (2003)
16. Lesaffre, E., Lawson, A.B.: Bayesian Biostatistics. Statistics in Practice. Wiley, Hoboken (2012)
17. Hussein, A., Selim, M.M.: Solution of the stochastic generalized shallow-water wave equation using RVT technique. Eur. Phys. J. Plus **139**, 249 (2015). https://doi.org/10.1140/epjp/i2015-15249-3
18. Hussein, A., Selim, M.M.: Solution of the stochastic radiative transfer equation with Rayleigh scattering using RVT technique. Appl. Math. Comput. **218**(13), 7193–7203 (2012). https://doi.org/10.1016/j.amc.2011.12.088
19. Angulo Pava, J.: Nonlinear Dispersive Equations. Existence of Stability of Solitary and Periodic Travelling Wave Solutions. Mathematical Surveys and Monographs, vol. 156. American Mathematical Society, Providence (2009)
20. Hussein, A., Selim, M.M.: New soliton solutions for some important nonlinear partial differential equations using a generalized Bernoulli method. Int. J. Math. Anal. Appl. **1**(1), 1–8 (2014)
21. Yang, X.-F., Deng, Z.-C., Wei, Y.: A Riccati-Bernoulli sub-ODE method for nonlinear partial differential equations and its application. Adv. Difference Equ. **2015**, 117 (2015). https://doi.org/10.1186/s13662-015-0452-4
22. Loève, M.: Probability Theory, vol. 1. Springer, New York (1977)
23. Braumann, C.A., Villafuerte, L., Cortés, J.-C., Jódar, L.: Random differential operational calculus: theory and applications. Comput. Math. Appl. **59**, 115–125 (2010). https://doi.org/10.1016/j.camwa.2009.08.061

Chapter 16
A Strong Averaging Principle for Lévy Diffusions in Foliated Spaces with Unbounded Leaves

Paulo Henrique da Costa, Michael A. Högele, and Paulo Regis Ruffino

Abstract This article extends a strong averaging principle for Lévy diffusions which live on the leaves of a foliated manifold subject to small transversal Lévy type perturbation to the case of non-compact leaves. The main result states that the existence of p-th moments of the foliated Lévy diffusion for $p \geqslant 2$ and an ergodic convergence of its coefficients in L^p implies the strong L^p convergence of the fast perturbed motion on the time scale t/ε to the system driven by the averaged coefficients. In order to compensate the non-compactness of the leaves we use an estimate of the dynamical system for each of the increments of the canonical Marcus equation derived in da Costa and Högele (Potential Anal 47(3):277–311, 2017), the boundedness of the coefficients in L^p and a nonlinear Gronwall-Bihari type estimate. The price for the non-compactness are slower rates of convergence, given as p-dependent powers of ε strictly smaller than $1/4$.

16.1 Introduction

The literature on averaging principles for deterministic and stochastic systems reaches far back to the eighteenth century and is enormously rich both in theory and applications. At this point, however, we would like to refrain from a more systematic review of the long and bifurcated history of the field and restrict ourselves to the references to some classical texts. Standard texts on the deterministic field

P. H. da Costa
Departamento de Matemática, Universidade de Brasília, Brasília, Brazil
e-mail: phcosta@unb.br

M. A. Högele (✉)
Departamento de Matemáticas, Universidad de los Andes, Bogotá, Colombia
e-mail: ma.hoegele@uniandes.edu.co

P. R. Ruffino
IMECC, Universidade Estadual de Campinas, Campinas, Brazil
e-mail: ruffino@ime.unicamp.br

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*,
Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_16

include [3, 25, 29, 30] and the references therein. For stochastic systems we refer to [4, 5, 7, 9, 13, 15, 21, 27] and the respective bibliographies.

Loosely speaking, an averaging principle describes the observation that in a coupled slow-fast system in the limit of infinite time scale separation, the slow system is close to a system, where the fast variable is replaced by the limiting measure of its ergodic time average. In the case of stochastic differential equations rescaling the time variable shows that this problem can be restated as a problem of an ergodic system perturbed by small perturbations.

The results of this article generalize recent approaches by the authors for diffusions on finite dimensional foliated manifolds. For properties of foliated spaces consult [6, 11, 28, 31]. Motivated by [20] Gargate and Ruffino studied in [10] the case of foliated Gaussian diffusions on compact leaves subject to deterministic Lipschitz transversal perturbation. In Högele and Ruffino [12] the authors treat the case of foliated Lévy jump diffusions with exponential moments but still with deterministic transversal perturbation and compact leaves. This type of processes is described in terms of canonical Marcus equations.

The recent work by da Costa and Högele [8] covers the case of a general class of foliated Lévy diffusions on compact leaves perturbed by a near optimally large class of Lévy diffusions. This is carried out with the help of a nonlinear comparison principle and a fine study of the individual jump increments. However in this case the compactness still allows global estimates of the horizontal components, for instance, in the force acting on the "vertical" component of the perturbed system.

The current article treats an averaging principle for the same type of foliated Lévy diffusions, however with non-compact leaves. The lack of compactness yields an almost unmitigated system of fully coupled SDEs. The strategies are once again non-linear Gronwall-Bihari type inequalities, using instead the L^p boundedness of the drift. However, this comes at the price of slower rates of convergence. Our main result, Theorem 16.4 states that locally the transversal behavior of $X^{\epsilon}_{t/\epsilon}$ can be approximated L^p uniformly in time by the Lévy stochastic differential equation in the transversal space with coefficients given by the average of the deterministic transversal component of the perturbation (with respect to the invariant measure on the leaves for the original unperturbed dynamics) and the diffusion component given by the projection of the original perturbation into the transversal space. We should mention that our results cover the results by [32] as the special case of uniformly bounded jumps.

In the Sect. 16.2 we present the dynamical and stochastic framework, the main hypotheses and the main result. In Sect. 16.3 we prove the key proposition which is the basis for the proof of the main theorem, proved in Sect. 16.4. Wherever possible in the exposition without lost of coherence we refer to the article [8] in order to avoid trivial repetition.

16.2 Object of Study and Main Results

16.2.1 The Setup

The following setup is a non-compact extension of the setup on [8] and [12].

The Foliated Manifold: Let M be a finite dimensional connected, smooth Riemannian manifold. It is known by the classical Nash theorem in [22] that any finite dimensional smooth manifold may be embedded in $\mathbb{R}^m$ with m sufficiently large. We assume that M is equipped with an n-dimensional foliation $\mathfrak{M}$ in the following sense. Let $\mathfrak{M} = (L_x)_{x\in M}$, with $M = \bigcup_{x\in M} L_x$ and the sets L_x are equivalence classes of the elements of M satisfying the following.

(a) Given $x_0 \in M$ there exist a neighborhood $U \subset M$ of the corresponding leaf L_{x_0} and a diffeomorphism $\varphi : U \to L_{x_0} \times V$, where $V \subset \mathbb{R}^d$ is a connected open set containing the origin $0 \in \mathbb{R}^d$.
(b) For any $L_{x_0} \in \mathfrak{M}$ the neighborhood $U \supset L_{x_0}$ can be taken small enough such that the coordinate map φ is uniformly Lipschitz continuous.

Remark 16.1 The second coordinate of a point $x \in U$, called the vertical coordinate, will be denoted with the help the projection $\pi : U \to V$ by $\varphi(x) = (\bar{x}, \pi(x))$ for some $\bar{x} \in L_x$. For any fixed $v \in V$, the preimage $\pi^{-1}(v)$ is the leaf L_x, where x is any point in U such that the vertical projection satisfies $\pi(x) = v$.

The Unperturbed Equation: We are interested in the ergodic behavior of the strong solution of a Lévy driven SDE with jump components which takes values in M and which respects the foliation. Intuitively, a straight line increment z does not cause the exit from the leaf of its current position if the entire line segment $(x_0 + \theta z)_{\theta\in[0,1]}$ is contained in it. Ordinary differential equations with a vector field F on the right-hand side generalize this concept in the following sense. By definition, their solutions follow F as "infinitesimal" tangents. If F itself is tangential to a given manifold the integral curves remain "infinitesimally tangential" to the manifold and hence will not leave it. Therefore a straight line jump increment z which is transformed in the stochastic integral into an integral curve following a tangential vector field F of a given leaf will remain on the leaf, that is, it respects the foliated structure of the space. This intuition is made rigorous in the notion of stochastic integration in the sense of a canonical Marcus equation in the sense of Kurtz et al. [19]. Those equations are the equivalent for Lévy jump diffusions to the Stratonovich equation for Brownian SDE in that they satisfy the Leibniz chain rule (cf. Proposition 4.2 in [19]). Their definition however is different since they treat discontinuous processes.

Let us consider the formal canonical Marcus stochastic differential equation

$$dX_t = F_0(X_t)dt + F(X_t) \diamond dZ_t + G(X_t) \circ dB_t, \qquad X_0 = x_0 \in M, \tag{16.1}$$

with the following components defined over a given filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbb{P})$ which satisfies the usual conditions in the sense of Protter [24].

1. Let $Z = (Z_t)_{t\geqslant 0}$ with $Z_t = (Z_t^1, \ldots, Z_t^r)$ be a Lévy process over Ω with values in $\mathbb{R}^r$ for some $r \in \mathbb{N}$ and characteristic triplet $(0, \nu, 0)$. It is a consequence of the Lévy-Itô decomposition of Z that Z is a pure jump process with respect to a Lévy measure $\nu : \mathscr{B}(\mathbb{R}^r) \to [0, \infty]$ satisfying

$$\int_{\mathbb{R}^r} (1 \wedge \|z\|^2)\, \nu(dz) < \infty \qquad \text{and} \qquad \nu(\{0\}) = 0. \tag{16.2}$$

   For details we refer to the overview article by Kunita [18] and the monographs of Sato [26] or Applebaum [2].
2. Let $F \in \mathscr{C}^2(M; L(\mathbb{R}^r; T\mathfrak{M}))$ satisfying the following. The function $x \mapsto F(x)$ is $\mathscr{C}^2$ and for each $x \in M$ the linear map $F(x)$ sends a vector $z \in \mathbb{R}^r \mapsto F(x)z \in T_x L_x$ to the tangent space of the respective leaf. Furthermore, let F and $(DF)F$ be globally Lipschitz continuous on M with common Lipschitz constant $\ell > 0$.
3. Let $B = (B^1, \ldots, B^r)$ be an Brownian motion on Ω with values in $\mathbb{R}^r$. For $G \in \mathscr{C}^2(M, L(\mathbb{R}^r, T\mathfrak{M}))$ we assume that G and $(DG)G$ are globally Lipschitz continuous on M with Lipschitz constant $\ell > 0$.

Following [19] a strong solution of the formal Eq. (16.1) is defined as a random map $X : [0, \infty) \times \Omega \to M$ satisfying almost surely for all $t \geqslant 0$

$$\begin{aligned} X_t = x_0 &+ \int_0^t F_0(X_s)ds + \int_0^t G(X_s)dB_s + \frac{1}{2}\int_0^t (DG(X_s))G(X_s)d\langle B\rangle_s \\ &+ \int_0^t F(X_{s-})dZ_s + \sum_{0<s\leqslant t} (\Phi^{F\Delta_s Z}(X_{s-}) - X_{s-} - F(X_{s-})\Delta_s Z), \end{aligned} \tag{16.3}$$

where $\langle B\rangle_\cdot$ stands for the quadratic variation process of B in $\mathbb{R}^r$ and the function $\Phi^{Fz}(x) = Y(1, x; Fz)$ and $Y(t, x; Fz)$ for the solution of the ordinary differential equation

$$\frac{d}{d\sigma}Y(\sigma) = F(Y(\sigma))z, \qquad Y(0) = x \in M, \quad z \in \mathbb{R}^r, \quad \sigma \geqslant 0. \tag{16.4}$$

The Perturbed Equation: This article studies the situation where an SDE in the sense of (16.3), which is invariant on the leaf of the initial condition x_0 is perturbed by a transversal smooth vector field $\varepsilon K dt$ and stochastic differentials $\varepsilon \tilde{G} \circ d\tilde{B}$ and $\varepsilon \tilde{K} \diamond d\tilde{Z}$, $\varepsilon > 0$, in the limit for $\varepsilon \searrow 0$. More precisely we denote by X^ε, $\varepsilon > 0$, the

analogous solution in the sense of (16.3) of the perturbed formal system

$$\begin{aligned} dX_t^\varepsilon &= F_0(X_t^\varepsilon)dt + F(X_t^\varepsilon) \diamond dZ_t + G(X_t^\varepsilon) \circ dB_t \\ &\quad + \varepsilon\Big(K(X_t^\varepsilon)dt + \tilde{K}(\pi(X_t^\varepsilon)) \diamond d\tilde{Z}_t + \tilde{G}(\pi(X_t^\varepsilon)) \circ d\tilde{B}_t\Big), \qquad (16.5) \\ X_0^\varepsilon &= x_0 \in M, \end{aligned}$$

where the additional coefficients are defined as follows.

4. The vector field $K : M \to TM$ is smooth and globally Lipschitz continuous.
5. Let $\tilde{Z} = (\tilde{Z}^1, \ldots, \tilde{Z}^r)$ be a Lévy process on Ω with values in $\mathbb{R}^r$ with Lévy triple $(0, \nu', 0)$, ν' being a given Lévy measure. The vector field $\tilde{K} \in \mathscr{C}^2(V, L(\mathbb{R}^r, TM))$ satisfies that $\tilde{K}$ and $(D\tilde{K})\tilde{K}$ are globally Lipschitz continuous with Lipschitz constant $\tilde{\ell} > 0$.
6. Let $\tilde{B} = (\tilde{B}^1, \ldots, \tilde{B}^r)$ be a Brownian motion on Ω with values in $\mathbb{R}^r$. We assume that $\tilde{G} \in \mathscr{C}^2(V, L(\mathbb{R}^r, TM))$ satisfies that $\tilde{G}$ and $(D\tilde{G})\tilde{G}$ are globally Lipschitz continuous with Lipschitz constant $\tilde{\ell} > 0$.
7. Assume that the stochastic processes Z, B, $\tilde{Z}$, $\tilde{B}$ are independent.

Theorem 16.2 ([19], Theorem 3.2 and 5.1)

1. *Under the preceding setup (items (a), (b), 1.–3. and 7.) there is a unique $(\mathscr{F}_t)_{t\geqslant 0}$ semimartingale X which is a strong global solution of (16.1) in the sense of Eq. (16.3). It has a càdlàg version and is a (strong) Markov process.*
2. *Under the preceding setup (in particular items a, b and 1.–7.) there is a unique semimartingale X^ε which is a strong global solution of Eq. (16.5) in the sense of Eq. (16.3), where F_0 is replaced by $F_0 + \varepsilon K$ and F by $(F, \varepsilon\tilde{K})$, G by $(G, \varepsilon\tilde{G})$, B by $(B, \tilde{B})$ and Z by $(Z, \tilde{Z})$. The perturbed solution X^ε has càdlàg paths almost surely and is a (strong) Markov process.*

The Support Theorem: We are now in the position to apply the crucial support theorem, Proposition 4.3, in Kurtz et al. [19]. The hypotheses of Theorem 16.2 imply for any $\varepsilon > 0$ and $x_0 \in M$ that $\mathbb{P}(X_t^\varepsilon(x_0) \in M \text{ for all } t \geqslant 0) = 1$. This result applied to the leaves of $\mathfrak{M}$ yields that each solution X of (16.1) is *foliated* in the sense that X stays on the leaf of its initial condition, i.e. for any $x_0 \in M$ we have $\mathbb{P}(X_t(x_0) \in L_{x_0} \text{ for all } t \geqslant 0) = 1$.

16.2.2 The Hypotheses and the Main Result

In the general setup of Sect. 16.2.1 we assume the following precise hypotheses.

Hypothesis 1: Integrability There is an exponent $p \geqslant 2$ such that the Lévy measures ν of Z and ν' of $\tilde{Z}$ satisfy

$$\int_{\mathbb{R}^r} \|z\|^p \, \nu(dz) < \infty \qquad \text{and} \qquad \int_{\mathbb{R}^r} \|z\|^{2p} \, \nu'(dz) < \infty.$$

Hypothesis 2: Foliated Invariant Measures

1. Each leaf $L_{x_0} \in \mathfrak{M}$ passing through $x_0 \in M$ has an associated unique invariant measure μ_{x_0} with $\operatorname{supp}(\mu_{x_0}) = L_{x_0}$ of the unperturbed foliated system (16.1) with initial condition x_0.
2. For $v_0 = \pi(x_0)$ the vertical coordinate of $x_0 \in M$ we define for $h : M \to TM$

$$Q^h(v_0) := \int_{L_{x_0}} h(y)\mu_{x_0}(dy). \tag{16.6}$$

We assume for any globally Lipschitz continuous map $h : M \to TM$ the function

$$\mathbb{R}^d \supset V \ni v \mapsto Q^h(v) \in \mathbb{R}^d \tag{16.7}$$

is globally Lipschitz continuous.

Remark 16.3 Note that L_{x_0} only depends on $v_0 = \pi(x_0)$. The same is true for μ_{x_0}.

Hypothesis 2 guarantees that for each $x_0 \in M$, $v_0 = \pi(x_0) \in V$ the stochastic differential equation

$$dw_t = Q^{\pi K}(w_t)\,dt + \tilde{K}(w_t) \diamond d\tilde{Z}_t + \tilde{G}(w_t) \circ d\tilde{B}_t, \qquad w_0 = v_0 \in V \tag{16.8}$$

has a unique strong solution $w = (w_t(v_0))_{t\in[0,\sigma)}$ on Ω, σ being the first exit time of w from V.

Hypothesis 3: Ergodic Convergence of the Vertical Coefficient in L^p Fix $p \geqslant 2$ from Hypothesis 1.

1. There are continuous functions $\eta^0 : [0,\infty) \to [0,\infty)$ and $\bar{\eta} : M \to [0,\infty)$, where η^0 is monotonically decreasing with $\eta^0(t) \to 0$ as $t \to \infty$ and $\bar{\eta}$ is globally Lipschitz continuous. For all $x_0 \in M$ and $t \geqslant 0$ we have

$$\left(\mathbb{E}\left|\frac{1}{t}\int_0^t \pi K(X_s(x_0))\,ds - Q^{\pi K}(\pi(x_0))\right|^p\right)^{\frac{1}{p}} \leqslant \bar{\eta}(x_0)\eta^0(t). \tag{16.9}$$

2. We assume for any $x_0 \in M$ that $\int \bar{\eta}(y)\mu_{x_0}(dy) < \infty$.

It is known in the literature that there is no standard rate of convergence [14, 16], which is why we assume an external rate of convergence, which decomposes by factors, see for instance [17].
For $\varepsilon > 0$ and $x_0 \in M$ let τ^ε being the first exit time of the solution $X^\varepsilon(x_0)$ of Eq. (16.5) from the foliated coordinate neighborhood U of item a) in Sect. 16.2.1.

The main result of this article is the following strong averaging principle.

Theorem 16.4 *Let Hypotheses 1, 2 and 3 be satisfied for some $p \geqslant 2$. Then for any $x_0 \in M$ and $\lambda \in (0, \frac{p-1}{p^2})$ there are constants $c, C > 0$ and $\varepsilon_0 \in (0,1]$ such that*

$\varepsilon \in (0, \varepsilon_0]$ and $T \in [0, 1]$ imply

$$\left(\mathbb{E}\left[\sup_{t\in[0,T\wedge\varepsilon\tau^\varepsilon\wedge\sigma]} |\pi\left(X^\varepsilon_{\frac{t}{\varepsilon}}(x_0)\right) - w_t(\pi(x_0))|^p\right]\right)^{\frac{1}{p}} \leqslant CT\left[\varepsilon^\lambda + \eta^0(cT|\ln(\varepsilon)|)\right]. \tag{16.10}$$

Remark 16.5 Our results focus on the case with only p-th moments, hence we set the coefficients G and $\tilde{G}$ to zero in the proofs.

16.3 The Transversal Perturbations

In order to prove the main theorem we need to control the error $X^\varepsilon - X$ in terms of L^p. This section is dedicated to the control of this error by the following result.

Proposition 16.6 *Let the assumptions of Sect. 16.2.1 and Hypotheses 1, 2 and 3 be satisfied for some $p \geqslant 2$. Then for any Lipschitz function $h : M \to \mathbb{R}$, $x_0 \in M$ and for all $T^\cdot : [0, 1] \to [1, \infty)$ satisfying $\varepsilon T^\varepsilon \to 0$ there exist positive constants $\varepsilon_0 \in (0, 1]$, $k_1, k_2, k_3 > 0$ such that $\varepsilon \in (0, \varepsilon_0]$ implies*

$$\left(\mathbb{E}\left[\sup_{t\in[0,T]} |h(X^\varepsilon_t(x_0)) - h(X_t(x_0))|^p\right]\right)^{\frac{1}{p}} \leqslant k_1\varepsilon^{\frac{p-1}{p^2}}\exp(k_2T). \tag{16.11}$$

In addition, we have $k_1(x_0) \leqslant k_3(1 + \bar{\eta}(x_0))$.

We apply this result for the following setting.

Corollary 16.7 *Let the assumptions of Proposition 16.6 be satisfied for some $p \geqslant 2$. Then for any $\lambda \in (0, \frac{p-1}{p^2})$ there exist positive constants c_λ, $\varepsilon_0 \in (0, 1]$, $k_4, k_5 > 0$ such that for $T_\varepsilon := c_\lambda|\ln(\varepsilon)|$, $\varepsilon \in (0, \varepsilon_0]$, satisfies*

$$\left(\mathbb{E}\left[\sup_{t\in[0,T_\varepsilon]} |h(X^\varepsilon_t(x_0)) - h(X_t(x_0))|^p\right]\right)^{\frac{1}{p}} \leqslant k_4\varepsilon^\lambda, \tag{16.12}$$

for the constant $k_4 = k_5k_1$.

Proof Plugging $T_\varepsilon = -c\ln(\varepsilon)$ in the right-hand side of (16.11) we obtain $k_1\varepsilon\exp(k_2T_\varepsilon) = k_1\varepsilon^{\frac{p-1}{p^2}-ck_2}$. Given $\lambda \in (0, \frac{p-1}{p^2})$ we fix $c_\lambda := \frac{1}{k_2}\left(\frac{p-1}{p^2} - \lambda\right)$ and infer the desired result.

The proof of Proposition 16.6 relies on the following lemma on positive invariant dynamical systems and the nonlinear comparison principle Corollary 16.14 given in the appendix. The main difficulty stems from the fact that the influence of the horizontal component in the vertical component cannot be estimated uniformly by

the "diameter" of the leaf but has to be taken fully into account, which leads to a non-linear comparison principle.

Lemma 16.8 *For $F \in \mathscr{C}^2(\mathbb{R}^{r+n}, L(\mathbb{R}^r, \mathbb{R}^{r+n}))$ being a globally Lipschitz continuous matrix-valued vector field and $z \in \mathbb{R}^r$ we denote by $(Y(t; x, Fz))_{t\geqslant 0}$ the unique global strong solution of the ordinary differential equation*

$$\frac{dY}{dt} = F(Y)z \qquad Y(0, x, Fz) = x \in \mathbb{R}^{r+n}.$$

1. *Then there exists $C > 0$ such that for any $z \in \mathbb{R}^r$ and $x, y \in M$ with $Y(t; x) = Y(t; x, Fz)$ we have*

$$\sup_{t\geqslant 0} |(DF(Y(t; x))z)F(Y(t; x))z - (DF(Y(t; y))z)F(Y(t; y))z| \leqslant C\ |x - y|\ \|z\|^2.$$

2. *For any $x \in M$ we have $\sup_{t\in[0,1]} \|DF(Y(t; x))F(Y(t; x))\| < \infty$.*

A proof is given in [8] under Lemma 3.1.

Proof (of Proposition 16.6) The first step of the proof yields the local orthogonality of the foliations and a transversal component by an appropriate change of coordinates. In a second step we estimate the transversal components with the help of the ergodic convergence of Hypothesis 3 and the nonlinear comparison principle Corollary 16.14. This is followed by the estimate of the horizontal component as the result of a classical Gronwall estimate before we conclude.

1. Change of Coordinates: We first rewrite X and X^ε, the solutions of Eqs. (16.1) and (16.5), in terms of the coordinates given by the diffeomorphism ϕ

$$(u_t, v_t) := \phi(X_t) \qquad \text{and} \qquad (u_t^\varepsilon, v_t^\varepsilon) := \phi(X_t^\varepsilon).$$

The Lipschitz regularities of h and ϕ yields for $C_0 := Lip(h \circ \phi^{-1})$ the estimate

$$|h(X_t^\varepsilon) - h(X_t)| \leqslant C_0(|u_t^\varepsilon - u_t| + |v_t^\varepsilon - v_t|). \tag{16.13}$$

The proof of the statement consists in calculating estimates for each summand on the right hand side of equation above. We define the

$$\mathfrak{F}_0 := (D\phi) \circ F_0 \circ \phi^{-1}, \qquad \mathfrak{F} := (D\phi) \circ F \circ \phi^{-1},$$
$$\mathfrak{K} := (D\phi) \circ K \circ \phi^{-1}, \qquad \tilde{\mathfrak{K}} := (D\phi) \circ \tilde{K} \circ \phi^{-1},$$

whose derivatives are uniformly bounded. Considering the components in the image of ϕ we have:

$$\mathfrak{K} = (\mathfrak{K}_H, \mathfrak{K}_V), \qquad \tilde{\mathfrak{K}} = (\tilde{\mathfrak{K}}_H, \tilde{\mathfrak{K}}_V)$$

with $\mathfrak{K}_H, \tilde{\mathfrak{K}}_H \in TL_{x_0}$ and $\mathfrak{K}_V, \tilde{\mathfrak{K}}_V \in TV \simeq \mathbb{R}^d$. The chain rule of the canonical Marcus equations mentioned in the introduction (Theorem 4.2 of [19]) yields for Eq. (16.5) the following form in ϕ coordinates

$$du_t^\varepsilon = \mathfrak{F}_0(u_t^\varepsilon, v_t^\varepsilon)dt + \mathfrak{F}(u_t^\varepsilon, v_t^\varepsilon) \diamond dZ_t + \varepsilon\, \mathfrak{K}_H(u_t^\varepsilon, v_t^\varepsilon)dt + \varepsilon \tilde{\mathfrak{K}}_H(v_t^\varepsilon) \diamond d\tilde{Z}_t, \tag{16.14}$$

$$dv_t^\varepsilon = \varepsilon\, \mathfrak{K}_V(u_t^\varepsilon, v_t^\varepsilon)dt + \varepsilon \tilde{\mathfrak{K}}_V(v_t^\varepsilon) \diamond d\tilde{Z}_t, \tag{16.15}$$

where $u_t^\varepsilon \in L_{x_0}$ and $v_t^\varepsilon \in V$ $\mathbb{P}$-a.s. for all $t \geqslant 0$.

2. Estimate of the Transversal Coordinate $\mathbb{E}[\sup |v^\varepsilon - v|^p]$: Identically to [8], we start with estimates on the transversal components $|v^\varepsilon - v|$. The change of variables formula $x \mapsto g(x) := |x|^p, x \in \mathbb{R}^{n+d}$ using $\langle Dg(x), u\rangle = p|x|^{p-2}\langle x, u\rangle$ yields almost surely for $t \geqslant 0$

$$\begin{aligned}
|v_t^\varepsilon - v_t|^p &= p\int_0^t |v_s^\varepsilon - v_s|^{p-2}\langle v_s^\varepsilon - v_s, \varepsilon\mathfrak{K}_V(u_s^\varepsilon, v_s^\varepsilon)\rangle ds \\
&\quad + p\int_0^t |v_{s-}^\varepsilon - v_{s-}|^{p-2}\langle v_{s-}^\varepsilon - v_{s-}, \varepsilon\tilde{\mathfrak{K}}_V(v_{s-}^\varepsilon) \diamond d\tilde{Z}_s\rangle \\
&\leqslant p\int_0^t |v_s^\varepsilon - v_s|^{p-1}|\varepsilon\mathfrak{K}_V(u_s^\varepsilon, v_s^\varepsilon) - \varepsilon\mathfrak{K}_V(u_s, v_s)|ds \qquad (H_1) \\
&\quad + p\int_0^t |v_s^\varepsilon - v_s|^{p-1}|\varepsilon\mathfrak{K}_V(u_s, v_s)|ds \qquad (H_2) \\
&\quad + p\int_0^t |v_{s-}^\varepsilon - v_{s-}|^{p-2}|\langle v_{s-}^\varepsilon - v_{s-}, \varepsilon(\tilde{\mathfrak{K}}_V(v_{s-}^\varepsilon) - \tilde{\mathfrak{K}}_V(v_{s-}))d\tilde{Z}_s\rangle| \\
&\qquad (H_3) \\
&\quad + p\int_0^t |v_{s-}^\varepsilon - v_{s-}|^{p-2}|\langle v_{s-}^\varepsilon - v_{s-}, \varepsilon\tilde{\mathfrak{K}}_V(v_{s-})d\tilde{Z}_s\rangle| \qquad (H_4) \\
&\quad + p\sum_{0<s\leqslant t} |v_{s-}^\varepsilon - v_{s-}|^{p-1}|\Phi^{\varepsilon\tilde{\mathfrak{K}}_V\Delta_s\tilde{Z}}(v_{s-}^\varepsilon) - \Phi^{\varepsilon\tilde{\mathfrak{K}}_V\Delta_s\tilde{Z}}(v_{s-}) \\
&\qquad - (v_{s-}^\varepsilon - v_{s-}) - \varepsilon(\tilde{\mathfrak{K}}_V(v_{s-}^\varepsilon) - \varepsilon\tilde{\mathfrak{K}}_V(v_{s-}))\Delta_s\tilde{Z}| \qquad (H_5) \\
&\quad + p\sum_{0<s\leqslant t} |v_{s-}^\varepsilon - v_{s-}|^{p-1}|\Phi^{\varepsilon\tilde{\mathfrak{K}}_V\Delta_s\tilde{Z}}(v_{s-}) - v_{s-} - \varepsilon\tilde{\mathfrak{K}}_V(v_{s-})\Delta_s\tilde{Z}| \\
&\qquad (H_6) \\
&=: H_1 + H_2 + H_3 + H_4 + H_5 + H_6. \qquad (16.16)
\end{aligned}$$

2.1 Pathwise Estimates $\mathbf{H_1}$: Clearly we have

$$H_1 \leqslant \varepsilon p \ell \int_0^t |v_s^\varepsilon - v_s|^p ds. \tag{16.17}$$

$\mathbf{H_2}$: Young's inequality for the conjugate indices p and $p/(p-1)$ yields

$$\begin{aligned} H_2 &= \varepsilon p \int_0^t |v_s^\varepsilon - v_s|^{p-1} |\mathfrak{K}_V(u_s, v_s)| ds \\ &\leqslant \varepsilon p \sup_{[0,t]} |v^\varepsilon - v|^{p-1} \int_0^t |\mathfrak{K}_V(u_s, v_s)| ds \\ &\leqslant \varepsilon \sup_{[0,t]} |v^\varepsilon - v|^p + \varepsilon (p-1) t^p \Big(\frac{1}{t} \int_0^t |\mathfrak{K}_V(u_s, v_s)| ds \Big)^p. \end{aligned} \tag{16.18}$$

$\mathbf{H_3}$ and $\mathbf{H_4}$: Switching to the Poisson random measure representation with respect to the compensated $\tilde{N}'$, for instance see Kunita [18], we obtain

$$\begin{aligned} H_3 \leqslant \varepsilon p \int_0^t \int_{\mathbb{R}^r} |v_{s-}^\varepsilon - v_{s-}|^{p-2} \langle v_{s-}^\varepsilon - v_{s-}, (\tilde{\mathfrak{K}}_V(v_{s-}^\varepsilon) - \tilde{\mathfrak{K}}_V(v_{s-}))z \rangle \tilde{N}'(dsdz) \\ + \varepsilon C_1 \int_0^t |v_s^\varepsilon - v_s|^p ds. \end{aligned} \tag{16.19}$$

and

$$\begin{aligned} H_4 \leqslant \varepsilon p \int_0^t \int_{\mathbb{R}^r} |v_{s-}^\varepsilon - v_{s-}|^{p-2} |\langle v_{s-}^\varepsilon - v_{s-}, \tilde{\mathfrak{K}}_V(v_{s-})z \rangle| \tilde{N}'(dsdz) \\ + \varepsilon C_2 \int_0^t |v_s^\varepsilon - v_s|^p ds. \end{aligned} \tag{16.20}$$

$\mathbf{H_5}$: For the canonical Marcus terms we apply Lemma 16.8, statement 1) which yields a positive constant such that

$$H_5 \leqslant \varepsilon^2 C_3 \int_0^t \int_{\mathbb{R}^r} |v_{s-}^\varepsilon - v_{s-}|^p \|z\|^2 \tilde{N}'(dsdz) + \varepsilon^2 C_4 \int_0^t |v_s^\varepsilon - v_s|^p \, ds. \tag{16.21}$$

The details can be found in [8].

$\mathbf{H_6}$: For the last term we apply Lemma 16.8, statement (2), and $\int_{\|z\|>1} \|z\|^4 \nu'(dz) < \infty$ and obtain a positive constant C_5 such that

$$H_5 \leqslant \varepsilon^2 C_5 \int_0^t \int_{\mathbb{R}^r} |v_{s-}^\varepsilon - v_{s-}|^{p-1} \|z\|^4 \tilde{N}'(dsdz) + \varepsilon^2 C_6 \int_0^t |v_s^\varepsilon - v_s|^{p-1} ds. \tag{16.22}$$

Combining the estimates (16.17)–(16.22) we obtain

$$|v_t^\varepsilon - v_t|^p \leqslant \varepsilon \sup_{[0,t]} |v^\varepsilon - v|^p + \varepsilon(p-1)t^p \Big(\frac{1}{t}\int_0^t |\mathfrak{K}_V(u_s, v_s)|ds\Big)^p \tag{16.23}$$

$$+ \varepsilon(C_1 + C_2)\int_0^t |v_s^\varepsilon - v_s|^p ds$$

$$+ \varepsilon^2 C_4 \int_0^t |v_s^\varepsilon - v_s|^p\, ds + \varepsilon^2 C_6 \int_0^t |v_s^\varepsilon - v_s|^{p-1} ds$$

$$+ \varepsilon p \int_0^t \int_{\mathbb{R}^r} |v_{s-}^\varepsilon - v_{s-}|^{p-2} \langle v_{s-}^\varepsilon - v_{s-}, \varepsilon \tilde{\mathfrak{K}}_V(v_{s-})z\rangle |\tilde{N}'(dsdz) \tag{16.24}$$

$$+ \varepsilon^2 C_3 \int_0^t \int_{\mathbb{R}^r} |v_{s-}^\varepsilon - v_{s-}|^p \|z\|^2 \tilde{N}'(dsdz) \tag{16.25}$$

$$+ \varepsilon^2 p C_5 \int_0^t \int_{\mathbb{R}^r} |v_{s-}^\varepsilon - v_{s-}|^{p-1} \|z\|^4 \tilde{N}'(dsdz). \tag{16.26}$$

2.2 Estimates on Average: The main difference to [8] is found in the treatment of term H_2. In the sequel we drop the superscript of $T = T^\varepsilon$ where $T^\varepsilon \in [1, \infty)$ satisfying $\varepsilon T^\varepsilon \to 0$. Taking the supremum $t \in [0, T]$ and taking the expectation yields that the term (16.23) can be bounded by

$$\varepsilon\, \mathbb{E}\Big[\sup_{[0,T]} |v^\varepsilon - v|^p\Big] + \varepsilon(p-1)C_\infty T^p,$$

where

$$C_\infty = C_\infty(x_0) = \sup_{t \geqslant 0} \mathbb{E}\Big[\Big(\frac{1}{t}\int_0^t |\mathfrak{K}_V(u_s(x_0), 0)|ds\Big)^p\Big] < \infty \tag{16.27}$$

due to the convergence

$$\mathbb{E}\Big[\Big(\frac{1}{t}\int_0^t |\mathfrak{K}_V(u_s(x_0), 0)|ds - \int |\mathfrak{K}_V(y, 0)|\mu_{x_0}(dy)\Big)^p\Big] \to 0, \quad \text{as } t \to \infty.$$

This implies in particular that

$$C_\infty(x_0) \leqslant \int |\mathfrak{K}_V(y, 0)|\mu_{x_0}(dy) + \eta^0(0)\bar{\eta}(x_0). \tag{16.28}$$

We obtain the integral inequality

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{[0,T]}|v^\varepsilon - v|^p\Big] \\
&\leqslant \varepsilon C_7 \mathbb{E}\Big[\sup_{[0,T]}|v^\varepsilon - v|^p\Big] + \varepsilon(p-1)C_\infty T^p + \varepsilon C_8 \int_0^T \mathbb{E}\Big[\sup_{[0,s]}|v^\varepsilon - v|^p\Big]ds \\
&\quad + \varepsilon C_9 \int_0^T \mathbb{E}\Big[\sup_{[0,s]}|v^\varepsilon - v|^{p-1}\Big]ds \\
&\leqslant \varepsilon \mathbb{E}\Big[\sup_{[0,T]}|v^\varepsilon - v|^p\Big] + \varepsilon(p-1)C_\infty T^p + \varepsilon C_8 \int_0^T \mathbb{E}\Big[\sup_{[0,s]}|v^\varepsilon - v|^p\Big]ds \\
&\quad + \varepsilon C_9 \int_0^T \mathbb{E}\Big[\sup_{[0,s]}|v^\varepsilon - v|^p\Big]^{\frac{p-1}{p}}ds.
\end{aligned}$$

Hence for any value $\varepsilon \in (0, \frac{1}{2}]$ we eliminate the first term

$$\begin{aligned}
\mathbb{E}\Big[\sup_{[0,T]}|v^\varepsilon - v|^p\Big] &\leqslant 2\varepsilon(p-1)C_\infty T^p + 2\varepsilon C_8 \int_0^T \mathbb{E}\Big[\sup_{[0,s]}|v^\varepsilon - v|^p\Big]ds \\
&\quad + 2\varepsilon C_9 \int_0^T \mathbb{E}\Big[\sup_{[0,s]}|v^\varepsilon - v|^p\Big]^{\frac{p-1}{p}}ds.
\end{aligned}$$

That is, for $\Psi(T) = \mathbb{E}\big[\sup_{[0,T]}|v^\varepsilon - v|^p\big]$ we have

$$\Psi(T) \leqslant \varepsilon C_{10} T^p + \varepsilon C_{11} \int_0^T \Psi(s)ds + \varepsilon C_{12} \int_0^T \Psi(s)^{\frac{p-1}{p}}ds.$$

Using the nonlinear extension of the Gronwall-Bihari inequality in Corollary 16.14 in the appendix essentially given by Pachpatte [23], Theorem 2.4.2, which we adapt to our case we obtain a global constant $C > 0$ such that using that $\varepsilon_0 T$ is sufficiently small implies for all $\varepsilon \in (0, \varepsilon_0]$

$$\Psi(T) \leqslant C\big(\varepsilon T^p + \varepsilon^{\frac{p-1}{p}} T^{p+\frac{p-1}{p}}\big). \tag{16.29}$$

3. Estimate of the Horizontal Component $\mathbb{E}[\sup|u^\varepsilon - u|^p]$: For convenience of notation we restart with the numbering of constants. Formally we obtain

$$\begin{aligned} u_t^\varepsilon - u_t &= \int_0^t (\mathfrak{F}_0(u_s^\varepsilon, v_s^\varepsilon) - \mathfrak{F}_0(u_s, v_s))ds + \int_0^t (\mathfrak{F}(u_{s-}^\varepsilon, v_{s-}^\varepsilon) - \mathfrak{F}(u_{s-}, v_{s-})) \diamond dZ_s \\ &\quad + \varepsilon \int_0^t \big(\mathfrak{K}_H(u_s^\varepsilon, v_s^\varepsilon) - \mathfrak{K}_H(u_s, v_s)\big)ds + \varepsilon \int_0^t \mathfrak{K}_H(u_s, v_s)ds \\ &\quad + \varepsilon \int_0^t \tilde{\mathfrak{K}}_H(v_{s-}^\varepsilon) \diamond d\tilde{Z}_s. \end{aligned} \tag{16.30}$$

For further details consult [8] where we obtain with the help of the change of variable formula for (16.30) the following equality in $\mathbb{R}^n$ almost surely for $t \geqslant 0$

$$\begin{aligned} &|u_t^\varepsilon - u_t|^p \\ &= p \int_0^t |u_s^\varepsilon - u_s|^{p-2}\langle u_s^\varepsilon - u_s, \mathfrak{F}_0(u_s^\varepsilon, v_s^\varepsilon) - \mathfrak{F}_0(u_s, v_s)\rangle ds && (I_1) \\ &\quad + p \int_0^t |u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, (\mathfrak{F}(u_{s-}^\varepsilon, v_{s-}^\varepsilon) - \mathfrak{F}(u_{s-}, v_{s-}))dZ_s\rangle && (I_2) \\ &\quad + p \sum_{0<s\leqslant t} |u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, \Phi^{\mathfrak{F}\Delta_s Z}(u_{s-}^\varepsilon, v_{s-}^\varepsilon) - \Phi^{\mathfrak{F}\Delta_s Z}(u_{s-}, v_{s-}) \\ &\qquad - (u_{s-}^\varepsilon - u_{s-}, v_{s-}^\varepsilon - v_{s-}) - (\mathfrak{F}(u_{s-}^\varepsilon, v_{s-}^\varepsilon) - \mathfrak{F}(u_{s-}, v_{s-}))\Delta_s Z\rangle && (I_3) \\ &\quad + \varepsilon\, p \int_0^t |u_s^\varepsilon - u_s|^{p-2}\langle u_s^\varepsilon - u_s, \mathfrak{K}_H(u_s^\varepsilon, v_s^\varepsilon) - \mathfrak{K}_H(u_s, v_s)\rangle ds && (I_4) \\ &\quad + \varepsilon p \int_0^t |u_s^\varepsilon - u_s|^{p-2}\langle u_s^\varepsilon - u_s, \mathfrak{K}_H(u_s, v_s)\rangle ds && (I_5) \\ &\quad + \varepsilon p \int_0^t |u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, \tilde{\mathfrak{K}}_H(v_{s-}^\varepsilon)d\tilde{Z}_s\rangle && (I_6) \\ &\quad + p \sum_{0<s\leqslant t} |u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, \Phi^{\varepsilon\tilde{\mathfrak{K}}_H\Delta_s\tilde{Z}}(v_{s-}^\varepsilon) - \Phi^{\varepsilon\tilde{\mathfrak{K}}_H\Delta_s\tilde{Z}}(v_{s-}) \\ &\qquad - (v_{s-}^\varepsilon - v_{s-}) - \varepsilon(\tilde{\mathfrak{K}}_H(v_{s-}^\varepsilon) - \tilde{\mathfrak{K}}_H(v_{s-}))\Delta_s\tilde{Z}\rangle && (I_7) \\ &\quad + p \sum_{0<s\leqslant t} |u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, \Phi^{\varepsilon\tilde{\mathfrak{K}}_H\Delta_s\tilde{Z}}(v_{s-}) - v_{s-} - \varepsilon\tilde{\mathfrak{K}}_H(v_{s-})\Delta_s\tilde{Z}\rangle \\ &&& (I_8) \\ &=: I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7 + I_8. && (16.31) \end{aligned}$$

In fact, we shall use the following estimate

$$|u_t^\varepsilon - u_t|^{2p} \leqslant 8^{p-1} \sum_{i=1}^{8} I_i^2. \tag{16.32}$$

We shall estimate each of the eight preceding summands on the right-hand side. The estimates of I_1 and I_4 are direct Lipschitz estimates. For the stochastic Itô terms we use the different kinds of maximal inequalities, see for instance [2] and [18]. The estimate of the canonical Marcus terms I_3, I_7 and I_8 is the most difficult task in which we use the result of Lemma 16.8. The term I_5 is straightforward.

3.1 Estimate of the Stochastic Itô Integral Terms I_2 and I_6: $\mathbf{I_2}$: Due to the existence of moments of order at least 1, I_2 has the following representation with respect to the compensated Poisson random measure associated to Z

$$\int_0^t |u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, \big(\mathfrak{F}(u_{s-}^\varepsilon, v_{s-}^\varepsilon) - \mathfrak{F}(u_{s-}, v_{s-})\big)dZ_s\rangle$$

$$= \int_0^t \int_{\mathbb{R}^r} |u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, \big(\mathfrak{F}(u_{s-}^\varepsilon, v_{s-}^\varepsilon) - \mathfrak{F}(u_{s-}, v_{s-})\big)z\rangle \tilde{N}(dsdz) \tag{16.33}$$

$$+ \int_0^t \int_{\|z\|>1} |u_s^\varepsilon - u_s|^{p-2}\langle u_s^\varepsilon - u_s, \big(\mathfrak{F}(u_s^\varepsilon, v_s^\varepsilon) - \mathfrak{F}(u_s, v_s)\big)z\rangle \nu(dz)ds. \tag{16.34}$$

For the first term (16.33) we exploit the embedding $L^2 \subset L^1$, Kunita's maximal inequality (see [2] or [18]) for exponent equal to 2, and the Young inequality for the exponents $p/2$ and $p/(p-2)$ combined with Inequality (16.29) and obtain

$$\mathbb{E}\left[\sup_{[0,T]}\Big|\int_0^\cdot \int_{\mathbb{R}^r} |u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, \big(\mathfrak{F}(u_{s-}^\varepsilon, v_{s-}^\varepsilon) - \mathfrak{F}(u_{s-}, v_{s-})\big)z\rangle \tilde{N}(dsdz)\Big|\right]^2$$

$$\leqslant \mathbb{E}\left[\sup_{[0,T]}\Big|\int_0^\cdot \int_{\mathbb{R}^r} |u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle \dots, \big(\mathfrak{F}(u_{s-}^\varepsilon, v_{s-}^\varepsilon) - \mathfrak{F}(u_{s-}, v_{s-})\big)z\rangle \tilde{N}(dsdz)\Big|^2\right]$$

$$= \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^r} |u_s^\varepsilon - u_s|^{2(p-2)}|\langle u_s^\varepsilon - u_s, \big(\mathfrak{F}(u_s^\varepsilon, v_s^\varepsilon) - \mathfrak{F}(u_s, v_s)\big)z\rangle|^2 \nu(dz)ds\right]$$

$$\leqslant C_1 \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^r} |u_s^\varepsilon - u_s|^{2(p-1)}\Big(|u_s^\varepsilon - u_s|^2 + |v_s^\varepsilon - v_s|^2\Big)\|z\|^2 \nu(dz)ds\right]$$

$$\leqslant C_1 \left(\int_{\mathbb{R}^r} \|z\|^2 \nu(dz)\right) \mathbb{E}\left[\int_0^T \big(|u_s^\varepsilon - u_s|^{2p} + |u_s^\varepsilon - u_s|^{p-2}|v_s^\varepsilon - v_s|^2\big)ds\right]$$

$$\leqslant C_2\Big(\int_0^T \mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon - u|^{2p}\Big]ds + \int_0^T \mathbb{E}\Big[|v_s^\varepsilon - v_s|^{2p}\Big]ds\Big)$$

$$\leqslant C_2\Big(\int_0^T \mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon - u|^{2p}\Big]ds\Big) + C\big(\varepsilon T^{2p} + \varepsilon^{\frac{2p-1}{2p}} T^{2(p+1)+1}\big). \tag{16.35}$$

The second term follows by Young's inequality and the Lipschitz continuity of $\mathfrak{F}$:

$$\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t\int_{\|z\|>1}|u_s^\varepsilon - u_s|^{p-2}\langle u_s^\varepsilon - u_s, \big(\mathfrak{F}(u_s^\varepsilon, v_s^\varepsilon) - \mathfrak{F}(u_s, v_s)\big)z\rangle\nu(dz)ds\Big]^2$$

$$\leqslant \Big(\ell\int_{\|z\|>1}\|z\|\nu(dz)\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t\big(|u_s^\varepsilon - u_s|^p + |u_s^\varepsilon - u_s|^{p-1}|v_s^\varepsilon - v_s|\big)ds\Big]\Big)^2$$

$$\leqslant \Big(\ell\int_{\|z\|>1}\|z\|\nu(dz)\Big(2\int_0^T\mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon - u|^p\Big]ds + \int_0^T\mathbb{E}\Big[|v_s^\varepsilon - v_s|^p\Big]ds\Big)\Big)^2$$

$$\leqslant C_3 T\int_0^T\mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon - u|^{2p}\Big]ds + C\big(\varepsilon T^{2p} + \varepsilon^{\frac{2p-1}{2p}}T^{2(p+1)+1}\big). \tag{16.36}$$

$\mathbf{I_6}$: We go over to the representation with the Poisson random measure $\tilde{N}'$ associated to the Lévy process $\tilde{Z}$ and obtain

$$\sup_{t\in[0,T]}\varepsilon\int_0^t|u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, \tilde{\mathfrak{K}}_H(v_{s-}^\varepsilon)d\tilde{Z}_s\rangle$$

$$= \sup_{t\in[0,T]}\varepsilon\int_0^t\int_{\mathbb{R}^r}|u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, (\tilde{\mathfrak{K}}_H(v_{s-}^\varepsilon) - \tilde{\mathfrak{K}}_H(v_{s-}))z\rangle\tilde{N}'(dsdz) \tag{J_1}$$

$$+ \sup_{t\in[0,T]}\varepsilon\int_0^t\int_{\|z\|>1}|u_s^\varepsilon - u_s|^{p-2}\langle u_s^\varepsilon - u_s, (\tilde{\mathfrak{K}}_H(v_s^\varepsilon) - \tilde{\mathfrak{K}}_H(v_s))z\rangle\nu'(dz)ds \tag{J_2}$$

$$+ \sup_{t\in[0,T]}\varepsilon\int_0^t\int_{\mathbb{R}^r}|u_{s-}^\varepsilon - u_{s-}|^{p-2}\langle u_{s-}^\varepsilon - u_{s-}, \tilde{\mathfrak{K}}_H(v_{s-})z\rangle\tilde{N}'(dsdz) \tag{J_3}$$

$$+ \sup_{t\in[0,T]}\varepsilon\int_0^t\int_{\|z\|>1}|u_s^\varepsilon - u_s|^{p-2}\langle u_s^\varepsilon - u_s, \tilde{\mathfrak{K}}_H(v_s)z\rangle\nu'(dz)ds. \tag{J_4}$$

The terms J_1 and J_2 are estimated analogously to (16.35) and (16.36) where $\mathfrak{F}$ is replaced by $\tilde{\mathfrak{K}}_H$, which yield the following estimates

$$\Big(\mathbb{E}\Big[\sup_{[0,T]}|\varepsilon\int_0^{\cdot}\int_{\mathbb{R}^r}|u^\varepsilon_{s-}-u_{s-}|^{p-2}\langle u^\varepsilon_{s-}-u_{s-},(\tilde{\mathfrak{K}}_H(v^\varepsilon_{s-})-\tilde{\mathfrak{K}}_H(v_{s-}))z\rangle\tilde{N}'(dsdz)|\Big]\Big)^2$$

$$\leqslant C_4\left(\int_0^T\mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon-u|^{2p}\Big]ds\right)^{\frac{1}{2}}+C\big(\varepsilon T^{2p}+\varepsilon^{\frac{2p-1}{2p}}T^{2(p+1)+1}\big)\quad\text{and}$$

$$\Big(\mathbb{E}\Big[\sup_{[0,T]}\varepsilon|\int_0^t\int_{\|z\|>1}|u^\varepsilon_s-u_s|^{p-2}\langle u^\varepsilon_s-u_s,(\tilde{\mathfrak{K}}_H(v^\varepsilon_s)-\tilde{\mathfrak{K}}_H(v_s))z\rangle\nu'(dz)ds|\Big]\Big)^2$$

$$\leqslant C_5\int_0^T\mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon-u|^{2p}\Big]ds+C\big(\varepsilon T^{2p}+\varepsilon^{\frac{2p-1}{2p}}T^{2(p+1)+1}\big).$$

For the term J_3 we observe that $v_s=0$ consequently $\mathfrak{K}_V(v_s)$ is constant. Applying Kunita's maximal inequality for the exponent 2, we obtain

$$\Big(\mathbb{E}\Big[\sup_{t\in[0,T]}\varepsilon|\int_0^t\int_{\mathbb{R}^r}|u^\varepsilon_{s-}-u_{s-}|^{p-2}\langle u^\varepsilon_{s-}-u_{s-},\tilde{\mathfrak{K}}_H(v_{s-})z\rangle\tilde{N}'(dsdz)|\Big]\Big)^2$$

$$\leqslant\varepsilon^2\mathbb{E}\Big[\sup_{t\in[0,T]}|\int_0^t\int_{\mathbb{R}^r}|u^\varepsilon_{s-}-u_{s-}|^{p-2}\langle u^\varepsilon_{s-}-u_{s-},\tilde{\mathfrak{K}}_H(v_{s-})z\rangle\tilde{N}'(dsdz)|^2\Big]$$

$$\leqslant\varepsilon^2C_6\int_0^T\int_{\mathbb{R}^r}\mathbb{E}\Big[|u^\varepsilon_s-u_s|^{2(p-1)}\Big]\|z\|^2\nu'(dz)ds$$

$$\leqslant\varepsilon^2C_6\left(\int_{\mathbb{R}^r}\|z\|^2\nu'(dz)\right)\left(\int_0^T\mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon-u|^{2p-2}\Big]ds\right)$$

$$\leqslant\varepsilon^2C_7\int_0^T\mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon-u|^{2p-2}\Big]ds$$

$$\leqslant\varepsilon^2C_7\left(\int_0^T\mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon-u|^{2p}\Big]ds+\frac{C_8}{p}T\right).$$

The term J_4 is again easier. Using $\varepsilon T<1$ and $\varepsilon<1$ we obtain

$$\Big(\mathbb{E}\Big[\sup_{[0,T]}\varepsilon\int_0^t\int_{\|z\|>1}|u^\varepsilon_s-u_s|^{p-2}\langle u^\varepsilon_s-u_s,\tilde{\mathfrak{K}}_H(v_s)z\rangle\nu'(dz)ds\Big]\Big)^2$$

$$\leqslant\left(\varepsilon\int_{\|z\|>1}\|z\|\nu'(dz)\|\tilde{\mathfrak{K}}_H(0)\|\int_0^T\mathbb{E}[\sup_{[0,s]}|u^\varepsilon-u|^{p-1}]ds\right)^2$$

$$\begin{aligned}
&\leqslant \varepsilon^2 C_9 \Big(\int_0^T \mathbb{E}\Big[\sup_{[0,s]} |u^\varepsilon - u|^p \Big] ds \Big)^2 + C_9 \varepsilon^{2p} T^2 \\
&\leqslant \varepsilon^2 T C_9 \int_0^T \mathbb{E}\Big[\sup_{[0,s]} |u^\varepsilon - u|^{2p} \Big] ds + C_9 \varepsilon^{2p} T^2 \\
&\leqslant C_9 \int_0^T \mathbb{E}\Big[\sup_{[0,s]} |u^\varepsilon - u|^{2p} \Big] ds + C_9 \varepsilon^{2p} T^2 .
\end{aligned}$$

Summing up we obtain

$$\mathbb{E}[\sup_{[0,T]} |I_6|^2] \leqslant C_{10} \Big(\int_0^T \mathbb{E}\Big[\sup_{[0,s]} |u^\varepsilon - u|^p \Big] ds + \varepsilon^{\frac{2p-1}{2p}} T^{2(p+1)+1} + \varepsilon^2 T \Big). \tag{16.37}$$

3.2 Estimate of the Canonical Marcus Terms I_3, I_7 and I_8 The estimate is identical to estimate (54) in [8] and yields a constant C_{11} such that

$$|I_3| \leqslant 2C_{11} \Big(\sum_{0<s\leqslant t} |u^\varepsilon_{s-} - u_{s-}|^p \|\Delta_s Z\|^2 + \sum_{0<s\leqslant t} |v^\varepsilon_{s-} - v_{s-}|^p \|\Delta_s Z\|^2 \Big). \tag{16.38}$$

Once again, the representation of this sum in terms of the Poisson random measure given in Kunita [18] tells us that

$$\begin{aligned}
&\sum_{0<s\leqslant t} |u^\varepsilon_{s-} - u_{s-}|^p \|\Delta_s Z\|^2 \\
&= \int_0^t \int_{\mathbb{R}^r} |u^\varepsilon_{s-} - u_{s-}|^p \|z\|^2 \tilde{N}(ds dz) + \int_0^t \int_{\|z\|>1} |u^\varepsilon_s - u_s|^p \|z\|^2 \, \nu(dz)\, ds.
\end{aligned} \tag{16.39}$$

The maximal inequality for integrals with respect to the compensated Poisson random measures and inequality (16.29) yield

$$\begin{aligned}
\mathbb{E}[\sup_{[0,T]} |I_3|^2] &\leqslant C_{12} \int_0^T \int_{\mathbb{R}^r} \Big(\mathbb{E}[\sup_{[0,s]} |u^\varepsilon - u|^{2p}] + \mathbb{E}[|v^\varepsilon_s - v_s|^{2p}] \Big) \|z\|^4 \, \nu(dz)\, ds \\
&= C_{12} \int_{\mathbb{R}^r} \|z\|^4 \nu(dz) \Big(\int_0^T \Big(\mathbb{E}[\sup_{[0,s]} |u^\varepsilon - u|^{2p}] + \mathbb{E}[|v^\varepsilon_s - v_s|^{2p}] \Big) ds \Big) \\
&\leqslant C_{13} \Big(\int_0^T \mathbb{E}[\sup_{[0,s]} |u^\varepsilon - u|^{2p} \, ds + C\big(\varepsilon T^{2p} + \varepsilon^{\frac{2p-1}{2p}} T^{2(p+1)+1} \big) \Big).
\end{aligned} \tag{16.40}$$

I₇: For I_7 we apply Lemma 16.8 statement 1) and Young's inequality and obtain the analogous result

$$\begin{aligned}
&\sum_{0<s\leqslant t} |u^\varepsilon_{s-} - u_{s-}|^{p-2}\langle u^\varepsilon_{s-} - u_{s-}, \Phi^{\varepsilon\tilde{\mathfrak{K}}_H\Delta_s\tilde{Z}}(v^\varepsilon_{s-}) - \Phi^{\varepsilon\tilde{\mathfrak{K}}_H\Delta_s\tilde{Z}}(v_{s-})\\
&\qquad - (v^\varepsilon_{s-} - v_{s-}) - \varepsilon(\tilde{\mathfrak{K}}_H(v^\varepsilon_{s-}) - \tilde{\mathfrak{K}}_H(v_{s-}))\Delta_s\tilde{Z}\rangle\\
&\leqslant \varepsilon^2 C_{14}\Big(\sum_{0<s\leqslant t}\big(|u^\varepsilon_{s-} - u_{s-}|^p + |v^\varepsilon_{s-} - v_{s-}|^p\big)\|\Delta_s\tilde{Z}\|^2\Big).
\end{aligned}$$

Rewriting the last expression in terms of the (compensated) Poisson random measure $\tilde{N}'$ we obtain

$$\begin{aligned}
&\sum_{0<s\leqslant t}\big(|u^\varepsilon_{s-} - u_{s-}|^p + |v^\varepsilon_{s-} - v_{s-}|^p\big)\|\Delta_s\tilde{Z}\|^2\\
&= \int_0^t\int_{\mathbb{R}^r}\big(|u^\varepsilon_{s-} - u_{s-}|^p + |v^\varepsilon_{s-} - v_{s-}|^p\big)\|z\|^2\tilde{N}'(dsdz) \qquad (16.41)\\
&\quad + \int_0^t\int_{\|z\|>1}\big(|u^\varepsilon_{s} - u_{s}|^p + |v^\varepsilon_{s} - v_{s}|^p\big)\|z\|^2\nu'(dz)ds. \qquad (16.42)
\end{aligned}$$

Kunita's maximal inequality for the exponent 2 yields

$$\begin{aligned}
&\mathbb{E}\Big[|\sup_{[0,T]}\int_0^t\int_{\mathbb{R}^r}\big(|u^\varepsilon_{s-} - u_{s-}|^p + |v^\varepsilon_{s-} - v_{s-}|^p\big)\|z\|^2\tilde{N}'(dsdz)|^2\Big]\\
&\leqslant C_{15}\int_0^T\int_{\mathbb{R}^r}\mathbb{E}\Big[|u^\varepsilon_s - u_s|^{2p} + |v^\varepsilon_s - v_s|^{2p}\Big]\|z\|^2\nu'(dz)ds\\
&\leqslant C_{16}\left(\int_{\mathbb{R}^r}\|z\|^2\nu'(dz)\int_0^T\mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon - u|^{2p}\Big]ds + \int_0^T\mathbb{E}\Big[|v^\varepsilon_s - v_s|^{2p}\Big]ds\right)\\
&\leqslant C_{16}\int_{\mathbb{R}^r}\|z\|^2\nu'(dz)\int_0^T\mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon - u|^{2p}\Big]ds + C_{17}\big(\varepsilon T^{2p} + \varepsilon^{\frac{2p-1}{2p}}T^{2(p+1)+1}\big),
\end{aligned}$$

where $C_{17} = C$ from (16.29). The term (16.42) is treated obviously such that

$$\begin{aligned}
\mathbb{E}[\sup_{[0,T]}|I_7|^2] &\leqslant \varepsilon^2 C_{18}\int_{\mathbb{R}^r}\|z\|^4\nu'(dz)\int_0^T\mathbb{E}\Big[\sup_{[0,s]}|u^\varepsilon - u|^{2p}\Big]ds\\
&\quad + C_{17}\big(\varepsilon T^{2p} + \varepsilon^{\frac{2p-1}{2p}}T^{2(p+1)+1}\big). \qquad (16.43)
\end{aligned}$$

I₈: For I_8 Lemma 16.8, statement 2), yields

$$\begin{aligned}
&\sum_{0<s\leqslant t} |u^{\varepsilon}_{s-} - u_{s-}|^{p-2}\langle u^{\varepsilon}_{s-} - u_{s-}, \Phi^{\varepsilon\tilde{\mathfrak{K}}_H\Delta_s\tilde{Z}}(v_{s-}) - v_{s-} - \varepsilon\tilde{\mathfrak{K}}_H(v_{s-})\Delta_s\tilde{Z}\rangle \\
&\leqslant \sum_{0<s\leqslant t} |u^{\varepsilon}_{s-} - u_{s-}|^{p-1}|\Phi^{\varepsilon\tilde{\mathfrak{K}}_H\Delta_s\tilde{Z}}(v_{s-}) - v_{s-} - \varepsilon\tilde{\mathfrak{K}}_H(v_{s-})\Delta_s\tilde{Z}| \\
&\leqslant \varepsilon^2 C_{19}\sum_{0<s\leqslant t} |u^{\varepsilon}_{s-} - u_{s-}|^{p-1}\|\Delta_s\tilde{Z}\|^2.
\end{aligned}$$

Therefore Kunita's inequality with exponent 2 and elementary Young's estimate for parameters $\frac{p}{p-2}$ and $\frac{p}{2}$ imply

$$\begin{aligned}
\mathbb{E}[\sup_{[0,T]}|I_8|^2] &\leqslant \varepsilon^2 C_{20}\int_{\mathbb{R}^r}\|z\|^4\nu'(dz)\int_0^T \mathbb{E}[\sup_{[0,s]}|u^{\varepsilon} - u|^{2p-2}\Big]ds \\
&\leqslant \varepsilon C_{21}\int_0^T \mathbb{E}[\sup_{[0,s]}|u^{\varepsilon} - u|^{2p}\Big]ds + C_{21}\varepsilon^{\frac{p}{2}}T. \qquad (16.44)
\end{aligned}$$

3.3 Estimate of I_5

$$\begin{aligned}
\int_0^T |u^{\varepsilon}_s - u_s|^{p-2}\langle u^{\varepsilon}_s - u_s, \varepsilon\mathfrak{K}_H(u_s, v_s)\rangle ds &\leqslant C_{22}\int_0^T \varepsilon|u^{\varepsilon}_s - u_s|^{p-1}ds \\
&\leqslant C_{22}\int_0^T \varepsilon|u^{\varepsilon}_s - u_s|^{p}ds + C_{22}\varepsilon^p T
\end{aligned}$$

such that

$$\mathbb{E}[\sup_{[0,T]}|I_5|^2] \leqslant \varepsilon C_{23}T\int_0^T \mathbb{E}\Big[\sup_{[0,s]}|u^{\varepsilon} - u|^{2p}\Big]ds + C_{23}\varepsilon^{2p}T^2. \qquad (16.45)$$

3.4 Linear Comparison Principle: Taking the supremum and the expectation of the left-hand side of Eq. (16.31) and combining the estimates of $\sum_{i=1}^{8}\mathbb{E}[\sup_{[0,T]}|I_i|]$ given by (16.35)–(16.37), (16.40), (16.43)–(16.45) we obtain

a positive constant C_{24}

$$\begin{aligned}
&\mathbb{E}\left[\sup_{[0,T]}|u^\varepsilon - u|^{2p}\right] \\
&\leqslant C_{24}\left(\int_0^T \mathbb{E}\left[\sup_{[0,s]}|u^\varepsilon - u|^{2p}\right]ds + \varepsilon T^{2p} + \varepsilon^{\frac{2p-1}{2p}}T^{2(p+1)+1} + (\varepsilon^p T)^2\right) \\
&\leqslant C_{24}\left(\int_0^T \mathbb{E}\left[\sup_{[0,s]}|u^\varepsilon - u|^{2p}\right]ds + \varepsilon T^{2p+3}\right).
\end{aligned}$$

Finally

$$\mathbb{E}\left[\sup_{[0,T]}|u^\varepsilon - u|^{p}\right] \leqslant \mathbb{E}\left[\sup_{[0,T]}|u^\varepsilon - u|^{2p}\right] \leqslant C_{25}\varepsilon^{\frac{2p-1}{2p}}T^{2p+3}e^{C_{25}T}.$$

4. Conclusion: The estimates of the sum of the vertical and the horizontal estimate yield

$$\begin{aligned}
\mathbb{E}\Big[\sup_{[0,T]}|h(X^\varepsilon) - h(X)|^p\Big] &\leqslant C_0\Big(\mathbb{E}\Big[\sup_{[0,T]}|u^\varepsilon - u|^p\Big] + \mathbb{E}\Big[\sup_{[0,T]}|v^\varepsilon - v|^p\Big]\Big) \\
&\leqslant C_{26}\varepsilon^{\frac{2p-1}{2p}}T^{2p+3}e^{C_{25}T} + C\big(\varepsilon T^p + \varepsilon^{\frac{p-1}{p}}T^{p+\frac{p-1}{p}}\big) \\
&\leqslant C_{27}\varepsilon^{\frac{p-1}{p}}e^{C_{27}T}.
\end{aligned}$$

We finally note that the only dependence on the initial conditions stems from C_∞ and hence by (16.28) the estimate $C_{27} \leqslant C_{28}(1 + \eta^0(0)\bar{\eta}(x_0))$. This finishes the proof.

16.4 The Averaging Error and the Proof of the Main Result

For convenience we fix the following notation. Given $h : M \to \mathbb{R}^n$ a globally Lipschitz continuous function and $Q^h : V \to \mathbb{R}^n$ its average on the leaves defined in definition (16.6). For $t \geqslant 0$, $x_0 \in M$ and $\varepsilon \in (0, 1]$ we write

$$\delta^h_{x_0}(\varepsilon, t) := \int_0^{t\wedge\varepsilon\tau^\varepsilon} h(X^\varepsilon_{\frac{s}{\varepsilon}}(x_0)) - Q^h(\pi(X^\varepsilon_{\frac{s}{\varepsilon}}(x_0)))ds.$$

Proposition 16.9 *Let the assumptions of Proposition 16.6 be satisfied for fixed $p \geqslant 2$. Then for any globally Lipschitz continuous function $h : M \to \mathbb{R}^n$, $\lambda \in (0, \frac{p-1}{p^2})$*

and $x_0 \in M$ there exist constants $b_1 > 0$ and $\varepsilon_0 \in (0,1]$ such that for $\varepsilon \in (0,\varepsilon_0]$ and $T \in [0,1]$ we have

$$\left(\mathbb{E}\left[\sup_{s\in[0,T]} |\delta^h_{x_0}(\varepsilon,s)|^p\right]\right)^{\frac{1}{p}} \leqslant b_1 T\left[\varepsilon^{\lambda} + \eta^0\left(cT|\ln\epsilon|\right)\right],$$

where $c := \frac{1}{k_2}\left(\frac{p-1}{p^2} - \lambda\right) \wedge \ell Lip(\phi^{-1})$ is given in Corollary 16.7 and η^0 is the temporal factor of the ergodic rate of convergence given in estimate (16.9) of Hypothesis 3.

Proof of Proposition (16.9) Fix $x_0 \in M$. For $\varepsilon \in (0,1)$ and $T > 0$ we define the partition

$$t_0 = 0 < t^\varepsilon_1 < \cdots < t^\varepsilon_{N^\varepsilon} \leqslant \frac{T}{\varepsilon} \wedge \tau^\varepsilon$$

with the step size

$$\Delta_\varepsilon := -cT\ln(\varepsilon) \qquad \text{for some } c > 0.$$

The grid points of the partition are given by $t^\varepsilon_n := n\Delta_\varepsilon \wedge \tau^\varepsilon$ for $0 \leqslant n \leqslant N_\varepsilon$ for $\varepsilon \in (0,1]$ with $N_\varepsilon = \lfloor \frac{1}{c\varepsilon|\ln(\varepsilon)|} \rfloor$. The term $\delta^h_{x_0}(\varepsilon,t)$ can be estimated by the following three sums

$$|\delta^h_{x_0}(\varepsilon,T)| \leqslant |A_1(T,\varepsilon)| + |A_2(T,\varepsilon)| + |A_3(T,\varepsilon)|, \tag{16.46}$$

where

$$A_1(T,\varepsilon) := \varepsilon \sum_{n=0}^{N_\varepsilon} \int_{t_n}^{t_{n+1}} [h(X^\varepsilon_s(x_0)) - h(X_{s-t_n}(X^\varepsilon_{t_n}(x_0)))]\, ds,$$

$$A_2(T,\varepsilon) := \varepsilon \sum_{n=0}^{N_\varepsilon} \int_{t_n}^{t_{n+1}} [h(X_{s-t_n}(X^\varepsilon_{t_n}(x_0))) - \Delta_\varepsilon Q(\pi(X^\varepsilon_{t_n}(x_0)))]\, ds,$$

$$A_3(T,\varepsilon) := \sum_{n=0}^{N_\varepsilon} \varepsilon\Delta_\varepsilon Q(\pi(X^\varepsilon_{t_n}(x_0))) - \int_0^{t_{N_\varepsilon+1}} \varepsilon Q(\pi(X^\varepsilon_{\frac{s}{\varepsilon}}(x_0)))\, ds.$$

The following lemmas estimate the preceding terms one-by-one. For convenience of the reader we number the constants C_i.

Lemma 16.10 *For any* $\lambda \in (0, \frac{p-1}{p})$ *there exist constants* $b_2 > 0$ *and* $\varepsilon_0 \in (0, 1]$ *such that for any* $\varepsilon \in (0, \varepsilon_0]$ *and* $T \geqslant 0$

$$\left(\mathbb{E}\left[\sup_{s\in[0,T]} |A_1(s,\varepsilon)|^p\right]\right)^{\frac{1}{p}} \leqslant b_2 T \varepsilon^{\lambda}.$$

Proof Using the Markov property analogously to [8] and Corollary 16.7 we obtain

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{[0,T]} A_1(s,\varepsilon)|^p\Big]^{\frac{1}{p}} \\
&= \varepsilon \sum_{n=0}^{N_\varepsilon-1} \mathbb{E}\Big[\mathbb{E}\Big[|\int_{t_n}^{t_{n+1}} |h(X_s^\varepsilon(x_0)) - h(X_{s-t_n}(X_{t_n}^\varepsilon(x_0)))ds|^p \mid \mathscr{F}_{t_n}\Big]\Big]^{\frac{1}{p}} \\
&= \varepsilon \sum_{n=0}^{N_\varepsilon-1} \mathbb{E}\Big[\mathbb{E}\Big[|\int_{t_n}^{t_{n+1}} |h(X_{s-t_n}^\varepsilon(y)) - h(X_{s-t_n}(y))ds|^p \mid y = X_{t_n}^\varepsilon(x_0)\Big]\Big]^{\frac{1}{p}} \\
&\leqslant \varepsilon(N_\varepsilon+1)\Delta_\varepsilon \max_{n=0,\dots,N_\varepsilon} \mathbb{E}\Big[\mathbb{E}\Big[|\sup_{s\in[0,t_1]} |h(X_s^\varepsilon(y)) - h(X_{s-t_n}(y))|^p \mid y = X_{t_n}^\varepsilon(x_0)\Big]\Big]^{\frac{1}{p}} \\
&\leqslant T\varepsilon^{\lambda} \max_{n=0,\dots,N_\varepsilon} \mathbb{E}\big[k_4(X_{t_n}^\varepsilon(x_0))\big].
\end{aligned}$$

Note that by Corollary (16.7) we have

$$\max_{n=0,\dots,N_\varepsilon} \mathbb{E}\big[k_4(X_{t_n}^\varepsilon(x_0))\big] \leqslant k_3k_5\big(1 + \max_{n=0,\dots,N_\varepsilon} \mathbb{E}\big[\bar{\eta}(X_{t_n}^\varepsilon(x_0))\big]\big).$$

It remains to bound the last summand. We estimate as follows for any $n \in \mathbb{N}$

$$\begin{aligned}
&\mathbb{E}\big[\bar{\eta}(X_{t_n}^\varepsilon(x_0))\big] \\
&\leqslant \mathbb{E}\big[\bar{\eta}(X_{t_n}^\varepsilon(x_0)) - \bar{\eta}(X_{t_n}(x_0))\big] + \mathbb{E}\Big[\bar{\eta}(X_{t_n}(x_0))\Big] \\
&\leqslant \bar{\ell}\,\mathbb{E}\big[|X_{t_n}^\varepsilon(x_0) - X_{t_n}(x_0)|\big] + \int \bar{\eta}(y)\mu_{x_0}(dy) + \sup_{t\geqslant 0} \mathbb{E}\Big[|\int \bar{\eta}(y)\mu_{x_0}(dy) - \bar{\eta}(X_t(x_0))|\Big] \\
&\leqslant \bar{\ell}\,\mathbb{E}\big[|X_{t_n}^\varepsilon(x_0) - X_{t_n}(x_0)|^p\big]^{\frac{1}{p}} + C_1.
\end{aligned}$$

For the first term in the preceding expression we derive a recursion formula. Using Theorem 3.2 in Kunita [18] it yields for the horizontal component

$$\mathbb{E}\Big[\sup_{s\in0,T]} |X_s(x_1) - X_s(x_2)|^p\Big] \leqslant e^{\ell Lip(\phi^{-1})T}|x_1 - x_2|^p,$$

which implies the inequality

$$\mathbb{E}\Big[|X_{t_1}(x_1) - X_{t_1}(x_2)|^p\Big] \leqslant C_2\varepsilon^\lambda |x_1 - x_2|^p.$$

We estimate

$$\begin{aligned}
&\mathbb{E}\big[|X^\varepsilon_{t_n}(x_0) - X_{t_n}(x_0)|^p\big]^{\frac{1}{p}}\\
&\leqslant \mathbb{E}\big[|X^\varepsilon_{t_n}(x_0) - X_{t_n-t_{n-1}}(X^\varepsilon_{t_{n-1}}(x_0))|^p\big]^{\frac{1}{p}} + \mathbb{E}\big[|X_{t_n}(X^\varepsilon_{t_{n-1}}(x_0)) - X_{t_n}(x_0)|^p\big]^{\frac{1}{p}}\\
&\leqslant C_3\varepsilon^\lambda \mathbb{E}\big[k_4(X^\varepsilon_{t_{n-1}}(x_0))\big] + C_2\varepsilon^\lambda \mathbb{E}\big[|X^\varepsilon_{t_{n-1}}(x_0) - X_{t_{n-1}}(x_0)|^p\big]^{\frac{1}{p}}\\
&\leqslant C_4\varepsilon^\lambda\Big(1 + \mathbb{E}\big[\bar\eta(X^\varepsilon_{t_{n-1}}(x_0))\big]\Big) + C_2\varepsilon^\lambda \mathbb{E}\big[|X^\varepsilon_{t_{n-1}}(x_0) - X_{t_{n-1}}(x_0)|^p\big]^{\frac{1}{p}}\\
&\leqslant C_4\varepsilon^\lambda\Big(C_1 + \bar\ell\,\mathbb{E}\big[|X^\varepsilon_{t_{n-1}}(x_0) - X_{t_{n-1}}(x_0)|^p\big]^{\frac{1}{p}}\Big) + C_2\varepsilon^\lambda \mathbb{E}\big[|X^\varepsilon_{t_{n-1}}(x_0) - X_{t_{n-1}}(x_0)|^p\big]^{\frac{1}{p}}\\
&= C_5\varepsilon^\lambda \mathbb{E}\big[|X^\varepsilon_{t_{n-1}}(x_0) - X_{t_{n-1}}(x_0)|^p\big]^{\frac{1}{p}} + C_6\varepsilon^\lambda.
\end{aligned}$$

That is, for $\psi_n := \mathbb{E}\big[|X^\varepsilon_{t_n}(x_0) - X_{t_n}(x_0)|^p\big]^{\frac{1}{p}}$ we have then

$$\psi_n \leqslant C_7\varepsilon^\lambda\psi_{n-1} + C_7\varepsilon^\lambda,$$

which gives the following estimate for any $n \in \mathbb{N}$ and $k \in \{1, \dots, n\}$

$$\psi_n \leqslant (C_7\varepsilon^\lambda)^{n-k} + \sum_{i=1}^{n-k}(C_7\varepsilon^\lambda)^i.$$

For $C_7\varepsilon_0^\lambda < \frac{1}{2}$ we obtain for any $n \in \mathbb{N}$ the estimate

$$\psi_n \leqslant C_7\varepsilon^\lambda + \sum_{i=1}^{\infty}(C_7\varepsilon^\lambda) \leqslant 3C_7\varepsilon^\lambda < \infty.$$

Under these assumptions, we obtain for all $\varepsilon \in (0, \varepsilon_0]$

$$\begin{aligned}
\max_{n=0,\dots,N_\varepsilon} \mathbb{E}\big[\bar\eta(X^\varepsilon_{t_n}(x_0))\big] &\leqslant \max_{n=0,\dots,N_\varepsilon}\Big(\bar\ell\,\mathbb{E}\big[|X^\varepsilon_{t_n}(x_0) - X_{t_n}(x_0)|^p\big]^{\frac{1}{p}} + C_1\Big)\\
&\leqslant C_7\varepsilon^\lambda + C_1 < \infty. \qquad (16.47)
\end{aligned}$$

Going back to our main estimate, we obtain $C_8 > 0$ such that

$$\mathbb{E}\Big[\sup_{s\in[0,T]} A_1(s,\varepsilon)|^p\Big]^{\frac{1}{p}} \leqslant T\varepsilon^\lambda k_3k_5\Big(1 + \max_{n=0,\dots,N_\varepsilon}\mathbb{E}\big[\bar\eta(X^\varepsilon_{t_n}(x_0))\big]\Big) \leqslant C_8T\varepsilon^\lambda.$$

Lemma 16.11 *For any* $\lambda \in (0, \frac{p-1}{p})$ *there exist constants* $b_3 > 0$ *and* $\varepsilon_0 \in (0, 1]$ *such that for any* $\varepsilon \in (0, \varepsilon_0]$ *and* $T \geqslant 0$

$$\left(\mathbb{E}\left[\sup_{s\in[0,T]} |A_2(s,\varepsilon)|^p\right]\right)^{\frac{1}{p}} \leqslant b_3 T \eta^0(cT|\ln(\varepsilon)|).$$

Proof We have

$$\begin{aligned}
&\left(\mathbb{E}\left[\sup_{s\in[0,T]} |A_2(s,\varepsilon)|^p\right]\right)^{\frac{1}{p}} \\
&\leqslant \epsilon \left[\mathbb{E}\left|\sum_{n=0}^{N_\varepsilon-1}\left[\int_{t_n}^{t_{n+1}} h(X_{s-t_n}(X^\epsilon_{t_n}(x_0)))ds - \Delta_\varepsilon Q^h(\pi(X^\epsilon_{t_n}(x_0)))\right]\right|^p\right]^{\frac{1}{p}} \\
&\leqslant \epsilon\Delta_\varepsilon \sum_{n=0}^{N_\varepsilon-1}\left[\mathbb{E}\left|\frac{1}{\Delta_\varepsilon}\int_{t_n}^{t_{n+1}} h(X_{s-t_n}(X^\epsilon_{t_n}(x_0)))ds - Q(\pi(X^\epsilon_{t_n}(x_0)))\right|^p\right]^{\frac{1}{p}}.
\end{aligned}$$

We apply the Markov property for all $n = 0, \ldots, N_\varepsilon$. By Hypothesis 3 the two terms inside the modulus converge to each other when Δ_ε goes to infinity with rate of convergence bounded by $\bar{\eta}(X^\varepsilon_{t_n}(x_0))\eta^0(\Delta_\varepsilon)$. Hence, for small ϵ we have

$$\begin{aligned}
\left(\mathbb{E}\left[\sup_{s\in[0,T]} |A_2(s,\varepsilon)|^p\right]\right)^{\frac{1}{p}} &\leqslant \varepsilon N_\varepsilon\, \Delta_\varepsilon\, \eta^0(\Delta_\varepsilon) \max_{n=0,\ldots,N_\varepsilon} \mathbb{E}[\bar{\eta}(X^\epsilon_{t_n}(x_0))] \\
&\leqslant T\eta^0(cT|\ln(\varepsilon)|) \max_{n=0,\ldots,N_\varepsilon} \mathbb{E}[\bar{\eta}(X^\varepsilon_{t_n}(x_0))].
\end{aligned}$$

Therefore, using (16.47), we obtain for $\varepsilon \in (0, \varepsilon_0]$ the estimate

$$\left(\mathbb{E}\left[\sup_{s\in[0,T]} |A_2(s,\varepsilon)|^p\right]\right)^{\frac{1}{p}} \leqslant C_8 T \eta^0(cT|\ln(\varepsilon)|).$$

Lemma 16.12 *For any* $\lambda \in (0, \frac{p-1}{p})$ *there exist positive constants* $b_4 > 0$ *and* $\varepsilon_0 \in (0, 1]$ *such that for any* $\varepsilon \in (0, \varepsilon_0]$ *and* $T \geqslant 0$

$$\left(\mathbb{E}\left[\sup_{s\in[0,T]} |A_3(s,\varepsilon)|^p\right]\right)^{\frac{1}{p}} \leqslant b_4 T \varepsilon^\lambda.$$

Proof We calculate

$$
\begin{aligned}
|A_3(T,\varepsilon)| &= \Big| \sum_{n=0}^{N_\varepsilon} \varepsilon \Delta_\varepsilon Q^{\pi K}(\pi(X^\varepsilon_{t_n})) - \int_0^{N_\varepsilon \Delta_\varepsilon} \varepsilon Q^{\pi K}(\pi(X^\varepsilon_{\frac{s}{\varepsilon}}))\, ds \Big| \\
&\leqslant \varepsilon \sum_{n=0}^{N_\varepsilon} \Delta_\varepsilon \sup_{t_n \leqslant s < t_{n+1}} |Q^{\pi K}(\pi(X^\varepsilon_s)) - Q^{\pi K}(\pi(X^\varepsilon_{t_n}))| \\
&\leqslant \varepsilon \Delta_\varepsilon C_1 \sum_{n=0}^{N_\varepsilon} \sup_{t_n \leqslant s < t_{n+1}} |v^\varepsilon_s - v^\varepsilon_{t_n}|. \qquad (16.48)
\end{aligned}
$$

By Minkowski's inequality, the Markov property, Proposition 16.9 and (16.47) (with the appropriate constant C_8) we have that

$$
\begin{aligned}
\mathbb{E}[\sup_{s\in[0,T]} |A_3(s,\varepsilon)|^p]^{\frac{1}{p}} &\leqslant TC_1 \max_{n\in\{0,\dots,N_\varepsilon\}} \mathbb{E}[\mathbb{E}[\sup_{t_n \leqslant s < t_{n+1}} |v^\varepsilon_{s-t_n}(y) - v^\varepsilon_0(y)|^p \mid y = \mathscr{F}_{t_n}]]^{\frac{1}{p}} \\
&\leqslant TC_1 \max_{n\in\{0,\dots,N_\varepsilon\}} \mathbb{E}[\mathbb{E}[\sup_{t_0 \leqslant s < t_1} |v^\varepsilon_s(y) - v^\varepsilon_0(y)|^p \mid y = X^\varepsilon_{t_n}(x_0)]]^{\frac{1}{p}} \\
&\leqslant TC_2\varepsilon^\lambda \mathbb{E}\Big[k_4(X^\varepsilon_{t_n}(x_0))\Big] \\
&\leqslant TC_2C_8\varepsilon^\lambda.
\end{aligned}
$$

This ends the proof of Proposition 16.9.

Proof of the Main Theorem 16.4: With the help of Proposition 16.9, the proof of Theorem 16.4 is identical the one given in Section 5 of [8].

Acknowledgements The author PHC would like to thank the Department of Mathematics of Brasilia University for providing support. The authors MAH and PRR would like express his gratitude for the hospitality received at the Departameto de Matemática at Universidade de Brasília and the IMECC at UNICAMP in February 2018. The funding of MAH by the FAPA project "Stochastic dynamics of Lévy driven systems" at the School of Science at Universidad de los Andes is greatly acknowledged. The author PRR is partially supported by Brazilian CNPq proc. nr. 305462/2016-4, by FAPESP proc. nr. 2015/07278-0 and 2015/50122-0.

Appendix

Proposition 16.13 (Pachpatte [23]) *Let u, f, g and h be nonnegative continuous functions defined on $\mathbb{R}^+$. Let v be a continuous non-decreasing subadditive and submultiplicative function defined on $\mathbb{R}^+$ and $v(u) > 0$ on $(0,\infty)$. Let e, ϕ be continuous and nondecreasing functions defined on $\mathbb{R}^+$ with p being strictly positive*

and $\phi(0) = 0$. *If*

$$u(t) \leqslant e(t) + g(t) \int_0^t f(s)u(s)ds + \phi\Big(\int_0^t h(s)v(u(s))ds \Big)$$

for all $t \geqslant 0$, *then for any* $0 \leqslant t \leqslant t_2$

$$u(t) \leqslant a(t)\Big[e(t) + \phi\Big(F^{-1}\big(F(A(t)) + \int_0^t h(s)v(a(s))ds\big)\Big)\Big],$$

where

$$a(t) := 1 + g(t) \int_0^t f(s) \exp\Big(\int_s^t g(\sigma) f(\sigma) d\sigma \Big) ds,$$

$$A(t) := \int_0^t h(s)v(a(s)e(s))ds,$$

$$F(t) := \int_0^t \frac{ds}{v(\phi(s))},$$

F^{-1} *is the inverse of* F *and* $t_2 \in \mathbb{R}^+$ *such that*

$$F(A(t)) + \int_0^t h(s)v(a(t))ds \in \mathrm{dom}(F^{-1}) \qquad \textit{for all } 0 \leqslant t \leqslant t_2.$$

In the following special case of coefficients it is possible to drop the continuity assumption on u.

Corollary 16.14 *Let* Ψ *a non-negative, measurable, increasing function and* h *be nonnegative, continuous, increasing function on the interval* $[0, T]$ *satisfying for* $p \geqslant 2$, $\varepsilon > 0$, $c > 0$ *and any* $t \in [0, T]$ *the inequality*

$$\Psi(t) \leqslant \varepsilon c t^p + \varepsilon c \Big(\int_0^t \Psi(s) + \Psi(s)^{\frac{p-1}{p}} ds \Big), \qquad t \in [0, T]. \tag{16.49}$$

Then there is a constant $k > 0$ *such that for any* $\varepsilon_0 \in (0, 1]$ *such that* $\varepsilon_0 T < k$ *we have for all* $t \in [0, T]$ *and* $\varepsilon \in (0, \varepsilon_0]$

$$\Psi(t) \leqslant C\Big(\varepsilon t^p + t^p (\varepsilon t)^{\frac{p-1}{p}}\Big).$$

Proof For $e(t) = c\varepsilon t^p$, $g \equiv 1$, $f, h \equiv \varepsilon c$, $\phi(t) = t$, $w(t) = t^{\frac{p-1}{p}}$ we calculate the coefficients of Proposition 16.13

$$a(t) := 1 + \varepsilon c \int_0^t \exp(\varepsilon c(t - s))ds = \exp(\varepsilon c t)$$

and in the limit of εt being small ($\varepsilon t \ll 1$) we have

$$\varepsilon \int_0^t a(s)^{\frac{p-1}{p}} ds = \varepsilon t \Big(\frac{\exp(c\frac{p-1}{p}\varepsilon t) - 1}{c\frac{p-1}{p}\varepsilon t} \Big) \leqslant_{\varepsilon t \ll 1} 2\varepsilon t.$$

Applying the change of parameter $r = \varepsilon s$ it follows that

$$\begin{aligned} A(t) &:= \int_0^t \exp(\varepsilon c \frac{p-1}{p} s)(e(s))^{\frac{p-1}{p}} ds = \int_0^t \exp(c\frac{p-1}{p}\varepsilon s)(c\varepsilon s^p)^{\frac{p-1}{p}} ds \\ &= \varepsilon^{\frac{p-1}{p}} \int_0^{\varepsilon t} \exp(c\frac{p-1}{p} r) c^{\frac{p-1}{p}} \Big(\frac{r}{\varepsilon}\Big)^{p-1} \frac{dr}{\varepsilon} \leqslant t \frac{1}{\varepsilon^p t} \int_0^{\varepsilon t} \exp(c\frac{p-1}{p} r)(cr)^{\frac{p-1}{p}} dr \\ &\leqslant_{\varepsilon t \ll 1} 2t \exp(c\frac{p-1}{p}\varepsilon t)(c\varepsilon t)^{\frac{p-1}{p}} \leqslant C_1 t \exp(c\frac{p-1}{p}\varepsilon t)(\varepsilon t)^{\frac{p-1}{p}}. \end{aligned}$$

Finally, we obtain

$$F(t) := \int_0^t s^{-\frac{p-1}{p}} ds = p r^{\frac{1}{p}} \qquad \text{and} \qquad F^{-1}(t) := \frac{t^p}{p^p}.$$

In the sequel we follow the proof of Theorem 2.4.2 in Pachpatte [23] and define the continuous, positive, non-decreasing function

$$n(t) := e(t) + \phi\Big(\int_0^t h(s) w(u(s)) ds \Big) = e(t) + \varepsilon c \int_0^t h(s) u(s)^{\frac{p-1}{p}} ds, \qquad t \geqslant 0,$$

such that inequality (16.49) can be restated as

$$u(t) \leqslant n(t) + g(t) \int_0^t f(s) u(s) ds = e(t) + \varepsilon c \int_0^t u(s) ds.$$

It is well-known, see for instance [1], that this integral estimate implies the following Gronwall-Bellmann inequality also in the case of u being merely positive measurable. The main reason is that the integral is absolutely continuous with a bounded density. This result yields

$$u(t) \leqslant a(t) n(t), \qquad t \geqslant 0.$$

The remainder of the proof of Theorem 2.4.2 in [23] does use the continuity of u and remains intact.

References

1. Amann, H.: Ordinary Differential Equations: An Introduction to Nonlinear Analysis. de Gruyter Studies in Mathematics, vol. 13. Walter de Gruyter & Co., Berlin (1990)
2. Applebaum, D.: Lévy Processes and Stochastic Calculus, 2nd edn. Cambridge University Press, Cambridge (2009)
3. Arnold, V.: Mathematical Methods in Classical Mechanics, 2nd edn. Springer, Berlin (1989)
4. Bakhtin, V., Kifer, Y.: Nonconvergence examples in averaging. Geometric and probabilistic structures in dynamics. Contemp. Math. **469**, 1–17 (2008)
5. Borodin, A., Freidlin, M.: Fast oscillating random perturbations of dynamical systems with conservation laws. Ann. Inst. H. Poincaré. Prob. Statist. **31**, 485–525 (1995)
6. Cannas, A.: Lectures on Symplectic Geometry. Lecture Notes in Mathematics, vol. 1764. Springer, Berlin (2008)
7. Cerrai, S.: A Khasminskii type averaging principle for stochastic reaction-diffusion equations. Ann. Probab. **19**(3), 899–948 (2009)
8. da Costa, P.H., Högele, M.A.: Strong averaging along foliated Lévy diffusions with heavy tails on compact leaves. Potential Anal. **47**(3), 277–311 (2017)
9. Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems. Springer, Berlin (1991)
10. Gargate, I.I.G., Ruffino, P.R.: An averaging principle for diffusions in foliated spaces. Ann. Probab. **44**(1), 567–588 (2016)
11. Garnett, L.: Foliation, the ergodic theorem and Brownian motion. J. Funct. Anal. **51**, 285–311 (1983)
12. Högele, M.A., Ruffino, P.R.: Averaging along foliated Lévy diffusions. Nonlinear Anal. **112**, 1–14 (2015)
13. Kabanov, Y., Pergamenshchikov, S.: Two-Scale Stochastic Systems: Asymptotic Analysis and Control. Springer, Berlin (2003)
14. Kakutani, S., Petersen, K.: The speed of convergence in the ergodic theorem. Monat. Mathematik **91**, 11–18 (1981)
15. Khasminski, R., Krylov, N.: On averaging principle for diffusion processes with null-recurrent fast component. Stoch Proc. Appl. **93**, 229–240 (2001)
16. Krengel, U.: On the speed of convergence of the ergodic theorem. Monat. Mathematik **86**, 3–6 (1978)
17. Kulik, A.: Exponential ergodicity of the solutions of SDEs with a jump noise. Stoch. Proc. Appl. **119**, 602–632 (2009)
18. Kunita, H.: Stochastic differential equations based on Lévy processes and stochastic flows of diffeomorphisms. In: Rao, M.M. (ed.) Real and Stochastic Analysis, pp. 305–373. Birkhäuser, Basel (2004)
19. Kurtz, T.G., Pardoux, E., Protter, Ph.: Stratonovich stochastic differential equations driven by general semimartingales. Ann. lH.I.P. Sect. B **31**(2), 351–377 (1995)
20. Li, X.-M.: An averaging principle for a completely integrable stochastic Hamiltonian systems. Nonlinearity **21**, 803–822 (2008)
21. Namachchvaya, S., Sowers, R.: Rigorous stochastic averaging at a center with additive noise. Meccanica **37**, 85–114 (2002)
22. Nash, J.: The imbedding problem for Riemannian manifolds. Ann. Math. **63**(1), 20–63 (1956)
23. Pachpatte, B.G.: Inequalities for Differential and Integral Equations. Academic, San Diego (1998)
24. Protter, Ph.: Stochastic Integration and Differential Equations. Springer, Berlin (2004)
25. Sanders, J.A., Verhulst, F., Murdock, J.: Averaging Methods in Nonlinear Dynamical Systems, 2nd edn. Springer, Berlin (2007)
26. Sato, K.-I.: Lévy processes and infinitely divisible distributions. Probab. Theory Relat. Fields **111**, 287321 (1998)

27. Sowers, R.: Stochastic averaging with a flattened Hamiltonian: a Markov process on a stratified space (a whiskered sphere). Trans. Am. Math. Soc. **354**, 853–900 (2002)
28. Tondeur, P.: Foliations on Riemannian Manifolds. Universitext. Springer, Berlin (1988)
29. Volsov, V.M.: Some types of calculation connected with averaging in the theory of non-linear vibrations. USSR Comput. Math. Math. Phys. **3**(1), 1–64 (1962)
30. Volsov, V.M., Morgunov, B.I.: Methods of calculating stationary resonance vibrational and rotational motions of certain non-linear systems. USSR Comput. Math. Math. Phys. **8**(2), 1–62 (1968)
31. Walcak, P.: Dynamics of Foliations, Groups and Pseudogroups. Birkhäuser, Basel (2004)
32. Xu, Y., Duan, J., Xu, W.: An averaging principle for stochastic dynamical systems with Lvy noise. Physica D **240**, 1395–1401 (2011)

Chapter 17
Young Differential Delay Equations Driven by Hölder Continuous Paths

Luu Hoang Duc and Phan Thanh Hong

Abstract In this paper we prove the existence and uniqueness of the solution of Young differential delay equations under weaker conditions than it is known in the literature. We also prove the continuity and differentiability of the solution with respect to the initial function and give an estimate for the growth of the solution. The proofs use techniques of stopping times, Shauder-Tychonoff fixed point theorem and a Gronwall-type lemma.

17.1 Introduction

In this paper we would like to study the deterministic delay equation of the differential form

$$\begin{aligned} dx(t) &= f(x_t)dt + g(x_t)d\omega(t), \qquad t \in [0,T] \\ x_0 &= \eta \in C_r := C([-r,0],\mathbb{R}^d) \end{aligned} \tag{17.1}$$

or in the integral form

$$\begin{aligned} x(t) &= x(0) + \int_0^t f(x_s)ds + \int_0^t g(x_s)d\omega(s), \qquad t \in [0,T] \\ x_0 &= \eta \in C_r \end{aligned} \tag{17.2}$$

L. H. Duc (✉)
Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

Institute of Mathematics, Vietnam Academy of Science and Technology, Hanoi, Vietnam
e-mail: duc.luu@mis.mpg.de; lhduc@math.ac.vn

P. T. Hong
Thang Long University, Hanoi, Vietnam
e-mail: hongpt@thanglong.edu.vn

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_17

for some fixed time interval $[0, T]$, where $C([a, b], \mathbb{R}^d)$ denote the space of all continuous paths $x : [a, b] \to \mathbb{R}^d$ equipped with sup norm $\|x\|_{\infty,[a,b]} = \sup_{t\in[a,b]} \|x(t)\|$, with $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$, $x_t \in C_r$ is defined by $x_t(u) := x(t + u)$ for all $u \in [-r, 0]$; $f, g : C_r \to \mathbb{R}^d$ are coefficient functions; and ω belongs to $C^{\nu-\text{Hol}}([0, T], \mathbb{R})$ - the space of Hölder continuous paths for index $\nu > \frac{1}{2}$. Such system appears, for example, while solving stochastic differential equations of the form

$$dx(t) = f(x_t)dt + g(x_t)dB^H(t), \qquad x_0 = \eta \in C_r, \tag{17.3}$$

where B^H is a fractional Brownian motion defined on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with the Hurst index $H \in (1/2, 1)$ [15]. Since B^H is neither Markov nor semimartingale if $H \neq \frac{1}{2}$, one cannot apply the classical Ito theory to solve (17.3). Instead, due to the fact that $B^H(\cdot)$ is Hölder continuous for almost surely all the realizations, one can define the stochastic integral w.r.t. the fBm as the integral driven by a Hölder continuous path using the so called *rough path theory* [8, 12–14], or *fractional calculus theory* [18, 21]. As a result, solving (17.3) leads to the deterministic equation (17.1) or (17.2), where the second integral in (17.2) is understood in the Young sense (see [11, 20]).

The theory of stochastic differential equations driven by the fBm B^H for $H > \frac{1}{2}$ has been well developed by many authors, especially results on existence and uniqueness of the pathwise solution, the generation of random dynamical systems (see e.g. [4, 5, 9–11, 16, 17, 20],... and the references therein). For studies on delay equations, we refer to [1–3, 6].

In the general case where f, g are functions of (t, x_t), under some regularity conditions, i.e. f is globally Lipschitz continuous and of linear growth, g is C^1 such that its Frechet derivative is bounded and globally Lipschitz continuous, there exists a unique solution $x(\cdot, \omega, \eta)$ of (17.1) (see [1] or [19]). These results are based on the tools of fractional calculus developed in [17, 21, 22].

In this paper, we reprove the existence and uniqueness theorem of (17.1) under the following assumptions.

($\mathbf{H}_f$) The function f is globally Lipschitz continuous and thus has linear growth, i.e there exist constants L_f such that for all $\xi, \eta \in C_r$

$$\|f(\xi) - f(\eta)\| \leq L_f \|\xi - \eta\|_{\infty,[-r,0]}$$

($\mathbf{H}_g$) The function g is C^1 such that its Frechet derivative is bounded and locally $\delta-$Hölder continuous with $1 \geq \delta > \frac{1-\nu}{\nu}$, i.e there exists L_g such that for all $\xi, \eta \in C_r$

$$\|Dg(\xi)\|_{L(C_r,\mathbb{R}^d)} \leq L_g$$

and for each $M > 0$, there exists L_M such that for all $\xi, \eta \in C_r$ that satisfy

$$\|\xi\|_{\infty,[-r,0]}, \|\eta\|_{\infty,[-r,0]} \leq M$$

one has

$$\|Dg(\xi) - Dg(\eta)\|_{L(C_r,\mathbb{R}^d)} \leq L_M \|\xi - \eta\|^{\delta}_{\infty,[-r,0]} \tag{17.4}$$

for some constant $1 > \delta > \frac{1-\nu}{\nu}$. Assumption (17.4) is weaker than the global Lipschitz continuity of Dg, as seen in [1, 6] or [19].

Furthermore, we show that the solution is differentiable with respect to the initial function η and give an estimate for the growth of the solution. Note that in order to define the second integral in (17.2) in the Young sense, one needs to consider the solution x and the initial function η in Hölder function spaces $C^{\beta\text{-Hol}}$ with $\beta+\nu > 1$.

To finish the introduction, we recall some facts about Young integral, more details can be seen in [8]. For $p \geq 1$ and $[a, b] \subset \mathbb{R}$, a continuous path $x : [a, b] \to \mathbb{R}^d$ is of finite p-variation if

$$|\!|\!| x |\!|\!|_{p\text{-var},[a,b]} := \left(\sup_{\Pi(a,b)} \sum_{i=1}^{n} \|x(t_{i+1}) - x(t_i)\|^p \right)^{1/p} < \infty, \tag{17.5}$$

where the supremum is taken over the whole class of finite partition of $[a, b]$. The subspace $C^{p-\text{var}}([a, b], \mathbb{R}^d) \subset C([a, b], \mathbb{R}^d)$ of all paths x with finite p-variation and equipped with the p-var norm

$$\|x\|_{p\text{-var},[a,b]} := \|x(a)\| + |\!|\!| x |\!|\!|_{p\text{-var},[a,b]},$$

is a nonseparable Banach space [8, Theorem 5.25, p. 92].

Also, for $0 < \beta \leq 1$ denote by $C^{\beta-\text{Hol}}([a, b], \mathbb{R}^d)$ the Banach space of all Hölder continuous paths $x : [a, b] \to \mathbb{R}^d$ with exponent β, equipped with the norm

$$\begin{aligned} \|x\|_{\infty,\beta,[a,b]} &:= \|x\|_{\infty,[a,b]} + |\!|\!| x |\!|\!|_{\beta,[a,b]} \text{ where} \\ |\!|\!| x |\!|\!|_{\beta,[a,b]} &:= \sup_{a\leq s<t\leq b} \frac{\|x(t) - x(s)\|}{(t-s)^\beta} < \infty. \end{aligned} \tag{17.6}$$

Note that the space is not separable. However, the closure of $C^\infty([a, b], \mathbb{R}^d)$ in the β-Holder norm denoted by $C^{0,\beta-\text{Hol}}([a, b], \mathbb{R}^d)$ is a separable space (see [8, Theorem 5.31, p. 96]), which can be defined as

$$C^{0,\beta-\text{Hol}}([a, b], \mathbb{R}^d) := \left\{ x \in C^{\beta\text{-Hol}}([a, b], \mathbb{R}^d) \,\middle|\, \lim_{h\to 0} \sup_{a\leq s<t\leq b, |t-s|\leq h} \frac{\|x(t) - x(s)\|}{(t-s)^\beta} = 0 \right\}.$$

Clearly, if $x \in C^{\beta-\text{Hol}}([a, b], \mathbb{R}^d)$ then for all $s, t \in [a, b]$ we have

$$\|x(t) - x(s)\| \leq |\!|\!| x |\!|\!|_{\beta,[a,b]} |t - s|^\beta.$$

Hence, for all p such that $p\beta \geq 1$ we have

$$\|\!|x|\!\|_{p\text{-var},[a,b]} \leq \|\!|x|\!\|_{\beta,[a,b]} (b-a)^{\beta} < \infty. \tag{17.7}$$

In particular, $C^{1/p\text{-Hol}}([a,b],\mathbb{R}^d) \subset C^{p-\text{var}}([a,b],\mathbb{R}^d)$.

For $a, b, c \in \mathbb{R}$ such that $a < b < c$ and $x \in C^{\beta\text{-Hol}}([a,c],\mathbb{R}^d)$, it is easy to see that

$$\|\!|x|\!\|_{\beta,[a,c]} \leq \|\!|x|\!\|_{\beta,[a,b]} + \|\!|x|\!\|_{\beta,[b,c]}.$$

Now consider $x \in C^{\beta\text{-Hol}}([a,b],\mathbb{R}^d)$ and $\omega \in C^{\nu\text{-Hol}}([a,b],\mathbb{R})$ with $\beta + \nu > 1$. Then by (17.7), $x \in C^{\frac{1}{\beta}\text{-var}}([a,b],\mathbb{R}^d)$ and $\omega \in C^{\frac{1}{\nu}\text{-var}}([a,b],\mathbb{R})$, thus it is well known that the Young integral $\int_a^b x(t)d\omega(t)$ exists (see [8, pp. 264–265]). Moreover, for all $s \leq t$ in $[a,b]$, due to the Young-Loeve estimate [8, Theorem 6.8, p. 116]

$$\begin{aligned}\left\| \int_s^t x(u)d\omega(u) - x(s)[\omega(t)-\omega(s)] \right\| &\leq K \|\!|\omega|\!\|_{\frac{1}{\nu}\text{-var},[s,t]} \|\!|x|\!\|_{\frac{1}{\beta}\text{-var},[s,t]} \\ &\leq K(t-s)^{\beta+\nu} \|\!|\omega|\!\|_{\nu,[s,t]} \|\!|x|\!\|_{\beta,[s,t]},\end{aligned}$$

where $K := \frac{1}{1-2^{1-(\beta+\nu)}}$. Hence

$$\left\| \int_s^t x(u)d\omega(u) \right\| \leq (t-s)^{\nu} \|\!|\omega|\!\|_{\nu,[s,t]} \left(\|x(s)\| + K(t-s)^{\beta} \|\!|x|\!\|_{\beta,[s,t]} \right). \tag{17.8}$$

17.2 Existence, Uniqueness and Continuity of the Solution

Since $\delta\nu + \nu > 1$, there exists $\beta < \nu$ such that

$$\beta + \nu > \beta\delta + \nu > 1.$$

By choosing a smaller $\nu' \in (\frac{1}{2}, \nu)$ if necessary, we can always assume without loss of generality that $\omega \in C^{0,\nu-\text{Hol}}([0,T],\mathbb{R})$. System (17.1) would then be considered for $\eta \in C^{\beta-\text{Hol}}([-r,0],\mathbb{R}^d)$, i.e. we consider the equation

$$\begin{aligned} dx(t) &= f(x_t)dt + g(x_t)d\omega(t), \qquad t \in [0,T] \\ x_0 &= \eta \in C^{\beta-\text{Hol}}([-r,0],\mathbb{R}^d). \end{aligned} \tag{17.9}$$

Lemma 17.1 *If* $x \in C^{\beta-\mathrm{Hol}}([a-r,b],\mathbb{R}^d)$ *then the function* $x_\cdot : [a,b] \to C_r$, $x_t(\cdot) = x(t+\cdot)$ *belongs to* $C^{\beta-\mathrm{Hol}}([a,b],C_r)$ *and satisfies*

$$i,\ |\!|\!| x_\cdot |\!|\!|_{\beta,[a,b]} \le |\!|\!| x |\!|\!|_{\beta,[a-r,b]} \tag{17.10}$$

$$ii,\ \|x_\cdot\|_{\infty,\beta,[a,b]} \le \|x\|_{\infty,\beta,[a-r,b]}. \tag{17.11}$$

Proof The fact that

$$\begin{aligned} |\!|\!| x_\cdot |\!|\!|_{\beta,[a,b]} &= \sup_{a\le s<t\le b} \frac{\|x_t - x_s\|_{\infty,[-r,0]}}{(t-s)^\beta} \\ &= \sup_{a\le s<t\le b} \sup_{-r\le u\le 0} \frac{\|x(t+u)-x(s+u)\|}{[(t+u)-(s+u)]^\beta} \\ &\le \sup_{a-r\le s'<t'\le b} \frac{\|x(t')-x(s')\|}{(t'-s')^\beta} = |\!|\!| x |\!|\!|_{\beta,[a-r,b]} \end{aligned}$$

proves (17.10). As a result,

$$\begin{aligned} \|x_\cdot\|_{\infty,\beta,[a,b]} &= \sup_{t\in[a,b]} \|x_t\|_{\infty,[-r,0]} + |\!|\!| x_\cdot |\!|\!|_{\beta,[a,b]} \\ &\le \|x(\cdot)\|_{\infty,[a-r,b]} + |\!|\!| x |\!|\!|_{\beta,[a-r,b]}, \end{aligned}$$

which proves (17.11).

Remark 17.1 The lemma is not true if we replace the Hölder continuous space by p-variation bounded space. Namely, if a function x belongs to $C^{p-\mathrm{var}}([a-r,b],\mathbb{R}^d)$, it does not follow that its translation function $x_\cdot$ belongs to $C^{p-\mathrm{var}}([a,b],C_r)$ with $p \ge 1$. As a counter example, consider the function $x(t) = |t|^\beta, t \in [-1,1], \beta p < 1$ then $x \in C^{p-\mathrm{var}}([-1,1],\mathbb{R})$. However, with the partition $\Pi = 0 < \frac{1}{n} < \frac{2}{n} < \cdots < \frac{n-1}{n} < 1$ we have

$$\begin{aligned} \left(\sum_i \|x_{\frac{i+1}{n}} - x_{\frac{i}{n}}\|^p_{\infty,[-1,0]}\right)^{1/p} &= \left(\sum_i \sup_{-1\le u\le 0} \left|x(\frac{i+1}{n}+u) - x(\frac{i}{n}+u)\right|^p\right)^{1/p} \\ &\ge \left(\sum_i \left|x(\frac{i+1}{n} - \frac{i}{n}) - x(\frac{i}{n} - \frac{i}{n})\right|^p\right)^{1/p} \\ &\ge \left(\sum_i \frac{1}{n^{\beta p}}\right)^{1/p} \\ &= n^{\frac{1-\beta p}{p}} \to \infty, \quad \text{as } n \to \infty. \end{aligned}$$

This shows that $x_\cdot$ is not of bounded p-variation.

Lemma 17.2 *Assume that g satisfies the condition $(\boldsymbol{H}_g)$. If $x \in C^{\beta-\mathrm{Hol}}([a-r,b],\mathbb{R}^d)$ then $g(x_\cdot) \in C^{\beta-\mathrm{Hol}}([a,b],\mathbb{R}^d)$ and*

$$|||g(x_\cdot)|||_{\beta,[a,b]} \leq L_g\, |||x|||_{\beta,[a-r,b]}. \tag{17.12}$$

Proof The proof is directed from the Lipschitz continuity of g and Lemma 17.1. Namely,

$$\begin{aligned}\sup_{a\leq s<t\leq b}\frac{\|g(x_t)-g(x_s)\|}{(t-s)^\beta} &\leq \sup_{a\leq s<t\leq b} L_g\frac{\|x_t-x_s\|_{\infty,[-r,0]}}{(t-s)^\beta}\\ &\leq L_g\, |||x|||_{\beta,[a-r,b]}.\end{aligned}$$

Remark 17.2 Since $\beta+\nu>1$ the integral $\int_a^b g(x_t)d\omega(t)$ is well defined.

Lemma 17.3 *Assume that g satisfies the condition $(\boldsymbol{H}_g)$. If $x, y \in C^{\beta-\mathrm{Hol}}([a-r,b],\mathbb{R}^d)$ are such that $\|x\|_{\infty,\beta,[a-r,b]}, \|y\|_{\infty,\beta,[a-r,b]} \leq M$, then*

$$\begin{aligned}|||g(x_\cdot)-g(y_\cdot)|||_{\delta\beta,[a,b]} &\leq L_g(b-a)^{\beta-\delta\beta}\,|||x-y|||_{\beta,[a-r,b]} + L_M M^\delta \|x-y\|_{\infty,[a-r,b]}\\ &\leq \left(L_g(b-a)^{\beta-\delta\beta}+L_M M^\delta\right)\|x-y\|_{\infty,\beta,[a-r,b]}\end{aligned} \tag{17.13}$$

Proof By the mean value theorem

$$\begin{aligned}&|g(x_t)-g(y_t)-g(x_s)+g(y_s)|\\ &= \left|\int_0^1 Dg(\theta x_t+(1-\theta y_t))(x_t-y_t)d\theta + \int_0^1 Dg(\theta x_s+(1-\theta)y_s)(x_s-y_s)d\theta\right|\\ &\leq \left|\int_0^1 Dg(\theta x_t+(1-\theta)y_t)[(x_t-y_t)-(x_s-y_s)]d\theta\right|\\ &\quad+\left|\int_0^1 [Dg(\theta x_t+(1-\theta)y_t)-Dg(\theta x_s+(1-\theta)y_s)](x_s-y_s)d\theta\right|\\ &\leq L_g\|(x_t-y_t)-(x_s-y_s)\|_{\infty,[-r,0]}\\ &\quad+L_M\|x_s-y_s\|_{\infty,[-r,0]}\int_0^1\left(\theta\|x_t-x_s\|^\delta_{\infty,[-r,0]}+(1-\theta)\|y_t-y_s\|^\delta_{\infty,[-r,0]}\right)d\theta\\ &\leq L_g(t-s)^\beta\,|||x_\cdot-y_\cdot|||_{\beta,[a,b]}\\ &\quad+L_M\|x_s-y_s\|_{\infty,[-r,0]}(t-s)^{\delta\beta}\max\left\{|||x_\cdot|||^\delta_{\beta,[a,b]},|||y_\cdot|||^\delta_{\beta,[a,b]}\right\}\\ &\leq L_g(t-s)^\beta\,|||x-y|||_{\beta,[a-r,b]}+L_M(t-s)^{\delta\beta}M^\delta\|x-y\|_{\infty,[a-r,b]}.\end{aligned}$$

This implies

$$|||g(x_\cdot) - g(y_\cdot)|||_{\delta\beta,[a,b]} \le L_g (b-a)^{\beta-\delta\beta} |||x-y|||_{\beta,[a-r,b]} + L_M M^{\delta} \|x-y\|_{\infty,[a-r,b]}.$$

Consider $x \in C^{\beta-\mathrm{Hol}}([t_0 - r, t_1], \mathbb{R}^d)$ with any interval $[t_0, t_1] \subset [0, T]$. Put

$$I(x)(t) := \int_{t_0}^{t} f(x_s) ds \quad \text{and} \quad J(x)(t) := \int_{t_0}^{t} g(x_s) d\omega(s), \quad t \in [t_0, t_1]$$

and define the map

$$F(x)(t) = \begin{cases} x(t_0) + I(x)(t) + J(x)(t) & \text{if } t \in [t_0, t_1] \\ x(t) & \text{if } t \in [t_0 - r, t_0] \end{cases}$$

Lemma 17.4 *If $x, y \in C^{\beta-\mathrm{Hol}}([a-r, b], \mathbb{R}^d)$ are such that $\|x\|_{\infty,\beta,[a-r,b]}$, $\|y\|_{\infty,\beta,[a-r,b]} \le M$, then there exists $L = L(b-a, M)$ satisfying*

$$|||F(x) - F(y)|||_{\beta,[a,b]} \le L \left((b-a)^{1-\beta} + (b-a)^{\nu-\beta} |||\omega|||_{\nu,[a,b]} \right) \|x-y\|_{\infty,\beta,[a-r,b]}. \tag{17.14}$$

Proof First, observe that

$$\begin{aligned} |||I(x) - I(y)|||_{\beta,[a,b]} &= \sup_{a \le s < t \le b} \frac{|I(x)(t) - I(y)(t) - I(x)(s) + I(y)(s)|}{(t-s)^{\beta}} \\ &\le \sup_{a \le s < t \le b} \frac{\int_s^t |f(x_u) - f(y_u)| du}{(t-s)^{\beta}} \\ &\le \sup_{a \le s < t \le b} \frac{L_f (t-s) \|x-y\|_{\infty,[a-r,b]}}{(t-s)^{\beta}} \\ &\le L_f (b-a)^{1-\beta} \|x-y\|_{\infty,[a-r,b]}. \end{aligned} \tag{17.15}$$

Secondly, since $\nu + \delta\beta > 1$, by assigning $K' = \frac{1}{1-2^{1-(\nu+\delta\beta)}}$ and applying Lemma 17.3 one has

$$\begin{aligned} &\sup_{a \le s < t \le b} \frac{|J(x)(t) - J(y)(t) - J(x)(s) + J(y)(s)|}{(t-s)^{\beta}} \\ \le &\sup_{a \le s < t \le b} \frac{|\int_s^t [g(x_u) - g(y_u)] d\omega(u)|}{(t-s)^{\beta}} \\ \le &\sup_{a \le s < t \le b} \frac{(t-s)^{\nu} |||\omega|||_{\nu,[s,t]} \left[\|g(x_s) - g(y_s)\| + K'(t-s)^{\delta\beta} |||g(x_\cdot) - g(y_\cdot)|||_{\delta\beta,[s,t]} \right]}{(t-s)^{\beta}} \end{aligned}$$

$$\leq (b-a)^{v-\beta} \|\omega\|_{v,[a,b]} \left[L_g \|x-y\|_{\infty,[a-r,b]} + L_g K'(b-a)^{\beta} \|x-y\|_{\beta,[a-r,b]} \right.$$
$$\left. + K' L_M M^{\delta} (b-a)^{\delta\beta} \|x-y\|_{\infty,[a-r,b]} \right]. \tag{17.16}$$

Now (17.14) is followed from (17.15) and (17.16) by choosing

$$L = L(b-a, M) := L_f + L_g + L_g K'(b-a)^{\beta} + K' L_M M^{\delta} (b-a)^{\delta\beta}. \tag{17.17}$$

We can now state the theorem on existence and uniqueness of solution of system (17.1).

Theorem 17.1 *Assume that* $(\boldsymbol{H}_f)$ *and* $(\boldsymbol{H}_g)$ *are satisfied. If* $\eta \in C^{\beta-\mathrm{Hol}}([-r,0], \mathbb{R}^d)$ *then there exists a unique solution to the Eq.* (17.9) *in* $C^{\beta-\mathrm{Hol}}([-r,T], \mathbb{R}^d)$. *Moreover, the solution is* v*-Hölder continuous on* $[0,T]$.

Proof The proof is divided into several steps.

Step 1: For any $a < b$ in $[0,T]$, one first proves that F is a mapping from $C^{\beta\text{-Hol}}([a-r,b], \mathbb{R}^d)$ into itself, or sufficiently

$$\|F(x)\|_{\beta,[a,b]} \leq \|I(x)\|_{\beta,[a,b]} + \|J(x)\|_{\beta,[a,b]} < \infty.$$

With $a \leq s < t \leq b$, using assumption $(\mathbf{H}_f)$ and assigning $L' := \max\{L_f, \|f(0)\|\}$ one has

$$\begin{aligned} \|I(x)(t) - I(x)(s)\| &= \left\| \int_s^t f(x_u) du \right\| \\ &\leq L'(t-s)(1 + \|x_\cdot\|_{\infty,[s,t]}) \\ &\leq L'(t-s)(1 + \|x\|_{\infty,[a-r,b]}) \end{aligned}$$

hence

$$\|I(x)\|_{\beta,[a,b]} \leq L'(b-a)^{1-\beta}(1 + \|x\|_{\infty,[a-r,b]}) \leq L'(b-a)^{1-\beta}(1 + \|x\|_{\infty,\beta,[a-r,b]}) < \infty.$$

On the other hand, using Lemma 17.2 with $K = \frac{1}{1-2^{1-(v+\beta)}}$, one has

$$\begin{aligned} &\|J(x)(t) - J(x)(s)\| \\ &= \left| \int_s^t g(x_u) d\omega(u) \right| \\ &\leq \|\omega\|_{v,[a,b]} (t-s)^{v} \left(\|g(x_s)\| + K(t-s)^{\beta} \|g(x_\cdot)\|_{\beta,[a,b]} \right) \\ &\leq \|\omega\|_{v,[a,b]} (t-s)^{v} \left(\|g(0)\| + L_g \|x\|_{\infty,[a-r,b]} + L_g K (t-s)^{\beta} \|x\|_{\beta,[a-r,b]} \right), \end{aligned}$$

which implies

$$\begin{aligned}
&\|\!| J(x) |\!\|_{\beta,[a,b]} \\
&\le (b-a)^{\nu-\beta} \|\!| \omega |\!\|_{\nu,[a,b]} \left(\|g(0)\| + L_g \|x\|_{\infty,[a-r,b]} + L_g K (b-a)^{\beta} \|\!| x |\!\|_{\beta,[a-r,b]} \right) \\
&\le (b-a)^{\nu-\beta} \|\!| \omega |\!\|_{\nu,[a,b]} \left[\|g(0)\| + L_g + L_g K (b-a)^{\beta} \right] (1 + \|x\|_{\infty,\beta,[a-r,b]}) < \infty.
\end{aligned}$$

Therefore $\|\!| F(x) |\!\|_{\beta,[a,b]}$ is finite. Moreover, by assigning $a := t_0, b := t_1$ it follows from the definition of F that

$$\begin{aligned}
&\|F(x)\|_{\infty,\beta,[t_0-r,t_1]} \\
&= \|F(x)\|_{\infty,[t_0-r,t_1]} + \|\!| F(x) |\!\|_{\beta,[t_0-r,t_1]} \\
&\le \max \left\{ \|F(x)\|_{\infty,[t_0-r,t_0]}, \|F(x)\|_{\infty,[t_0,t_1]} \right\} + \|\!| F(x) |\!\|_{\beta,[t_0-r,t_0]} + \|\!| F(x) |\!\|_{\beta,[t_0,t_1]} \\
&\le \max \left\{ \|F(x)\|_{\infty,[t_0-r,t_0]}, \|F(x)(t_0)\| + (t_1-t_0)^{\beta} \|\!| F(x) |\!\|_{\beta,[t_0,t_1]} \right\} \\
&\quad + \|\!| F(x) |\!\|_{\beta,[t_0-r,t_0]} + \|\!| F(x) |\!\|_{\beta,[t_0,t_1]} \\
&\le \|x\|_{\infty,[t_0-r,t_0]} + \|\!| x |\!\|_{\beta,[t_0-r,t_0]} + [1 + (t_1-t_0)^{\beta}] \|\!| F(x) |\!\|_{\beta,[t_0,t_1]} \\
&\le \|x\|_{\infty,\beta,[t_0-r,t_0]} + C' \left[(t_1-t_0)^{1-\beta} + (t_1-t_0)^{\nu-\beta} \|\!| \omega |\!\|_{\nu,[t_0,t_1]} \right] (1 + \|x\|_{\infty,\beta,[t_0-r,t_1]}),
\end{aligned} \tag{17.18}$$

where

$$C' = C'(t_1 - t_0) := [1 + (t_1 - t_0)^{\beta}](\|g(0)\| + L_g + L_g K (t_1 - t_0)^{\beta} + L'). \tag{17.19}$$

Furthermore, for $0 < \epsilon \le \nu - \beta$ small enough,

$$\begin{aligned}
&\|\!| F(x) |\!\|_{(\beta+\epsilon),[t_0,t_1]} \\
&\le C' \left((t_1 - t_0)^{1-\beta-\epsilon} + (t_1 - t_0)^{\nu-\beta-\epsilon} \|\!| \omega |\!\|_{\nu,[t_0,t_1]} \right) (1 + \|x\|_{\infty,\beta,[t_0-r,t_1]}).
\end{aligned} \tag{17.20}$$

Step 2: Following [5] and [7] , assign

$$C := 2(\|g(0)\| + L' + L_g(K+1)) \tag{17.21}$$

and fix $\mu < \min\{1, C\}$. We construct a sequence t_i in $[0, \infty)$ such that $t_0 = 0$ and

$$t_{i+1} = \sup\{t \ge t_i : C \left[(t - t_i)^{1-\beta} + (t - t_i)^{\nu-\beta} \|\!| \omega |\!\|_{\nu,[t_i,t]} \right] \le \mu\}.$$

Since $\omega \in C^{0,\nu-\mathrm{Hol}}([0,T],\mathbb{R})$,

$$\left| \|\omega\|_{\nu,[0,\tau]} - \|\omega\|_{\nu,[0,\tau\pm h]} \right| \le \max\left\{ \|\omega\|_{\nu,[\tau,\tau+h]}, \|\omega\|_{\nu,[\tau-h,\tau]} \right\} \to 0 \text{ as } h \to 0^+,$$

the function $\tau^{1-\beta} + \tau^{\nu-\beta} \|\omega\|_{\nu,[0,\tau]}$ is then continuous due to the continuity of each component in τ. Hence

$$(t_{i+1}-t_i)^{1-\beta} + (t_{i+1}-t_i)^{\nu-\beta} \|\omega\|_{\nu,[t_i,t_{i+1}]} = \frac{\mu}{C}, \ \forall i \ge 0. \tag{17.22}$$

If $t_\infty := \sup t_i < \infty$, then by choosing k such that $k(\nu-\beta) \ge 1$, one has

$$\begin{aligned} n(\mu/C)^k &\le \sum_{i=0}^{n-1} \left[(t_{i+1}-t_i)^{1-\beta} + (t_{i+1}-t_i)^{\nu-\beta} \|\omega\|_{\nu,[t_i,t_{i+1}]} \right]^k \\ &\le 2^{k-1} \sum_{i=0}^{n-1} \left[(t_{i+1}-t_i)^{k(1-\beta)} + (t_{i+1}-t_i)^{k(\nu-\beta)} \|\omega\|^k_{\nu,[0,t_\infty]} \right] \\ &\le 2^{k-1} \left[\sum_{i=0}^{n-1} (t_{i+1}-t_i)^{k(1-\beta)} + \sum_{i=0}^{n-1} (t_{i+1}-t_i)^{k(\nu-\beta)} \|\omega\|^k_{\nu,[0,t_\infty]} \right] \\ &\le 2^{k-1} t_\infty^{k(1-\beta)} + t_\infty^{k(\nu-\beta)} \|\omega\|^k_{\nu,[0,t_\infty]} < \infty \end{aligned}$$

for all $n \in \mathbb{N}$, which is contradiction. Hence $\{t_i\}$ is increasing to infinity and it makes sense to define

$$N(T,\omega) := \max\{i : t_i \le T\}.$$

Moreover,

$$N(T,\omega) \le 2^{k-1} \left(\frac{C}{\mu}\right)^k \left(T^{k(1-\beta)} + T^{k(\nu-\beta)} \|\omega\|^k_{\nu,[0,T]} \right). \tag{17.23}$$

Step 3: In this step one shows the *local existence* of solution on $[t_0, t_1]$ constructed as above. From definition of stopping times, $|t_1 - t_0| < 1$ and $C'(t_1 - t_0) \le C$, hence it follows that

$$F : C^{\beta\text{-Hol}}([t_0-r,t_1],\mathbb{R}^d) \to C^{\beta\text{-Hol}}([t_0-r,t_1],\mathbb{R}^d)$$

satisfying

$$\|F(x)\|_{\infty,\beta,[t_0-r,t_1]} \le \|x\|_{\infty,\beta,[t_0-r,t_0]} + \mu(1 + \|x\|_{\infty,\beta,[t_0-r,t_1]}) \tag{17.24}$$

Introducing the set

$$B := \left\{ x \in C^{\beta\text{-Hol}}([t_0 - r, t_1], \mathbb{R}^d) |\ x_{t_0} = \eta, \|x\|_{\infty,\beta,[t_0-r,t_1]} \leq R := \frac{\|\eta\|_{\infty,\beta,[t_0-r,t_0]} + \mu}{1-\mu} \right\},$$

then $F : B \to B$. By Lemma 17.4 and the definition of F, the following estimate

$$\begin{aligned}
\|F(x) - F(y)\|_{\infty,\beta,[t_0-r,t_1]} &= \|F(x) - F(y)\|_{\infty,[t_0,t_1]} + \|\!|F(x) - F(y)|\!\|_{\beta,[t_0,t_1]} \\
&\leq \left[1 + (t_1 - t_0)^{\beta}\right] \|\!|F(x) - F(y)|\!\|_{\beta,[t_0,t_1]} \\
&\leq L(t_1 - t_0, R)\left[1 + (t_1 - t_0)^{\beta}\right] \|x - y\|_{\infty,\beta,[t_0-r,t_1]}.
\end{aligned}$$

proves the continuity of F on B.
Observe that F is a compact operator on B. Indeed, take the sequence $y^n = F(x^n)$, $x^n \in B$, by (17.20)

$$\|\!|y^n|\!\|_{(\beta+\epsilon),[t_0,t_1]} \leq C\left((t_1 - t_0)^{1-\beta-\epsilon} + (t_1 - t_0)^{\nu-\beta-\epsilon} \|\!|\omega|\!\|_{\nu,[t_0,t_1]}\right)(1 + R).$$

By Proposition 5.28 of [8], there exists a subsequence $y^{n_k} 1_{[t_0,t_1]}$ which converges in $C^{\beta\text{-Hol}}([t_0, t_1], \mathbb{R}^d)$. Additionally, for all k, $y^{n_k}(t) = \eta(t), \forall t \in [t_0 - r, t_0]$, hence

$$\|y^{n_k} - y^{n_{k'}}\|_{\infty,\beta,[t_0-r,t_1]} = \|y^{n_k} - y^{n_{k'}}\|_{\infty,\beta,[t_0,t_1]} \to 0 \ \text{ as } \ k, k' \to \infty.$$

Since $C^{\beta\text{-Hol}}([t_0 - r, t_1], \mathbb{R}^d)$ is Banach one concludes that there is a subsequence of y^n that converges in $C^{\beta\text{-Hol}}([t_0 - r, t_1], \mathbb{R}^d)$.
To sum up, $F : B \to B$ is a compact operator on the non empty, closed, bounded, convex subset of Banach space $C^{\beta\text{-Hol}}([t_0 - r, t_1], \mathbb{R}^d)$. By Schauder-Tychonoff fixed point theorem (see e.g [23, Theorem 2.A, p. 56]), there exists a function $x^* \in B$ such that $F(x^*) = x^*$, i.e x^* is a local solution of (17.9) on $[t_0 - r, t_1]$. The fact that $x^* \in C^{\nu\text{-Hol}}([t_0 - r, t_1], \mathbb{R}^d)$ is then obvious.

Step 4: The local solution is unique.
Assuming that x and y are solutions to (17.9) on $[t_0 - r, t_1]$ with the same initial condition η , bounded by $M > 0$. Put $z = x - y$ then $F(x) - F(y) = z$. By virtue of Lemma 17.4, for $t_0 \leq s < t \leq t_1$,

$$\begin{aligned}
\|\!|z|\!\|_{\beta,[s,t]} &\leq L(t_1 - t_0, M)\left[(t - s)^{1-\beta} + (t - s)^{\nu-\beta} \|\!|\omega|\!\|_{\nu,[s,t]}\right] \|z\|_{\infty,\beta,[s-r,t]} \\
&\leq L(t_1 - t_0, M)\left[(t - s)^{1-\beta} + (t - s)^{\nu-\beta} \|\!|\omega|\!\|_{\nu,[s,t]}\right] \left(\|z\|_{\infty,[s-r,t]} + \|\!|z|\!\|_{\beta,[s-r,t]}\right) \\
&\leq L(t_1 - t_0, M)\left[(t - s)^{1-\beta} + (t - s)^{\nu-\beta} \|\!|\omega|\!\|_{\nu,[s,t]}\right] \\
&\quad \times \left(\max\{\|z\|_{\infty,[s-r,s]}, \|z\|_{\infty,[s,t]}\} + \|\!|z|\!\|_{\beta,[s-r,s]} + \|\!|z|\!\|_{\beta,[s,t]}\right).
\end{aligned}$$

Since $\|z\|_{\infty,[s,t]} \leq \|z(s)\| + (t-s)^{\beta} \|\!|z|\!\|_{\beta,[s,t]} \leq \|z\|_{\infty,[s-r,s]} + \|\!|z|\!\|_{\beta,[s,t]}$, it follows that

$$\begin{aligned} &\|\!|z|\!\|_{\beta,[s,t]} \\ &\leq 2L(t_1 - t_0, M)\left[(t-s)^{1-\beta} + (t-s)^{\nu-\beta} \|\!|\omega|\!\|_{\nu,[s,t]}\right]\left(\|z\|_{\infty,\beta,[s-r,s]} + \|\!|z|\!\|_{\beta,[s,t]}\right). \end{aligned} \tag{17.25}$$

Construct similarly to **Step 2** a finite sequence $\{s_i\}$ on $[t_0, t_1]$ such that $s_0 = t_0$ and

$$(s_{i+1} - s_i)^{1-\beta} + (s_{i+1} - s_i)^{\nu-\beta} \|\!|\omega|\!\|_{\nu,[s_i,s_{i+1}]} = \frac{\mu}{2L(t_1 - t_0, M)}.$$

It follows from (17.25) that

$$\|\!|z|\!\|_{\beta,[s_0,s_1]} \leq \mu\left(\|z\|_{\infty,\beta,[s_0-r,s_0]} + \|\!|z|\!\|_{\beta,[s_0,s_1]}\right) = \mu \|\!|z|\!\|_{\beta,[s_0,s_1]}. \tag{17.26}$$

Consequently, $\|\!|z|\!\|_{\beta,[s_0,s_1]} = 0$. By induction, one can prove that $\|\!|z|\!\|_{\beta,[t_0,t_1]} = 0$. Therefore, $z(u) \equiv 0$, $\forall u \in [t_0 - r, t_1]$, i.e. $x \equiv y$ on $[t_0 - r, t_1]$.

Step 5: By induction, there exists a unique solution of (17.9) on each $[t_i - r, t_{i+1}]$. Finally, due to the unboundedness of $\{t_i\}$ the solution of (17.9) can be extended to the whole $[-r, T]$ by concatenation.

Theorem 17.2 *Under the assumptions of Theorem 17.1, one has*

$$\sup_{t\in[t_{N(t,\omega)},t_{N(t,\omega)+1}]} \|x_t\|_{\infty,\beta,[-r,0]} \leq e^{-[N(t,\omega)+1]\log(1-\mu)}\left[\|x_{t_0}\|_{\infty,\beta,[-r,0]} + 1\right], \tag{17.27}$$

where $N(t,\omega)$-the number of stopping times (17.22) in $(0,t]$, can be approximated by (17.23).

Proof From the proof of Theorem 17.1, in particular (17.18) and (17.24), it follows that for any $i \geq 0$

$$\|x\|_{\infty,\beta,[t_i-r,t]} \leq \|x\|_{\infty,\beta,[t_i-r,t_i]} + \mu(1 + \|x\|_{\infty,\beta,[t_i-r,t]}),\ \forall t \in [t_i, t_{i+1}].$$

In other words,

$$\|x\|_{\infty,\beta,[t_i-r,t]} \leq \frac{\mu}{1-\mu} + \frac{1}{1-\mu}\|x\|_{\infty,\beta,[t_i-r,t_i]},\ \forall t \in [t_i, t_{i+1}]. \tag{17.28}$$

On the other hand,

$$\|x\|_{\infty,\beta,[t-r,t]} = \|x\|_{\infty,[t-r,t]} + \|\!|x|\!\|_{\beta,[t-r,t]} \leq \|x\|_{\infty,\beta,[t_i-r,t]},\ \forall t \in [t_i, t_{i+1}]. \tag{17.29}$$

Hence it follows from (17.28) and (17.29) that

$$\|x\|_{\infty,\beta,[t-r,t]} \le \frac{\mu}{1-\mu} + \frac{1}{1-\mu}\|x\|_{\infty,\beta,[t_i-r,t_i]},\ \forall t \in [t_i, t_{i+1}].$$

which implies that

$$\sup_{t\in[t_i,t_{i+1}]} \|x_t\|_{\infty,\beta,[-r,0]} \le \frac{\mu}{1-\mu} + \frac{1}{1-\mu}\|x_{t_i}\|_{\infty,\beta,[-r,0]}. \tag{17.30}$$

In particular, for any $i \ge 0$,

$$\|x_{t_{i+1}}\|_{\infty,\beta,[-r,0]} \le \frac{\mu}{1-\mu} + \frac{1}{1-\mu}\|x_{t_i}\|_{\infty,\beta,[-r,0]},$$

or equivalently

$$\|x_{t_{i+1}}\|_{\infty,\beta,[-r,0]} + 1 \le \frac{1}{1-\mu}\Big[\|x_{t_i}\|_{\infty,\beta,[-r,0]} + 1\Big].$$

By induction arguments, one can conclude that

$$\|x_{t_i}\|_{\infty,\beta,[-r,0]} \le \left[\frac{1}{1-\mu}\right]^i \Big[\|x_{t_0}\|_{\infty,\beta,[-r,0]} + 1\Big] - 1,\ \forall i \ge 0. \tag{17.31}$$

Equation (17.27) is then a direct consequence of (17.30) and (17.31).

The arguments in the proof of Theorem 17.2 help us to derive a type of Gronwall lemma for Hölder norms.

Lemma 17.5 (Gronwall-Type Lemma) *Assume that $z : [-r, T] \to \mathbb{R}^d$ satisfies for any $0 \le s \le t \le T$*

$$|||z|||_{\beta,[s,t]} \le A + C\left((t-s)^{1-\beta} + (t-s)^{\nu-\beta}\,|||\omega|||_{\nu,[s,t]}\right)\|z\|_{\infty,\beta,[s-r,t]} \tag{17.32}$$

with some constants $A, C > 0$. Then for $\mu < \min\{\frac{1}{2}, C\}$ the following estimate holds

$$\|z_t\|_{\infty,\beta,[-r,0]} \le e^{-[N(t,\omega)+1]\log(1-2\mu)}\Big[\frac{A}{\mu} + \|z\|_{\infty,\beta,[-r,0]}\Big],\ \forall t \in [0, T]. \tag{17.33}$$

Proof Using the construction of stopping times in (17.22), one has

$$|||z|||_{\beta,[t_i,t]} \le A + \mu\|z\|_{\infty,\beta,[t_i-r,t]},\ \forall t \in [t_i, t_{i+1}],$$

hence

$$\begin{aligned}\|z\|_{\infty,\beta,[t_i-r,t]} &\le \max\{\|z\|_{\infty,[t_i-r,t_i]}, \|z\|_{\infty,[t_i,t]}\} + \|\!|z|\!\|_{\beta,[t_i-r,t_i]} + \|\!|z|\!\|_{\beta,[t_i,t]} \\ &\le \|z\|_{\infty,\beta,[t_i-r,t_i]} + (1+(t_{i+1}-t_i)^{\beta})\,\|\!|z|\!\|_{\beta,[t_i,t]} \\ &\le \|z\|_{\infty,\beta,[t_i-r,t_i]} + 2[A+\mu\|z\|_{\infty,\beta,[t_i-r,t]}]\end{aligned}$$

due to the fact that $\frac{\mu}{C} < 1$. It follows that

$$\|z\|_{\infty,\beta,[t_i-r,t]} \le \frac{2A}{1-2\mu} + \frac{1}{1-2\mu}\|z\|_{\infty,\beta,[t_i-r,t_i]},\ \forall t \in [t_i, t_{i+1}],$$

(provided that $\mu < \frac{1}{2}$), which has similar form to (17.28). As a consequence, by following the same arguments as in Theorem 17.2, one has

$$\|z_{t_i}\|_{\infty,\beta,[-r,0]} \le \left[\frac{1}{1-2\mu}\right]^i \left[\|z_{t_0}\|_{\infty,\beta,[-r,0]} + \frac{A}{\mu}\right] - \frac{A}{\mu},\ \forall i \ge 0,$$

which proves (17.33).

Denote by $x(\cdot,\omega,\eta)$ the solution of (17.1) with initial function η. We prove in the following the continuity of the solution with respect to the initial condition.

Theorem 17.3 *Under the assumptions of Theorem 17.1, the solution $x_t(\cdot,\omega,\eta)$ is continuous with respect to η.*

Proof For any $\eta^1, \eta^2 \in C^{\beta\text{-Hol}}([-r,0],\mathbb{R}^d)$ denote $x^i(\cdot) = x(\cdot,\omega,\eta^i)$, $i = 1,2$. Fix η^1, by (17.27) one can choose M large enough such that $\|x(\cdot,\omega,\eta^2)\|_{\infty,\beta,[-r,T]} \le M$ for all η^2 such that $\|\eta^2 - \eta^1\|_{\infty,\beta,[-r,0]} \le 1$. From (17.14) in Lemma 17.4, one has for all $0 \le a \le b \le T$,

$$\left\|\!\left|x^1 - x^2\right|\!\right\|_{\beta,[a,b]} \le L(T,M)\left((b-a)^{1-\beta} + (b-a)^{\nu-\beta}\,\|\!|\omega|\!\|_{\nu,[a,b]}\right)\|x^1 - x^2\|_{\infty,\beta,[a-r,b]},$$

which has the form (17.32) with $A = 0$ and $C = L(T,M)$. Therefore,

$$\|x_t(\cdot,\omega,\eta^2) - x_t(\cdot,\omega,\eta^1)\|_{\infty,\beta,[-r,0]} \le e^{-[N(t,\omega)+1]\log(1-2\mu)}\|\eta^1 - \eta^2\|_{\infty,\beta,[-r,0]}, \forall t \in [0,T],$$

in which $N(t,\omega)$ is defined in (17.23) with $C = L(T,M)$ and $\mu < \min\{1/2, L(T,M)\}$ and N depends on $L(T,M)$ - the local constant in the vicinity of η_1. That proves the continuity of $x_t(\cdot,\omega,\eta)$ w.r.t. the initial function η.

Remark 17.3 It can be seen that for $x \in C^{\beta-\text{Hol}}([-r,T],\mathbb{R}^d)$ there exists $C(T,r)$ such that

$$\|x(\cdot)\|_{\infty,\beta,[-r,T]} \le C(T,r)\sup_{t\in[0,T]}\|x_t(\cdot)\|_{\infty,\beta,[-r,0]}.$$

Indeed, since $\|x(\cdot)\|_{\infty,[-r,T]} = \sup_{t\in[0,T]} \|x_t(\cdot)\|_{\infty,[-r,0]}$, for $s, t \in [-r, T]$ one can construct a finite sequence s_i as follow: $s = s_0, s_1 = s_0 + r, s_2 = s_1 + r, \ldots$, until $s_n + r \geq t$ and assign $s_{n+1} := t$. Then

$$\begin{aligned}
\frac{\|x(t) - x(s)\|}{|t-s|^\beta} &\leq \sum_{i=0}^{n} \frac{\|x(s_{i+1}) - x(s_i)\|}{|s_{i+1} - s_i|^\beta} \\
&\leq \sum_{i=0}^{n} \frac{\|x_{s_{i+1}}(s_i - s_{i+1}) - x_{s_{i+1}}(0)\|}{|s_{i+1} - s_i|^\beta} \\
&\leq \sum_{i=0}^{n} \left|\!\left|\!\left| x_{s_{i+1}} \right|\!\right|\!\right|_{\beta,[-r,0]} \\
&\leq (1 + T/r) \sup_{t\in[0,T]} |\!|\!| x_t |\!|\!|_{\beta,[-r,0]}.
\end{aligned}$$

Hence, from Theorem 17.3 one concludes that

$$\|x(\cdot,\omega,\eta^2) - x(\cdot,\omega,\eta^1)\|_{\infty,\beta,[-r,T]} \leq C(T,r) e^{-[N(T,\omega)+1]\log(1-2\mu)} \|\eta^1 - \eta^2\|_{\infty,\beta,[-r,0]}.$$

Next, assuming that f is C^1, we fix a solution $x(\cdot,\omega,\eta)$ of (17.1) and consider the linearized equation

$$y(t) = \eta^1(0) - \eta(0) + \int_0^t Df(x_s) y_s ds + \int_0^t Dg(x_s) y_s d\omega(s), \tag{17.34}$$

with initial function $\eta^1 - \eta \in C^{\beta-\mathrm{Hol}}([-r, 0], \mathbb{R}^d)$. Since $y \in C^{\delta\beta-\mathrm{Hol}}([-r, T], \mathbb{R}^d)$ and

$$\begin{aligned}
\|Dg(x_t) - Dg(x_s)\|_{L(C_r,\mathbb{R}^d)} &\leq L_M \|x_t - x_s\|^\delta_{\infty,[-r,0]} \\
&\leq L_M \, |\!|\!| x |\!|\!|^\delta_{\beta,[-r,T]} \, (t-s)^{\delta\beta}, \\
|\!|\!| Dg(x_\cdot) |\!|\!|_{\delta\beta,[a,b]} &\leq L_M M^\delta
\end{aligned} \tag{17.35}$$

with $M \geq \|x\|_{\infty,\beta,[-r,T]}$, the integrals $\int_0^t Df(x_s) y_s ds$ and $\int_0^t Dg(x_s) y_s d\omega(s)$ are well defined. We need to prove the following lemma

Lemma 17.6 *The Eq. (17.34) has unique solution y in $C^{\delta\beta-\mathrm{Hol}}([-r, T], \mathbb{R}^d)$. Moreover, the solution is Hölder continuous with exponent ν on $[0, T]$.*

Proof The proof is similar to that of Theorem 17.1. Note that $\|Df(\xi)\| \leq L_f$ for all $\xi \in C_r$

Define the map

$$G(y)(t) = \begin{cases} y(t_0) + \int_0^t Df(x_s)y_s ds + \int_0^t Dg(x_s)y_s d\omega(s), & \text{if } t \in [t_0, t_1] \\ y(t), & \text{if } t \in [t_0 - r, t_0], \end{cases}$$

then for $s, t \in [0, T]$

$$\begin{aligned} &\|G(y)(t) - G(y)(s)\| \\ &\le L_f \|y\|_{\infty,[s-r,t]}(t-s) + |||\omega|||_{\nu,[s,t]}(t-s)^{\nu}\Big[L_g\|y\|_{\infty,[s-r,t]} \\ &\quad + K'(t-s)^{\delta\beta}\|y\|_{\infty,[s-r,t]}|||Dg(x_\cdot)|||_{\delta\beta,[s,t]} + K'L_g(t-s)^{\delta\beta}|||y|||_{\delta\beta,[s-r,t]}\Big], \end{aligned}$$

with $K' = \frac{1}{1-2^{1-(\nu+\delta\beta)}}$. Combining with (17.35), it follows that

$$|||G(y)|||_{\delta\beta,[s,t]} \le C\left((t-s)^{1-\delta\beta} + |||\omega|||_{\nu,[s,t]}(t-s)^{\nu-\delta\beta}\right)\|y\|_{\infty,\delta\beta,[s-r,t]}$$

Repeat the arguments in Theorem 17.1, one can prove the existence of solution to (17.35). Since G is linear, the uniqueness of the solution is derived by a contraction mapping argument. Finally, it is obvious that the solution depends linearly on the initial function.

Theorem 17.4 *Assuming that f, g satisfy conditions (H_f) and (H_g) and f is a C^1 function. Then the solution $x_t(\cdot, \omega, \eta)$ of (17.9) is differentiable with respect to initial function η.*

Proof Consider two solutions $x(\cdot) = x(\cdot, \omega, \eta)$ and $x^1(\cdot) = x(\cdot, \omega, \eta^1)$ of (17.9)

$$\begin{aligned} x^1(t) &= \eta^1(0) + \int_0^t f(x_s^1)ds + \int_0^t g(x_s^1)d\omega(s) \\ x(t) &= \eta(0) + \int_0^t f(x_s)ds + \int_0^t g(x_s)d\omega(s), \end{aligned}$$

and the solution $y(\cdot) = y(\cdot, \omega, \eta^1 - \eta)$ of (17.34). Define

$$z(\cdot) = x^1(\cdot) - x(\cdot) - y(\cdot)$$

then $z \equiv 0$ on $[-r, 0]$. By the assumptions, there exists F^*, G^*- the nonlinear remaining terms of f, g such that

$$\begin{aligned} f(x_s^1) - f(x_s) &= Df(x_s)(x_s^1 - x_s) + F^*(x_s^1 - x_s) \\ g(x_s^1) - g(x_s) &= Dg(x_s)(x_s^1 - x_s) + G^*(x_s^1 - x_s). \end{aligned}$$

Since f, g are C^1, there exist a number $h > 0$ and continuous functions $p,\ q : [0, h] \to \mathbb{R}_+$, $p(0) = q(0) = 0$ and $\lim_{u\to 0} p(u) = \lim_{u\to 0} q(u) = 0$, such that

$$\begin{aligned}\|F^*(x_s^1 - x_s)\| &= \left\| \int_0^1 [Df(\theta x_s^1 + (1-\theta)x_s) - Df(x_s)](x_s^1 - x_s)d\theta \right\| \\ &\le \|x^1(\cdot) - x(\cdot)\|_{\infty,\beta,[-r,T]}\ p\big(\|x^1(\cdot) - x(\cdot)\|_{\infty,\beta,[-r,T]}\big)\end{aligned}$$

and

$$\begin{aligned}\|G^*(x_s^1 - x_s)\| &= \left\| \int_0^1 [Dg(\theta x_s^1 + (1-\theta)x_s) - Dg(x_s)](x_s^1 - x_s)d\theta \right\| \\ &\le \|x^1(\cdot) - x(\cdot)\|_{\infty,\beta,[-r,T]}\ q\big(\|x^1(\cdot) - x(\cdot)\|_{\infty,\beta,[-r,T]}\big).\end{aligned}$$

whenever $\|x^1(\cdot) - x(\cdot)\|_{\infty,\beta,[-r,T]} \le h$. Similar to Lemma 17.3, we estimate the Hölder norm of G^*. Specifically, for $0 \le s < t \le T$,

$$\begin{aligned}&\|G^*(x_t^1 - x_t) - G^*(x_s^1 - x_s)\| \\ &= \left\| \int_0^1 [Dg(\theta x_t^1 + (1-\theta)x_t) - Dg(x_t)](x_t^1 - x_t)d\theta \right. \\ &\qquad \left. - \int_0^1 [Dg(\theta x_s^1 + (1-\theta)x_s) - Dg(x_s)](x_s^1 - x_s)d\theta \right\| \\ &\le \left\| \int_0^1 [Dg(\theta x_t^1 + (1-\theta)x_t) - Dg(x_t) - Dg(\theta x_s^1 + (1-\theta)x_s) \right. \\ &\qquad \left. + Dg(x_s)](x_t^1 - x_t)d\theta \right\| \\ &+ \left\| \int_0^1 [Dg(\theta x_s^1 + (1-\theta)x_s) - Dg(x_s)](x_t^1 - x_t - x_s^1 + x_s)d\theta \right\|. \qquad (17.36)\end{aligned}$$

From the assumption of g the second integral in (17.36) is less than or equal

$$L_M \|x_t^1 - x_t - x_s^1 + x_s\| . \|x_s^1 - x_s\|_{\infty,[-r,0]}^{\delta},$$

where M is a upper bound of $\|x\|_{\infty,\beta,[-r,T]}$ and $\|x_1\|_{\infty,\beta,[-r,T]}$. It follows from Lemma 17.1 that

$$\begin{aligned}&\left\| \int_0^1 [Dg(\theta x_s^1 + (1-\theta)x_s) - Dg(x_s)](x_t^1 - x_t - x_s^1 + x_s)d\theta \right\| \\ &\le L_M (t-s)^{\beta} \|x^1 - x\|_{\infty,\beta,[-r,T]}^{1+\delta}. \qquad (17.37)\end{aligned}$$

Since $\delta\beta + \nu > 1$, one can choose $0 < \gamma < 1$ such that $\gamma\delta\beta + \nu > 1$. Then

$$\begin{aligned}
&\|Dg(\theta x_t^1 + (1-\theta)x_t) - Dg(x_t) - Dg(\theta x_s^1 + (1-\theta)x_s) + Dg(x_s)\|^{\gamma}_{L(C_r,\mathbb{R}^d)}\\
&\le \|Dg(\theta x_t^1 + (1-\theta)x_t) - Dg(\theta x_s^1 - (1-\theta)x_s)\|^{\gamma}_{L(C_r,\mathbb{R}^d)}\\
&\quad + \|Dg(x_t) - Dg(x_s)\|^{\gamma}_{L(C_r,\mathbb{R}^d)}\\
&\le 2L_M^{\gamma}\left(\|x_t^1 - x_s^1\|^{\gamma\delta}_{\infty,[-r,0]} + |x_t - x_s\|^{\gamma\delta}_{\infty,[-r,0]}\right)\\
&\le 2L_M^{\gamma}(t-s)^{\gamma\delta\beta}\left(\|x_\cdot^1\|^{\gamma\delta}_{\beta,[0,T]} + \|x_\cdot\|^{\gamma\delta}_{\beta,[0,T]}\right)\\
&\le 2L_M^{\gamma}(t-s)^{\gamma\delta\beta}\left(\|x^1\|^{\gamma\delta}_{\infty,\beta,[-r,T]} + \|x\|^{\gamma\delta}_{\infty,\beta,[-r,T]}\right)\\
&\le 4M^{\gamma\delta}L_M^{\gamma}(t-s)^{\gamma\delta\beta}. \qquad (17.38)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\|Dg(\theta x_t^1 + (1-\theta)x_t) - Dg(x_t) - Dg(\theta x_s^1 + (1-\theta)x_s) + Dg(x_s)\|^{1-\gamma}_{L(C_r,\mathbb{R}^d)}\\
&\le \|Dg(\theta x_t^1 + (1-\theta)x_t) - Dg(x_t)\|^{1-\gamma}_{L(C_r,\mathbb{R}^d)}\\
&\quad + \|Dg(x_s^1 + (1-\theta)x_s) - Dg(x_s)\|^{1-\gamma}_{L(C_r,\mathbb{R}^d)}\\
&\le 2L_M^{1-\gamma}\|x^1 - x\|^{(1-\gamma)\delta}_{\infty,\beta,[-r,T]}. \qquad (17.39)
\end{aligned}$$

Therefore, the first integral in (17.36) does not exceed $8M^{\gamma\delta}L_M(t-s)^{\gamma\delta\beta}\|x^1 - x\|^{1+(1-\gamma)\delta}_{\infty,\beta,[-r,T]}$. Combining this with (17.36) and (17.37), one obtains

$$\|G^*(x_t^1 - x_t) - G^*(x_s^1 - x_s)\| \le (t-s)^{\gamma\delta\beta}C(T,M)\|x^1 - x\|^{1+(1-\gamma)\delta}_{\infty,\beta,[-r,T]},$$

which implies

$$\left\|\left\|G^*(x_\cdot^1 - x_\cdot)\right\|\right\|_{\gamma\delta\beta,[0,T]} \le C(G^*)\|x^1 - x\|^{1+(1-\gamma)\delta}_{\infty,\beta,[-r,T]}, \qquad (17.40)$$

Rewrite the equation of z in the form

$$\begin{aligned}
z(t) &= \int_0^t \Big(f(x_s^1) - f(x_s) - Df(x_s)y_s\Big)ds + \int_0^t \Big(g(x_s^1) - g(x_s) - Dg(x_s)y_s\Big)d\omega(s)\\
&= \int_0^t \Big(Df(x_s)z_s + F^*(x_s^1 - x_s)\Big)ds + \int_0^t \Big(Dg(x_s)z_s + G^*(x_s^1 - x_s)\Big)d\omega(s)\\
&= \Big[\int_0^t F^*(x_s^1 - x_s)ds + \int_0^t G^*(x_s^1 - x_s)d\omega(s)\Big] + \int_0^t Df(x_s)z_s ds + \int_0^t Dg(x_s)z_s d\omega(s).
\end{aligned}$$

By similar estimates as in Theorem 17.1, there exist constants K_1, K_2 depending on $\nu, \beta, \delta, \gamma$ and generic constants C_1, C_2 such that for all $0 \le s < t \le T$

$$\begin{aligned}
|\!|\!| z |\!|\!|_{\beta,[s,t]} &\le \Big(|t-s|^{1-\beta} \|F^*(x_\cdot^1 - x_\cdot)\|_{\infty,[s,t]} \\
&\quad + |t-s|^{\nu-\beta} K(1+T^{\gamma\delta\beta})\, |\!|\!| \omega |\!|\!|_{\nu,[s,t]}\, \|G^*(x_\cdot^1 - x_\cdot)\|_{\infty,\gamma\delta\beta,[s,t]} \Big) \\
&\quad + \Big(\|Df(x_\cdot) z_\cdot\|_{\infty,[s,t]} |t-s|^{1-\beta} \\
&\quad + |t-s|^{\nu-\beta} K(1+T^{\delta\beta})\, |\!|\!| \omega |\!|\!|_{\nu,[s,t]}\, \|Dg(x_\cdot) z_\cdot\|_{\infty,\delta\beta,[s,t]} \Big) \\
&\le C_1 \Big(\|F^*(x_\cdot^1 - x_\cdot)\|_{\infty,[0,T]} + \|G^*(x_\cdot^1 - x_\cdot)\|_{\infty,\gamma\delta\beta,[0,T]} \Big) \\
&\quad + C_2 \Big((t-s)^{1-\beta} + (t-s)^{\nu-\beta}\, |\!|\!| \omega |\!|\!|_{\nu,[s,t]} \Big)\, \|z\|_{\infty,\beta,[s-r,t]} \\
&\le C_1 \Big(\|x^1 - x\|_{\infty,\beta,[-r,T]}\, P(\|x^1 - x\|_{\infty,\beta,[-r,T]}) \Big) \\
&\quad + C_2 \Big((t-s)^{1-\beta} + (t-s)^{\nu-\beta}\, |\!|\!| \omega |\!|\!|_{\nu,[s,t]} \Big)\, \|z\|_{\infty,\beta,[s-r,t]} \qquad (17.41)
\end{aligned}$$

where $P(u) = p(u) + q(u) + u^{(1-\gamma)\delta}$. Due to Remark 17.3, there exist constants $A(T,\eta)$, $C(T,\eta)$, a number $h_1 > 0$ and a function $p_1 : [0, h_1] \to \mathbb{R}_+$ with $p_1(0) = 0$, $\lim_{u\to 0} p_1(u) = 0$, such that

$$\begin{aligned}
|\!|\!| z |\!|\!|_{\beta,[s,t]} &\le A(T,\eta) \|\eta^1 - \eta\|_{\infty,\beta,[-r,0]} p_1\big(\|\eta^1 - \eta\|_{\infty,\beta,[-r,0]}\big) \\
&\quad + C(T,\eta) \Big((t-s)^{1-\beta} + (t-s)^{\nu-\beta}\, |\!|\!| \omega |\!|\!|_{\nu,[s,t]} \Big)\, \|z\|_{\infty,\beta,[s-r,t]},
\end{aligned}$$

whenever $\|\eta^1 - \eta\|_{\infty,\beta,[-r,0]} \le h_1$. Applying Lemma 17.5, one concludes that there exists a generic constant C such that for all t in $[0, T]$

$$\begin{aligned}
\|z_t\|_{\infty,\beta,[-r,0]} &\le C \Big[\|\eta^1 - \eta\|_{\infty,\beta,[-r,0]} p_1\big(\|\eta^1 - \eta\|_{\infty,\beta,[-r,0]}\big) + \|z\|_{\infty,\beta,[-r,0]} \Big] \\
&\le C \|\eta^1 - \eta\|_{\infty,\beta,[-r,0]} p_1\big(\|\eta^1 - \eta\|_{\infty,\beta,[-r,0]}\big)
\end{aligned}$$

for all $\|\eta^1 - \eta\|_{\infty,\beta,[-r,0]} \le h_1$, since $z \equiv 0$ on $[-r, 0]$. Therefore,

$$\begin{aligned}
&\|x_t(\cdot,\omega,\eta^1) - x_t(\cdot,\omega,\eta) - y_t(\cdot,\omega,\eta^1 - \eta)\|_{\infty,\beta,[-r,0]} \\
&\le \|\eta^1 - \eta\|_{\infty,\beta,[-r,0]} C p_1\big(\|\eta^1 - \eta\|_{\infty,\beta,[-r,0]}\big) \qquad (17.42)
\end{aligned}$$

for any η^1 in the vicinity of η such that $\|\eta^1 - \eta\|_{\infty,\beta,[-r,0]} \le h_1$. Finally, (17.42) implies that $x_t(\cdot,\omega,\eta)$ is differentiable with respect to η, with its derivative to be $y_t(\cdot,\omega,\cdot)$.

Acknowledgements We would like to thank the anonymous referees for their careful reading and insightful remarks which led to improvement of our manuscript. This research is partly funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) and by Thang Long University.

References

1. Boufoussi, B., Hajji, S.: Functional differential equations driven by a fractional Brownian motion. Comput. Math. Appl. **62**(1), 746–754 (2011)
2. Boufoussi, B., Hajji, S.: Neutral stochastic functional differential equations driven by a fractional Brownian motion in Hilbert space. Stat. Probab. Lett. **82**, 1549–1558 (2012)
3. Boufoussi, B., Hajji, S., Lakhel, E.H.: Functional differential equations in Hilbert spaces driven by a fractional Brownian motion. Afr. Math. **23**(2), 173–194 (2012)
4. Chen, Y., Gao, H., Garrido-Arienza, M.J., Schmalfuß, B.: Pathwise solutions of stochastic partial differential equations driven by Hölder-continuous integrators with exponent larger than $\frac{1}{2}$ and random dynamical systems. Discrete Contin. Dyn. Syst. **34**(1), 79–98 (2013)
5. Cong, N.D., Duc, L.H., Hong, P.T.: Nonautonomous Young differential equations revisited. J. Dyn. Diff. Equat. **262**, 1–23 (2017). https://doi.org/10.1007/s10884-017-9634-y
6. Duc, L.H., Schmalfuß, B., Siegmund, S.: A note on the generation of random dynamical systems from fractional stochastic delay differential equations. Stoch. Dyn. **15**(3), 1550018 (2015). https://doi.org/10.1142/S0219493715500185
7. Duc, L.H., Garrido-Atienza, M.J., Neuenkirch, A., Schmalfuß, B.: Exponential stability of stochastic evolution equations driven by small fractional Brownian motion with Hurst parameter in $(\frac{1}{2}, 1)$. J. Differ. Equ. **264**(2), 1119–1145 (2018)
8. Friz, P., Victoir, N.: Multidimensional Stochastic Processes as Rough Paths: Theory and Applications. Cambridge Studies in Advanced Mathematics, vol. 120. Cambridge University Press, Cambridge (2010)
9. Garrido-Atienza, M.J., Lu, K., Schmalfuß, B.: Random dynamical systems for stochastic partial differential equations driven by a fractional Brownian motion. Discrete Contin. Dyn. Syst. B **14**(2), 473–493 (2010)
10. Lejay, A.: Controlled differential equations as Young integrals: a simple approach. J. Differ. Equ. **249**, 1777–1798 (2010)
11. Lyons, T.: Differential equations driven by rough signals, I, An extension of an inequality of L.C. Young. Math. Res. Lett. **1**, 45–464 (1994)
12. Lyons, T.: Differential equations driven by rough signals. Rev. Mat. Iberoam. **14**(2), 215–310 (1998)
13. Lyons, T., Qian, Zh.: System Control and Rough Paths. Oxford Mathematical Monographs. Clarendon Press, Oxford (2002)
14. Lyons, T., Caruana, M., Lévy, T.: Differential Equations Driven by Rough Paths. Lecture Notes in Mathematics, vol. 1908. Springer, Berlin (2007)
15. Mandelbrot, B., van Ness, J.: Fractional Brownian motion, fractional noises and applications. SIAM Rev. **4**(10), 422–437 (1968)
16. Maslowski, B., Nualart, D.: Evolution equations driven by a fractional Brownian motion. J. Funct. Anal. **202**(1), 277–305 (2003)
17. Nualart, D., Răşcanu, A.: Differential equations driven by fractional Brownian motion. Collect. Math. **53**(1), 55–81 (2002)
18. Samko, S., Kilbas, A., Marichev, O.: Fractional Integrals and Derivatives: Theory and Application. Gordon and Breach Science Publishers, Yvendon (1993)
19. Shevchenko, G.: Mixed stochastic delay differential equations. Theory Probab. Math. Stat. **89**, 181–195 (2014)

20. Young, L.C.: An integration of Hölder type, connected with Stieltjes integration. Acta Math. **67**, 251–282 (1936)
21. Zähle, M.: Integration with respect to fractal functions and stochastic calculus. I. Probab. Theory Related Fields. **111**(3), 333–374 (1998)
22. Zähle, M.: Integration with respect to fractal functions and stochastic calculus. II. Math. Nachr. **225**, 145–183 (2001)
23. Zeidler, E.: Nonlinear Functional Analysis and Its Applications I. Springer, Berlin (1986)

Chapter 18
Uniform Strong Law of Large Numbers for Random Signed Measures

O. I. Klesov and I. Molchanov

Abstract We prove a strong law of large numbers for random signed measures on Euclidean space that holds uniformly over a family of arguments (sets) scaled by diagonal matrices. Applications to random measures generated by sums of random variables, marked point processes and stochastic integrals are also presented.

18.1 Introduction

Set-indexed stochastic processes naturally appear in many areas of probability theory and mathematical statistics, e.g., as empirical measures [26], set-indexed martingales [15], point processes [7, 8], and random measures [17].

Both empirical and partial sum processes are special cases of marked point processes or random measures. They can be described via the pairs $(\mathbf{x}_i, m_i)$, where $\mathbf{x}_i$ are locations and m_i is the mass located at $\mathbf{x}_i$, also called the mark of $\mathbf{x}_i$.

Empirical processes assign the same nonrandom mass to each random location. More precisely, based on d-dimensional sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ the empirical measure is defined for any Borel $A \subset \mathbb{R}^d$ by

$$F_n(A) = \frac{1}{n}\#\{j = 1, \dots, n : \mathbf{X}_j \in A\}.$$

O. I. Klesov
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Department of Mathematical Analysis and Probability Theory, Kyiv, Ukraine
e-mail: klesov@matan.kpi.ua

I. Molchanov (✉)
University of Bern, Institute of Mathematical Statistics and Actuarial Science, Bern, Switzerland
e-mail: ilya.molchanov@stat.unibe.ch

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_18

Partial-sum processes are defined by assigning i.i.d. random masses to fixed locations on a grid in the space $\mathbb{R}^d$. If $\mathbb{Z}^d$ stands for the set of integer points in $\mathbb{R}^d$, then the partial-sum process is the normalized version of

$$S(A) = \sum_{\mathbf{j} \in A} X_{\mathbf{j}} \tag{18.1}$$

being the sum of i.i.d. random variables $X_{\mathbf{j}}$, $\mathbf{j} \in \mathbb{Z}^d$, with $\mathbf{j} \in A$. We set $S(A) = 0$ if $\{\mathbf{j} : \mathbf{j} \in A\} = \emptyset$.

The partial sum processes in dimension one are extensively studied in the classical probability theory as cumulative sums of random variables. We discuss below the higher-dimensional setting and allow for a richer family of sets A than in the classical case of multiple sums, see, e.g., [19]. There are three main types of asymptotic results for partial-sum processes indexed by sets, namely,

- strong laws of large numbers,
- central limit theorems,
- laws of the iterated logarithms.

Perhaps the first strong law of large numbers appeared in the paper by Bass and Pyke [3]. The central limit theorem for partial-sum processes is obtained by Kuelbs [22], while the law of the iterated logarithm is due to Wichura [29]. Further references can be found in the survey papers by Pyke [25] and Gaenssler and Ziegler [11]. From now on, we concentrate on the strong law of large numbers.

The paper is organised as follows. First, we recall the Bass–Pyke theorem (the uniform strong law of large numbers) in Sect. 18.2. It is generalised for signed measures in Sect. 18.3 and proved in the subsequent Sect. 18.4. The main feature is the general scaling of the argument set using diagonal matrices with the determinant converging to infinity. The case of stationary measures is considered in Sect. 18.5. The most important special cases concern random measures generated by marked point processes and by sums of random variables on a grid. Section 18.6 describes an application to stochastic integrals. Section 18.7 concludes and mentions a number of further related references.

18.2 The Bass–Pyke Theorem

Let $\mathbb{N}^d$ be the set of d-dimensional vectors with positive integer coordinates. Consider a family of independent identically distributed random variables $\{X_{\mathbf{j}}, \mathbf{j} \in \mathbb{N}^d\}$. If A is a Borel measurable subset of $\mathbb{R}^d$ define $S(A)$ by (18.1). Let $|A|$ denote the Lebesgue measure of A and $tA = \{tx : x \in A\}$ for $t > 0$, and let B be the open

unit Euclidean ball centered at the origin. For $r \geq 0$ and Borel $A \subset \mathbb{R}^d$,

$$A^r = \{x \in \mathbb{R}^d : \ x + rB \cap A \neq \emptyset\}$$

denotes the outer r-parallel set of A and

$$A^{-r} = \{x : \ x + rB \subset A\}$$

is the inner r-parallel set. Therefore,

$$A(r) = A^r \setminus A^{-r} = \{x : \rho(x, \partial A) < r\},$$

where ρ is the Euclidean distance and ∂A is the boundary of A.

Theorem 18.1 (See [3, Th. 1]) *Assume that the expectation* $\mu = \mathbf{E}\left[X_j\right]$ *exists. Let* $\mathscr{A}$ *be a collection of Borel measurable subsets of* $[0, 1]^d$. *If*

$$\sup_{A \in \mathscr{A}} |A(\delta)| \to 0 \qquad \text{as } \ \delta \downarrow 0, \tag{18.2}$$

then

$$\lim_{n \to \infty} \sup_{A \in \mathscr{A}} \left| \frac{S(nA)}{n^d} - \mu |A| \right| = 0 \qquad \text{a.s.} \tag{18.3}$$

To appreciate some peculiarities of this theorem we briefly discuss below its simplest case, where $\mathscr{A}$ consists of a single set A.

Example 18.1 If $A = [0, 1]^d$, then nA is the cube in $\mathbb{N}^d$ with a side of length n and thus $mA \subseteq nA$ if $m \leq n$. Therefore $S(nA)$ is, in fact, a subsequence of sums of independent identically distributed random variables with the expectation μ. In this case, (18.3) follows from the Kolmogorov strong law of large numbers for independent identically distributed random variables.

Example 18.2 Let A be the set of points with rational coordinates in $[0, 1]^d$. Clearly (18.2) fails. On the other hand, $S(nA)$ is the same as in the case of $[0, 1]^d$ but $|A| = 0$. Therefore, (18.3) holds if $\mu = 0$ and it fails otherwise.

Example 18.3 Let A be the set of points of $[0, 1]^d$ with irrational coordinates. Clearly (18.2) fails. Since $S(nA) = 0$, strong law of large numbers (18.3) fails if $\mu \neq 0$. Otherwise (18.3) holds.

The situation is even more complicated if $\mathscr{A}$ becomes richer.

18.3 Uniform Law of Large Numbers for Random Signed Measures

Let $\xi(A)$, $A \in \mathscr{B}$, be a random signed measure defined on the family $\mathscr{B}$ of Borel sets in $\mathbb{R}^d$, see, e.g., [17]. Denote $|\mathbf{t}| = \prod_{i=1}^d t_i$ and $[\mathbf{t}, \mathbf{s}] = \times_{i=1}^d [t_i, s_i]$ for $\mathbf{t} = (t_1, \dots, t_d)$ and $\mathbf{s} = (s_1, \dots, s_d)$ from $\mathbb{R}^d$. For $\mathbf{t} \in \mathbb{R}^d_+$ and $A \subset \mathbb{R}^d$, write

$$\mathbf{t} \cdot A = \{(t_1 x_1, \dots, t_d x_d) : \mathbf{x} = (x_1, \dots, x_d) \in A\}.$$

Assume that $\xi(A)$ is integrable for each bounded Borel A and let

$$\Lambda(A) = \mathbf{E}\,[\xi(A)]\,, \qquad A \in \mathscr{B},$$

be the first moment measure of ξ.

The signed measure ξ is said to satisfy the multiparameter strong law of large numbers if

$$\lim_{|\mathbf{t}| \to \infty} \frac{\xi(\mathbf{t} \cdot I) - \mathbf{E}\,[\xi(\mathbf{t} \cdot I)]}{|\mathbf{t}|} = 0 \qquad \text{a.s.}, \tag{18.4}$$

where $I = (\mathbf{0}, \mathbf{1}]$. Note that $\mathbf{t}$ converges to infinity in a rather arbitrary manner, it is only essential that the volume of the rectangle $[\mathbf{0}, \mathbf{t}]$ converges to infinity.

Let $\mathscr{A}$ be a subfamily of Borel sets in I. For $m \geq 1$, denote

$$C_m(\mathbf{k}) = \frac{1}{m}(\mathbf{k} - \mathbf{1}, \mathbf{k}], \qquad k \in \mathbb{N}^d.$$

Here $\mathbf{1} = (1, \dots, 1)$ and $\mathbf{k} - \mathbf{1} = (k_1 - 1, \dots, k_d - 1)$ for $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$. For every $A \in \mathscr{A}$,

$$A = \bigcup_{C_m(\mathbf{k}) \subseteq A} C_m(\mathbf{k}), \qquad A''_m = \bigcup_{C_m(\mathbf{k}) \cap A \neq \varnothing} C_m(\mathbf{k})$$

are discrete analogues of the inner and outer parallel sets to A.

The following result generalizes Theorem 18.1.

Theorem 18.2 *Let ξ be a random signed measure that satisfies the multiparameter strong law of large numbers. Assume that*

$$\lim_{m \to \infty} \limsup_{|\mathbf{t}| \to \infty} \sup_{A \in \mathscr{A}} \left| \frac{\mathbf{E}\left[\xi(\mathbf{t} \cdot (A \setminus A'_m))\right]}{|\mathbf{t}|} \right| = 0 \tag{18.5}$$

and $|\xi(A)| \le \eta(A)$ *for all Borel sets* A *and a random measure* η *that satisfies the multiparameter strong law of large numbers and such that*

$$\lim_{m\to\infty} \limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \frac{\mathbf{E}\left[\eta(\mathbf{t}\cdot(A''_m \setminus A'_m))\right]}{|\mathbf{t}|} = 0. \tag{18.6}$$

Then ξ *satisfies the uniform strong law of large numbers, that is,*

$$\lim_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\xi(\mathbf{t}\cdot A) - \mathbf{E}\,[\xi(\mathbf{t}\cdot A)]}{|\mathbf{t}|} \right| = 0 \qquad a.s. \tag{18.7}$$

Corollary 18.1 *Assume that* ξ *is a random (non-negative) measure that satisfies the multiparameter strong law of large numbers. If*

$$\lim_{m\to\infty} \limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \frac{\mathbf{E}\left[\xi(\mathbf{t}\cdot(A''_m \setminus A'_m))\right]}{|\mathbf{t}|} = 0,$$

then (18.7) *holds.*

Even in the setting of partial sums, there are several differences with Theorem 18.1. First, the growth parameter $\mathbf{t}$ is continuous. This allows one to treat the cases where some of the coordinates of $\mathbf{t}$ approach the axes or are constant, while all coordinates are separated from zero and grow in the setting of [3], so that the set nA increases to the whole $\mathbb{R}^d_+$ in the limit if A contains a neighborhood of the origin.

Second, we deal with signed measures rather than with sums of random variables over sets in $\mathbb{R}^d$. Even if we restrict our setting and consider a particular case where ξ is constructed in the same manner as in [3], we still are in a more general situation, since we do not impose the independence assumption on the auxiliary random variables, e.g., it is applicable to orthogonal random variables. Of course, one should be aware of appropriate conditions for the strong law of large numbers (18.4) for every particular dependence scheme. Various examples are presented in [19]. Therefore, we provide a universal method for obtaining the uniform strong law of large numbers (18.7) from (18.4).

18.4 Proof of Theorem 18.2

For $\mathbf{x} = (x_1, \dots, x_d) \in I$, we have $\mathbf{x}\cdot I = (\mathbf{0}, \mathbf{x}]$. Then

$$\lim_{|\mathbf{t}|\to\infty} \frac{\xi(\mathbf{t}\cdot A) - \mathbf{E}\,[\xi(\mathbf{t}\cdot A)]}{|\mathbf{t}|} = 0 \qquad \text{a.s.} \tag{18.8}$$

holds with $A = \mathbf{x} \cdot I$ for any fixed $\mathbf{x} \in I$, since $\mathbf{t} \cdot (\mathbf{x} \cdot I) = \mathbf{s} \cdot I$ for $\mathbf{s} = \mathbf{t} \cdot \mathbf{x} = (t_1 x_1, \dots, t_d x_d)$ and $|\mathbf{t}| \to \infty$ is equivalent to $|\mathbf{s}| \to \infty$.

Since ξ is a signed measure, condition (18.8) also holds for every set A being a difference of $\mathbf{x}_1 \cdot I$ and $\mathbf{x}_2 \cdot I$. Thus, (18.8) holds for sets A being a finite union of differences $(\mathbf{x}_1 \cdot I) \setminus (\mathbf{x}_2 \cdot I)$.

Turning to the general set $A \in \mathscr{A}$, fix $m \geq 1$ and write

$$\begin{aligned}
&\limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\xi(\mathbf{t}\cdot A) - \mathbf{E}\,[\xi(\mathbf{t}\cdot A)]}{|\mathbf{t}|} \right| \\
&\qquad \leq \limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\xi(\mathbf{t}\cdot A) - \xi(\mathbf{t}\cdot A'_m)}{|\mathbf{t}|} \right| \\
&\qquad\quad + \limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\xi(\mathbf{t}\cdot A'_m) - \mathbf{E}\left[\xi(\mathbf{t}\cdot A'_m)\right]}{|\mathbf{t}|} \right| \qquad (18.9) \\
&\qquad\quad + \limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\mathbf{E}\left[\xi(\mathbf{t}\cdot A'_m)\right] - \mathbf{E}\,[\xi(\mathbf{t}\cdot A)]}{|\mathbf{t}|} \right|.
\end{aligned}$$

Since ξ is a signed measure, $\mathbf{E}\left[\xi(\mathbf{t}\cdot A'_m)\right] - \mathbf{E}\,[\xi(\mathbf{t}\cdot A)] = -\mathbf{E}\left[\xi(\mathbf{t}\cdot (A \setminus A'_m))\right]$, hence,

$$\limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\mathbf{E}\left[\xi(\mathbf{t}\cdot A'_m)\right] - \mathbf{E}\,[\xi(\mathbf{t}\cdot A)]}{|\mathbf{t}|} \right| = \limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\mathbf{E}\left[\xi(\mathbf{t}\cdot (A \setminus A'_m))\right]}{|\mathbf{t}|} \right|.$$

Passing to the second term on the right hand side of (18.9), note that there is only a finite number of possible combinations of the cubes $C_m(\mathbf{k})$ belonging to I (this number depends on m, of course). Since A'_m is constructed from the cubes $C_m(\mathbf{k})$, there is only a finite number of possible values for A'_m if $A \in \mathscr{A}$. From the strong law of large numbers (18.8) we conclude that

$$\limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\xi(\mathbf{t}\cdot A'_m) - \mathbf{E}\left[\xi(\mathbf{t}\cdot A'_m)\right]}{|\mathbf{t}|} \right| = 0 \qquad \text{a.s.}$$

Now we proceed with the first term on the right-hand side of (18.9). Since

$$|\xi(\mathbf{t}\cdot A) - \xi(\mathbf{t}\cdot A'_m)| = |\xi(\mathbf{t}\cdot (A \setminus A'_m))| \leq \eta(\mathbf{t}\cdot (A \setminus A'_m)) \leq \eta(\mathbf{t}\cdot (A''_m \setminus A'_m)),$$

we get

$$\limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\xi(\mathbf{t}\cdot A) - \xi(\mathbf{t}\cdot A'_m)}{|\mathbf{t}|} \right|$$
$$\le \limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\eta(\mathbf{t}\cdot(A''_m \setminus A'_m)) - \mathbf{E}\left[\eta(\mathbf{t}\cdot(A''_m \setminus A'_m))\right]}{|\mathbf{t}|} \right|$$
$$+ \limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \frac{\mathbf{E}\left[\eta(\mathbf{t}\cdot(A''_m \setminus A'_m))\right]}{|\mathbf{t}|}.$$

Since η is assumed to satisfy the multiparameter strong law of large numbers, (18.8) holds for η instead of ξ and with A being a finite union of the cubes $C_m(\mathbf{k})$. The set $A''_m \setminus A'_m$ belongs to $(0, 1+(1/m)]^d$. Since only a finite number of configurations of the cubes $C_m(\mathbf{k}) \subseteq (0, 1+(1/m)]^d$ exists, the strong law of large numbers (18.8) for η implies that

$$\limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\eta(\mathbf{t}\cdot(A''_m \setminus A'_m)) - \mathbf{E}\left[\eta(\mathbf{t}\cdot(A''_m \setminus A'_m))\right]}{|\mathbf{t}|} \right| = 0 \qquad \text{a.s.}$$

Therefore,

$$\limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\xi(\mathbf{t}\cdot A) - \mathbf{E}\left[\xi(\mathbf{t}\cdot A)\right]}{|\mathbf{t}|} \right|$$
$$\le \limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \left| \frac{\mathbf{E}\left[\xi(\mathbf{t}\cdot(A \setminus A'_m))\right]}{|\mathbf{t}|} \right|$$
$$+ \limsup_{|\mathbf{t}|\to\infty} \sup_{A\in\mathscr{A}} \frac{\mathbf{E}\left[\eta(\mathbf{t}\cdot(A''_m \setminus A'_m))\right]}{|\mathbf{t}|}.$$

Passing to the limit as $m \to \infty$ and using assumptions (18.5) and (18.6), we complete the proof of the uniform strong law of large numbers (18.7).

18.5 Homogeneous Random Fields and Stationary Measures

A random signed measure ξ in $\mathbb{R}^d$ is said to be *stationary* if $\xi(\cdot)$ shares the finite-dimensional distributions with $\xi(\cdot + \mathbf{t})$ for each $\mathbf{t} \in \mathbb{R}^d$. If the first moment $\mathbf{E}\,[\xi(\cdot)]$ is finite, then the first moment measure Λ is proportional to the Lebesgue measure.

The ergodic theorem of Zygmund [31] implies that, if ξ is stationary with

$$\mathbf{E}\left[|\xi(A)|(\log^+|\xi(A)|)^{d-1}\right] < \infty \tag{18.10}$$

for all bounded Borel A, then

$$|\mathbf{t}|^{-1} \int_{[\mathbf{0},\mathbf{t}]} \xi(A + \mathbf{x}) d\mathbf{x}$$

converges almost surely as $\mathbf{t} \to \infty$ to a finite random variable, being the conditional expectation of $\xi(A)$ with respect to the invariant σ-algebra. The limit is deterministic and equals $\mathbf{E}\,[\xi(A)]$ if ξ is ergodic. Here $\log^+ z = \log(e + z)$ for $z \geq 0$. Note that all results of Sect. 18.3 can be amended for the convergence $\mathbf{t} \to \infty$ instead of $|\mathbf{t}| \to \infty$. The notation $\mathbf{t} \to \infty$ means that all coordinates of $\mathbf{t}$ tend to infinity, while $|\mathbf{t}| \to \infty$ means that at least one of them tends to infinity.

Theorem 18.3 *Let $\mathscr{A}$ be a family of Borel sets in I that satisfies* (18.2)*. Assume that ξ is a stationary ergodic random measure such that* (18.10) *holds for $A = I$. Then ξ satisfies the uniform strong law of large numbers as $\mathbf{t} \to \infty$, that is,* (18.7) *holds with $\mathbf{t} \to \infty$.*

Proof Note that

$$\int_{[\mathbf{0},\mathbf{t}]} \xi(I + \mathbf{x}) d\,\mathbf{x} = \int_{[\mathbf{0},\mathbf{t}]} \int_{I+\mathbf{x}} \xi(d\,\mathbf{u}) d\,\mathbf{x} = \int_{[\mathbf{0},\mathbf{t}+\mathbf{1}]} |(-I + \mathbf{u}) \cap [\mathbf{0}, \mathbf{t}]| \xi(d\,\mathbf{u}).$$

Further, $|(-I + \mathbf{u}) \cap [\mathbf{0}, \mathbf{t}]|$ is less than or equal to one for all $\mathbf{u} \in [\mathbf{0}, \mathbf{t} + \mathbf{1}]$ and is exactly one if $\mathbf{u} \in [\mathbf{1}, \mathbf{t}]$, whence

$$\xi([\mathbf{1}, \mathbf{t}]) \leq \int_{[\mathbf{0},\mathbf{t}]} \xi(I + \mathbf{x})\, d\mathbf{x} \leq \xi([\mathbf{0}, \mathbf{t} + \mathbf{1}]), \tag{18.11}$$

since ξ is nonnegative. If $d = 1$, then

$$\lim_{\mathbf{t}\to\infty} \frac{\xi([\mathbf{0}, \mathbf{t}]) - \xi([\mathbf{1}, \mathbf{t}])}{|\mathbf{t}|} = 0 \qquad \text{a.s.} \tag{18.12}$$

which together with (18.11) and ergodic theorem (Zygmund's theorem [31] for $d = 1$) yields for $d = 1$

$$\lim_{\mathbf{t}\to\infty} \frac{\xi(\mathbf{t} \cdot I)}{|\mathbf{t}|} = \mathbf{E}\,[\xi(I)] \qquad \text{a.s.} \tag{18.13}$$

Now let $d > 1$ and assume that (18.13) holds for all dimensions less than d. Then (18.12) holds for the dimension d. This together with (18.11) combined with Zygmund's theorem [31] yields (18.13) for the dimension d. Since ξ is stationary, (18.5) and (18.6) follow from (18.2). The result follows from a variant of Theorem 18.2 for the convergence $\mathbf{t} \to \infty$.

Let $X_{\mathbf{j}}$, $\mathbf{j} \in \mathbb{N}^d$, be a homogeneous random field, that is, $(X_{\mathbf{j}_1}, \ldots, X_{\mathbf{j}_m})$ coincides in distribution with $(X_{\mathbf{j}_1+\mathbf{s}}, \ldots, X_{\mathbf{j}_m+\mathbf{s}})$ for all $m \in \mathbb{N}$, and $\mathbf{s}, \mathbf{j}_1, \ldots, \mathbf{j}_m \in$

$\mathbb{N}^d$. For $\mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d$, denote $S_{\mathbf{n}} = S([\mathbf{0}, \mathbf{n}])$ from (18.1). In other words,

$$S_{\mathbf{n}} = \sum_{\mathbf{k} \preceq \mathbf{n}} X_{\mathbf{k}}$$

where $\preceq$ is a partial order in $\mathbb{N}^d$ defined by

$$\mathbf{k} \preceq \mathbf{n} \quad \Longleftrightarrow \quad k_1 \leq n_1, \ldots, k_d \leq n_d$$

for $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{N}^d$ and $\mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d$.

Dunford [9] proved that if

$$\mathbf{E}\left[|X_{\mathbf{j}}| \left(\log^+ |X_{\mathbf{j}}|\right)^{d-1}\right] < \infty, \tag{18.14}$$

then the limit of the averages

$$\frac{S_{\mathbf{n}}}{|\mathbf{n}|} \tag{18.15}$$

exists almost surely as $|\mathbf{n}| = n_1 \times \cdots \times n_n \to \infty$. Smythe [27] provides a probabilistic statement and proof of this result for independent identically distributed random variables $X_{\mathbf{j}}$. Etemadi [10] obtains the same result for pairwise independent identically distributed random variables. The limit of $S_{\mathbf{n}}/|\mathbf{n}|$ in the latter case coincides with the expectation $\mu = \mathbf{E}\left[X_{\mathbf{j}}\right]$. This property requires the ergodicity if random variables are not necessarily pairwise independent and identically distributed.

Note that $S(A)$ given by (18.1) is not a stationary random measure and it may be also signed, so Theorem 18.3 is not directly applicable. The following result follows from Theorem 18.2.

Corollary 18.2 *Let $\{X_{\mathbf{j}}, \mathbf{j} \in \mathbb{N}^d\}$ be a homogeneous random field and the moment condition (18.14) holds. Further let $\mathscr{A}$ be a family of subsets of the unit cube I that satisfies (18.2). If $\{X_{\mathbf{j}}\}$ is ergodic, then*

$$\lim_{|\mathbf{t}| \to \infty} \sup_{A \in \mathscr{A}} \left| \frac{S(\mathbf{t} \cdot A)}{|\mathbf{t}|} - \mu |A| \right| = 0 \qquad a.s.$$

Proof Note that the expectations in (18.5) and (18.6) are dominated by a constant times $|A''_m \setminus A'_m|$.

Remark 18.1 Condition (18.14) is necessary for the almost sure convergence of (18.15) in the case of independent identically distributed random variables.

Another particularly important family of random signed measures is generated by marked point processes. Let $N = \{(\mathbf{x}_i, m_i), i \geq 1\}$ be a point process in $\mathbb{R}^d \times \mathbb{R}$,

where the second coordinate m_i represents the mark of the point $\mathbf{x}_i$. Then

$$\xi(A) = \sum_{\mathbf{x}_i \in A} m_i \tag{18.16}$$

is a random signed measure. The process is called independently marked if the marks are i.i.d. random variables and independent of locations. The first order moment measure

$$\Lambda(A \times C) = \mathbf{E}\,[\#\{i : (\mathbf{x}_i, m_i) \in A \times C\}]$$

is the measure on Borel sets in $\mathbb{R}^d \times \mathbb{R}$, and we assume that $\Lambda(A \times \mathbb{R})$ is finite for each bounded Borel set A.

The marked point process is stationary if its distribution does not change if the locations $\mathbf{x}_i$ are all translated by any vector $\mathbf{t}$; then $\Lambda(A \times C) = \Lambda((A + \mathbf{t}) \times C)$ for all $\mathbf{t} \in \mathbb{R}^d$.

Theorem 18.4 *Assume that ξ is given by (18.16) for an ergodic independently marked point process satisfying*

$$\mathbf{E}\left[|m_1|(\log^+ |m_1|)^{d-1}\right] < \infty,$$

and the random variable $N = \operatorname{card}\{i : \mathbf{x}_i \in I\}$ is square integrable. Then (18.7) holds for any family $\mathscr{A}$ satisfying (18.2).

The proof of Theorem 18.4 is based on the following elementary upper bound for the function $x(\log x)^r$.

Lemma 18.1 *Let $r > 0$, $n \geq 1$ and $a_1, \dots, a_n \geq e^{r-1}$. Put $A_n = a_1 + \cdots + a_n$. Then*

$$A_n(\log A_n)^r \leq \sum_{i=1}^n a_i(\log a_i)^r + r\sum_{i=1}^n (A_n - a_i)(\log a_i)^{r-1}.$$

Proof It is clear that

$$A_n(\log A_n)^r = \sum_{i=1}^n a_i(\log A_n)^r = \sum_{i=1}^n a_i(\log a_i)^r + \sum_{i=1}^n a_i\left((\log A_n)^r - (\log a_i)^r\right).$$

By the mean value theorem,

$$(\log A_n)^r - (\log a_i)^r = (A_n - a_i) \cdot r\frac{(\log \xi)^{r-1}}{\xi}$$

for some $a_i \le \xi \le A_n$. Since the right hand side is a decreasing function in ξ,

$$(\log A_n)^r - (\log a_i)^r \le (A_n - a_i) \cdot r \frac{(\log a_i)^{r-1}}{a_i}.$$

Therefore,

$$A_n (\log A_n)^r \le \sum_{i=1}^{n} a_i (\log a_i)^r + r \sum_{i=1}^{n} (A_n - a_i)(\log a_i)^{r-1}.$$

Proof (of Theorem 18.4) In order to apply Theorem 18.2, we only need to show that

$$\bar{\xi}(I) = \sum_{\mathbf{x}_i \in I} |m_i|$$

satisfies (18.14). Without loss of generality, we assume that $|m_i| \ge e^{d-2}$ almost surely, since $\bar{\xi}(I) = \bar{\xi}_1(I) + \bar{\xi}_2(I)$, where $\bar{\xi}_1(I)$ and $\bar{\xi}_2(I)$ are constructed from $m_i \mathbb{1}_{\{|m_i| < e^{d-2}\}}$ and $m_i \mathbb{1}_{\{|m_i| \ge e^{d-2}\}}$, respectively, with $m_i \mathbb{1}_{\{|m_i| < e^{d-2}\}}$ being bounded. By Lemma 18.1 with $r = d - 1$, and $a_i = |m_i|$

$$\mathbf{E}\left[A_N (\log A_N)^{d-1}\right] \le \mathbf{E}\left[\sum_{i=1}^{N} a_i (\log a_i)^r\right] + (d-1)\mathbf{E}\left[\sum_{i=1}^{N} (A_N - a_i)(\log a_i)^{r-1}\right].$$

Since N and $\{m_i\}$ are independent, Wald's equality implies

$$\mathbf{E}\left[\sum_{i=1}^{N} a_i (\log a_i)^r\right] = \mathbf{E}\,[N] \cdot \mathbf{E}\left[a_i (\log a_i)^r\right].$$

The total expectation formula yields that

$$\begin{aligned}
\mathbf{E}\left[\sum_{i=1}^{N} (A_N - a_i)(\log a_i)^{r-1}\right] &= \sum_{n=1}^{\infty} \mathbf{P}(N = n)\mathbf{E}\left[\sum_{i=1}^{n} (A_n - a_i)(\log a_i)^{r-1}\right] \\
&= \sum_{n=1}^{\infty} \mathbf{P}(N = n) \sum_{i=1}^{n} \mathbf{E}\left[(A_n - a_i)(\log a_i)^{r-1}\right] \\
&= \sum_{n=1}^{\infty} \mathbf{P}(N = n) \sum_{i=1}^{n} \mathbf{E}\,[(A_n - a_i)] \cdot \mathbf{E}\left[(\log a_i)^{r-1}\right]
\end{aligned}$$

$$= \sum_{n=1}^{\infty} \mathbf{P}(N = n) n(n-1) \mathbf{E}\,[|m_1|] \cdot \mathbf{E}\left[(\log|m_1|)^{r-1}\right]$$

$$= \mathbf{E}\,[|m_1|] \cdot \mathbf{E}\left[(\log|m_1|)^{r-1}\right] \mathbf{E}\,[N(N-1)]\,.$$

This together with the latter bound proves the desired result.

18.6 Stochastic Integrals

Stochastic integrals with respect to the Brownian sheet have been intensively studied since the 70s. Treated as signed measures, stochastic integrals fit very well the framework of Theorem 18.2. Although the construction of stochastic integrals can be done for any dimension, we restrict ourselves to the case of $d = 2$ as in [6].

Let W be a white noise in the plane, that is a finitely additive set function defined on the Borel subsets of $\mathbb{R}^2_+$ such that $W(A)$ is a normal random variable with parameters 0 and $|A|$ and $W(A)$ and $W(B)$ are independent for disjoint Borel subsets A and B.

If R_{st} denotes the rectangle $[0,s] \times [0,t]$, then $W_{st} = W(R_{st})$ is called the Brownian sheet. By $\mathscr{F}_{st}$, $(s,t) \in \mathbb{R}^2_+$, we denote the σ-algebra generated by the random variables W_{uv}, $(u,v) \preceq (s,t)$.

Let A be a closed rectangle with the lower left-hand corner z_0. Introduce the function ϕ_z, $z \in \mathbb{R}^2_+$, as follows

$$\phi_z = \phi_0 \mathbb{1}_A(z), \qquad z \in \mathbb{R}^2_+, \tag{18.17}$$

where ϕ_0 is a $\mathscr{F}_{z0}$ measurable random variable. Then, by definition,

$$\int_{R_z} \phi \, dW = \phi_0 W(A \cap R_z), \qquad z \in \mathbb{R}^2_+.$$

The integral is extended by linearity to simple ϕ, i.e. to finite linear combinations of "step" functions of the form (18.17). In general, let ϕ be such that

(a) ϕ_z is $\mathscr{F}_z$-measurable,
(b) $(z,\omega) \mapsto \phi_z(\omega)$, $z \in \mathbb{R}^2_+$, $\omega \in \Omega$, is $\mathscr{B} \times \mathscr{F}$-measurable where $\mathscr{B}$ is the family of Borel subsets in the plane and $\mathscr{F}$ is the σ-algebra of the probability space $(\Omega, \mathscr{F}, \mathbf{P})$, and
(c) $\int_{R_z} \mathbf{E}\left[\phi_\zeta^2\right] d\zeta < \infty$ for all $z \in \mathbb{R}^2_+$.

Then one can find a sequence of simple random functions $\{\phi_n\}$ for which

$$\lim_{n\to\infty} \int_{R_z} \mathbf{E}\left[(\phi_n - \phi)^2\right] d\zeta = 0 \qquad \text{for all } z \in \mathbb{R}^2_+.$$

The integrals $\int_{R_z} \phi_n \, dW$ converge in the mean square sense, the limit is denoted by $\int_{R_z} \phi \, dW$. The defined integral is

- continuous as a function of z,
- a two-parameter martingale, and
- for all $z \in \mathbb{R}^2_+$,

$$\mathbf{E}\left[\left(\int_{R_z} \phi \, dW\right)^2\right] = \int_{R_z} \mathbf{E}\left[\phi_\zeta^2\right] \, d\zeta. \tag{18.18}$$

Recall the three defining properties for an arbitrary two-parameter martingale M_z, $z \in \mathbb{R}^2_+$, with respect to the family of σ-algebras $\mathscr{F}_z$, $z \in \mathbb{R}^2_+$ (see [5]):

(I) $\mathbf{E}\,[|M_z|] < \infty$ for all $z \in \mathbb{R}^2_+$;
(II) M_z is $\mathscr{F}_z$-measurable;
(III) if $z \preceq z'$, then $\mathbf{E}\left[M_{z'} \middle| \mathscr{F}_z\right] = M_z$.

The final step in the construction of the integral is to pass to a general bounded Borel set A by letting

$$\int_A \phi \, dW = \int_R \mathbb{1}_A \, \phi \, dW,$$

where R is a rectangle containing A. Then, for each fixed ϕ,

$$\xi(A) = \int_A \phi \, dW$$

is a signed measure.

Now we define the two-parameter discrete time martingale associated with the stochastic integral. For $(m, n) \in \mathbb{N}^2$, define $r_{mn} = R_{mn} \setminus (R_{m-1,n} \cup R_{m,n-1})$ and put

$$X_{mn} = \int_{r_{mn}} \phi \, dW, \qquad S_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij} = \int_{R_{mn}} \phi \, dW.$$

Then S_{mn} is a two-parameter discrete time martingale with respect to the family of σ-algebras $\mathscr{F}_{mn}$. It follows from [20] that if

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{\mathbf{E}\left[X_{mn}^2\right]}{(mn)^2} < \infty, \tag{18.19}$$

then the strong law of large numbers holds for $\{S_{mn}\}$,

$$\lim_{mn\to\infty} \frac{S_{mn}}{mn} = 0 \qquad \text{a.s.} \tag{18.20}$$

In view of (18.18), condition (18.19) is equivalent to

$$\int_1^\infty \int_1^\infty \frac{\mathbf{E}\left[\phi_{st}^2\right]}{(st)^2}\, ds\, dt < \infty. \tag{18.21}$$

The strong law of large numbers (18.20) is easily extended to the continuous time limit result by using the Cairoli maximal inequality [5]. Thus (18.4) holds with the signed measure $\xi(A) = \int_A \phi\, dW$. Now Theorem 18.2 implies the following corollary.

Corollary 18.3 *Let ϕ_z, $z \in \mathbb{R}_+^2$, satisfy conditions (a)–(c). Let $\mathscr{A}$ be a family of subsets of the square $[0,1] \times [0,1]$ such that conditions (18.5) and (18.6) hold. Then (18.21) implies*

$$\lim_{\substack{st\to\infty \\ s\ge 1,\ t\ge 1}} \sup_{A\in\mathscr{A}} \left| \frac{1}{st} \int_{st\cdot A} \phi\, dW \right| = 0 \qquad a.s.$$

18.7 Concluding Remarks

The assumptions for the uniform strong law of large numbers imposed on the family $\mathscr{A}$ (either (18.2) in Theorem 18.1 or (18.5)–(18.6) in Theorem 18.2) do not involve any entropy type restriction needed for both the central limit theorem [1] and law of the iterated logarithm [2]. For the both latter results, one needs to assume that the entropy is integrable, that is

$$\int_0^1 \sqrt{\frac{H(u)}{u}}\, du < \infty,$$

where $H(u)$ is the entropy of the family $\mathscr{A}$ being the logarithm of the cardinality of a minimal u-net.

Krengel and Pyke [21] provide the strong law of large numbers for multiparameter subadditive processes rather than for signed measures as in our Theorem 18.2. It is worthwhile mentioning that they do not get a uniform version. Liu, Rio, and Rouault [23] treat the uniform strong law of large numbers for random measures, which is a partial case of signed measures with a one-dimensional growth parameter. A version of Theorem 18.1 for random product measures is considered by Kil and Kwon [18]. Jang and Kwon [16] obtain a generalization of Theorem 18.1 for fuzzy random variables. Bing [4] extends Theorem 18.1 for the α-mixing case. Note that

this result follows from Theorem 18.2 by referring to the usual strong law of large numbers available in this case. Ziegler [30] investigates the uniform law of large numbers for triangular arrays extending Theorem 18.1 to the case of non-identically distributed random variables.

Considering random sets as measurable mappings from a probability space into the set of compact convex subsets of a Banach space, Jang and Kwon [16] prove a uniform strong law of large numbers for sequences of independent and identically distributed random sets, which is another direct generalization of Theorem 18.1.

Under a mild assumption, Giné and Zinn [12] show that condition (18.2) is necessary and sufficient for the uniform strong law of large numbers (18.3) if $\mu = 0$ (see also Hong and Kwon [13]). However, the case of $\mu \neq 0$ is different.

Ivanoff [14] discusses the uniform strong law of large numbers in connection to possible generalizations of the definitions of a stochastic process indexed by $\mathbb{R}_+$ to processes indexed by a multidimensional time parameter or a class of sets.

Müller and Song [24] apply the uniform strong law of large numbers for partial-sum process to investigate the problem of edge estimation in a two-region image in the setting of a fixed design regression model. Terán and López-Díaz [28] use Theorem 18.1 to study some aspects of the approximation of mappings taking values in a special class of upper semicontinuous functions and to obtain some Korovkin type theorems for positive linear operators.

Acknowledgements The authors are grateful to Andrii Ilienko for criticism and valuable comments that allow them to fill some gaps in the preliminary version of the manuscript.

This research was supported by the Swiss National Science Foundation Grant IZ7320_152292.

O.I. Klesov was supported by the grant 0118U003614 from Ministry of Education and Science of Ukraine (project N 2105 Φ).

References

1. Alexander, K.S.: Central limit theorems for stochastic processes under random entropy conditions. Probab. Theory Relat. Fields **75**, 351–378 (1987)
2. Bass, R.F., Pyke, R.: Functional law of the iterated logarithm and uniform central limit theorem for partial-sum processes indexed by sets. Ann. Probab. **12**, 13–34 (1984)
3. Bass, R.F., Pyke, R.: A strong law of large numbers for partial-sum processes indexed by sets. Ann. Probab. **12**, 268–271 (1984)
4. Bing, X.: Strong laws for α-mixing sequence processes indexed by sets. Appl. Math. J. Chinese Univ. **10**, 45–48 (1995)
5. Cairoli, R.: Une inégalité pour martingales à indices multiples et ses applications. In: Séminaire de Probabilités, IV (Univ. Strasbourg, 1968/69). Lecture Notes in Mathematics, vol. 124, pp. 1–27. Springer, Berlin (1970)
6. Cairoli, R., Walsh, J.B.: Stochastic integrals in the plane. Acta Math. **134**, 111–183 (1975)
7. Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods, 2nd edn. Springer, New York (2003)
8. Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure, 2nd edn. Springer, New York (2008)

9. Dunford, N.: An individual ergodic theorem for non-commutative transformations. Acta Sci. Math. Szeged **14**, 1–4 (1951)
10. Etemadi, N.: An elementary proof of the strong law of large numbers. Z. Wahrscheinlichkeitstheorie verw. Gebiete **55**, 119–122 (1981)
11. Gaenssler, P., Ziegler, K.: On function-indexed partial-sum processes. In: Probability Theory and Mathematical Statistics (Vilnius, 1993), pp. 285–311. TEV, Vilnius (1994)
12. Giné, E., Zinn, J.: The law of large numbers for partial sum processes indexed by sets. Ann. Probab. **15**, 154–163 (1987)
13. Hong, D.H., Kwon, J.S.: Laws of large numbers for products of some measures and partial sum processes indexed by sets. J. Korean Math. Soc. **30**, 79–91 (1993)
14. Ivanoff, B.G.: Set-indexed processes: distributions and weak convergence. In: Topics in Spatial Stochastic Processes (Martina Franca, 2001). Lecture Notes in Mathematics, vol. 1802, pp. 85–125. Springer, Berlin (2003)
15. Ivanoff, G., Merzbach, E.: Set-Indexed Martingales. Chapman & Hall/CRC, Boca Raton (2000)
16. Jang, L.-C., Kwon, J.-S.: A uniform strong law of large numbers for partial sum processes of fuzzy random variables indexed by sets. Fuzzy Set. Syst. **99**, 97–103 (1998)
17. Kallenberg, O.: Random Measures, Theory and Applications. Springer, Berlin (2017)
18. Kil, B.M., Kwon, J.S.: A uniform law of large numbers for product random measures. Bull. Korean Math. Soc. **32**, 221–231 (1995)
19. Klesov, O.: Limit Theorems for Multi-indexed Sums of Random Variables, vol. 71. Springer, Heidelberg (2014)
20. Klesov, O.I.: The Hájek-Rényi inequality for random fields and the strong law of large numbers. Teor. Veroyatnost. Mat. Statist. **22**, 58–66, 163 (1980)
21. Krengel, U., Pyke, R.: Uniform pointwise ergodic theorems for classes of averaging sets and multiparameter subadditive processes. Stoch. Process. Appl. **26**, 289–296 (1987)
22. Kuelbs, J.: The invariance principle for a lattice of random variables. Ann. Math. Stat. **39**, 382–389 (1968)
23. Liu, Q., Rio, E., Rouault, A.: Limit theorems for multiplicative processes. J. Theor. Probab. **16**, 971–1014 (2003)
24. Müller, H.-G., Song, K.-S.: A set-indexed process in a two-region image. Stoch. Process. Appl. **62**, 87–101 (1996)
25. Pyke, R.: Asymptotic results for empirical and partial-sum processes: a review. Canad. J. Statist. **12**, 241–264, 283–287 (1984)
26. Shorack, G.R., Wellner, J.A.: Empirical Processes with Applications to Statistics. Wiley, New York (1986)
27. Smythe, R.T.: Strong laws of large numbers for r-dimensional arrays of random variables. Ann. Probab. **1**, 164–170 (1973)
28. Terán, P., López-Díaz, M.: Approximation of mappings with values which are upper semicontinuous functions. J. Approx. Theory **113**, 245–265 (2001)
29. Wichura, M.J.: Some Strassen-type laws of the iterated logarithm for multiparameter stochastic processes with independent increments. Ann. Probab. **1**, 272–296 (1973)
30. Ziegler, K.: Uniform laws of large numbers for triangular arrays of function-indexed processes under random entropy conditions. Results Math. **39**, 374–389 (2001)
31. Zygmund, A.: An individual ergodic theorem for non-commutative transformations. Acta Sci. Math. Szeged **14**, 103–110 (1951)

Chapter 19
On Comparison Results for Neutral Stochastic Differential Equations of Reaction-Diffusion Type in $L_2(\mathbb{R}^d)$

Oleksandr M. Stanzhytskyi, Viktoria V. Mogilova, and Alisa O. Tsukanova

Abstract In the present paper, we establish a comparison result for solutions to the Cauchy problems for two stochastic integro-differential equations of reaction-diffusion type with delay. On this subject number of authors have obtained their comparison results. We deal with the Cauchy problems for two stochastic integro-differential equations of reaction-diffusion type with delay. Except drift and diffusion coefficients, our equations include also one integro-differential term. Basic difference of our case from the case of all earlier investigated problems is presence of this term. Presence of this term turns this equation into a nonlocal neutral stochastic equation of reaction-diffusion type. Nonlocal deterministic equations of this type are well known in literature and have wide range of applications. Such equations arise, for instance, in mechanics, electromagnetic theory, heat flow, nuclear reactor dynamics, and population dynamics. These equations are used in modeling of phytoplankton growth, distant interactions in epidemic models and nonlocal consumption of resources. We introduce a concept of mild solutions to our problems and state and prove a comparison theorem for them. According to our result, under certain assumptions on coefficients of equations under consideration, their solutions depend on the transient coefficients in a monotone way.

O. M. Stanzhytskyi · A. O. Tsukanova (✉)
Taras Shevchenko National University of Kiev, Kiev, Ukraine

V. V. Mogilova
National Technical University of Ukraine "Igor Sikorsky Kiev Polytechnic Institute", Kiev, Ukraine

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_19

19.1 Introduction

In the given paper we study the following Cauchy problems for neutral partial stochastic integro-differential equations of reaction-diffusion type

$$d\Big(u_i(t,x)+\int\limits_{\mathbb{R}^d} b_i(t,x,u_i(t-r,\xi),\xi)d\xi\Big)=\big(\Delta_x u_i(t,x)+f_i(t,u_i(t-r,x),x)\big)dt$$

$$+\sigma(t,x)dW(t,x),\ 0<t\le T,\ x\in\mathbb{R}^d,\ i\in\{1,2\},\tag{19.1}$$

$$u_i(t,x)=\phi_i(t,x),\ -r\le t\le 0,\ x\in\mathbb{R}^d,\ r>0,\ i\in\{1,2\},\tag{19.2}$$

where $T>0$ is fixed, $\Delta_x\equiv\sum\limits_{i=1}^{d}\frac{\partial^2}{\partial x_i^2}$ is d-measurable Laplacian in the space variables, W is a Q-Wiener process, $f_i\colon[0,T]\times\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}$, $i\in\{1,2\}$, $\sigma\colon[0,T]\times\mathbb{R}^d\to\mathbb{R}$ and $b_i\colon[0,T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}$, $i\in\{1,2\}$, are some given functions to be specified later, $\phi_i\colon[-r,0]\times\mathbb{R}^d\to\mathbb{R}$, $i\in\{1,2\}$, are initial-datum functions. For solutions u_1 and u_2 of these problems we prove a comparison theorem. According to our result, if $f_1\ge f_2$, then $u_1\ge u_2$ with probability one.

A problem of comparison of solutions to stochastic differential equations in finite-dimensional case has firstly arised in [10]. A comparison theorem for equation of the form $d\xi(t)=f(t,\xi(t))dt+\sigma(t,\xi(t))d\beta(t)$, where β is standard one-dimensional Brownian motion, has been obtained in this work. According to this theorem, under certain assumptions, a solution of the equation above is monotonously non-decreasing function from "drift" coefficient f. A more general presentation of the comparison theorem is given in [11, 12]. Variations of the result from [10] have been the proposed in [2, 3, 5, 6, 8, 9, 13]. In [4] this theorem for solutions to stochastic differential equations with a multidimensional Wiener process and stochastic partial differential equations has been obtained. In [7] a comparison result for solutions to the Cauchy problem for stochastic differential equations with a Q-Wiener processes in Hilbert space is presented. The main goal of the given work is to prove a comparison theorem for solutions of problem (19.1)–(19.2), using the idea from this work. This result plays an important role when studying the existence of solutions to the Cauchy problem for stochastic differential equations of reaction-diffusion type with non-Lipschitz conditions on "drift" coefficients.

This paper is organised as follows. Firstly, in Sect. 19.2, we introduce a statement of the problem and formulate our main result. Then we represent a few necessary facts, needed for the treatment in the subsequent sections. These auxiliary results of independent interest are gathered in Sect. 19.3. Section 19.4 is devoted to the proof of the main theorem.

19.2 Problem Definition

Throughout the paper let $(\Omega, \mathscr{F}, \mathbf{P})$ be a complete probability space, $L_2(\mathbb{R}^d)$ denotes real Hilbert space with the norm $\|g\|_{L_2(\mathbb{R}^d)} = \left(\int\limits_{\mathbb{R}^d} g^2(x)dx \right)^{\frac{1}{2}}$. Let $\{e_n(x), n \in \{1, 2, \dots\}\}$ be an orthonormal basis on $L_2(\mathbb{R}^d)$ such that

$$\sup_{n\in\{1,2,\dots\}} ess\sup_{x\in\mathbb{R}^d} |e_n(x)| \le 1.$$

We now define $L_2(\mathbb{R}^d)$-valued Q-Wiener process $W(t, x) = W(t, \cdot), t \ge 0, x \in \mathbb{R}^d$, as follows

$$W(t, \cdot) = \sum_{n=1}^{\infty} \sqrt{\lambda_n} e_n(\cdot)\beta_n(t), t \ge 0,$$

where $\{\beta_n(t), n \in \{1, 2, \dots\}\} \subset \mathbb{R}$ are independent standard one-dimensional Brownian motions on $t \ge 0$, $\{\lambda_n, n \in \{1, 2, \dots\}\}$ is a sequence of positive numbers such that $\lambda = \sum\limits_{n=1}^{\infty} \lambda_n < \infty$. Let $\{\mathscr{F}_t, t \ge 0\}$ be a normal filtration on $\mathscr{F}$. We assume that $W(t, \cdot), t \ge 0$, is a Q-Wiener process with respect to a filtration $\{\mathscr{F}_t, t \ge 0\}$, i.e.,

- $W(t, \cdot), t \ge 0$, is $\mathscr{F}_t$-measurable;
- the increments $W(t + h, \cdot) - W(t, \cdot)$ are independent of $\mathscr{F}_t$ for all $h > 0$ and $t \ge 0$.

Let the following conditions be true

(1) $f_i : [0, T]\times\mathbb{R}\times\mathbb{R}^d \to \mathbb{R}, i \in \{1, 2\}, \sigma : [0, T]\times\mathbb{R}^d \to \mathbb{R}, b_i : [0, T]\times\mathbb{R}^d\times\mathbb{R}\times \times \mathbb{R}^d \to \mathbb{R}, i \in \{1, 2\}$, are measurable functions with respect to all of their variables.

(2) The initial-data functions $\phi_i(t, x, \omega) : [-r, 0] \times \mathbb{R}^d \times \Omega \to L_2(\mathbb{R}^d), i \in \{1, 2\}$, are $\mathscr{F}_0$-measurable random functions, independent of $W(t, x), t \ge 0, x \in \mathbb{R}^d$, with almost surely continuous paths and such that

$$\sup_{-r\le t\le 0} \mathbf{E}\|\phi_i(t, \cdot)\|^2_{L_2(\mathbb{R}^d)} < \infty, i \in \{1, 2\}.$$

(3) $b_i, i \in \{1, 2\}$, are uniformly continuous in the first argument and satisfy the Lipschitz condition in the third argument of the form

$$|b_i(t, x, u, \xi) - b_i(t, x, v, \xi)| \le l(t, x, \xi)|u - v|,$$
$$0 \le t \le T, \{x, \xi\} \subset \mathbb{R}^d, \{u, v\} \subset \mathbb{R}, i \in \{1, 2\},$$

where $l\colon [0,T]\times\mathbb{R}^d\times\mathbb{R}^d\to[0,\infty)$ is such that

$$\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\sqrt{\int\limits_{\mathbb{R}^d} l^2(t,x,\xi)d\xi dx}<\infty,$$

$$\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(t,x,\xi)d\xi dx<\frac{1}{4}. \tag{19.3}$$

(4) There exists a function $\chi\colon\mathbb{R}^d\times\mathbb{R}^d\to[0,\infty)$, satisfying the following conditions

$$\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\chi(x,\xi)d\xi dx<\infty,$$

$$\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\chi(x,\xi)d\xi\right)^2 dx<\infty,$$

such that

$$\sup_{0\le t\le T}|b_i(t,x,0,\xi)|\le\chi(x,\xi),\ 0\le t\le T,\ \{x,\xi\}\subset\mathbb{R}^d,\ i\in\{1,2\}. \tag{19.4}$$

(5) There exists a function $\eta\colon[0,T]\times\mathbb{R}^d\to[0,\infty)$ with

$$\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\eta^2(t,x)dx<\infty,$$

such that the following linear-growth and Lipschitz conditions are valid for f_i, $i\in\{1,2\}$,

$$|f_i(t,u,x)|\le\eta(t,x)+L|u|,\ 0\le t\le T,\ u\in\mathbb{R},\ x\in\mathbb{R}^d,\ i\in\{1,2\}, \tag{19.5}$$

$$|f_i(t,u,x)-f_i(t,v,x)|\le L|u-v|,\ 0\le t\le T,\ u\in\mathbb{R},\ x\in\mathbb{R}^d,\ i\in\{1,2\}.$$

(6) The next condition holds true for σ

$$\sup_{0\le t\le T}\|\sigma(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}<\infty.$$

(7) For $\nabla_x b_i$ and $D_x^2 b_i$, $i \in \{1,2\}$, the following linear-growth condition with respect to the third argument is true

$$|\nabla_x b_i(t,x,u,\xi)| + \|D_x^2 b_i(t,x,u,\xi)\| \le \psi(t,x,\xi)\big(1+|u|\big),$$
$$0 \le t \le T, \{x,\xi\} \subset \mathbb{R}^d, u \in \mathbb{R}, i \in \{1,2\},$$

and for $D_x^2 b_i$, $i \in \{1,2\}$, – the following Lipschitz condition

$$\|D_x^2 b_i(t,x,u,\xi) - D_x^2 b_i(t,x,v,\xi)\| \le \psi(t,x,\xi)|u-v|,$$
$$0 \le t \le T, \{x,\xi\} \subset \mathbb{R}^d, \{u,v\} \subset \mathbb{R}, i \in \{1,2\}, \tag{19.6}$$

where $\psi : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty)$ is such that

$$\sup_{0\le t\le T} \int\limits_{\mathbb{R}^d} \left(\int\limits_{\mathbb{R}^d} \psi(t,x,\xi) d\xi \right)^2 dx < \infty,$$

$$\sup_{0\le t\le T} \int\limits_{\mathbb{R}^d} \int\limits_{\mathbb{R}^d} \psi^2(t,x,\xi) d\xi dx < \infty,$$

and besides for any point $x_0 \in \mathbb{R}^d$ there is its neighborhood $B_\delta(x_0)$ and a nonnegative function φ such that

$$\sup_{0\le t\le T} \varphi(t,\cdot,x_0,\delta) \in L_2(\mathbb{R}^d) \cap L_1(\mathbb{R}^d), \delta > 0,$$

$$|\psi(t,x,\xi) - \psi(t,x_0,\xi)| \le \varphi(t,\xi,x_0,\delta)|x-x_0|,$$
$$0 \le t \le T, |x-x_0| < \delta, \xi \in \mathbb{R}^d.$$

Next, we introduce the notion of a mild solution to the problem (19.1)–(19.2).

Remark 19.1 From now on we use the notation $S(t)g(\cdot)$, $g \in L_2(\mathbb{R}^d)$, to denote the convolution

$$\big(S(t)g(\cdot)\big)(x) = \int\limits_{\mathbb{R}^d} \mathscr{K}(t,x-\xi) g(\cdot) d\xi, x \in \mathbb{R}^d, g \in L_2(\mathbb{R}^d).$$

It is known from semi-group theory that

$$\|(S(t)g(\cdot))(x)\|^2_{L_2(\mathbb{R}^d)} \le \|g(x)\|^2_{L_2(\mathbb{R}^d)}.$$

Here

$$\mathscr{K}(t,x)=\begin{cases}\frac{1}{(4\pi t)^{\frac{d}{2}}}\exp\big\{-\frac{|x|^2}{4t}\big\}, & t>0,\ x\in\mathbb{R}^d,\\ 0, & t<0,\ x\in\mathbb{R}^d,\end{cases}$$

denotes the fundamental solution (source function, diffusion kernel) of the heat equation.

For convenience denote $u\equiv u_i, \phi\equiv\phi_i, b\equiv b_i, f\equiv f_i, i\in\{1,2\}$.

Definition 19.1 A continuous random process $u(t,x,\omega)\colon[-r,T]\times\mathbb{R}^d\times\Omega\to L_2(\mathbb{R}^d)$ is called a **mild solution (solution)** to (19.1)–(19.2) provided

1. It is $\mathscr{F}_t$-measurable for almost all $-r\le t\le T$.
2. It satisfies the integral equation

$$\begin{aligned}u(t,x)=&\int_{\mathbb{R}^d}\mathscr{K}(t,x-\xi)\Big(\phi(0,\xi)+\int_{\mathbb{R}^d}b(0,\xi,\phi(-r,\zeta),\zeta)d\zeta\Big)d\xi\\&-\int_{\mathbb{R}^d}b(t,x,u(t-r,\xi),\xi)d\xi\\&-\int_0^t\Big(\Delta_x\int_{\mathbb{R}^d}\mathscr{K}(t-s,x-\xi)\Big(\int_{\mathbb{R}^d}b(s,\xi,u(s-r,\zeta),\zeta)d\zeta\Big)d\xi\Big)ds\\&+\int_0^t\int_{\mathbb{R}^d}\mathscr{K}(t-s,x-\xi)f(s,u(s-r,\xi),\xi)d\xi ds\\&+\int_0^t\sum_{n=1}^{\infty}\sqrt{\lambda_n}\Big(\int_{\mathbb{R}^d}\mathscr{K}(t-s,x-\xi)\sigma(s,\xi)e_n(\xi)d\xi\Big)d\beta_n(s),\\&0<t\le T,\ x\in\mathbb{R}^d,\end{aligned}\tag{19.7}$$

$$u(t,x)=\phi(t,x),\ -r\le t\le 0,\ x\in\mathbb{R}^d,\ r>0.\tag{19.8}$$

3. It satisfies the condition

$$\mathbf{E}\int_0^T\|u(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}dt<\infty.$$

Remark 19.2 It is assumed in the definition above that all the integrals from (19.7) are well defined.

The following is the comparison theorem.

Theorem 19.1 (Comparison Theorem) *Suppose assumptions* **(1)**–**(7)** *are satisfied. Let*

1) the initial-datum functions satisfy the condition

$$\phi_1(t,x) \geq \phi_2(t,x),\ 0 \leq t \leq T,\ x \in \mathbb{R}^d;$$

2) the functions b_i, $i \in \{1,2\}$, *satisfy the conditions*

$$b_1(0,x,\phi_2(-r,\xi),\xi) = b_2(0,x,\phi_2(-r,\xi),\xi),\ \{x,\xi\} \subset \mathbb{R}^d,$$
$$b_1(0,x,\phi_1(-r,\xi),\xi) = b_2(0,x,\phi_1(-r,\xi),\xi),\ \{x,\xi\} \subset \mathbb{R}^d,$$
$$b_1(0,x,\phi_1(-r,\xi),\xi) = b_1(0,x,\phi_2(-r,\xi),\xi),\ \{x,\xi\} \subset \mathbb{R}^d,$$
$$b_1(t,x,u,\xi) \leq b_2(t,x,u,\xi),\ 0 \leq t \leq T,\ \{x,\xi\} \subset \mathbb{R}^d,\ u \in \mathbb{R};$$

3) the functions f_i, $i \in \{1,2\}$, *satisfy the conditions*

$$f_1(t,u,x) \geq f_2(t,u,x),\ 0 \leq t \leq T,\ u \in \mathbb{R},\ x \in \mathbb{R}^d.$$

Let one of the following conditions be true

M1) b_1 *is monotonously non-increasing,* f_1 *is monotonously non-decreasing with respect to* u*;*

M2) b_2 *is monotonously non-increasing,* f_2 *is monotonously non-decreasing with respect to* u.

Then for all $0 \leq t \leq T$ *the solutions of* (19.1)–(19.2) *satisfy the inequality*

$$u_1(t,x) \geq u_2(t,x),\ x \in \mathbb{R}^d,$$

with probability one.

19.3 Preliminaries

This section is the toolbox of the results that will be used in the proof of Theorem 19.1.

19.3.1 Comparison Theorem for Finite-Dimensional Case

In order to prove our main result we need a finite-dimensional comparison theorem for the following Cauchy problems for two neutral stochastic integro-differential equations

$$d\left(u_i(t,x)+\int\limits_{\mathbb{R}^d} b_i(t,x,u_i(\alpha(t),\xi),\xi)d\xi\right)=f_i(t,u_i(\alpha(t),x),x)dt$$

$$+\sigma(t,x)d\beta(t),\ 0<t\le T,\ x\in\mathbb{R}^d,\ i\in\{1,2\},\tag{19.9}$$

$$u_i(t,x)=\phi_i(t,x),\ -r\le t\le 0,\ x\in\mathbb{R}^d,\ r>0,\ i\in\{1,2\},\tag{19.10}$$

where β is one-dimensional real-valued Brownian motion, $\alpha\colon[0,T]\to[-r,\infty)$ is a delay function.

Concerning coefficients of this problem we impose the following conditions

(1) $\alpha\colon[0,T]\to[-r,\infty)$ belongs to $C^1([0,T])$ with $\alpha'\ge 1$, $\alpha(t)\le t$;

(2) $f_i\colon[0,T]\times\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}, i\in\{1,2\}, \sigma\colon[0,T]\times\mathbb{R}^d\to\mathbb{R}, b_i\colon[0,T]\times\mathbb{R}^d\times\mathbb{R}\times\times\mathbb{R}^d\to\mathbb{R}$, $i\in\{1,2\}$, are measurable with respect to all of their variables functions;

(3) the initial-datum functions $\phi_i(t,x,\omega)\colon[-r,0]\times\mathbb{R}^d\times\Omega\to L_2(\mathbb{R}^d), i\in\{1,2\}$, are $\mathscr{F}_0$-measurable random functions and such that

$$\sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}<\infty, i\in\{1,2\};$$

(4) $b_i, i\in\{1,2\}$, satisfy the Lipshitz condition in the third argument of the form

$$|b_i(t,x,u,\xi)-b_i(t,x,v,\xi)|\le l(t,x,\xi)|u-v|,$$

$$0\le t\le T,\{x,\xi\}\subset\mathbb{R}^d,\{u,v\}\subset\mathbb{R}, i\in\{1,2\},$$

where $l\colon[0,T]\times\mathbb{R}^d\times\mathbb{R}^d\to[0,\infty)$ is such that

$$\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(t,x,\xi)d\xi dx<\frac{1}{4};$$

(5) there exists a function $\chi:\mathbb{R}^d\times\mathbb{R}^d\to[0,\infty)$, satisfying the following condition

$$\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\chi(x,\xi)d\xi\right)^2 dx<\infty,$$

such that

$$\sup_{0\le t\le T}|b_i(t,x,0,\xi)|\le\chi(x,\xi),\ 0\le t\le T,\ \{x,\xi\}\subset\mathbb{R}^d,\ i\in\{1,2\}.$$

For notational simplicity denote $u\equiv u_i$, $\phi\equiv\phi_i$, $b\equiv b_i$, $f\equiv f_i$, $i\in\{1,2\}$.

Definition 19.2 A continuous random process $u(t,x,\omega)\colon[-r,T]\times\mathbb{R}^d\times\Omega\to$ $\to L_2(\mathbb{R}^d)$ is called a **solution** to (19.9)–(19.10) provided

1. It is $\mathscr{F}_t$-measurable for almost all $-r\le t\le T$.
2. It satisfies the following integral equation

$$u(t,x)=\phi(0,x)+\int_{\mathbb{R}^d}b(0,x,\phi(-r,\xi),\xi)d\xi-\int_{\mathbb{R}^d}b(t,x,u(\alpha(t),\xi),\xi)d\xi$$
$$+\int_0^t f(s,u(\alpha(s),x),x)ds+\int_0^t\sigma(s,x)d\beta(s),\ 0\le t\le T,\ x\in\mathbb{R}^d, \tag{19.11}$$

$$u(t,x)=\phi(t,x),\ -r\le t\le 0,\ x\in\mathbb{R}^d,\ r>0. \tag{19.12}$$

3. It satisfies the condition

$$\mathbf{E}\int_0^T\|u(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}dt<\infty.$$

Earlier we have stated and proved the following two theorems.

Theorem 19.2 (Existence Theorem) *Suppose assumptions* **(1)–(5)** *and conditions* **(5)**, **(6)** *from Sect. 19.2 are valid. Then* (19.11)–(19.12) *has a unique solution.*

Theorem 19.3 (Finite-Dimensional Comparison Theorem) *Suppose conditions of existence theorem above are valid and*

1) the initial-datum functions satisfy the condition

$$\phi_1(t,x)\ge\phi_2(t,x),\ 0\le t\le T,\ x\in\mathbb{R}^d;$$

2) the functions b_i, $i\in\{1,2\}$, satisfy the conditions

$$b_1(0,x,\phi_2(-r,\xi),\xi)=b_2(0,x,\phi_2(-r,\xi),\xi),\ \{x,\xi\}\subset\mathbb{R}^d,$$
$$b_1(0,x,\phi_1(-r,\xi),\xi)=b_2(0,x,\phi_1(-r,\xi),\xi),\ \{x,\xi\}\subset\mathbb{R}^d,$$
$$b_1(0,x,\phi_1(-r,\xi),\xi)=b_1(0,x,\phi_2(-r,\xi),\xi),\ \{x,\xi\}\subset\mathbb{R}^d,$$
$$b_1(t,x,u,\xi)\le b_2(t,x,u,\xi),\ 0\le t\le T,\ \{x,\xi\}\subset\mathbb{R}^d,\ u\in\mathbb{R};$$

3) the functions f_i, $i \in \{1, 2\}$, *satisfy the conditions*

$$f_1(t, u, x) \geq f_2(t, u, x),\ 0 \leq t \leq T,\ u \in \mathbb{R},\ x \in \mathbb{R}^d.$$

Let one of the following conditions be true

M1) b_1 *is monotonously non-increasing,* f_1 *is monotonously non-decreasing with respect to* u*;*

M2) b_2 *is monotonously non-increasing,* f_2 *is monotonously non-decreasing with respect to* u.

Then for all $0 \leq t \leq T$ *the solutions of* (19.9)–(19.10) *satisfy the inequality*

$$u_1(t, x) \geq u_2(t, x),\ x \in \mathbb{R}^d,$$

with probability one.

19.3.2 Approximation Properties

During the proof we will need also auxiliary results of independent interest for the following Cauchy problem

$$\frac{\partial z(t, x)}{\partial t} = Az(t, x),\ t > 0,\ x \in \mathbb{R}^d, \tag{19.13}$$

$$z(0, x) = g(x),\ x \in \mathbb{R}^d, \tag{19.14}$$

where $A\colon L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ is a monotone operator. The next theorem is true.

Theorem 19.4 ([1], p. 25) *For any* $g \in L_2(\mathbb{R}^d)$ *there exists a unique solution* z *to* (19.13)–(19.14)*, belonging to* $C^1([0, \infty) \times \mathbb{R}^d) \cap ([0, \infty) \times \mathbb{R}^d)$*, and besides for* $t > 0$

$$\begin{aligned} &\|z(t, \cdot)\|_{L_2(\mathbb{R}^d)} \leq \|g(\cdot)\|_{L_2(\mathbb{R}^d)}, \\ &\left\| \frac{\partial z(t, \cdot)}{\partial t} \right\|_{L_2(\mathbb{R}^d)} = \|Az(t, \cdot)\|_{L_2(\mathbb{R}^d)} \leq \|Ag(\cdot)\|_{L_2(\mathbb{R}^d)}. \end{aligned} \tag{19.15}$$

Lemma 19.1 ([1], p. 22) *Let* $z_N \in C^1([0, \infty) \times \mathbb{R}^d)$ *be a solution to the following Yosida approximating equation*

$$\frac{\partial z_N(t, x)}{\partial t} = A_N z_N(t, x),\ t > 0,\ x \in \mathbb{R}^d,$$

where A_N, $N \in \{1, 2, \ldots\}$, is Yosida approximation of operator A. Then $\|z_N(t, \cdot)\|_{L_2(\mathbb{R}^d)}$ and $\left\|\frac{\partial z_N(t,\cdot)}{\partial t}\right\|_{L_2(\mathbb{R}^d)}$ are monotonously non-increasing on $t > 0$.

The following approximative property is valid.

Lemma 19.2 *There exists Yosida approximation of operator $A = \Delta_x$ by a sequence $\{A_N, N \in \{1, 2, \ldots\}\}$ of linear bounded operators $A_N \colon L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ and the following conditions are true*

1) for each $N \in \{1, 2, \ldots\}$ there exists a constant $C_N > 0$ such that

$$\|A_N\|^2_{\mathscr{L}(L_2(\mathbb{R}^d), L_2(\mathbb{R}^d))} \le C_N; \tag{19.16}$$

2) for each $g \in L_2(\mathbb{R}^d)$ the following equality is true

$$\lim_{N\to\infty} \left\| \big((A_N - A)\, g(\cdot)\big)(x) \right\|^2_{L_2(\mathbb{R}^d)} = 0,\ x \in \mathbb{R}^d; \tag{19.17}$$

3) operators A_N, $N \in \{1, 2, \ldots\}$, generate semigroup $\{S_N(t), N \in \{1, 2, \ldots\}\}$ of operators $S_N(t) \colon L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ with the following properties

a) for an arbitrary $x \in \mathbb{R}^d$ there exists $N_0 = N_0(x) \in \{1, 2, \ldots\}$ such that for all $N \ge N_0(x)$ $\big(S_N(t-s)g(\cdot)\big)(x) \ge 0$, $0 \le s \le t \le T$, $x \in \mathbb{R}^d$, $g \in L_2(\mathbb{R}^d)$, $g \ge 0$;

b)

$$\lim_{N\to\infty} \sup_{0\le s\le t\le T} \left\| \big((S_N(t-s) - S(t-s))\, g(\cdot)\big)(x) \right\|^2_{L_2(\mathbb{R}^d)} = 0,$$
$$x \in \mathbb{R}^d,\ g \in L_2(\mathbb{R}^d). \tag{19.18}$$

19.4 Proof of Theorem 19.1

1. From now on $x \in \mathbb{R}^d$ is supposed to be fixed. Let fix an arbitrary $M \in \{1, 2, \ldots\}$ and define by $W_M(t, \cdot)$ Q_M-Wiener process

$$W_M(t, \cdot) = \sum_{j=1}^{M} \sqrt{\lambda_j} e_j(\cdot)\beta_j(t),\ 0 \le t \le T.$$

Let us consider the following Cauchy problems

$$u_i^{N,M}(t,\cdot) = \phi_i(0,\cdot) + \int\limits_{\mathbb{R}^d} b_i(0,\cdot,\phi_i(-r,\xi),\xi)d\xi - \int\limits_{\mathbb{R}^d} b_i(t,\cdot,u_i^{N,M}(t-r,\xi),\xi)d\xi$$
$$+ \int\limits_0^t \big(A_N u_i^{N,M}(s,\cdot) + f_i(s,u_i^{N,M}(s-r,\cdot),\cdot)\big)ds + \int\limits_0^t \sigma(s,\cdot)dW_M(s,\cdot),$$
$$0 < t \le T,\ i \in \{1,2\},\ N \in \{1,2,\ldots\}, \tag{19.19}$$

$$u_i^{N,M}(t,\cdot) = \phi_i(t,\cdot),\ -r \le t \le 0, r > 0, i \in \{1,2\}, N \in \{1,2,\ldots\}, \tag{19.20}$$

where $\{A_N, N \in \{1,2,\ldots\}\}$ are operators from Lemma 19.2. Denote $u \equiv u_i$, $\phi \equiv \phi_i$, $b \equiv b_i$, $f \equiv f_i$, $i \in \{1,2\}$, for simplicity. A continuous $\mathscr{F}_t$-measurable for almost all $-r \le t \le T$ random process $u^{N,M} : [-r,T] \times \Omega \to L_2(\mathbb{R}^d)$ is called a **solution** to (19.19)–(19.20) provided

$$u^{N,M}(t,\cdot) = S_N(t)\left(\phi(0,\cdot) + \int\limits_{\mathbb{R}^d} b(0,\cdot,\phi(-r,\xi),\xi)d\xi\right)$$
$$- \int\limits_{\mathbb{R}^d} b(t,\cdot,u^{N,M}(t-r,\xi),\xi)d\xi$$
$$- \int\limits_0^t A_N S_N(t-s)\left(\int\limits_{\mathbb{R}^d} b(s,\cdot,u^{N,M}(s-r,\xi),\xi)d\xi\right)ds$$
$$+ \int\limits_0^t S_N(t-s) f(s,u^{N,M}(s-r,\cdot),\cdot)ds + \int\limits_0^t \sigma(s,\cdot)dW_M(s,\cdot),$$
$$0 < t \le T,\ N \in \{1,2,\ldots\}, \tag{19.21}$$

$$u^{N,M}(t,\cdot) = \phi(t,\cdot),\ -r \le t \le 0, r > 0, N \in \{1,2,\ldots\}, \tag{19.22}$$

and $\mathbf{E}\int\limits_0^T \|u(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}dt < \infty$. Since operators $\{A_N, N \in \{1,2,\ldots\}\}$ are bounded, (19.21)–(19.22) possesses a unique up to equivalence solution. Fix additionally $N \in \{1,2,\ldots\}$ and write u_i instead of $u_i^{N,M}$, $i \in \{1,2\}$, for notational simplicity. Let us prove that $u_1(t,\cdot) \ge u_2(t,\cdot)$, $0 \le t \le T$, almost surely.

2. Let us fix $n \in \{1, 2, \ldots\}$, put $t_k = \frac{kT}{N}$, $k \in \{0, \ldots, n\}$, with $t_{k+1} - t_k = \frac{T}{n} < r$, $-r < 0 \le t_1 \le r \le 2t_1 \le 2r \le \ldots$, and consider the next equations

$$z_i^{0,n}(t, \cdot) = \phi_i(0, \cdot) + \int_{\mathbb{R}^d} b_i(0, \cdot, \phi_i(-r, \xi), \xi) d\xi - \int_{\mathbb{R}^d} b_i(t, \cdot, z_i^{0,n}(t-r, \cdot), \xi) d\xi$$

$$+ \int_0^t \sigma(s, \cdot) dW_M(s, \cdot), 0 \le t \le t_1, i \in \{1, 2\}, \tag{19.23}$$

$$v_i^{0,n}(t, \cdot) = z_i^{0,n}(t_1, \cdot) + \int_0^t \big(A_N v_i^{0,n}(s, \cdot) + f_i(s, v_i^{0,n}(s-r, \cdot), \cdot)\big) ds,$$

$$0 \le t \le t_1, i \in \{1, 2\}, \tag{19.24}$$

and

$$z_i^{k,n}(t, \cdot) = v_i^{k-1,n}(t_k, \cdot) + \int_{\mathbb{R}^d} b_i(t_k, \cdot, z_i^{k-1,n}(t_k - r, \xi), \xi) d\xi$$

$$- \int_{\mathbb{R}^d} b_i(t, \cdot, z_i^{k,n}(t-r, \xi), \xi) d\xi$$

$$+ \int_{t_k}^t \sigma(s, \cdot) dW_M(s, \cdot), t_k \le t \le t_{k+1}, k \in \{1, \ldots, n-1\}, i \in \{1, 2\}, \tag{19.25}$$

$$v_i^{k,n}(t, \cdot) = z_i^{k,n}(t_{k+1}, \cdot) + \int_{t_k}^t \big(A_N v_i^{k,n}(s, \cdot) + f_i(s, v_i^{k,n}(s-r, \cdot), \cdot)\big) ds,$$

$$t_k \le t \le t_{k+1}, k \in \{1, \ldots, n-1\}, i \in \{1, 2\}, \tag{19.26}$$

$$z_i^{k,n}(t, \cdot) = v_i^{k,n}(t, \cdot) = \phi_i(t, \cdot), \ -r \le t \le 0, k \in \{0, \ldots, n-1\}, i \in \{1, 2\}.$$

3. Define $z_i^n\colon [0,T]\times\Omega\to L_2(\mathbb{R}^d)$, $v_i^n\colon [0,T]\times\Omega\to L_2(\mathbb{R}^d)$, $i\in\{1,2\}$, as follows

$$z_i^n(t,\cdot)=z_i^{k,n}(t,\cdot),\ t_k\le t<t_{k+1},\ k\in\{0,\dots,n-1\},\ i\in\{1,2\},\tag{19.27}$$

$$v_i^n(t,\cdot)=v_i^{k,n}(t,\cdot),\ t_k<t\le t_{k+1},\ k\in\{0,\dots,n-1\},\ i\in\{1,2\},\tag{19.28}$$

$$z_i^n(T,\cdot)=v_i^n(T,\cdot),\ i\in\{1,2\},$$

$$z_i^n(t,\cdot)=v_i^n(t,\cdot)=\phi_i(t,\cdot),\ -r\le t\le 0,\ i\in\{1,2\}.$$

Taking into account identities for $v_i^{k-1,n}(t_k,\cdot)$, $z_i^{k-1,n}(t_k,\cdot)$, $k\in\{1,\dots,n-1\}$, $i\in\{1,2\}$, and $z_i^{0,n}(t_1,\cdot)$, $i\in\{1,2\}$, one easily verifies

$$v_i^{k-1,n}(t_k,\cdot)=\underbrace{z_i^{k-1,n}(t_k,\cdot)}+\int_{t_{k-1}}^{t_k}\big(A_N v_i^{k-1,n}(s,\cdot)+f_i(s,v_i^{k-1,n}(s-r,\cdot),\cdot)\big)ds$$

$$=\underbrace{\overbrace{v_i^{k-2,n}(t_{k-1},\cdot)}+\int_{\mathbb{R}^d}b_i(t_{k-1},\cdot,z_i^{k-2,n}(t_{k-1}-r,\xi),\xi)d\xi}$$

$$\underbrace{-\int_{\mathbb{R}^d}b_i(t_k,\cdot,z_i^{k-1,n}(t_k-r,\xi),\xi)d\xi+\int_{t_{k-1}}^{t_k}\sigma(s,\cdot)dW_M(s,\cdot)}$$

$$+\int_{t_{k-1}}^{t_k}\big(A_N v_i^{k-1,n}(s,\cdot)+f_i(s,v_i^{k-1,n}(s-r,\cdot),\cdot)\big)ds=\overbrace{z_i^{k-2,n}(t_{k-1},\cdot)}$$

$$+\overbrace{\int_{t_{k-2}}^{t_{k-1}}\big(A_N v_i^{k-2,n}(s,\cdot)+f_i(s,v_i^{k-2,n}(s-r,\cdot),\cdot)\big)ds}$$

$$+\int_{\mathbb{R}^d}b_i(t_{k-1},\cdot,z_i^{k-2,n}(t_{k-1}-r,\xi),\xi)d\xi$$

$$-\int_{\mathbb{R}^d}b_i(t_k,\cdot,z_i^{k-1,n}(t_k-r,\xi),\xi)d\xi+\int_{t_{k-1}}^{t_k}\sigma(s,\cdot)dW_M(s,\cdot)$$

$$+\int_{t_{k-1}}^{t_k}\big(A_N v_i^{k-1,n}(s,\cdot)+f_i(s,v_i^{k-1,n}(s-r,\cdot),\cdot)\big)ds=\dots$$

$$= v_i^{0,n}(t_1, \cdot) + \int\limits_{\mathbb{R}^d} b_i(t_1, \cdot, \phi_i(t_1 - r, \xi), \xi) d\xi - \int\limits_{\mathbb{R}^d} b_i(t_k, \cdot, z_i^{k-1,n}(t_k - r, \xi), \xi) d\xi$$

$$+ \int\limits_{t_1}^{t_2} \sigma(s, \cdot) dW_M(s, \cdot) + \int\limits_{t_2}^{t_3} \sigma(s, \cdot) dW_M(s, \cdot) + \ldots + \int\limits_{t_{k-2}}^{t_{k-1}} \sigma(s, \cdot) dW_M(s, \cdot)$$

$$+ \int\limits_{t_{k-1}}^{t_k} \sigma(s, \cdot) dW_M(s, \cdot) + \int\limits_{t_1}^{t_2} \big(A_N v_i^{1,n}(s, \cdot) + f_i(s, v_i^{1,n}(s - r, \cdot), \cdot)\big) ds$$

$$+ \int\limits_{t_2}^{t_3} \big(A_N v_i^{2,n}(s, \cdot) + f_i(s, v_i^{2,n}(s - r, \cdot), \cdot)\big) ds + \ldots$$

$$+ \int\limits_{t_{k-2}}^{t_{k-1}} \big(A_N v_i^{k-2,n}(s, \cdot) + f_i(s, v_i^{k-2,n}(s - r, \cdot), \cdot)\big) ds$$

$$+ \int\limits_{t_{k-1}}^{t_k} \big(A_N v_i^{k-1,n}(s, \cdot) + f_i(s, v_i^{k-1,n}(s - r, \cdot), \cdot)\big) ds = \ldots$$

$$= \phi_i(0, \cdot) + \int\limits_{\mathbb{R}^d} b_i(0, \cdot, \phi_i(-r, \xi), \xi) d\xi - \int\limits_{\mathbb{R}^d} b_i(t_1, \cdot, \phi_i(t_1 - r, \xi), \xi) d\xi$$

$$+ \int\limits_0^{t_1} \sigma(s, \cdot) dW_M(s, \cdot) + \int\limits_0^{t_1} \big(A_N v_i^{0,n}(s, \cdot) + f_i(s, v_i^{0,n}(s - r, \cdot), \cdot)\big) ds$$

$$+ \int\limits_{\mathbb{R}^d} b_i(t_1, \cdot, \phi_i(t_1 - r, \xi), \xi) d\xi - \int\limits_{\mathbb{R}^d} b_i(t_k, \cdot, z_i^{k-1,n}(t_k - r, \xi), \xi) d\xi$$

$$+ \int\limits_{t_1}^{t_2} \sigma(s, \cdot) dW_M(s, \cdot) + \int\limits_{t_2}^{t_3} \sigma(s, \cdot) dW_M(s, \cdot) + \ldots + \int\limits_{t_{k-2}}^{t_{k-1}} \sigma(s, \cdot) dW_M(s, \cdot)$$

$$+ \int\limits_{t_{k-1}}^{t_k} \sigma(s, \cdot) dW_M(s, \cdot) + \int\limits_{t_1}^{t_2} \big(A_N v_i^{1,n}(s, \cdot) + f_i(s, v_i^{1,n}(s - r, \cdot), \cdot)\big) ds$$

$$+ \int\limits_{t_2}^{t_3} \big(A_N v_i^{2,n}(s, \cdot) + f_i(s, v_i^{2,n}(s - r, \cdot), \cdot)\big) ds + \ldots$$

$$+\int_{t_{k-2}}^{t_{k-1}}\big(A_N v_i^{k-2,n}(s,\cdot)+f_i(s,v_i^{k-2,n}(s-r,\cdot),\cdot)\big)ds$$

$$+\int_{t_{k-1}}^{t_k}\big(A_N v_i^{k-1,n}(s,\cdot)+f_i(s,v_i^{k-1,n}(s-r,\cdot),\cdot)\big)ds,\ i\in\{1,2\}. \tag{19.29}$$

After substitution (19.29) into (19.25) and, using (19.27), we obtain for $t_k \le t < t_{k+1}$, $k \in \{1,\dots,n-1\}$,

$$z_i^{k,n}(t,\cdot)=z_i^n(t,\cdot)=\phi_i(0,\cdot)+\int_{\mathbb{R}^d} b_i(0,\cdot,\phi_i(-r,\xi),\xi)d\xi$$

$$-\int_{\mathbb{R}^d} b_i(t_k,\cdot,z_i^n(t_k-r,\xi),\xi)d\xi+\int_0^{t_k}\sigma(s,\cdot)dW_M(s,\cdot)$$

$$+\int_0^{t_k}\big(A_N v_i^n(s,\cdot)+f_i(s,v_i^n(s-r,\cdot),\cdot)\big)ds+\int_{\mathbb{R}^d} b_i(t_k,\cdot,z_i^n(t_k-r,\xi),\xi)d\xi$$

$$-\int_{\mathbb{R}^d} b_i(t,\cdot,z_i^n(t-r,\xi),\xi)d\xi+\int_{t_k}^{t}\sigma(s,\cdot)dW_M(s,\cdot)$$

$$=\phi_i(0,\cdot)+\int_{\mathbb{R}^d} b_i(0,\cdot,\phi_i(-r,\xi),\xi)d\xi-\int_{\mathbb{R}^d} b_i(t,\cdot,z_i^n(t-r,\xi),\xi)d\xi$$

$$+\int_0^{t_k}\big(A_N v_i^n(s,\cdot)+f_i(s,v_i^n(s-r,\cdot),\cdot)\big)ds+\int_0^{t}\sigma(s,\cdot)dW_M(s,\cdot),\ i\in\{1,2\}.$$

Similarly, taking into account (19.28), we get from (19.26) for $t_k < t \le t_{k+1}$, $k \in \{1,\dots,n-1\}$,

$$v_i^n(t,\cdot)=v_i^{k,n}(t,\cdot)=\phi_i(0,\cdot)+\int_{\mathbb{R}^d} b_i(0,\cdot,\phi_i(-r,\xi),\xi)d\xi$$

$$-\int_{\mathbb{R}^d} b_i(t_{k+1},\cdot,z_i^n(t_{k+1}-r,\xi),\xi)d\xi+\int_0^{t_k}\big(A_N v_i^n(s,\cdot)+f_i(s,v_i^n(s-r,\cdot),\cdot)\big)ds$$

$$+\int_0^{t_{k+1}}\sigma(s,\,\cdot)dW_M(s,\cdot)+\int_{t_k}^{t}\big(A_N v_i^n(s,\,\cdot)+f_i(s,v_i^n(s-r,\,\cdot),\,\cdot)\big)ds$$

$$=\phi_i(0,\,\cdot)+\int_{\mathbb{R}^d}b_i(0,\,\cdot,\phi_i(-r,\xi),\xi)d\xi-\int_{\mathbb{R}^d}b_i(t_{k+1},\,\cdot,z_i^n(t_{k+1}-r,\xi),\xi)d\xi$$

$$+\int_0^{t}\big(A_N v_i^n(s,\,\cdot)+f_i(s,v_i^n(s-r,\,\cdot),\,\cdot)\big)ds+\int_0^{t_{k+1}}\sigma(s,\,\cdot)dW_M(s,\cdot),\ i\in\{1,2\}.$$

Thus, one easily verifies for $z_i^n(t,\,\cdot)$, $v_i^n(t,\,\cdot)$, $i\in\{1,2\}$,

$$z_i^n(t,\,\cdot)=\phi_i(0,\,\cdot)+\int_{\mathbb{R}^d}b_i(0,\,\cdot,\phi_i(-r,\xi),\xi)d\xi-\int_{\mathbb{R}^d}b_i(t,\,\cdot,z_i^n(t-r,\xi),\xi)d\xi$$

$$+\int_0^{t_k}\big(A_N v_i^n(s,\,\cdot)+f_i(s,v_i^n(s-r,\,\cdot),\,\cdot)\big)ds+\int_0^{t}\sigma(s,\,\cdot)dW_M(s,\cdot),$$

$$t_k\le t<t_{k+1},\,k\in\{0,\dots,n-1\},\,i\in\{1,2\},\tag{19.30}$$

$$v_i^n(t,\,\cdot)=\phi_i(0,\,\cdot)+\int_{\mathbb{R}^d}b_i(0,\,\cdot,\phi_i(-r,\xi),\xi)d\xi-\int_{\mathbb{R}^d}b_i(t_{k+1},\,\cdot,z_i^n(t_{k+1}-r,\xi),\xi)d\xi$$

$$+\int_0^{t}\big(A_N v_i^n(s,\,\cdot)+f_i(s,v_i^n(s-r,\,\cdot),\,\cdot)\big)ds+\int_0^{t_{k+1}}\sigma(s,\,\cdot)dW_M(s,\cdot),$$

$$t_k<t\le t_{k+1},\,k\in\{0,\dots,n-1\},\,i\in\{1,2\},\tag{19.31}$$

$$z_i^n(t,\,\cdot)=v_i^n(t,\,\cdot)=\phi_i(t,\,\cdot),\ -r\le t\le 0,\,i\in\{1,2\}.$$

4. Now let us show that

$$z_1^n(t,\,\cdot)\ge z_2^n(t,\,\cdot),\tag{19.32}$$

$$v_1^n(t,\,\cdot)\ge v_2^n(t,\,\cdot),\tag{19.33}$$

almost surely for any $0\le t\le T$.

Let us prove (19.32) for $0 \le t \le t_1$. Invoking Theorem 19.3, one obtains

$$z_1^n(t,\cdot) = \phi_1(0,\cdot) + \int_{\mathbb{R}^d} b_1(0,\cdot,\phi_1(-r,\xi),\xi)d\xi - \int_{\mathbb{R}^d} b_1(t,\cdot,z_1^n(t-r,\xi),\xi)d\xi$$

$$+ \sum_{j=1}^{M} \sqrt{\lambda_j} e_j(\cdot) \int_0^t \sigma(s,\cdot)d\beta_j(s) \ge \phi_2(0,\cdot) + \int_{\mathbb{R}^d} b_2(0,\cdot,\phi_2(-r,\xi),\xi)d\xi$$

$$- \int_{\mathbb{R}^d} b_2(t,\cdot,z_2^n(t-r,\xi),\xi)d\xi + \sum_{j=1}^{M} \sqrt{\lambda_j} e_j(\cdot) \int_0^t \sigma(s,\cdot)d\beta_j(s) = z_2^n(t,\cdot),$$

$$0 \le t < t_1. \tag{19.34}$$

Similarly we obtain for $z_i^{0,n}(t_1,\cdot)$, $i \in \{1,2\}$,

$$z_1^{0,n}(t_1,\cdot) = \phi_1(0,\cdot) + \int_{\mathbb{R}^d} b_1(0,\cdot,\phi_1(-r,\xi),\xi)d\xi - \int_{\mathbb{R}^d} b_1(t_1,\cdot,z_1^{0,n}(t_1-r,\xi),\xi)d\xi$$

$$+ \sum_{j=1}^{M} \sqrt{\lambda_j} e_j(\cdot) \int_0^{t_1} \sigma(s,\cdot)d\beta_j(s) \ge \phi_2(0,\cdot) + \int_{\mathbb{R}^d} b_2(0,\cdot,\phi_2(-r,\xi),\xi)d\xi$$

$$- \int_{\mathbb{R}^d} b_2(t_1,\cdot,z_2^{0,n}(t_1-r,\xi),\xi)d\xi + \sum_{j=1}^{M} \sqrt{\lambda_j} e_j(\cdot) \int_0^{t_1} \sigma(s,\cdot)d\beta_j(s) = z_2^{0,n}(t_1,\cdot).$$

Now let us prove (19.33) for $0 \le t \le t_1$. Since this inequality is obvious for $t = 0$, we will show it for $0 < t \le t_1$. Since we have $-r < t - r \le t_1 - r \le 0$ for $0 < t \le t_1$, then $z_i^n(t-r,\cdot) = \phi_i(t-r,\cdot)$, $0 < t \le t_1$, $i \in \{1,2\}$, and $f_2(s,\phi_1(s-r,\cdot),\cdot) \ge f_2(s,\phi_2(s-r,\cdot),\cdot)$, according to **M2**. Then we get for $v_1^n(t,\cdot) - v_2^n(t,\cdot)$, $0 < t \le t_1$, from (19.31), taking into account conditions of the theorem, the following identity

$$v_1^n(t,\cdot) - v_2^n(t,\cdot) = \big(z_1^{0,n}(t_1,\cdot) - z_2^{0,n}(t_1,\cdot)\big) + \int_0^t A_N\big(v_1^n(s,\cdot) - v_2^n(s,\cdot)\big)ds$$

$$+ \int_0^t \big(f_1(s,\phi_1(s-r,\cdot),\cdot) - f_2(s,\phi_1(s-r,\cdot),\cdot)\big)ds$$

$$+ \int_0^t \big(f_2(s,\phi_1(s-r,\cdot),\cdot) - f_2(s,\phi_2(s-r,\cdot),\cdot)\big)ds.$$

Since S_N supposed to be positivity preserving, the equality above can be rewritten in the following manner

$$v_1^n(t,\,\cdot\,) - v_2^n(t,\,\cdot\,) = S_N(t)\big(z_1^{0,n}(t_1,\,\cdot\,) - z_2^{0,n}(t_1,\,\cdot\,)\big)$$
$$+\int_0^t S_N(t-s)\big(f_1(s,\phi_1(s-r,\,\cdot\,),\,\cdot\,) - f_2(s,\phi_1(s-r,\,\cdot\,),\,\cdot\,)\big)ds$$
$$+\int_0^t S_N(t-s)\big(f_2(s,\phi_1(s-r,\,\cdot\,),\,\cdot\,) - f_2(s,\phi_2(s-r,\,\cdot\,),\,\cdot\,)\big)ds \geq 0.$$

Thus we have

$$v_1^n(t,\,\cdot\,) \geq v_2^n(t,\,\cdot\,),\, 0 < t \leq t_1. \tag{19.35}$$

This estimate implies (19.33) for $0 \leq t \leq t_1$.

It remains to show that $z_1^n(t_1,\,\cdot\,) \geq z_2^n(t_1,\,\cdot\,)$. Since

$$v_i^n(t_1,\,\cdot\,) = z_i^n(t_1,\,\cdot\,), i \in \{1,2\},$$

then we obviously obtain from (19.35)

$$z_1^n(t_1,\,\cdot\,) \geq z_2^n(t_1,\,\cdot\,).$$

This estimate and relation (19.34), in turn, give (19.32) for any $0 \leq t \leq t_1$.

Let us prove (19.32) for $t_1 \leq t \leq t_2$. First let estimate $z_1^n(t,\,\cdot\,) - z_2^n(t,\,\cdot\,)$, $t_1 \leq t < t_2$. Since we have $-r \leq t_1 - r \leq t - r < t_2 - r = 2t_1 - r \leq t_1$, i.e. $-r \leq t - r \leq t_1$ for $t_1 \leq t < t_2$, then $z_1^n(t-r,\,\cdot\,) \geq z_2^n(t-r,\,\cdot\,)$, $v_1^n(t-r,\,\cdot\,) \geq v_2^n(t-r,\,\cdot\,)$, $t_1 \leq t < t_2$, and, according to **M2**, $f_2(t, v_1^n(t-r,\,\cdot\,),\,\cdot\,) \geq f_2(t, v_2^n(t-r,\,\cdot\,),\,\cdot\,)$, $b_2(t,\,\cdot\,, z_2^n(t-r,\xi),\xi) \geq b_2(t,\,\cdot\,, z_1^n(t-r,\xi),\xi)$, $t_1 \leq t < t_2$, $\xi \in \mathbb{R}^d$. Then we obtain for $z_1^n(t,\,\cdot\,) - z_2^n(t,\,\cdot\,)$, $t_1 \leq t < t_2$, from (19.30)

$$z_1^n(t,\,\cdot\,) - z_2^n(t,\,\cdot\,) = \big(\phi_1(0,\,\cdot\,) - \phi_2(0,\,\cdot\,)\big)$$
$$+\int_{\mathbb{R}^d} \big(b_1(0,\,\cdot\,,\phi_1(-r,\xi),\xi) - b_2(0,\,\cdot\,,\phi_2(-r,\xi),\xi)\big)d\xi$$
$$+\int_{\mathbb{R}^d} \big(b_2(t,\,\cdot\,,z_2^n(t-r,\xi),\xi) - b_2(t,\,\cdot\,,z_1^n(t-r,\xi),\xi)\big)d\xi$$

$$+\int_{\mathbb{R}^d}\big(b_2(t,\,\cdot\,,z_1^n(t-r,\xi),\xi)-b_1(t,\,\cdot\,,z_1^n(t-r,\xi),\xi)\big)d\xi$$

$$+\int_0^{t_1}A_N\big(v_1^n(s,\,\cdot\,)-v_2^n(s,\,\cdot\,)\big)ds$$

$$+\int_0^{t_1}\big(f_1(s,v_1^n(s-r,\,\cdot\,),\,\cdot\,)-f_2(s,v_1^n(s-r,\,\cdot\,),\,\cdot\,)\big)ds$$

$$+\int_0^{t_1}\big(f_2(s,v_1^n(s-r,\,\cdot\,),\,\cdot\,)-f_2(s,v_2^n(s-r,\,\cdot\,),\,\cdot\,)\big)ds.$$

Since S_N supposed to be positivity preserving, the equality above can be rewritten in the following manner

$$z_1^n(t,\,\cdot\,)-z_2^n(t,\,\cdot\,)=S_N(t)\big(\phi_1(0,\,\cdot\,)-\phi_2(0,\,\cdot\,)\big)$$

$$+S_N(t)\int_{\mathbb{R}^d}\big(b_2(t,\,\cdot\,,z_2^n(t-r,\xi),\xi)-b_2(t,\,\cdot\,,z_1^n(t-r,\xi),\xi)\big)d\xi$$

$$+S_N(t)\int_{\mathbb{R}^d}\big(b_2(t,\,\cdot\,,z_1^n(t-r,\xi),\xi)-b_1(t,\,\cdot\,,z_1^n(t-r,\xi),\xi)\big)d\xi$$

$$+\int_0^{t_1}S_N(t-s)\big(f_1(s,v_1^n(s-r,\,\cdot\,),\,\cdot\,)-f_2(s,v_1^n(s-r,\,\cdot\,),\,\cdot\,)\big)ds$$

$$+\int_0^{t_1}S_N(t-s)\big(f_2(s,v_1^n(s-r,\,\cdot\,),\,\cdot\,)-f_2(s,v_2^n(s-r,\,\cdot\,),\,\cdot\,)\big)ds\geq 0.$$

The last inequality holds because of the conditions of the theorem.

Hence it follows that

$$z_1^n(t,\,\cdot\,)\geq z_2^n(t,\,\cdot\,),t_1\leq t<t_2. \tag{19.36}$$

Now let us prove (19.33) for $t_1 \le t \le t_2$. Since estimate for $t = t_1$ follows from (19.35), we will show it for $t_1 < t \le t_2$. We derive from (19.31)

$$
\begin{aligned}
v_1^n(t,\cdot) - v_2^n(t,\cdot) &= \big(\phi_1(0,\cdot) - \phi_2(0,\cdot)\big) \\
&+ \int_{\mathbb{R}^d} \big(b_1(0,\cdot,\phi_1(-r,\xi),\xi) - b_2(0,\cdot,\phi_2(-r,\xi),\xi)\big)d\xi \\
&+ \int_{\mathbb{R}^d} \big(b_2(t_2,\cdot,z_2^n(t_2-r,\xi),\xi) - b_2(t_2,\cdot,z_1^n(t_2-r,\xi),\xi)\big)d\xi \\
&+ \int_{\mathbb{R}^d} \big(b_2(t_2,\cdot,z_1^n(t_2-r,\xi),\xi) - b_1(t_2,\cdot,z_1^n(t_2-r,\xi),\xi)\big)d\xi \\
&+ \int_0^t \big(f_1(s,v_1^n(s-r,\cdot),\cdot) - f_2(s,v_1^n(s-r,\cdot),\cdot)\big)ds \\
&+ \int_0^t \big(f_2(s,v_1^n(s-r,\cdot),\cdot) - f_2(s,v_2^n(s-r,\cdot),\cdot)\big)ds.
\end{aligned}
$$

Since S_N supposed to be positivity preserving, the equality above can be rewritten as

$$
\begin{aligned}
v_1^n(t,\cdot) - v_2^n(t,\cdot) &= S_N(t)\big(\phi_1(0,\cdot) - \phi_2(0,\cdot)\big) \\
&+ S_N(t)\int_{\mathbb{R}^d} \big(b_2(t_2,\cdot,z_2^n(t_2-r,\xi),\xi) - b_2(t_2,\cdot,z_1^n(t_2-r,\xi),\xi)\big)d\xi \\
&+ S_N(t)\int_{\mathbb{R}^d} \big(b_2(t_2,\cdot,z_1^n(t_2-r,\xi),\xi) - b_1(t_2,\cdot,z_1^n(t_2-r,\xi),\xi)\big)d\xi \\
&+ \int_0^t S_N(t-s)\big(f_1(s,v_1^n(s-r,\cdot),\cdot) - f_2(s,v_1^n(s-r,\cdot),\cdot)\big)ds \\
&+ \int_0^t S_N(t-s)\big(f_2(s,v_1^n(s-r,\cdot),\cdot) - f_2(s,v_2^n(s-r,\cdot),\cdot)\big)ds \ge 0.
\end{aligned}
$$

Hence,

$$
v_1^n(t,\cdot) \ge v_2^n(t,\cdot), t_1 < t \le t_2.
$$

Thus, (19.33) is proved for $t_1 \le t \le t_2$.

It remains to show that $z_1^n(t_2,\,\cdot\,) \geq z_2^n(t_2,\,\cdot\,)$. Since

$$v_i^n(t_2,\,\cdot\,) = z_i^n(t_2,\,\cdot\,),\, i \in \{1,2\},$$

then

$$z_1^n(t_2,\,\cdot\,) \geq z_2^n(t_2,\,\cdot\,).$$

The inequality above and estimate (19.36) give (19.32) for $t_1 \leq t \leq t_2$.

5. We will prove here that there exists $C_n > 0$ such that

$$\sup_{0\leq t\leq T} \mathbf{E}\|z_i^n(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} \leq C_n,\, i \in \{1,2\}, \tag{19.37}$$

$$\sup_{0\leq t\leq T} \mathbf{E}\|v_i^n(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} \leq C_n,\, i \in \{1,2\}, \tag{19.38}$$

where $v_i^n,\, i \in \{1,2\}$, are defined from (19.31), $z_i^n,\, i \in \{1,2\}$, – from (19.30). In order to prove (19.37) it is sufficient to show that

$$\sup_{t_j\leq t\leq t_{j+1}} \mathbf{E}\|z_i^{j,n}(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} \leq c_n,\, j \in \{0,\ldots,n-1\}, i \in \{1,2\}, \tag{19.39}$$

for some $c_n > 0$. It is sufficient to prove (19.39) for $j \in \{0,1\}$, because for $j \in$ $\in \{2,\ldots,n-1\}$ the proof is similar.

5.1. Let us estimate $\sup\limits_{0\leq t\leq t_1} \mathbf{E}\|z_i^{0,n}(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}$, $i \in \{1,2\}$. From (19.23) we derive

$$\begin{aligned}
\sup_{0\leq t\leq t_1} \mathbf{E}\|z_i^{0,n}(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} &\leq 4\mathbf{E}\|\phi_i(0,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} \\
&+ 4\mathbf{E}\left\|\int\limits_{\mathbb{R}^d} |b_i(0,\,\cdot\,,\phi_i(-r,\xi),\xi)|d\xi\right\|^2_{L_2(\mathbb{R}^d)} \\
&+ 4\sup_{0\leq t\leq t_1} \mathbf{E}\left\|\int\limits_{\mathbb{R}^d} |b_i(t,\,\cdot\,,\phi_i(t-r,\xi),\xi)|d\xi\right\|^2_{L_2(\mathbb{R}^d)} \\
&+ 4\sup_{0\leq t\leq t_1} \mathbf{E}\left\|\int\limits_0^t \sigma(s,\,\cdot\,)dW_M(s,\,\cdot\,)\right\|^2_{L_2(\mathbb{R}^d)} \\
&= 4\mathbf{E}\|\phi_i(0,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} + \sum_{j=1}^3 S_j^{(Z^0)},\, i \in \{1,2\}.
\end{aligned} \tag{19.40}$$

Let us now estimate each of $S_j^{(Z^0)}$, $j \in \{1,2,3\}$, separately. We have, taking into account conditions of the theorem,

$$S_1^{(Z^0)} = 4\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b(0,x,\phi_i(-r,\xi),\xi)|d\xi\right)^2 dx$$

$$\le 8\left(\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(0,x,\xi)d\xi dx\right)\mathbf{E}\|\phi_i(-r,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$

$$+8\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\chi(x,\xi)d\xi\right)^2 dx, i \in \{1,2\},$$

$$S_2^{(Z^0)} = 4\sup_{0\le t\le t_1}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b_i(t,x,\phi_i(t-r,\xi),\xi)|d\xi\right)^2 dx$$

$$\le 8\left(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(t,x,\xi)d\xi dx\right)$$

$$\times \sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} + 8\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\chi(x,\xi)d\xi\right)^2 dx, i \in \{1,2\},$$

$$S_3^{(Z^0)} = 4\sup_{0\le t\le t_1}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\sum_{j=1}^M\sqrt{\lambda_j}\left(\int\limits_0^t\sigma(s,x)d\beta_j(s)\right)e_j(x)\right)^2 dx$$

$$\le 4M\int\limits_{\mathbb{R}^d}\left(\sum_{j=1}^M\lambda_j\int\limits_0^{t_1}\sigma^2(s,x)ds\right)dx$$

$$\le 4M\left(\sum_{j=1}^M\lambda_j\right)\frac{T}{n}\sup_{0\le t\le T}\|\sigma(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}.$$

With the help of these estimates we get from (19.40)

$$\sup_{0\le t\le t_1}\mathbf{E}\|z_i^{0,n}(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} \le 4\mathbf{E}\|\phi_i(0,\cdot)\|^2_{L_2(\mathbb{R}^d)} + 8\left(\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(0,x,\xi)d\xi dx\right)$$

$$\times\,\mathbf{E}\|\phi_i(-r,\cdot)\|^2_{L_2(\mathbb{R}^d)} + 16\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\chi(x,\xi)d\xi\right)^2 dx$$

$$+8\Big(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(t,x,\xi)d\xi dx\Big)\sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}$$

$$+4M\Big(\sum_{j=1}^{M}\lambda_j\Big)\frac{T}{n}\sup_{0\le t\le T}\|\sigma(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}=C_n^{(Z^0)},\ i\in\{1,2\}. \tag{19.41}$$

5.2. Let us estimate $\mathbf{E}\|v_i^{0,n}(t_1,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}$, $i\in\{1,2\}$. We derive from (19.24), taking into account (19.41),

$$\mathbf{E}\|v_i^{0,n}(t_1,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}\le 3\mathbf{E}\|z_i^{0,n}(t_1,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}+3\mathbf{E}\Big\|\int\limits_0^{t_1}A_N v_i^{0,n}(s,\,\cdot\,)ds\Big\|^2_{L_2(\mathbb{R}^d)}ds$$

$$+3\mathbf{E}\Big\|\int\limits_0^{t_1}|f_i(s,\phi_i(s-r,\,\cdot\,),\,\cdot\,)|ds\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$\le 3\big(C_n^{(Z^0)}+S_1^{(V^0)}+S_2^{(V^0)}\big),\ i\in\{1,2\}. \tag{19.42}$$

In order to estimate $S_1^{(V^0)}$ let us take (19.16) into account. We obtain

$$S_1^{(V^0)}=\mathbf{E}\int\limits_{\mathbb{R}^d}\Big(\int\limits_0^{t_1}A_N v_i^{0,n}(s,x)ds\Big)^2dx\le\frac{T}{n}\mathbf{E}\int\limits_0^{t_1}\|A_N v_i^{0,n}(s,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}ds$$

$$\le\frac{C_N T}{n}\int\limits_0^{t_1}\mathbf{E}\|v_i^{0,n}(s,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)},\ i\in\{1,2\}.$$

In order to estimate $S_2^{(V^0)}$ let us take (19.5) into account. We obtain

$$S_2^{(V^0)}=\mathbf{E}\int\limits_{\mathbb{R}^d}\Big(\int\limits_0^{t_1}|f_i(s,\phi_i(s-r,x),x)|ds\Big)^2dx$$

$$\le\frac{2T}{n}\mathbf{E}\int\limits_0^{t_1}\int\limits_{\mathbb{R}^d}\big(\eta^2(s,x)+L^2\phi_i^2(s-r,x)\big)dxds$$

$$\le \frac{2T}{n}\left(T\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\eta^2(t,x)dx + L^2\int\limits_{-r}^{t_1-r}\mathbf{E}\|\phi_i(s-r,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}d(s-r)\right)$$

$$\le \frac{2T}{n}\left(T\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\eta^2(t,x)dx + rL^2\sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}\right), i\in\{1,2\}.$$

With the help of the obtained estimates it follows from (19.42) and from Bellman-Gronwalls inequality

$$\mathbf{E}\|v_i^{0,n}(t_1,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} \le \left[3C_n^{(Z^0)} + \frac{6T}{n}\left(T\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\eta^2(t,x)dx + rL^2\sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}\right)\right]$$
$$\times\exp\left\{\frac{3C_NT}{n}\cdot\frac{T}{n}\right\}, i\in\{1,2\}. \tag{19.43}$$

5.3. Let estimate $\sup\limits_{t_1\le t\le t_2}\mathbf{E}\|z_i^{1,n}(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}$, $i\in\{1,2\}$. Applying (19.25) and (19.43), we conclude

$$\sup_{t_1\le t\le t_2}\mathbf{E}\|z_i^{1,n}(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}$$
$$\le 6\mathbf{E}\|v_i^{0,n}(t_1,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} + 6\mathbf{E}\left\|\int\limits_{\mathbb{R}^d}|b_i(t_1,\,\cdot\,,\phi_i(t_1-r,\xi),\xi)|d\xi\right\|^2_{L_2(\mathbb{R}^d)}$$
$$+2\sup_{t_1\le t\le t_2}\left\|\int\limits_{\mathbb{R}^d}|b_i(t,\,\cdot\,,z_i^{1,n}(t-r,\xi),\xi)|d\xi\right\|^2_{L_2(\mathbb{R}^d)}$$
$$+6\sup_{t_1\le t\le t_2}\mathbf{E}\left\|\int\limits_{t_1}^{t}\sigma(s,\,\cdot\,)dW_M(s,\,\cdot\,)\right\|^2_{L_2(\mathbb{R}^d)}$$
$$\le 6\left[3C_n^{(Z^0)}+\frac{6T}{n}\left(T\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\eta^2(t,x)dx + rL^2\sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}\right)\right]$$
$$\times\exp\left\{\frac{3C_NT}{n}\cdot\frac{T}{n}\right\}+\sum_{j=1}^{3}S_j^{(Z^1)}, i\in\{1,2\}. \tag{19.44}$$

Let us estimate each of $S_j^{(Z^1)}$, $j \in \{1, 2, 3\}$, separately. We conclude

$$
\begin{aligned}
S_1^{(Z^1)} &= 6\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b_i(t_1,x,\phi_i(t_1-r,\xi),\xi)|d\xi\right)^2 dx\\
&\le 12\left(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(t,x,\xi)d\xi dx\right)\\
&\times \sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} + 12\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\chi(x,\xi)d\xi\right)^2 dx,\ i\in\{1,2\},\\
S_2^{(Z^1)} &= 2\sup_{t_1\le t\le t_2}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b_i(t,x,z_i^{1,n}(t-r,\xi),\xi)|d\xi\right)^2 dx\\
&\le 4\left(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(t,x,\xi)d\xi dx\right)\\
&\times\left(\sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} + \sup_{0\le t\le t_1}\mathbf{E}\|z_i^{0,n}(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\right)\\
&+ 4\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\chi(x,\xi)d\xi\right)^2 dx\\
&\le 4\left(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(t,x,\xi)d\xi dx\right)\left(\sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} + C_n^{(Z^0)}\right)\\
&+ 4\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\chi(x,\xi)d\xi\right)^2 dx,\ i\in\{1,2\},\\
S_3^{(Z^1)} &= 6\sup_{t_1\le t\le t_2}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\sum_{j=1}^M\sqrt{\lambda_j}\left(\int\limits_{t_1}^t\sigma(s,x)d\beta_j(s)\right)e_j(x)\right)^2 dx\\
&\le 6M\left(\sum_{j=1}^M\lambda_j\right)\frac{T}{n}\sup_{0\le t\le T}\|\sigma(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}.
\end{aligned}
$$

These three estimates and (19.44) finally give

$$\sup_{t_1\le t\le t_2}\mathbf{E}\|z_i^{1,n}(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\le 6\Bigg[3C_n^{(Z^0)}+\frac{6T}{n}\Bigg(T\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\eta^2(t,x)dx$$

$$+rL^2\sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\Bigg)\Bigg]$$

$$\times\exp\left\{\frac{3C_NT}{n}\cdot\frac{T}{n}\right\}+16\Bigg(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\Bigg)\sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$

$$+16\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\chi(x,\xi)d\xi\right)^2dx+4\Bigg(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\Bigg)C_n^{(Z^0)}$$

$$+6M\Bigg(\sum_{j=1}^M\lambda_j\Bigg)\frac{T}{n}\sup_{0\le t\le T}\|\sigma(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}=C_n^{(Z^1)},\ i\in\{1,2\}.$$

Equation (19.39) and, obviously, (19.37) is proved. Equation (19.38) is proved in a similar way.

6. Next we will prove that there exists some $C^{(U)}(T)>0$ such that

$$\sup_{0\le t\le T}\mathbf{E}\|u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\le C^{(U)}(T),\ i\in\{1,2\}. \tag{19.45}$$

In order to do it we need to estimate $\sup\limits_{0\le t\le T}\mathbf{E}\|u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$, $i\in\{1,2\}$, from (19.19). We have

$$\sup_{0\le t\le T}\mathbf{E}\|u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\le 10\mathbf{E}\|\phi_i(0,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$

$$+10\mathbf{E}\left\|\int\limits_{\mathbb{R}^d}|b_i(0,\cdot,\phi_i(-r,\xi),\xi)|d\xi\right\|^2_{L_2(\mathbb{R}^d)}$$

$$+10\sup_{0\le t\le T}\mathbf{E}\left\|\int\limits_0^t A_Nu_i(s,\cdot)ds\right\|^2_{L_2(\mathbb{R}^d)}$$

$$+10\sup_{0\le t\le T}\mathbf{E}\left\|\int\limits_0^t|f_i(s,u_i(s-r,\cdot),\cdot)|ds\right\|^2_{L_2(\mathbb{R}^d)}$$

$$+10\sup_{0\le t\le T}\mathbf{E}\left\|\int\limits_0^t\sigma(s,\cdot)dW_M(s,\cdot)\right\|^2_{L_2(\mathbb{R}^d)}$$

$$+2\sup_{0\le t\le T}\mathbf{E}\left\|\int_{\mathbb{R}^d}|b_i(t,\,\cdot\,,u_i(t-r,\xi),\xi)|d\xi\right\|^2_{L_2(\mathbb{R}^d)}$$

$$=10\mathbf{E}\|\phi_i(0,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}+\sum_{j=1}^{5}S_j^{(U)},\ i\in\{1,2\}. \tag{19.46}$$

Taking into account previous calculations, we obtain

$$S_1^{(U)}=10\mathbf{E}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}|b_i(0,x,\phi_i(-r,\xi),\xi)|d\xi\right)^2dx$$

$$\le 20\left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}l^2(0,x,\xi)d\xi dx\right)\mathbf{E}\|\phi_i(-r,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}$$

$$+20\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\chi(x,\xi)d\xi\right)^2dx,\ i\in\{1,2\},$$

$$S_2^{(U)}=10\sup_{0\le t\le T}\mathbf{E}\int_{\mathbb{R}^d}\left(\int_0^t A_N u_i(s,x)ds\right)^2dx$$

$$\le 10C_NT\int_0^T\sup_{0\le s\le t}\mathbf{E}\|u_i(s,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}dt,\ i\in\{1,2\},$$

$$S_3^{(U)}=10\sup_{0\le t\le T}\mathbf{E}\int_{\mathbb{R}^d}\left(\int_0^t|f_i(s,u_i(s-r,x),x)|ds\right)^2dx\le 20T\left(T\sup_{0\le t\le T}\int_{\mathbb{R}^d}\eta^2(t,x)dx\right.$$

$$\left.+rL^2\sup_{-r\le t\le 0}\mathbf{E}\|\phi_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}+L^2\int_0^T\sup_{0\le s\le t}\mathbf{E}\|u_i(s,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}dt\right),\ i\in\{1,2\},$$

$$S_4^{(U)}=10\sup_{0\le t\le T}\mathbf{E}\int_{\mathbb{R}^d}\left(\sum_{j=1}^M\sqrt{\lambda_j}\left(\int_0^t\sigma(s,x)d\beta_j(s)\right)e_j(x)\right)^2dx$$

$$\le 10M\left(\sum_{j=1}^M\lambda_j\right)T\sup_{0\le t\le T}\|\sigma(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)},$$

$$S_5^{(U)}=2\sup_{0\le t\le T}\mathbf{E}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}|b_i(t,x,u_i(t-r,\xi),\xi)|d\xi\right)^2dx$$

$$\le 4\left(\sup_{0\le t\le T}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\right)$$

$$\times \left(\sup_{-r\le t\le 0} \mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} + \sup_{0\le t\le T} \mathbf{E}\|u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} \right)$$

$$+ 4 \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \chi(x,\xi) d\xi \right)^2 dx,\ i \in \{1,2\}.$$

Put

$$c^{(U)}(T) = 10\mathbf{E}\|\phi_i(0,\cdot)\|^2_{L_2(\mathbb{R}^d)} + 20\left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} l^2(0,x,\xi) d\xi dx \right) \mathbf{E}\|\phi_i(-r,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$

$$+ 24 \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \chi(x,\xi) d\xi \right)^2 dx + 20T \left(T \sup_{0\le t\le T} \int_{\mathbb{R}^d} \eta^2(t,x) dx \right.$$

$$\left. + rL^2 \sup_{-r\le t\le 0} \mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} \right)$$

$$+ 10M \left(\sum_{j=1}^{M} \lambda_j \right) T \sup_{0\le t\le T} \|\sigma(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} + 4 \left(\sup_{0\le t\le T} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} l^2(t,x,\xi) d\xi dx \right)$$

$$\times \sup_{-r\le t\le 0} \mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)},\ i \in \{1,2\}.$$

Invoking Bellman-Gronwalls lemma and condition (19.3), we obtain from (19.46) estimate (19.45) of the form

$$\sup_{0\le t\le T} \mathbf{E}\|u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} \le \left(1 - 4 \sup_{0\le t\le T} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} l^2(t,x,\xi) d\xi dx \right)^{-1} c^{(U)}(T)$$

$$\times \exp\left\{ \left(1 - 4 \sup_{0\le t\le T} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} l^2(t,x,\xi) d\xi dx \right)^{-1} \left(10C_N T + 20L^2 T \right) \cdot T \right\} = C^{(U)}(T),$$

$i \in \{1,2\}$.

7. Due to (19.37), (19.38) and (19.45), there exists a constant $C_n > 0$ such that

$$\sup_{0\le t\le T} \mathbf{E}\|v_i^n(t,\cdot) - u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} + \sup_{0\le t\le T} \mathbf{E}\|z_i^n(t,\cdot) - u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$

$$\le 2 \sup_{0\le t\le T} \mathbf{E}\|v_i^n(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} + 4 \sup_{0\le t\le T} \mathbf{E}\|u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$

$$+ 2 \sup_{0\le t\le T} \mathbf{E}\|z_i^n(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} \le C_n,\ i \in \{1,2\}. \tag{19.47}$$

Now let us prove

$$\lim_{n\to\infty}\sup_{t_k\le t<t_{k+1}}\mathbf{E}\|z_i^n(t,\,\cdot\,)-u_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}=0,\ i\in\{1,2\}. \tag{19.48}$$

7.1. Let estimate $\mathbf{E}\|v_i^n(t,\,\cdot\,)-u_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}$, $t_k<t\le t_{k+1}$, $k\in\{0,\dots,n-1\}$, $i\in\{1,2\}$. We have

$$\begin{aligned}
\mathbf{E}\|v_i^n(t,\,\cdot\,)-u_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}&\le 2\mathbf{E}\Big\|\int_{\mathbb{R}^d}|b_i(t_{k+1},\,\cdot\,,u_i(t_{k+1}-r,\xi),\xi)\\
&-b_i(t_{k+1},\,\cdot\,,z_i^n(t_{k+1}-r,\xi),\xi)|d\xi\Big\|^2_{L_2(\mathbb{R}^d)}+10\mathbf{E}\Big\|\int_{\mathbb{R}^d}|b_i(t,\,\cdot\,,u_i(t-r,\xi),\xi)\\
&-b_i(t_{k+1},\,\cdot\,,u_i(t-r,\xi),\xi)|d\xi\Big\|^2_{L_2(\mathbb{R}^d)}+10\mathbf{E}\Big\|\int_{\mathbb{R}^d}|b_i(t_{k+1},\,\cdot\,,u_i(t-r,\xi),\xi)\\
&-b_i(t_{k+1},\,\cdot\,,u_i(t_{k+1}-r,\xi),\xi)|d\xi\Big\|^2_{L_2(\mathbb{R}^d)}+10\mathbf{E}\Big\|A_N\big(v_i^n(s,\,\cdot\,)-u_i(s,\,\cdot\,)\big)ds\Big\|^2_{L_2(\mathbb{R}^d)}\\
&+10\mathbf{E}\Big\|\int_0^t|f_i(s,v_i^n(s-r,\,\cdot\,),\,\cdot\,)-f_i(s,u_i(s-r,\,\cdot\,),\,\cdot\,)|ds\Big\|^2_{L_2(\mathbb{R}^d)}\\
&+10\mathbf{E}\Big\|\int_t^{t_{k+1}}\sigma(s,\,\cdot\,)dW_M(s,\,\cdot\,)\Big\|^2_{L_2(\mathbb{R}^d)}=\sum_{j=1}^6 S_j^{(V-U)},\ i\in\{1,2\}.
\end{aligned}\tag{19.49}$$

Let us estimate each of $S_j^{(V-U)}$, $j\in\{1,\dots,6\}$, from (19.49) separately. We conclude

$$\begin{aligned}
S_1^{(V-U)}&=2\mathbf{E}\int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d}|b_i(t_{k+1},x,u_i(t_{k+1}-r,\xi),\xi)-b_i(t_{k+1},x,z_i^n(t_{k+1}-r,\xi),\xi)|d\xi\Big)^2dx\\
&\le 2\Big(\sup_{0\le t\le T}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\Big)\mathbf{E}\|z_i^n(t_{k+1}-r,\,\cdot\,)-u_i(t_{k+1}-r,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)},\\
&\quad i\in\{1,2\}.
\end{aligned}$$

In order to estimate $S_2^{(V-U)}$ we will use the unform continuity of b_i, $i\in\{1,2\}$, and Lebesgue's dominated convergence theorem. Finally we get

$$\begin{aligned}
S_2^{(V-U)}&=10\mathbf{E}\int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d}|b_i(t,x,u_i(t-r,\xi),\xi)-b_i(t_{k+1},x,u_i(t-r,\xi),\xi)|d\xi\Big)^2dx\\
&=\epsilon_1(n),\ i\in\{1,2\},\ \lim_{n\to\infty}\epsilon_1(n)=0.
\end{aligned}$$

Taking into account continuity of u_i, $i \in \{1,2\}$, we get for $S_3^{(V-U)}$

$$S_3^{(V-U)} = 10\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b_i(t_{k+1},x,u_i(t-r,\xi),\xi)-b_i(t_{k+1},x,u_i(t_{k+1}-r,\xi),\xi)|d\xi\right)^2 dx$$

$$\le 10\left(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(t,x,\xi)d\xi dx\right)\mathbf{E}\|u_i(t-r,\cdot)$$

$$-u_i(t_{k+1}-r,\cdot)\|^2_{L_2(\mathbb{R}^d)} = \epsilon_2(n),$$

$$i\in\{1,2\},\ \lim_{n\to\infty}\epsilon_2(n)=0.$$

Considering $S_j^{(V-U)}$, $j \in \{4,5,6\}$, let us take into account proceeding analysis. We conclude

$$S_4^{(V-U)} = 10\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_0^t A_N\big(u_i(s,x)-v_i^n(s,x)\big)ds\right)^2 dx$$

$$\le 10C_N T\int\limits_0^t \mathbf{E}\|u_i(s,\cdot)-v_i^n(s,\cdot)\|^2_{L_2(\mathbb{R}^d)}ds,\ i\in\{1,2\},$$

$$S_5^{(V-U)} = 10\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_0^t |f_i(s,v_i^n(s-r,x),x)-f_i(s,u_i(s-r,x),x)|ds\right)^2 dx$$

$$\le 10L^2T\int\limits_{-r}^{t-r}\mathbf{E}\|u_i(s-r,\cdot)-v_i^n(s-r,\cdot)\|^2_{L_2(\mathbb{R}^d)}d(s-r)$$

$$\le 10L^2T\int\limits_0^t \mathbf{E}\|u_i(s,\cdot)-v_i^n(s,\cdot)\|^2_{L_2(\mathbb{R}^d)}ds,\ i\in\{1,2\},$$

$$S_6^{(V-U)} = 10\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\sum_{j=1}^M\sqrt{\lambda_j}\left(\int\limits_t^{t_{k+1}}\sigma(s,x)d\beta_j(s)\right)e_j(x)\right)^2 dx$$

$$\le 10M\left(\sum_{j=1}^M\lambda_j\right)\frac{T}{n}\sup_{0\le t\le T}\|\sigma(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}.$$

Taking into account obtained estimates, we apply Bellman-Gronwalls inequality to (19.49) (it is applicated due to (19.47)) and, summing up, obtain

the following estimate

$$\sup_{t_k<t\le t_{k+1}} \mathbf{E}\|v_i^n(t,\,\cdot\,)-u_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} \le \beta_n^{(V-U)}(t_{k+1}) \times \exp\{10\big(C_N+L^2\big)T^2\},\ i\in\{1,2\}, \tag{19.50}$$

with

$$\begin{aligned}\beta_n^{(V-U)}(t_{k+1}) &= \epsilon_3(n)+2\Big(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} l^2(t,x,\xi)d\xi dx\Big)\mathbf{E}\|z_i^n(t_{k+1}-r,\,\cdot\,)\\ &\quad - u_i(t_{k+1}-r,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}\\ &\quad + 10M\Big(\sum_{j=1}^M\lambda_j\Big)\frac{T}{n}\sup_{0\le t\le T}\|\sigma(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)},\ i\in\{1,2\},\\ \epsilon_3(n) &= \min\{\epsilon_1(n),\epsilon_2(n)\}.\end{aligned}$$

7.2. Now let us estimate $\sup\limits_{t_k\le t<t_{k+1}} \mathbf{E}\|z_i^n(t,\,\cdot\,)-u_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}$, $k\in\{0,\dots,n-1\}$, $i\in\{1,2\}$. Since we have for difference

$$\begin{aligned}v_i^n(t_k,\,\cdot\,)-u_i(t_k,\,\cdot\,) &= -\int\limits_{\mathbb{R}^d} b_i(t_k,\,\cdot\,,z_i^n(t_k-r,\xi),\xi)d\xi\\ &\quad + \int\limits_{\mathbb{R}^d} b_i(t_k,\,\cdot\,,u_i(t_k-r,\xi),\xi)d\xi\\ &\quad + \int\limits_0^{t_k}\big(A_Nv_i^n(s,\,\cdot\,)+f_i(s,v_i^n(s-r,\,\cdot\,),\,\cdot\,)\big)ds - \int\limits_0^{t_k}\big(A_Nu_i(s,\,\cdot\,)\\ &\quad + f_i(s,u_i(s-r,\,\cdot\,),\,\cdot\,)\big)ds,\\ &\quad i\in\{1,2\},\end{aligned}$$

we observe

$$\begin{aligned}\sup_{t_k\le t<t_{k+1}} \mathbf{E}\|z_i^n(t,\,\cdot\,)-u_i(t,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} &= \sup_{t_k\le t<t_{k+1}}\mathbf{E}\Big\|v_i^n(t_k,\,\cdot\,)-u_i(t_k,\,\cdot\,)\\ &\quad + \int\limits_{\mathbb{R}^d}\big(b_i(t,\,\cdot\,,u_i(t-r,\xi),\xi)-b_i(t_k,\,\cdot\,,u_i(t_k-r,\xi),\xi)\big)d\xi\end{aligned}$$

$$+\int\limits_{\mathbb{R}^d}\big(b_i(t_k,\,\cdot\,,z_i^n(t_k-r,\xi),\xi)-b_i(t,\,\cdot\,,z_i^n(t-r,\xi),\xi)\big)d\xi$$

$$-\int\limits_{t_k}^{t}\big(A_Nu_i(s,\,\cdot\,)+f_i(s,u_i(s-r,\,\cdot\,),\,\cdot\,)\big)ds\Big\|^2_{L_2(\mathbb{R}^d)}\le 2\mathbf{E}\|v_i^n(t_k,\,\cdot\,)-u_i(t_k,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}$$

$$+8\sup_{t_k\le t<t_{k+1}}\mathbf{E}\Big\|\int\limits_{\mathbb{R}^d}|b_i(t,\,\cdot\,,u_i(t-r,\xi),\xi)-b_i(t_k,\,\cdot\,,u_i(t_k-r,\xi),\xi)|d\xi\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$+8\sup_{t_k\le t<t_{k+1}}\mathbf{E}\Big\|\int\limits_{\mathbb{R}^d}|b_i(t_k,\,\cdot\,,z_i^n(t_k-r,\xi),\xi)-b_i(t,\,\cdot\,,z_i^n(t-r,\xi),\xi)|d\xi\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$+8\sup_{t_k\le t<t_{k+1}}\mathbf{E}\Big\|\int\limits_{t_k}^{t}A_Nu_i(s,\,\cdot\,)ds\Big\|^2_{L_2(\mathbb{R}^d)}+8\sup_{t_k\le t<t_{k+1}}\mathbf{E}\Big\|\int\limits_{t_k}^{t}|f_i(s,u_i(s-r,\,\cdot\,),\,\cdot\,)|ds\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$=2\mathbf{E}\|v_i^n(t_k,\,\cdot\,)-u_i(t_k,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)}+8\sum_{j=1}^{4}S_j^{(Z-U)},\ i\in\{1,2\}. \tag{19.51}$$

Let us estimate each of $S_j^{(Z-U)}$, $j\in\{1,\dots,4\}$, from (19.51) separately. Estimating as before, we obtain

$$S_1^{(Z-U)}=\sup_{t_k\le t<t_{k+1}}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b_i(t,\,\cdot\,,u_i(t-r,\xi),\xi)-b_i(t_k,\,\cdot\,,u_i(t_k-r,\xi),\xi)|d\xi\right)^2dx$$

$$\le 2\sup_{t_k\le t<t_{k+1}}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b_i(t,\,\cdot\,,u_i(t-r,\xi),\xi)-b_i(t,\,\cdot\,,u_i(t_k-r,\xi),\xi)|d\xi\right)^2dx$$

$$+2\sup_{t_k\le t<t_{k+1}}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b_i(t,\,\cdot\,,u_i(t_k-r,\xi),\xi)-b_i(t_k,\,\cdot\,,u_i(t_k-r,\xi),\xi)|d\xi\right)^2dx$$

$$=\epsilon_4(n),\ i\in\{1,2\},\ \lim_{n\to\infty}\epsilon_4(n)=0, \tag{19.52}$$

$$S_2^{(Z-U)}=\sup_{t_k\le t<t_{k+1}}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b_i(t_k,x,z_i^n(t_k-r,\xi),\xi)-b_i(t,x,z_i^n(t-r,\xi),\xi)|d\xi\right)^2dx$$

$$\le 2\sup_{t_k\le t<t_{k+1}}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b_i(t_k,x,z_i^n(t_k-r,\xi),\xi)-b_i(t_k,x,z_i^n(t-r,\xi),\xi)|d\xi\right)^2dx$$

$$+2\sup_{t_k\le t<t_{k+1}}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}|b_i(t_k,x,z_i^n(t-r,\xi),\xi)-b_i(t,x,z_i^n(t-r,\xi),\xi)|d\xi\right)^2dx$$

$$=\epsilon_5(n),\ i\in\{1,2\},\ \lim_{n\to\infty}\epsilon_5(n)=0, \tag{19.53}$$

$$S_3^{(Z-U)} = \sup_{t_k \le t < t_{k+1}} \mathbf{E} \int_{\mathbb{R}^d} \left(\int_{t_k}^{t} A_N u_i(s,x) ds \right)^2 dx \le \frac{C_N T^2}{n} \sup_{0 \le t < T} \mathbf{E} \| u_i(t,\cdot) \|^2_{L_2(\mathbb{R}^d)},$$

$$i \in \{1,2\}, \tag{19.54}$$

$$S_4^{(Z-U)} = \sup_{t_k \le t < t_{k+1}} \mathbf{E} \int_{\mathbb{R}^d} \left(\int_{t_k}^{t} |f_i(s, u_i(s-r,x),x)| ds \right)^2 dx \le \frac{2T}{n} \Bigg(T \sup_{0 \le t \le T} \int_{\mathbb{R}^d} \eta^2(t,x) dx$$

$$+ rL^2 \sup_{-r \le t \le 0} \mathbf{E} \| \phi_i(t,\cdot) \|^2_{L_2(\mathbb{R}^d)} + L^2 T \sup_{0 \le t \le T} \mathbf{E} \| u_i(t,\cdot) \|^2_{L_2(\mathbb{R}^d)} \Bigg),$$

$$i \in \{1,2\}. \tag{19.55}$$

It follows from (19.50)

$$\mathbf{E} \| v_i^n(t_k,\cdot) - u_i(t_k,\cdot) \|^2_{L_2(\mathbb{R}^d)} \le \sup_{t_{k-1} < t \le t_k} \mathbf{E} \| v_i^n(t,\cdot) - u_i(t,\cdot) \|^2_{L_2(\mathbb{R}^d)}$$

$$\le \beta_n^{(V-U)}(t_k) \cdot \exp\{10\big(C_N + L^2\big)T^2\}, i \in \{1,2\}, \tag{19.56}$$

with

$$\beta_n^{(V-U)}(t_k) = \epsilon_3(n) + 2 \Bigg(\sup_{0 \le t \le T} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} l^2(t,x,\xi) d\xi dx \Bigg) \mathbf{E} \| z_i^n(t_k - r,\cdot)$$

$$- u_i(t_k - r,\cdot) \|^2_{L_2(\mathbb{R}^d)}$$

$$+ 10M \Bigg(\sum_{j=1}^{M} \lambda_j \Bigg) \frac{T}{n} \sup_{0 \le t \le T} \| \sigma(t,\cdot) \|^2_{L_2(\mathbb{R}^d)} \le \epsilon_3(n)$$

$$+ 2 \Bigg(\sup_{0 \le t \le T} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} l^2(t,x,\xi) d\xi dx \Bigg)$$

$$\times \sup_{t_k \le t < t_{k+1}} \mathbf{E} \| z_i^n(t,\cdot) - u_i(t,\cdot) \|^2_{L_2(\mathbb{R}^d)}$$

$$+ 10M \Bigg(\sum_{j=1}^{M} \lambda_j \Bigg) \frac{T}{n} \sup_{0 \le t \le T} \| \sigma(t,\cdot) \|^2_{L_2(\mathbb{R}^d)},$$

$$i \in \{1,2\}.$$

Thus, we finally obtain from (19.51), using estimates (19.52)–(19.56),

$$\sup_{t_k\le t<t_{k+1}}\mathbf{E}\|z_i^n(t,\cdot)-u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\le C_n^{(Z-U)}(T)$$
$$\times\left(1-4\exp\{10\big(C_N+L^2\big)T^2\}\Big(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\Big)\right)^{-1},$$
$$i\in\{1,2\},\tag{19.57}$$

with

$$C_n^{(Z-U)}(T)=\epsilon(n)+\frac{8C_NT^2}{n}\sup_{0\le t<T}\mathbf{E}\|u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$
$$+20M\Big(\sum_{j=1}^M\lambda_j\Big)\frac{T}{n}\sup_{0\le t\le T}\|\sigma(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$
$$\times\exp\{10\big(C_N+L^2\big)T^2\}+\frac{16T}{n}\Big(T\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\eta^2(t,x)dx$$
$$+rL^2\sup_{-r\le t\le0}\mathbf{E}\|\phi_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$
$$+L^2T\sup_{0\le t\le T}\mathbf{E}\|u_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\Big),\ i\in\{1,2\},$$
$$\epsilon(n)=\min\{2\exp\{10\big(C_N+L^2\big)T^2\}\epsilon_3(n),\epsilon_4(n),\epsilon_5(n)\},\ \lim_{n\to\infty}\epsilon(n)=0.$$

Since $\lim\limits_{n\to\infty}C_n^{(Z-U)}(T)=0$, then (19.57) clearly implies (19.48).

8. For any $0\le t\le T$ a sequence $\{z_i^n(t,\cdot),n\in\{1,2,\ldots\}\}$, $i\in\{1,2\}$, contains a subsequence $\{z_i^{n_m}(t,\cdot),m\in\{1,2,\ldots\}\}$, $i\in\{1,2\}$, converging to $u_i(t,\cdot)$, $i\in\{1,2\}$, in $L_2(\mathbb{R}^d)$) almost surely. This implies

$$u_1(t,\cdot)\ge u_2(t,\cdot)$$

almost surely for $0\le t\le T$.

9. Denote $u \equiv u_i$, $\phi \equiv \phi_i$, $b \equiv b_i$, $f \equiv f_i$, $i \in \{1,2\}$, for brevity. Let $u^M : [-r,T] \times \Omega \to L_2(\mathbb{R}^d)$ be a continuous $\mathscr{F}_t$-measurable for almost all $-r \le t \le T$ process, defined as a unique solution to the following integral equation

$$u^M(t,\cdot) = S(t)\Big(\phi(0,\cdot) + \int\limits_{\mathbb{R}^d} b(0,\cdot,\phi(-r,\xi),\xi)d\xi\Big)$$
$$- \int\limits_0^t AS(t-s)\Big(\int\limits_{\mathbb{R}^d} b(s,\cdot,u^M(s-r,\xi),\xi)d\xi\Big)ds$$
$$+ \int\limits_0^t S(t-s)f(s,u^M(s-r,\cdot),\cdot)ds + \int\limits_0^t \sigma(s,\cdot)dW_M(s,\cdot),$$
$$0 < t \le T, \tag{19.58}$$
$$u^M(t,\cdot) = \phi(t,\cdot),\ -r \le t \le 0, r > 0, \tag{51*}$$

satisfying the condition

$$\mathbf{E}\int\limits_0^T \|u^M(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}dt < \infty.$$

It remains to show that

$$\lim_{N\to\infty}\sup_{0\le t\le T}\mathbf{E}\|u_i^{N,M}(t,\cdot) - u_i^M(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} = 0, i \in \{1,2\}, \tag{19.59}$$

$$\lim_{M\to\infty}\sup_{0\le t\le T}\mathbf{E}\|u_i^{M}(t,\cdot) - U_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)} = 0, i \in \{1,2\}, \tag{19.60}$$

where $U_i : [-r,T] \times \Omega \to L_2(\mathbb{R}^d)$, $i \in \{1,2\}$, is a unique solution to (19.1).

9.1. Let us prove (19.59), where $u_i^{N,M}$, $i \in \{1,2\}$, are defined from (19.19)–(19.20). In order to do it we will estimate $\sup\limits_{0\le t\le T}\mathbf{E}\|u_i^{N,M}(t,\cdot) - u_i^M(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$, $i \in \{1,2\}$. We get

$$\sup_{0\le t\le T}\mathbf{E}\|u_i^{N,M}(t,\cdot) - u_i^M(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$
$$\le 4\sup_{0\le t\le T}\mathbf{E}\Big\|(S_N(t) - S(t))\Big(\phi_i(0,\cdot) + \int\limits_{\mathbb{R}^d} b_i(0,\cdot,\phi_i(-r,\xi),\xi)d\xi\Big)\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$+4\sup_{0\le t\le T}\mathbf{E}\Big\|\int_{\mathbb{R}^d}|b_i(t,\cdot,u_i^M(t-r,\xi),\xi)-b_i(t,\cdot,u_i^{N,M}(t-r,\xi),\xi)|d\xi\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$+4\sup_{0\le t\le T}\mathbf{E}\Big\|\int_0^t AS(t-s)\Big(\int_{\mathbb{R}^d}b_i(s,\cdot,u_i^M(s-r,\xi),\xi)d\xi\Big)ds$$

$$-\int_0^t A_N S_N(t-s)\Big(\int_{\mathbb{R}^d}b_i(s,\cdot,u_i^{N,M}(s-r,\xi),\xi)d\xi\Big)ds\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$+4\sup_{0\le t\le T}\mathbf{E}\Big\|\int_0^t\big(S_N(t-s)f_i(s,u^{N,M}(s-r,\cdot),\cdot)$$

$$-S(t-s)f_i(s,u^M(s-r,\cdot),\cdot)\big)ds\Big\|^2_{L_2(\mathbb{R}^d)}=\sum_{j=1}^4 S_j^{(U^N-U)},\ i\in\{1,2\}. \tag{19.61}$$

Let estimate $S_j^{(U^N-U)}$, $j\in\{1,\dots,4\}$, from (19.61) separately. Taking into account belonging $\phi_i(0,\cdot)+\int_{\mathbb{R}^d}b_i(0,\cdot,\phi_i(-r,\xi),\xi)d\xi$, $i\in\{1,2\}$, to $L_2(\mathbb{R}^d)$ and property (19.18), we conclude

$$S_1^{(U^N-U)}=4\sup_{0\le t\le T}\mathbf{E}\Big\|\big(S_N(t)-S(t)\big)\Big(\phi_i(0,\cdot)+\int_{\mathbb{R}^d}b_i(0,\cdot,\phi_i(-r,\xi),\xi)d\xi\Big)\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$=\epsilon_1(N),\ i\in\{1,2\},\ \lim_{N\to\infty}\epsilon_1(N)=0. \tag{19.62}$$

For $S_2^{(U^N-U)}$ we get, estimating as before,

$$S_2^{(U^N-U)}=4\sup_{0\le t\le T}\mathbf{E}\int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d}|b_i(t,x,u_i^M(t-r,\xi),\xi)-b_i(t,x,u_i^{N,M}(t-r,\xi),\xi)|d\xi\Big)^2dx$$

$$\le 4\Big(\sup_{0\le t\le T}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\Big)\sup_{0\le t\le T}\mathbf{E}\|u_i^M(t-r,\cdot)-u_i^{N,M}(t-r,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$

$$\le 4\Big(\sup_{0\le t\le T}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\Big)\sup_{0\le t\le T}\mathbf{E}\|u_i^M(t,\cdot)-u_i^{N,M}(t,\cdot)\|^2_{L_2(\mathbb{R}^d)},$$

$$i\in\{1,2\}. \tag{19.63}$$

For $S_3^{(U^N-U)}$ we conclude

$$
\begin{aligned}
S_3^{(U^N-U)} &= 4 \sup_{0\le t\le T} \mathbf{E}\Big\| \int_0^t AS(t-s)\Big(\int_{\mathbb{R}^d} b_i(s,\,\cdot\,,u_i^M(s-r,\xi),\xi)d\xi\Big)ds \\
&\quad - \int_0^t A_N S_N(t-s)\Big(\int_{\mathbb{R}^d} b_i(s,\,\cdot\,,u_i^M(s-r,\xi),\xi)d\xi\Big)ds\Big\|^2_{L_2(\mathbb{R}^d)} \\
&\quad + \int_0^t A_N S_N(t-s)\Big(\int_{\mathbb{R}^d} \big(b_i(s,\,\cdot\,,u_i^M(s-r,\xi),\xi) \\
&\quad - b_i(s,\,\cdot\,,u_i^{N,M}(s-r,\xi),\xi)\big)d\xi\Big)ds\Big\|^2_{L_2(\mathbb{R}^d)} \\
&\le 8 \sup_{0\le t\le T} \mathbf{E}\Big\| \int_0^t AS(t-s)\Big(\int_{\mathbb{R}^d} b_i(s,\,\cdot\,,u_i^M(s-r,\xi),\xi)d\xi\Big)ds \\
&\quad - \int_0^t A_N S_N(t-s)\Big(\int_{\mathbb{R}^d} b_i(s,\,\cdot\,,u_i^M(s-r,\xi),\xi)d\xi\Big)ds\Big\|^2_{L_2(\mathbb{R}^d)} \\
&\quad + 8 \sup_{0\le t\le T} \mathbf{E}\Big\| \int_0^t A_N S_N(t-s)\Big(\int_{\mathbb{R}^d} \big(b_i(s,\,\cdot\,,u_i^M(s-r,\xi),\xi) \\
&\quad - b_i(s,\,\cdot\,,u_i^{N,M}(s-r,\xi),\xi)d\xi\big)\Big)ds\Big\|^2_{L_2(\mathbb{R}^d)}, i\in\{1,2\}.
\end{aligned}
\tag{19.64}
$$

For estimating the first term in (19.64) we will use Lemma 19.1. It follows from this lemma uniform in t convergence $z_N(t,\,\cdot\,)$ to $z(t,\,\cdot\,)$ as $N\to\infty$, where $z_N(t,\,\cdot\,) = S_N(t-s)\Big(\int_{\mathbb{R}^d} b_i(s,\,\cdot\,,u_i^M(s-r,\xi),\xi)d\xi\Big)$, $t\ge s$, $i\in\{1,2\}$, is a solution to the problem (19.13)–(19.14) of the form

$$
\begin{aligned}
\frac{\partial z_N(t,\,\cdot\,)}{\partial t} &= A_N z_N(t,\,\cdot\,), t>s, \\
z_N(s,\,\cdot\,) &= \int_{\mathbb{R}^d} b_i(s,\,\cdot\,,u_i^M(s-r,\xi),\xi)d\xi, i\in\{1,2\},
\end{aligned}
$$

and $z(t,\cdot) = S(t-s)\Big(\int\limits_{\mathbb{R}^d} b_i(s,\cdot,u_i^M(s-r,\xi),\xi)d\xi\Big)$, $t \geq s$, $i \in \{1,2\}$, solves the problem (19.13)–(19.14) of the form

$$\frac{\partial z(t,\cdot)}{\partial t} = Az(t,\cdot), t > s,$$

$$z(s,\cdot) = \int\limits_{\mathbb{R}^d} b_i(s,\cdot,u_i^M(s-r,\xi),\xi)d\xi, i \in \{1,2\}.$$

Therefore we have from (19.64)

$$8 \sup_{0\leq t\leq T} \mathbf{E}\Big\| \int\limits_0^t AS(t-s)\Big(\int\limits_{\mathbb{R}^d} b_i(s,\cdot,u_i^M(s-r,\xi),\xi)d\xi\Big)ds$$

$$- \int\limits_0^t A_N S_N(t-s)\Big(\int\limits_{\mathbb{R}^d} b_i(s,\cdot,u_i^M(s-r,\xi),\xi)d\xi\Big)ds\Big\|^2_{L_2(\mathbb{R}^d)} = \epsilon_2(N),$$

$$i \in \{1,2\}, \lim_{N\to\infty} \epsilon_2(N) = 0.$$

For estimating the second term in (19.64) we will use Lemma 19.1, (19.15) from Theorem 19.4 and property (19.6). We conclude

$$8 \sup_{0\leq t\leq T} \mathbf{E}\Big\| \int\limits_0^t A_N S_N(t-s)\Big(\int\limits_{\mathbb{R}^d} \big(b_i(s,\cdot,u_i^M(s-r,\xi),\xi)$$

$$- b_i(s,\cdot,u_i^{N,M}(s-r,\xi),\xi)\big)d\xi\Big)ds\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$\leq 8 \sup_{0\leq t\leq T} \mathbf{E}\Big\| \int\limits_0^t A_N\Big(\int\limits_{\mathbb{R}^d} \big(b_i(s,\cdot,u_i^M(s-r,\xi),\xi)$$

$$- b_i(s,\cdot,u_i^{N,M}(s-r,\xi),\xi)\big)d\xi\Big)ds\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$\leq 8 \sup_{0\leq t\leq T} \mathbf{E}\int\limits_{\mathbb{R}^d}\Big(\int\limits_0^t \Delta_x\Big(\int\limits_{\mathbb{R}^d} \big(b_i(s,x,u_i^M(s-r,\xi),\xi)$$

$$- b_i(s,x,u_i^{N,M}(s-r,\xi),\xi)\big)d\xi\Big)ds\Big)^2 dx$$

$$\leq 8T\mathbf{E}\int\limits_0^T\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\|D_x^2 b_i(s,x,u_i^M(s-r,\xi),\xi)$$

$$-D_x^2 b_i(s,x,u_i^{N,M}(s-r,\xi),\xi)\|d\xi\right)^2 dxds$$

$$\leq 8T\mathbf{E}\int\limits_0^T\int\limits_{\mathbb{R}^d}\left(\int\limits_{\mathbb{R}^d}\psi(s,x,\xi)|u_i^M(s-r,\xi)-u_i^{N,M}(s-r,\xi)|d\xi\right)^2 dxds$$

$$\leq 8T\left(\sup_{0\leq t\leq T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\psi^2(t,x,\xi)d\xi dx\right)\int\limits_{-r}^{T-r}\mathbf{E}\|u_i^M(s-r,\cdot)$$

$$-u_i^{N,M}(s-r,\cdot)\|^2_{L_2(\mathbb{R}^d)}d(s-r)$$

$$\leq 8T\left(\sup_{0\leq t\leq T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\psi^2(t,x,\xi)d\xi dx\right)\int\limits_0^T\sup_{0\leq s\leq t}\mathbf{E}\|u_i^M(s,\cdot)$$

$$-u_i^{N,M}(s,\cdot)\|^2_{L_2(\mathbb{R}^d)}dt,$$

$i\in\{1,2\}$.

Thus we obtain from (19.64)

$$S_3^{(U^N-U)}\leq\epsilon_2(n)+8T\left(\sup_{0\leq t\leq T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\psi^2(t,x,\xi)d\xi dx\right)$$

$$\times\int\limits_0^T\sup_{0\leq s\leq t}\mathbf{E}\|u_i^M(s,\cdot)-u_i^{N,M}(s,\cdot)\|^2_{L_2(\mathbb{R}^d)}dt,\ i\in\{1,2\}.\tag{19.65}$$

For estimating $S_4^{(U^N-U)}$ let apply (19.18) and belonging $\eta(t,\cdot)+|u_i^{N,M}(t-r,\cdot)|$,
$0\leq t\leq T$, $i\in\{1,2\}$, to $L_2(\mathbb{R}^d)$. We have

$$S_4^{(U^N-U)}=4\sup_{0\leq t\leq T}\mathbf{E}\int\limits_{\mathbb{R}^d}\left(\int\limits_0^t\big(S_N(t-s)-S(t-s)\right.$$

$$\left.+S(t-s)\big)f_i(s,u_i^{N,M}(s-r,\cdot),\cdot)ds\right.$$

$$
\begin{aligned}
&- \int_0^t S(t-s) f_i(s, u_i^M(s-r,\,\cdot\,),\,\cdot\,)ds\Big)^2 dx \\
&\le 8 \sup_{0\le t\le T} \mathbf{E} \int_{\mathbb{R}^d} \Big(\int_0^t \big(S_N(t-s) - S(t-s)\big) | f_i(s, u_i^{N,M}(s-r,\,\cdot\,),\,\cdot\,)|ds\Big)^2 dx \\
&+ 8 \sup_{0\le t\le T} \mathbf{E} \int_{\mathbb{R}^d} \Big(\int_0^t S(t-s) | f_i(s, u_i^{N,M}(s-r,x),x) \\
&- f_i(s, u_i^M(s-r,x),x)|ds\Big)^2 dx \\
&\le 8 \sup_{0\le t\le T} \mathbf{E} \int_{\mathbb{R}^d} \Big(\int_0^t \big(S_N(t-s) - S(t-s)\big)\big(\eta(s,x) \\
&\quad + L|u_i^{N,M}(s-r,x)|\big)ds\Big)^2 dx \\
&+ 8L^2 \sup_{0\le t\le T} \mathbf{E} \int_{\mathbb{R}^d} \Big(\int_0^t S(t-s)|u_i^{N,M}(s-r,x) - u_i^M(s-r,x)|ds\Big)^2 dx \\
&\le 8T \sup_{0\le t\le T} \mathbf{E} \int_0^t \Big\| \big(S_N(t-s) - S(t-s)\big)\big(\eta(s,\,\cdot\,) \\
&\quad + L|u_i^{N,M}(s-r,\,\cdot\,)|\big)\Big\|^2_{L_2(\mathbb{R}^d)} ds \\
&+ 8L^2T \sup_{0\le t\le T} \mathbf{E} \int_0^t \Big\| S(t-s)\big(u_i^{N,M}(s-r,\,\cdot\,) - u_i^M(s-r,\,\cdot\,)\big)\Big\|^2_{L_2(\mathbb{R}^d)} ds \\
&\le 8T \sup_{0\le t\le T} \mathbf{E} \int_0^t \Big\| \big(S_N(t-s) - S(t-s)\big)\big(\eta(s,\,\cdot\,) + L|u_i^{N,M}(s-r,\,\cdot\,)|\big)\Big\|^2_{L_2(\mathbb{R}^d)} ds \\
&+ 8L^2T \sup_{0\le t\le T} \int_{-r}^{T-r} \mathbf{E}\|u_i^{N,M}(s,\,\cdot\,) - u_i^M(s,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} ds \le \epsilon_3(n) \\
&+ 8L^2T \int_0^T \sup_{0\le s\le t} \mathbf{E}\|u_i^{N,M}(s,\,\cdot\,) - u_i^M(s,\,\cdot\,)\|^2_{L_2(\mathbb{R}^d)} dt, \ \lim_{N\to\infty} \epsilon_3(N) = 0.
\end{aligned}
\tag{19.66}
$$

Taking into account (19.62), (19.63), (19.65) and (19.66), we conclude

$$\sup_{0\le t\le T}\mathbf{E}\|u_i^{N,M}(t,\cdot)-u_i^M(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\le\left(1-4\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\right)^{-1}$$

$$\times\left[\epsilon(N)+8T\left(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\psi^2(t,x,\xi)d\xi dx+L^2\right)\right.$$

$$\left.\times\int\limits_0^T\sup_{0\le s\le t}\mathbf{E}\|u_i^{N,M}(s,\cdot)-u_i^M(s,\cdot)\|^2_{L_2(\mathbb{R}^d)}dt\right],\ i\in\{1,2\},$$

$$\epsilon(N)=\min\{\epsilon_1(N),\epsilon_2(N),\epsilon_3(N)\},\ \lim_{N\to\infty}\epsilon(N)=0.$$

An application of Bellman-Gronwalls inequality yields

$$\sup_{0\le t\le T}\mathbf{E}\|u_i^{N,M}(t,\cdot)-u_i^M(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\le\epsilon(N)\left(1-4\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\right)^{-1}$$

$$\times\exp\left\{\left(1-4\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\right)^{-1}\right.$$

$$\left.\times 8T\left(\sup_{0\le t\le T}\int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d}\psi^2(t,x,\xi)d\xi dx+L^2\right)\cdot T\right\},\ i\in\{1,2\}.\tag{19.67}$$

Since $\lim_{N\to\infty}\epsilon(N)=0$, then (19.67) obviously implies (19.59).

9.2. Finally let us prove (19.60). In order to do it we will estimate $\sup_{0\le t\le T}\mathbf{E}\|u_i^M(t,\cdot)-U_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$, $i\in\{1,2\}$. We have

$$\sup_{0\le t\le T}\mathbf{E}\|u_i^M(t,\cdot)-U_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}$$

$$\le 4\sup_{0\le t\le T}\mathbf{E}\left\|\int\limits_{\mathbb{R}^d}|b_i(t,\cdot,U_i(t-r,\xi),\xi)-b_i(t,\cdot,u_i^M(t-r,\xi),\xi)|d\xi\right\|^2_{L_2(\mathbb{R}^d)}$$

$$+4\sup_{0\le t\le T}\mathbf{E}\left\|\int\limits_0^t A_N\left(\int\limits_{\mathbb{R}^d}\big(b_i(s,\cdot,u_i^M(s-r,\xi),\xi)\right.\right.$$

$$\left.\left.-b_i(s,\cdot,U_i(s-r,\xi),\xi)\big)d\xi\right)ds\right\|^2_{L_2(\mathbb{R}^d)}$$

$$+4\sup_{0\le t\le T}\mathbf{E}\left\|\int\limits_0^t S(t-s)|f_i(s,u_i^M(s-r,\cdot),\cdot)-f_i(s,U_i(s-r,\cdot),\cdot)|ds\right\|^2_{L_2(\mathbb{R}^d)}$$

$$+4\sup_{0\le t\le T}\mathbf{E}\Big\|\int_0^t S(t-s)\sigma(s,\cdot)dW_M(s,\cdot)-\int_0^t S(t-s)\sigma(s,\cdot)dW(s,\cdot)\Big\|^2_{L_2(\mathbb{R}^d)}$$

$$=\sum_{j=1}^4 S_j^{(u-U)},\ i\in\{1,2\}. \tag{19.68}$$

Using preceding calculations, we obtain

$$S_1^{(u-U)}=4\sup_{0\le t\le T}\mathbf{E}\int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d}|b_i(t,x,U_i(t-r,\xi),\xi)-b_i(t,x,u_i^M(t-r,\xi),\xi)|d\xi\Big)^2dx$$

$$\le 4\Big(\sup_{0\le t\le T}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\Big)\sup_{0\le t\le T}\mathbf{E}\|U_i(t,\cdot)-u_i^M(t,\cdot)\|^2_{L_2(\mathbb{R}^d)},$$

$$i\in\{1,2\}, \tag{19.69}$$

$$S_2^{(u-U)}=4\sup_{0\le t\le T}\mathbf{E}\int_{\mathbb{R}^d}\Big(\int_0^t\Delta_x S(t-s)\Big(\int_{\mathbb{R}^d}\big(b_i(s,x,u_i^M(s-r,\xi),\xi)$$

$$-b_i(s,x,U_i(s-r,\xi),\xi)\big)d\xi\Big)ds\Big)^2dx$$

$$\le 4T\sup_{0\le t\le T}\mathbf{E}\int_0^t\int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d}\|D_x^2b_i(s,x,u_i^M(s-r,\xi),\xi)$$

$$-D_x^2b_i(s,x,U_i(s-r,\xi),\xi)\|d\xi\Big)^2dxds$$

$$\le 4T\Big(\sup_{0\le t\le T}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\psi^2(t,x,\xi)d\xi dx\Big)\int_0^T\sup_{0\le s\le t}\mathbf{E}\|u_i^M(s,\cdot)-U_i(s,\cdot)\|^2_{L_2(\mathbb{R}^d)}dt,$$

$$i\in\{1,2\}. \tag{19.70}$$

$$S_3^{(u-U)}=4\sup_{0\le t\le T}\mathbf{E}\int_{\mathbb{R}^d}\Big(\int_0^t S(t-s)|f_i(s,u_i^M(s-r,x),x)-f_i(s,U_i(s-r,x),x)|ds\Big)^2dx$$

$$\le 4L^2T\int_0^T\sup_{0\le s\le t}\mathbf{E}\|u_i^M(s,\cdot)-U_i(s,\cdot)\|^2_{L_2(\mathbb{R}^d)}dt,\ i\in\{1,2\}. \tag{19.71}$$

$$
\begin{aligned}
S_4^{(u-U)} &= 4 \sup_{0\le t\le T} \mathbf{E} \int_{\mathbb{R}^d} \Bigg(\sum_{j=1}^{M} \sqrt{\lambda_j} \Bigg(\int_0^t S(t-s)\sigma(s,x)d\beta_j(s) \Bigg) e_j(x) \\
&\quad - \sum_{j=1}^{\infty} \sqrt{\lambda_j} \Bigg(\int_0^t S(t-s)\sigma(s,x)d\beta_j(s) \Bigg) e_j(x) \Bigg)^2 dx \\
&= 4 \sup_{0\le t\le T} \mathbf{E} \int_{\mathbb{R}^d} \Bigg(\sum_{j=M+1}^{\infty} \sqrt{\lambda_j} \Bigg(\int_0^t S(t-s)\sigma(s,x)d\beta_j(s) \Bigg) e_j(x) \Bigg)^2 dx \\
&\le 4 \Bigg(\sum_{j=M+1}^{\infty} \lambda_j \Bigg) \sup_{0\le t\le T} \mathbf{E} \int_{\mathbb{R}^d} \int_0^t \big(S(t-s)\sigma(s,x) \big)^2 ds dx \\
&= 4 \Bigg(\sum_{j=M+1}^{\infty} \lambda_j \Bigg) \sup_{0\le t\le T} \mathbf{E} \int_0^t \| S(t-s)\sigma(s,\cdot) \|^2_{L_2(\mathbb{R}^d)} ds \\
&\le 4T \Bigg(\sum_{j=M+1}^{\infty} \lambda_j \Bigg) \sup_{0\le t\le T} \mathbf{E} \| \sigma(t,\cdot) \|^2_{L_2(\mathbb{R}^d)}. \qquad (19.72)
\end{aligned}
$$

Estimates (19.69)–(19.72) with property (19.3) help us conclude from (19.68) that

$$
\begin{aligned}
\sup_{0\le t\le T} \mathbf{E} \| u_i^M(t,\cdot) - U_i(t,\cdot) \|^2_{L_2(\mathbb{R}^d)} &\le \Bigg(1 - 4 \sup_{0\le t\le T} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} l^2(t,x,\xi) d\xi dx \Bigg)^{-1} \\
&\times \Bigg(4T \Bigg(\sup_{0\le t\le T} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \psi^2(t,x,\xi) d\xi dx \Bigg) \int_0^T \sup_{0\le s\le t} \mathbf{E} \| u_i^M(s,\cdot) - U_i(s,\cdot) \|^2_{L_2(\mathbb{R}^d)} dt \\
&+ 4L^2 T \int_0^T \sup_{0\le s\le t} \mathbf{E} \| u_i^M(s,\cdot) - U_i(s,\cdot) \|^2_{L_2(\mathbb{R}^d)} dt \\
&+ 4T \Bigg(\sum_{j=M+1}^{\infty} \lambda_j \Bigg) \sup_{0\le t\le T} \mathbf{E} \| \sigma(t,\cdot) \|^2_{L_2(\mathbb{R}^d)} \Bigg), \; i \in \{1,2\}.
\end{aligned}
$$

Finally Bellman-Gronwalls inequality leads to

$$\sup_{0\le t\le T}\mathbf{E}\|u_i^M(t,\cdot)-U_i(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\le\left(1-4\sup_{0\le t\le T}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\right)^{-1}$$

$$\times 4T\left(\sum_{j=1}^{M}\lambda_j\right)\sup_{0\le t\le T}\mathbf{E}\|\sigma(t,\cdot)\|^2_{L_2(\mathbb{R}^d)}\exp\left\{\left(1-4\sup_{0\le t\le T}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}l^2(t,x,\xi)d\xi dx\right)^{-1}\right.$$

$$\left.\times 4T\left(\sup_{0\le t\le T}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\psi^2(t,x,\xi)d\xi dx+L^2\right)\cdot T\right\},\ i\in\{1,2\}.$$

Recall that $\sum_{n=1}^{\infty}\lambda_n<\infty$. This fact along with estimate above implies (19.60), thereby completing the proof of the theorem. □

References

1. Apartim, De.: Hille Yosida Theorem and Some Applications. CEU eTD. Collection, Department of Mathematics and its Applications, Central European University, Budapest, 86 pp.
2. Curtain, R.F., Pritchard, A.J.: Infinite Dimensional Linear Systems Theory, vol. 8. Lecture Notes in Control and Information Sciences. Springer, Berlin (1978)
3. Galcuk, L.I., Davis M.H.A.: A note on a comparison theorem for equations with different diffusions. Stochastics **6**, 147–149 (1982)
4. Geib, C., Manthey, R.: Comparison theorems for stochastic differential equations in finite and infinite dimensions. Stoch. Process. Appl. **53**(1), 23–35 (1994)
5. Huang, Z.Y.: A comparison theorem for solutions of stochastic differential equations and its applications. Proc. Proc. Am. Math. Soc. **91**(4), 611–617 (1984)
6. Kotelenz, P.: Comparison methods for a class of function valued stochastic partial differential equations. Probab. Theory Relat. Fields **93**, 1–19 (1992)
7. Monthey, R.: Stochastic evolution equations in $L^{\rho}_{2\nu}(\mathbb{R}^d)$. Stoch. Rep. **66**, 37–85 (1998)
8. OBrien, G.L.: A new comparison theorem for solutions of stochastic differential equations. Stochastics **3**, 245–249 (1980)
9. Ouknine, Y.: Comparison et non-confluence des solutions dequations differentielles stochasuques unidimensionnelles. Probab. Math. Stat. **11**(1), 37–46 (1990)
10. Skorokhod, A.V.: Issledovaniya po teorii sluchainyh processov [Research on the Theory of Random Processes], 216 pp. Kiev University, Kiev (1961)
11. Watanabe, S., Ikeda, N.: Stokhasticheskie differencial'nye uravneniya i diffusionnye processy [Stochastic Differential Equations and Diffusional Processes], 445 pp. Nauka, Moscow (1986)
12. Yamada, T.: On a comparison theorem for solutions of stochastic differential equations and its applications. J. Math. Kyoto Univ. **13**(3), 497–512 (1973)
13. Yamada, T.: On the strong comparison theorems for solutions of stochastic differential equations. Z. Vahrsch. Verw. Gebiete **56**, 3–19 (1981)

Chapter 20
Maximum Sets of Initial Conditions in Practical Stability and Stabilization of Differential Inclusions

Volodymyr V. Pichkur

Abstract In this work we consider the problem of practical stability of differential inclusion solutions on the basis of the maximum sets of practical stability concept. On one hand we propose results concerning nonlinear differential inclusion including both topological properties of the maximum sets of initial conditions for four types of practical stability (internal, weak internal, external, weak external) and the necessary and sufficient conditions of internal practical stability using the optimal Lyapunov function. On the other hand we offer the analytical forms of the maximum sets of initial conditions representation in the linear differential inclusion case. In the last section we consider the problem of practical stabilization.

20.1 Introduction

Practical stability studies behavior of dynamic systems under state constrains on a finite time interval. The concept of this theory was for the first time introduced in [1, 2]. The main direction of research consisted in developing stability technique taking into account features of the problem. The second Lyapunov method has been generalized due to [3–5]. Sufficient conditions of practical stability have been obtained for different types of practical stability and different classes of dynamic systems. In particular cases the necessary conditions have been also proved.

In [4] the concept of extremal sets of initial conditions on the basis of directional stability has been proposed and effective numerical methods for the problem of charge particles beam optimization have been developed. In works [6–10] this approach has been generalized. The topological properties of maximum sets of practical stability have been studied including the systems with set-valued righthand part. In the case of linear systems and inclusions under convex phase constraints

V. V. Pichkur (✉)
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*,
Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_20

different techniques (based on Minkowski functions, support functions etc.) in order to describe the optimal sets have been used. In [11] the necessary and sufficient conditions of practical stability via Lyapunov functions has been obtained. The problem of practical stabilization have been studied in [3, 4, 11–13]. Different problems concerning qualitative analysis and control for differential inclusions and other classes of dynamic systems have been investigated in [14–23].

In this paper we introduce the following basic notations: $\mathbb{R}^n$ is an n-dimensional Euclidean space; $\langle x, y\rangle$ is the usual inner product of $x, y \in \mathbb{R}^n$, $\|x\| = \sqrt{\langle x, x\rangle}$; $K_r(a)$ is the ball in $\mathbb{R}^n$, $K_r(a) = \{x \in \mathbb{R}^n : \|x - a\| \le r\}$, $S = \{x \in \mathbb{R}^n : \|x\| = 1\}$, and $E(a, Q)$ is the ellipsoid in $\mathbb{R}^n$,

$$E(a, Q) = \left\{x \in \mathbb{R}^n : \left\langle Q^{-1}(x - a), x - a\right\rangle \le 1\right\},$$

where Q is symmetric positive definite $n \times n$-matrix, $a \in \mathbb{R}^n$, $r > 0$; M^* is the transpose of $n \times m$-matrix M; $graph\ f$ is the graph of a mapping f; $\overline{co}_\psi\ f$ is the convex hull of a function f with respect to ψ; $intA$, ∂A are respectively the set of inner points and the boundary, $A \subset \mathbb{R}^n$; $comp(\mathbb{R}^n)$ $(conv(\mathbb{R}^n))$ is the set of nonempty (convex) compact subsets of $\mathbb{R}^n$; $\alpha(A, B)$ is the Hausdorf distance for $A, B \subset \mathbb{R}^n$, $\|A\| = \alpha(A, 0)$, $c(A, \psi) = \sup_{x\in A}\langle x, \psi\rangle$ is the support function, $\psi \in \mathbb{R}^n$. Let $A \subset \mathbb{R}^n$ be a star shaped set, $0 \in intA$. We denote by $m(A, x) = \inf\{\lambda > 0 : x \in \lambda A\}$ the Minkowski function, and by $k(A, x) = \sup\{\lambda > 0 : \lambda x \in A\}$ the inverse Minkowski function of the set A, $x \in \mathbb{R}^n$. If the inverse Minkowski function is defined only on the unit sphere S, we will call it the deformation function of the set A.

We consider differential inclusion

$$\frac{dx}{dt} \in F(x, t), \tag{20.1}$$

where $x \in \mathbb{R}^n$ is an n-dimensional vector of phase coordinate, $(x, t) \in D$, $D \subset \mathbb{R}^{n+1}$ is a bounded domain. The set-valued mapping $F : D \to conv(\mathbb{R}^n)$ is measurable with respect to variable t and satisfies the Lipschitz condition

$$\alpha(F(x, t), F(y, t)) \le L(t)\|x - y\|.$$

Here $L(t)$ is a positive integrable function, $(x, t) \in D$, $(y, t) \in D$, $F(0, t) = 0$, $(0, t) \in D$. The map F is integrably bounded. It means that there exists an integrable positive function $\lambda(\cdot)$ so that

$$F(x, t) \subseteq \lambda(t)K_1(0),\ (x, t) \in D.$$

Let $x(t, z, s)$ be a solution of (20.1) corresponding to the Cauchy condition $x(s) = z$, $\Omega(\cdot, z, s)$ be a solutions set, and $X(t, z, s)$ be an attainability set of (20.1), where $x(s) = z$. The attainability set $X(t, z, s)$ generates set-valued mappings $X(\cdot, z, s) : t \mapsto X(t, z, s)$, $X(t, \cdot, s) : z \mapsto X(t, z, s)$, $(z, s) \in D$. A multifunction

$\Phi : [t_0, T] \to comp(\mathbb{R}^n)$ prescribes state constraints, graph of the mapping Φ belongs to D, $0 \in int\Phi(t)$, $t \in [t_0, T]$, $G_0 \subset \mathbb{R}^n$ is a set of initial conditions, $0 \in G_0$.

We define four modes of practical stability of differential inclusion (20.1): internal, external, weakly internal, and weakly external.

Definition 20.1 We say, that the equilibrium position $x(t) = 0$ of differential inclusion (20.1) is $\{G_0, \Phi(t), t_0, T\}$—internally stable, if arbitrary solution $x(t, x_0, t_0)$ of (20.1) belongs to $\Phi(t)$ for any point $x_0 \in G_0$, and for all $t \in [t_0, T]$.

Definition 20.2 The solution $x(t) = 0$ of (20.1) is called weakly $\{G_0, \Phi(t), t_0, T\}$-internally stable if for arbitrary $x_0 \in G_0$ there exists a solution $x(\cdot, x_0, t_0)$ of differential inclusion (20.1) such, that $x(t, x_0, t_0) \in \Phi(t)$, $t \in [t_0, T]$.

Let $E_0 \subseteq \mathbb{R}^n$ be a set containing zero point.

Definition 20.3 If for any $x_0 \in E_0$ and for all solutions $x(\cdot, x_0, t_0)$ of differential inclusion (20.1) there exists a point $t \in [t_0, T]$ so that $x(t, x_0, t_0) \in \Phi(t)$ then the equilibrium position $x(t) = 0$ of differential inclusion (20.1) is called $\{E_0, \Phi(t), t_0, T\}$-externally stable.

Definition 20.4 The equilibrium position $x(t) = 0$ of (20.1) is said to be $\{E_0, \Phi(t), t_0, T\}$-weakly external stable if for arbitrary $x_0 \in E_0$ there exist $t \in [t_0, T]$ and a solution $x(\cdot, x_0, t_0)$ of inclusion (20.1) such that $x(t, x_0, t_0)$ belongs to $\Phi(t)$.

On the basis of Definition 20.1 we state the following problems:

1. given G_0, Φ, and $[t_0, T]$. Verify if the equilibrium position $x(t) = 0$ of differential inclusion (20.1) is $\{G_0, \Phi(t), t_0, T\}$—internally stable;
2. given Φ, $[t_0, T]$. Find the maximum set $G_* \subseteq \Phi(t_0)$ such that the equilibrium position $x(t) = 0$ of differential inclusion (20.1) is $\{G_*, \Phi(t), t_0, T\}$—internally stable;
3. given Φ, $[t_0, T]$, and the set of initial conditions $G_0(a) \subseteq \Phi(t_0)$ depends on some parameter a. Find all parameter a values so that the zero solution $x(t) = 0$ of differential inclusion (20.1) is $\{G_0(a), \Phi(t), t_0, T\}$—internally stable;
4. given G_0, $[t_0, T]$, the states constraints $\Phi(t, b)$ depends on some parameter b, $t \in [t_0, T]$. Find all parameter b values so that he equilibrium position $x(t) = 0$ of differential inclusion (20.1) is $\{G_0(a), \Phi(t, b), t_0, T\}$—internally stable;
5. given G_0, Φ, t_0. Find the maximum T such that the equilibrium position $x(t) = 0$ of differential inclusion (20.1) is $\{G_0, \Phi(t), t_0, T\}$—internally stable.

One can state the same problems for weak internal, external, and weak external modes of practical stability by a similar way. It turns out that the problem of finding the maximum set is central in the following sense: the solution of other problems is based on the properties of the maximum set.

Bellow we discuss some topological properties of the maximum sets of initial conditions for four types of practical stability (internal, weak internal, external, weak external) and the necessary and sufficient conditions of inner practical stability

using the optimal Lyapunov function. Further we offer the analytical forms of the maximum sets of initial conditions representation in the linear differential inclusion case. In the last section we consider the problem of practical stabilization.

20.2 Maximum Set of Initial Conditions: Nonlinear Case

20.2.1 Internal Practical Stability

In this subsection we describe properties of a set $G_* \subseteq \Phi(t_0)$. The set G_* consists of all points $x_0 \in \Phi(t_0)$ such that any solution $x(\cdot, x_0, t_0)$ of differential inclusion (20.1) belongs to $\Phi(t)$ for all $t \in [t_0, T]$. We will call G_* the maximum set of internal practical stability [6, 7, 9]. One can observe that Definition 20.1 takes place if $G_0 \subseteq G_*$. Consider the following assertions [6, 7, 9].

Theorem 20.1 *Suppose that the set-valued mapping Φ is closed; then G_* is compact.*

Theorem 20.2 *Assume Φ is upper semicontinuous. Then $x_0 \in \partial G_*$ if and only if $X(t, x_0, t_0) \subseteq \Phi(t)$, $t \in [t_0, T]$ and $tube X(\cdot, x_0, t_0) \cap tube\Phi \neq \emptyset$.*

Here $tube\Phi$ denotes a tube of the mapping Φ. It consists of all points $z \in \partial graph\Phi$ such that there exists a sequence $z_k \in \mathbb{R}^n \times [t_0, T]$, $z_k \notin graph\Phi$ tending to z as $k \to \infty$.

Corollary 20.1 *Let the set-valued mapping Φ be upper semicontinuous and quasi-open; then $x_0 \in \partial G_*$ if and only if $X(t, x_0, t_0) \subseteq \Phi(t)$ for all $t \in [t_0, T]$ and there exists a solution $x(\cdot, x_0, t_0)$ of differential inclusion (20.1) and $s \in [t_0, T]$ so that $x(s, x_0, t_0) \in \partial\Phi(s)$.*

Notice that under the corollary conditions $x_0 \in intG_*$ iff $X(t, x_0, t_0) \subset int\Phi(t)$, $t \in [t_0, T]$. Suppose that the mapping Φ is continuous and quasi-open, $t \in [t_0, T]$. Consider a continuous function $\varphi \in C(\mathbb{R}^n \times [t_0, T])$ such that $\varphi(y, t) < 1$, $y \in int\Phi(t)$; $\varphi(y, t) = 1$ if $y \in \partial\Phi(t)$; $\varphi(y, t) > 1$, $y \notin \Phi(t)$.

Theorem 20.3 *The function*

$$\xi(z) = \max_{t \in [t_0, T]} \max_{y \in X(t, z, t_0)} \varphi(y, t)$$

is continuous, $\xi(z) < 1$, $z \in intG_*$; $\xi(z) = 1$ if $z \in \partial G_*$; $\xi(z) > 1$, $z \notin G_*$, $t \in [t_0, T]$.

Define the set-valued function Φ on a set $[\bar{t}, \bar{T}]$. We assume as above that Φ is continuous and quasi-open. It is clear that the optimal set of internal practical stability depends on points t_0, T. Denote by θ the map of $[\bar{t}, \bar{T}] \times [\bar{t}, \bar{T}]$ to $comp(\mathbb{R}^n)$ such that $\theta(t_0, T) = G_*$ for all $t_0, T \in [\bar{t}, \bar{T}]$, $t_0 < T$.

Theorem 20.4 *The map θ is continuous.*

Consider a parametric class of multifunctions $\Phi(u)$ taking each $t \in [t_0, T]$ to a compact set $\Phi(t, u)$ in $\mathbb{R}^n$, $u \in U$. Here $U \subseteq \mathbb{R}^m$ is a closed domain, $graph\Phi(u) \subset D$. In this case the optimal set of practical stability under the phase constraints given by $\Phi(u)$ depends on $u \in U$. We denote this set by $G_*(u)$. Suppose that the set-valued function $(t, u) \mapsto \Phi(t, u)$ is continuous in u and upper semicontinuous with respect to t, $0 \in \Phi(t, u)$, $t \in [t_0, T]$, $u \in U$. If $graph\Phi(u) \subseteq graph\Phi(v)$, then $G_*(u) \subseteq G_*(v)$, $u \in U$, $v \in U$. We propose the following statements.

Theorem 20.5 *The set-valued mapping $F : u \mapsto G_*(u)$ is continuous, $u \in U$. If there exists $\varepsilon > 0$ such that $\Phi(t, u) + \varepsilon K_1(0) \subseteq \Phi(t, v)$, $t \in [t_0, T]$, then $G_*(u) \subset intG_*(v)$, $u \in U$, $v \in U$.*

Theorem 20.6 *Suppose $graph\Phi(u) \subseteq graph\Phi(v)$, $tube\Phi(u) \cap tube\Phi(v) = \emptyset$; then $G_*(u) \subset intG_*(v)$, $u \in U$, $v \in U$.*

Now we discuss the necessary and sufficient condition of internal practical stability via the Lyapunov function [11].

Theorem 20.7 *The zero solution of differential inclusion (20.1) is $\{G_0, \Phi(t), t_0, T\}$—internally stable, if and only if there exists a continuous function $V : D \to \mathbb{R}^1$ such that the following conditions take place:*

1)

$$G_0 \subseteq \left\{x \in \mathbb{R}^n : V(x, t_0) \leq 1\right\};$$

2)

$$\left\{x \in \mathbb{R}^n : V(x, t) \leq 1\right\} \subseteq \Phi(t),\ t \in [t_0, T];$$

3) $V(x(t), t)$ is nonincreasing function, where $x(t)$ is a solution of differential inclusion (20.1).

The function $V(x, t)$ in Theorem 20.7 is called the Lyapunov function. One of the possibilities to construct the Lyapunov function is as follows [11]. Consider a function $g : \mathbb{R}^n \to \mathbb{R}^1$ having the following properties: $g(x) < 1$, if $x \in intG_*$; $g(x) = 1$, if $x \in \partial G_*$; $g(x) > 1$, $x \notin G_*$. For instance, $g(x) = 1 - \rho(x)$, if $x \in G_*$; $1 + \rho(x)$, if $x \notin G_*$ satisfies the requirements. Here $\rho(x) = \rho(x, \partial G_*)$ is the distance from $x \in \mathbb{R}^n$ to ∂G_*. The function

$$V(y, t) = \min_{x(\cdot) \in \Omega(\cdot, y, t)} g(x(t_0)) = \min_{z \in X(t_0, y, t)} g(z)$$

satisfies conditions 1–3 of Theorem 20.7. Since we use the optimal set of practical stability G_* to construct $V(y, t)$, we call such a function the optimal Lyapunov function.

Another way to find $V(y,t)$ satisfying Theorem 20.7 conditions consists of using the set G_0 instead of G_*. Let $G_0 \in comp(\mathbb{R}^n)$. Define a continuous function $g_0 : \mathbb{R}^n \to \mathbb{R}^1$ such that $g_0(x) < 1$, if $x \in intG_0$; $g_0(x) = 1$, if $x \in \partial G_0$; $g_0(x) > 1$, $x \notin G_0$. To prove the necessity of Theorem 20.7 we can use the function

$$V(y,t) = \min_{x(\cdot)\in\Omega(\cdot,y,t)} g_0(x(t_0)) = \min_{z\in X(t_0,y,t)} g_0(z).$$

For instance, if $G_0 = K_r(0)$, then

$$V(y,t) = \min_{z\in X(t_0,y,t)} \|z\| + r - 1$$

Corollary 20.2 *Suppose that there exists a continuously differentiable function $V : D \to \mathbb{R}^1$ such that conditions 1, 2 of Theorem 20.7 hold and upper derivative due to differential inclusion (20.1)*

$$\left(\frac{dV}{dt}\right)_{(20.1)} = \frac{\partial V(x,t)}{\partial t} + \max_{v\in F(x,t)} \langle grad_x V(x,t), v\rangle \le 0$$

on $\{(x,t) \in D : V(x,t) \ge 1\}$. Then the zero solution of differential inclusion (20.1) is $\{G_0, \Phi(t), t_0, T\}$—internally stable.

20.2.2 Weak Internal Practical Stability

In this subsection we offer the main properties of optimal set of weak internal practical stability $I_* \subseteq \Phi(t_0)$ [6, 7, 9]. This set consists of all initial points $x_0 \in \Phi(t_0)$ such that there exists a solution $x(\cdot, x_0, t_0)$ of differential inclusion (20.1) belonging to $\Phi(t)$ for all $t \in [t_0, T]$. Obviously if $G_0 \subseteq I_*$, then Definition 20.2 is true.

Assume that the mapping Φ is upper semicontinuous [6, 7, 9].

Theorem 20.8 *The set I_* is contained in $comp(\mathbb{R}^n)$.*

Theorem 20.9 *Suppose that $x_0 \in \partial I_*$ and $x = x(\cdot, x_0, t_0) \in X(\cdot, x_0, t_0)$ is a solution of differential inclusion (20.1) such that $x(t, x_0, t_0) \in \Phi(t)$, $t \in [t_0, T]$; then $graph\, x \cap tube\Phi \ne \emptyset$.*

Definition 20.5 We will say that $x = x(\cdot, z_0, t_0) \in X(\cdot, z_0, t_0)$ is consistent with Φ if $graph\, x \subset graph\Phi$ and there exists a sequence $z_k \to z_0$, $k = 1, 2, \ldots$ such that $graph\, x_k / graph\Phi \ne \emptyset$ for any $x_k \in \Omega(\cdot, z_k, t_0)$, $k = 1, 2, \ldots$.

Theorem 20.10 *Suppose $x_0 \in I_*$ and the following conditions hold:*

1. *$graph\, x \cap tube\Phi \ne \emptyset$ for all $x = x(\cdot, x_0, t_0) \in \Omega(\cdot, x_0, t_0)$ such that $x(t, x_0, t_0) \in \Phi(t)$, $t \in [t_0, T]$;*

2. *there exists a consistent with* Φ *solution* $x(\cdot, x_0, t_0) \in \Omega(\cdot, x_0, t_0)$ *of differential inclusion* (20.1)*;*

then $x_0 \in \partial I_*$.

20.2.3 Weak External Practical Stability

We consider the set $E_* \subseteq \mathbb{R}^n$ of all points $x_0 \in \mathbb{R}^n$ such that there is a solution $x(\cdot, x_0, t_0)$ to differential inclusion (20.1) and a time $t \in [t_0, T]$ such that $x(t, x_0, t_0)$ lies in $\Phi(t)$. The set E_* is called the maximum set of weak external practical stability [6, 7, 9].

Theorem 20.11 *Let* Φ *be upper semicontinuous; then the set* E_* *is compact.*

Theorem 20.12 *Suppose* Φ *is upper semicontinuous and* $X(t, x_0, t_0) \cap int\Phi\ (t) \neq \emptyset$ *in a time* $t \in [t_0, T]$*; then* $x_0 \in int E_*$.

Denote $I(\Phi) = graph\Phi / tube\Phi$ and assume that for arbitrary $z \in tube\Phi$ any neighborhood of z contains a point from $I(\Phi)$.

Theorem 20.13 *Let* Φ *be upper semicontinuous and quasi-open. If* $x_0 \in int E_*$*, then* $graph\, x(\cdot) \cap I(\Phi) \neq \emptyset$ *for all* $x(\cdot) \in \Omega(\cdot, x_0, t_0)$.

Corollary 20.3 *Let* Φ *be upper semicontinuous and quasi-open. The initial point* $x_0 \in int E_*$ *if and only if* $graph\, x(\cdot) \cap I(\Phi) \neq \emptyset$ *for all* $x(\cdot) \in \Omega(\cdot, x_0, t_0)$.

20.2.4 External Practical Stability

We denote by D_* the maximum set of external practical stability [6, 7, 9]. The set D_* consists of all points $x_0 \in \mathbb{R}^n$ such that for any solution $x(\cdot, x_0, t_0)$ of differential inclusion (20.1) there exists a time $t \in [t_0, T]$ such that $x(t, x_0, t_0) \in \Phi(t)$. Let Φ be upper semicontinuous

Theorem 20.14 *The set* D_* *is compact. If* $x_0 \in \partial D_*$*, then there exists a solution* $x(\cdot) \in \Omega(\cdot, x_0, t_0)$ *to differential inclusion* (20.1) *such that* $graph\, x(\cdot) \cap graph\Phi \neq \emptyset$*,* $graph\, x(\cdot) \cap I(\Phi) = \emptyset$.

Definition 20.6 We say that $x = x(\cdot, z_0, t_0) \in \Omega(\cdot, z_0, t_0)$ is consistent with the mapping Φ exterior if $graph\, x \cap graph\Phi \neq \emptyset$ and there exists a sequence $x_k \in \Omega(\cdot, z_k, t_0)$ tending uniformly to x, $k = 1, 2, \ldots$, and $graph\, x_k \cap graph\Phi = \emptyset$.

Theorem 20.15 *Assume that* $x_0 \in D_*$ *and there exists a consistent with the mapping* Φ *exterior solution* $x = x(\cdot, x_0, t_0) \in \Omega(\cdot, x_0, t_0)$ *of differential inclusion* (20.1) *such that* $graph\, x(\cdot) \cap graph\Phi \neq \emptyset$, $graph\, x(\cdot) \cap I(\Phi) = \emptyset$; *then* $x_0 \in \partial D_*$.

20.3 Maximum Set of Initial Conditions: Linear Case

In this section we consider linear differential inclusion

$$\frac{dx}{dt} \in A(t)x + U(t), \tag{20.2}$$

where $x \in \mathbb{R}^n$ is an n-dimensional vector of phase coordinate, $A(t)$ is a measurable integrally bounded $n \times n$-matrix, $U : [t_0, T] \to comp(\mathbb{R}^n)$ is a measurable bounded set-valued mapping, $0 \in intU(t)$. We denote by $\Theta(t, s)$ a fundamental matrix of the linear system $\frac{dx}{dt} = A(t)x$ normalized at the point s so that $\Theta(s, s) = I$, where I is the identity $n \times n$-matrix.

The attainability set of (20.2) satisfies the Cauchy-like formula [21] $X(t, x_0, t_0) = \Theta(t, t_0)x_0 + Q(t)$, and it is convex and compact. In (20.2) $Q(t) = \int_{t_0}^{t} \Theta(t, s)U(s)ds$, where integral is considered in Aumann meaning. A continuous multifunction $\Phi : [t_0, T] \to conv(\mathbb{R}^n)$ prescribes phase constraints, $Q(t) \subset int\Phi(t)$, $t \in [t_0, T]$.

Let us denote by $\Sigma(U)$ the set of measurable integrable selections of the mapping U, $q(t, u) = \int_{t_0}^{t} \Theta(t, s)u(s)ds$, $u \in \Sigma(U)$, $P(\ell) = \{\psi \in S : \langle \Theta^*(t, t_0)\psi, \ell \rangle > 0\}$, $\ell \in S$. The following assertion is true [6, 7, 9].

Theorem 20.16 *The maximum set of internal stability* G_* *is convex and* $0 \in intG_*$. *The support function, the Minkowski function, and the deformation function of the set* G_* *are given as follows:*

$$c(G_*, \psi) = \overline{co}_{\psi} \min_{t \in [t_0, T]} c(\Phi(t) \overset{*}{-} Q(t), z(t, \psi)), \psi \in \mathbb{R}^n,$$

$$\frac{dz(t, \psi)}{dt} = -A^*(t)z(t, \psi), \; z(t_0, \psi) = \psi, \; t \in [t_0, T]; \tag{20.3}$$

$$m(G_*, x_0) = \max_{t \in [t_0, T]} \max_{\psi \in S} \frac{\langle \Theta^*(t, t_0)\psi, x_0 \rangle}{c(\Phi(t), \psi) - c(Q(t), \psi)}, \; x_0 \in \mathbb{R}^n;$$

$$k(G_*, \ell) = \min_{t \in [t_0, T]} \min_{\psi \in P(\ell)} \frac{c(\Phi(t), \psi) - c(Q(t), \psi)}{\langle \Theta^*(t, t_0)\psi, \ell \rangle}, \; \ell \in S.$$

Here $\Phi(t) \stackrel{*}{-} Q(t)$ is the Minkowski difference between $\Phi(t)$ and $Q(t)$. Notice that G_* is symmetric if $U(t)$ and $\Phi(t)$ are symmetric for all $t \in [t_0, T]$.

Theorem 20.16 gives possibility to represent the maximum set of internal practical stability as follows [6, 7, 9]

$$G_* = \cap_{\psi \in S} L(\psi) = \left\{x \in \mathbb{R}^n : m(G_*, x) \leq 1\right\} = \cup_{\ell \in S} I(\ell)$$

where $L(\psi) = \{x \in \mathbb{R}^n : \langle x, \psi\rangle \leq c(G_*, \psi)\}$, $I(\ell) = \cup_{k \in [0, k(G_*, \ell)]} k\ell$. Consider a particular case $U(t) = E(0, H(t))$. Here $H(t)$ is a continuous symmetric positive definite $n \times n$-matrix. The following assertion holds [11].

Theorem 20.17 *We suppose that* $G_0 = E(0, R_0)$,

$$\min_{t \in [t_0, T]} \min_{\psi \in S} \left(c(\Phi(t), \psi) - \sqrt{\langle R(t)\psi, \psi\rangle}\right) \geq 0,$$

an $n \times n$*- matrix* $R(t)$ *is a positive definite solution of the matrix differential equation*

$$\frac{dR(t)}{dt} = A(t)R(t) + R(t)A^*(t) + qR(t) + q^{-1}H(t),\ R(t_0) = R_0,\ t \in [t_0, T],$$

$q > 0$, R_0 *is a symmetric positive definite* $n \times n$*-matrix. Then the trivial solution of differential inclusion (20.2) is* $\{G_0, \Phi(t), t_0, T\}$*—internally stable.*

Let the righthand part of differential inclusion (20.2) be time-independent. It means that $A(t) = A$, $H(t) = H$, $t \in [t_0, T]$, where A is $n \times n$—matrix, H is a symmetric positive definite $n \times n$-matrix. In this case the following statement is true [11].

Theorem 20.18 *Suppose that*

$$\min_{t \in [t_0, T]} \min_{\psi \in S} \left(c(\Phi(t), \psi) - \sqrt{\langle R\psi, \psi\rangle}\right) \geq 0,$$

an $n \times n$*- matrix* R *is positive definite and satisfies the matrix equation*

$$AR + RA^* + qR + q^{-1}H = 0,\ q > 0,$$

and $R - R_0^{-1}$ *is a nonnegative definite matrix. Here* R_0 *is a symmetric positive definite* $n \times n$*-matrix,* $G_0 = E(0, R_0)$*. Then the trivial solution of differential inclusion (20.2) is* $\{G_0, \Phi(t), t_0, T\}$*—internally stable.*

Now we discuss some properties of maximum sets of initial conditions for weak internal practical stability, weak external practical stability, and external practical stability of the trivial solution (20.2) under the state constraints given by the multifunction Φ [6, 7, 9].

Theorem 20.19 *The maximum set of weak internal stability I_* is convex and*

$$c(I_*, \psi) = \overline{co}_\psi \max_{u\in\Sigma(U)} \min_{t\in[t_0,T]} c(\Phi(t) - q(t,u), z(t,\psi)),\ \psi \in \mathbb{R}^n;$$

$$m(I_*, x_0) = \min_{u\in\Sigma(U)} \max_{t\in[t_0,T]} \max_{\psi\in S} \frac{\langle \Theta^*(t,t_0)\psi, x_0\rangle}{c(\Phi(t) - q(t,u), \psi)},\ x_0 \in \mathbb{R}^n;$$

$$k(I_*, \ell) = \max_{u\in\Sigma(U)} \min_{t\in[t_0,T]} \min_{\psi\in P(\ell)} \frac{c(\Phi(t) - q(t,u), \psi)}{\langle \Theta^*(t,t_0)\psi, \ell\rangle},\ \ell \in S.$$

Here $z(t,\psi)$ satisfies (20.3).

If $U(t)$ and $\Phi(t)$ are symmetric, $t \in [t_0, T]$, then I_* is symmetric.

Theorem 20.20 *The maximum set of weak external practical stability E_* is star shaped. The support function, the Minkowski function, and the deformation function of E_* are given as follows:*

$$c(E_*, \psi) = \overline{co}_\psi \max_{t\in[t_0,T]} c(\Phi(t) + (-1)\cdot Q(t), z(t,\psi)),\ \psi \in \mathbb{R}^n;$$

$$m(E_*, x_0) = \min_{t\in[t_0,T]} \max_{\psi\in S} \frac{\langle \Theta^*(t,t_0)\psi, x_0\rangle}{c(\Phi(t),\psi) - c(Q(t),\psi)},\ x_0 \in \mathbb{R}^n;$$

$$k(E_*, \ell) = \max_{t\in[t_0,T]} \min_{\psi\in P(\ell)} \frac{c(\Phi(t),\psi) - c(Q(t),\psi)}{\langle \Theta^*(t,t_0)\psi, x_0\rangle},\ \ell \in S.$$

Here $z(t,\psi)$ satisfies (20.3).

As it was in the previous case if $U(t)$ and $\Phi(t)$ are symmetric, $t \in [t_0, T]$, then E_* is symmetric.

Theorem 20.21 *The maximum set of external practical stability D_* is star shaped and*

$$c(D_*, \psi) = \overline{co}_\psi \min_{u\in\Sigma(U)} \max_{t\in[t_0,T]} c(\Phi(t) - q(t,u), z(t,\psi)),\ \psi \in \mathbb{R}^n;$$

$$m(D_*, x_0) = \max_{u\in\Sigma(U)} \min_{t\in[t_0,T]} \max_{\psi\in S} \frac{\langle \Theta^*(t,t_0)\psi, x_0\rangle}{c(\Phi(t) - q(t,u), \psi)},\ x_0 \in \mathbb{R}^n;$$

$$k(D_*, \ell) = \min_{u\in\Sigma(U)} \max_{t\in[t_0,T]} \min_{\psi\in P(\ell)} \frac{c(\Phi(t) - q(t,u), \psi)}{\langle \Theta^*(t,t_0)\psi, \ell\rangle},\ \ell \in S.$$

Here $z(t,\psi)$ satisfies (20.3).

20.4 Internal Practical Stabilization

We consider differential inclusion

$$\frac{dx}{dt} \in F(x,t) + G(t)u(x,t), \tag{20.4}$$

where $(x,t) \in D$, $D \subset \mathbb{R}^{n+1}$ is a bounded domain, a set-valued mapping $F : D \to conv(\mathbb{R}^n)$ is measurable in t, upper semicontinuous in x, $F(0,t) = 0$, $(0,t) \in D$ and integrably bounded on D. Further, $G(t)$ is integrable $n \times m$-matrix, $u(x,t)$ is an m- dimensional control function, $u(0,t) = 0$. We assume that $u(x,t)$ is integrably bounded on D, continuous with respect to variable x being measurable with respect to t.

A multifunction $\Phi : [t_0, T] \to comp(\mathbb{R}^n)$ prescribes phase constraints, the graph of the mapping Φ belongs to D, $0 \in int\Phi(t)$, $t \in [t_0, T]$, $G_0 \subset \Phi(t_0)$. Suppose, that there exists a continuously differentiable function $V : D \to \mathbb{R}^1$ such that

$$\Phi(t) = \left\{x \in \mathbb{R}^n : V(x,t) \le 1\right\}, \ t \in [t_0, T],$$

and $V(0,t) = 0$, $grad_x V(0,t) = 0$ on $[t_0, T]$. We assume that the maximum set of internal practical stability of the zero solution of (20.1) does not contain G_0. The problem of $\{G_0, \Phi(t), t_0, T\}$—internal stabilization for differential inclusion (20.4) consists of finding the admissible control function $u(x,t)$ such that the zero solution to (20.4) is $\{G_0, \Phi(t), t_0, T\}$—internally stable.

Theorem 20.22 ([11]) *Suppose that $W(x,t)$ is a continuous nonnegative function on D and*

$$|W(x,t) + \frac{\partial V(x,t)}{\partial t} + c(F(x,t), grad_x V(x,t))| \le C \left\| G^*(t) grad_x V(x,t) \right\|^2,$$

$C > 0$. Then the control function

$$u(x,t) = \begin{cases} (k(x,t)I + P)G^*(t)grad_x V(x,t), & \text{if } G^*(t)grad_x V(x,t) \neq 0, \\ 0, & \text{if } G^*(t)grad_x V(x,t) = 0 \end{cases} \tag{20.5}$$

solves the problem of $\{G_0, \Phi(t), t_0, T\}$—internal stabilization for differential inclusion (20.4). Here $P = -P^*$ is an arbitrary $m \times m$—matrix,

$$k(x,t) = -\frac{W(x,t) + \frac{\partial V(x,t)}{\partial t} + c(F(x,t), grad_x V(x,t))}{\|G^*(t) grad_x V(x,t)\|^2}. \tag{20.6}$$

Consider linear differential inclusion

$$\frac{dx}{dt} \in \Omega(t)x + G(t)u(x,t). \tag{20.7}$$

Here $\Omega : [t_0, T] \to comp(\mathbb{R}^{n\times n})$ is a measurable integrably bounded multifunction. In other words, there exists an integrable positive function $\lambda(\cdot)$ so that $\Omega(t) \subseteq \lambda(t)B$, $t \in [t_0, T]$, where B is the unit ball in $\mathbb{R}^{n\times n}$, $\mathbb{R}^{n\times n}$ is space of $n \times n$- matrices with real components.

The solution of $\{G_0, \Phi(t), t_0, T\}$—internal stabilization problem for differential inclusion (20.7) is given by Theorem 20.22, where

$$|W(x,t) + \frac{\partial V(x,t)}{\partial t} + c(\Omega(t), grad_x V(x,t)x^*)| \le C \left\| G^*(t) grad_x V(x,t) \right\|^2,$$

$$k(x,t) = -\frac{W(x,t) + \frac{\partial V(x,t)}{\partial t} + c(\Omega(t), grad_x V(x,t)x^*)}{\|G^*(t) grad_x V(x,t)\|^2}. \tag{20.8}$$

Here $c(\Omega(t), \Psi) = \max_{A\in\Omega(t)} tr(A^*\Psi)$ is a support function, $\Psi \in \mathbb{R}^{n\times n}$.

Example 20.1 Suppose that in (20.7) $G(t) = I$, $V(x,t) = \frac{1}{2}\langle Mx, x\rangle$, $W(x,t) = \langle Nx, x\rangle$,

$$\Omega(t) = \left\{ A \in \mathbb{R}^{n\times n} : tr(A^* Q^{-1} A) \le 1 \right\},$$

where Q, M, N are symmetric positive definite $n \times n$—matrices. Then from (20.5), (20.8) it follows that

$$u(x) = \begin{cases} 2(k(x)I + P)G^* Mx, & \text{if } x \ne 0, \\ 0, & \text{if } x = 0 \end{cases} \tag{20.9}$$

gives a solution of $\{G_0, \Phi(t), t_0, T\}$—internal stabilization problem for differential inclusion (20.7) where

$$k(x) = -\frac{\langle Nx, x\rangle + \|x\|\sqrt{\langle M^* QMx, x\rangle}}{\|Mx\|^2}.$$

Example 20.2 Let under Example 20.1 conditions

$$\Omega(t) = \left\{ A = (a_{ij})_{i,j=1}^n \in \mathbb{R}^{n\times n} : |a_{ij}| \le r_{ij} \right\},$$

$r_{ij} > 0$. Then the control function (20.9) solves $\{G_0, \Phi(t), t_0, T\}$—internal stabilization problem for differential inclusion (20.7) if

$$k(x) = -\frac{\langle Nx, x\rangle + \sum_{i,j=1}^{n} r_{ij}|x_j m_i^* x|}{\|Mx\|^2}.$$

Here $m_i \in \mathbb{R}^n$ is the i-th row of the matrix M, $i = 1, 2, \ldots, n$.

References

1. Chetaev, N.G.: On certain questions related to the problem of the stability of unsteady motion. J. Appl. Math. Mech. **24**, 6–19 (1960)
2. Lasalle, J., Lefshetz, S.: Stability by Lyapunov Direct Method and Application. Academic, Boston (1961)
3. Kirichenko, N.F.: Introduction to Stability Theory. Vyshcha Shkola, Kyiv (1978)
4. Bublik, B.N., Garashchenko, F.G., Kirichenko, N.F.: Structural - Parametric Optimization and Stability of Bunch Dynamics. Naukova dumka, Kyiv (1985)
5. Lakshmikantham, V., Leela, S., Martynyuk, A.A.: Practical Stability of Nonlinear Systems. World Scientific, Singapore (1990)
6. Garashchenko, F.G., Pichkur, V.V.: Analysis of optimal properties of practical stability of dynamic systems. Cybern. Syst. Anal. **38**(5), 703–715 (2002)
7. Pichkur, V.V.: Study of Practical Stability of Differential Inclusions. Taras Shevchenko National University of Kyiv, Kyiv (2005)
8. Garashchenko, F.G., Pichkur, V.V.: Properties of Optimal Sets of Practical Stability of Differential Inclusions. Part I. Part II. J. Autom. Inf. Sci. **38**(3), 1–19 (2006)
9. Bashnyakov, O.M., Garashchenko, F.G., Pichkur, V.V.: Practical Stability, Estimations and Optimization. Taras Shevchenko National University of Kyiv, Kyiv (2008)
10. Pichkur, V.V., Sasonkina, M.S.: Maximum set of initial conditions for the problem of weak practical stability of a discrete inclusion. J. Math. Sci. **194**, 414–425 (2013)
11. Pichkur, V.: On practical stability of differential inclusions using Lyapunov functions. Discrete Contin. Dyn. Syst. Ser. B **22**, 1977–1986 (2017)
12. Pichkur, V.V., Sasonkina, M.S.: Practical stabilization of discrete control systems. Int. J. Pure Appl. Math. **81**(6), 877–884 (2012)
13. Tairova, M.S., Strakhov, Y.M.: Reverse algorithm for practical stabilization of discrete inclusions. Int. J. Pure Appl. Math. **116**(2), 89–499 (2017)
14. Angeli, D., Ingalls, B., Sontag, E.D., Wang, Y.: Uniform global asymptotic stability of differential inclusions. J. Dyn. Control Syst. **10**, 391–412 (2004)
15. Aubin, J.P., Cellina, A.: Differential Inclusions. Set-Valued Maps and Viability Theory. Springer, Berlin (1984)
16. Bacciotti, A., Rosier, L.: Liapunov Functions and Stability in Control Theory. Springer, Berlin (2005)
17. Gama, R., Smirnov, G.: Quasimonotonicity, regularity and duality for nonlinear systems of partial differential equations stability and optimality of solutions to differential inclusions via averaging method. Set-Val. Var. Anal. **22**, 349–374 (2014)
18. Filippov, A.F.: Differential equations with discontinuous righthand sides and differential inclusions. In: Trenogin, V.A., Filippov, A.F. (eds.) Nonlinear Analysis and Nonlinear Differential Equations, pp. 265–288. FIZMATLIT, Moscow (2003)
19. Kapustyan, A.V., Mel'nik, V.S.: On global attractors of multivalued semidynamical systems and their approximations. Doklady Academii Nauk. **366**(4), 445–448 (1999)

20. Kapustyan, O.V., Kapustian, O.A., Sukretna, A.V.: Approximate stabilization for a nonlinear parabolic boundary-value problem. Ukr. Math. J. **63**(5), 759–767 (2011)
21. Perestyuk, N.A., Plotnikov, V.A., Samoilenko, A.M., Skripnik, N.V.: Differential Equations with Impulse Effects: Multivalued Right-hand Sides with Discontinuities. Walter De Gruyter, Berlin (2011)
22. Smirnov, G.: Introduction to the Theory of Differential Inclusions. American Mathematical Society, Providence (2002)
23. Veliov, V.: Stability-like properties of differential inclusions. Set-Valued Anal. **5**, 73–88 (1997)

Part IV
Modern Methods of Optimization and Control Sciences for Continuum Mechanics

Chapter 21
Asymptotic Translation Uniform Integrability and Multivalued Dynamics of Solutions for Non-autonomous Reaction-Diffusion Equations

Michael Z. Zgurovsky, Pavlo O. Kasyanov, Nataliia V. Gorban, and Liliia S. Paliichuk

Abstract In this note we introduce asymptotic translation uniform integrability condition for a function acting from a positive semi-axes of time-line to a Banach space. We prove that this condition is equivalent to uniform integrability condition. As a result, we obtain the corollaries for the multivalued dynamics (as time $t \to +\infty$) of solutions for non-autonomous reaction-diffusion equations.

21.1 Introduction

Let $\mathbb{R} = [0, +\infty)$, $\gamma \geq 1$, and $\mathscr{E}$ be a real separable Banach space. As $L^{\mathrm{loc}}_{\gamma}(\mathbb{R}_+; \mathscr{E})$ we consider the Fréchet space of all locally integrable functions with values in $\mathscr{E}$, that is, $\varphi \in L^{\mathrm{loc}}_{\gamma}(\mathbb{R}_+; \mathscr{E})$ if and only if for any finite interval $[\tau, T] \subset \mathbb{R}_+$ the restriction of φ on $[\tau, T]$ belongs to the space $L_{\gamma}(\tau, T; \mathscr{E})$. If $\mathscr{E} \subseteq L_1(\Omega)$, then any function φ from $L^{\mathrm{loc}}_{\gamma}(\mathbb{R}_+; \mathscr{E})$ can be considered as a measurable mapping that acts from $\Omega \times \mathbb{R}_+$ into $\mathbb{R}$. Further, we write $\varphi(x, t)$, when we consider this mapping as a function from $\Omega \times \mathbb{R}_+$ into $\mathbb{R}$, and $\varphi(t)$, if this mapping is considered as an element from $L^{\mathrm{loc}}_{\gamma}(\mathbb{R}_+; \mathscr{E})$; cf. Gajewski et al. [5, Chapter III]; Temam [10]; Babin and Vishik [1]; Chepyzhov and Vishik [3]; Zgurovsky et at. [12] and references therein.

M. Z. Zgurovsky
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

P. O. Kasyanov · N. V. Gorban · L. S. Paliichuk (✉)
Institute for Applied System Analysis, National Technical University of Ukraine, Igor Sikorsky Kyiv Politechnic Institute, Kyiv, Ukraine
e-mail: kasyanov@i.ua; nata_gorban@i.ua

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_21

A function $\varphi \in L_\gamma^{\mathrm{loc}}(\mathbb{R}_+; \mathscr{E})$ is called *translation bounded* in $L_\gamma^{\mathrm{loc}}(\mathbb{R}_+; \mathscr{E})$, if

$$\sup_{t\geq 0} \int_t^{t+1} \|\varphi(s)\|_{\mathscr{E}}^{\gamma} ds < +\infty; \tag{21.1}$$

Chepyzhov and Vishik [4, p. 105].

Let $N = 1, 2, \ldots$ and $\Omega \subset \mathbb{R}^N$ be a *bounded domain*. A function $\varphi \in L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$ is called *translation uniform integrable one (t.u.i.)* in $L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$, if

$$\lim_{K\to+\infty} \sup_{t\geq 0} \int_t^{t+1} \int_\Omega |\varphi(x,s)| \chi_{\{|\varphi(x,s)|\geq K\}} dx ds = 0; \tag{21.2}$$

Gorban et al. [6–9]. Dunford-Pettis compactness criterion provides that a function $\varphi \in L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$ is t.u.i. in $L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$ if and only if for every sequence of elements $\{\tau_n\}_{n\geq 1} \subset \mathbb{R}_+$ the sequence $\{\varphi(\cdot + \tau_n)\}_{n\geq 1}$ contains a subsequence which converges weakly in $L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$. We note that for any $\gamma > 1$ Hölder's and Chebyshev's inequalities imply that every translation bounded in $L_\gamma^{\mathrm{loc}}(\mathbb{R}_+; L_\gamma(\Omega))$ function is t.u.i. in $L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$, because

$$\int_t^{t+1} \int_\Omega |\varphi(x,s)| \chi_{\{|\varphi(x,s)|\geq K\}} dx ds \leq \frac{1}{K^{\gamma-1}} \sup_{t\geq 0} \int_t^{t+1} \int_\Omega |\varphi(x,s)|^\gamma dx ds \to 0 \text{ as } K \to +\infty.$$

Let us introduce the definition of asymptotic translation uniform integrable function.

Definition 21.1 A function $\varphi \in L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$ is called *asymptotic translation uniform integrable one (a.t.u.i.)* in $L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$, if

$$\lim_{K\to+\infty} \overline{\lim_{t\to+\infty}} \int_t^{t+1} \int_\Omega |\varphi(x,s)| \chi_{\{|\varphi(x,s)|\geq K\}} dx ds = 0. \tag{21.3}$$

Remark 21.1 The limit (as $K \to +\infty$) in (21.2) ((21.3)) exists because the function

$$K \mapsto \sup_{t\geq 0} \left(\overline{\lim_{t\to+\infty}} \right) \int_t^{t+1} \int_\Omega |\varphi(x,s)| \chi_{\{|\varphi(x,s)|\geq K\}} dx ds \tag{21.4}$$

is nonincreasing in $K > 0$.

The main result of this note has the following formulation.

Theorem 21.1 *Let $\varphi \in L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$. Then there exists $\tilde{T} \geq 0$ such that $\varphi(\cdot + \tilde{T})$ is t.u.i. in $L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$ iff φ is a.t.u.i. in $L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$.*

In Sect. 21.3 we apply Theorem 21.1 to non-autonomous nonlinear reaction-diffusion system.

21.2 Proof of Theorem 21.1

Let us prove Theorem 21.1. The t.u.i. of $\varphi(\cdot + \tilde{T})$ for some $\tilde{T} \geq 0$ implies a.u.t.i. of $\varphi(\cdot)$ because for each sequence $\{a_n\}_{n=1,2,\ldots} \subset \overline{\mathbb{R}}$ its limit superior is no greater than its supremum, that is, (21.2) implies (21.3). Let us prove the converse statement: if $\varphi(\cdot)$ is a.t.u.i., then $\varphi(\cdot + \tilde{T})$ is t.u.i. for some $\tilde{T} \geq 0$. We provide the proof in several steps.

Step 1 The following equalities hold:

$$\begin{aligned} 0 &= \lim_{K\to+\infty} \overline{\lim_{t\to+\infty}} \int_t^{t+1}\int_\Omega |\varphi(x,s)|\chi_{\{|\varphi(x,s)|\geq K\}}dxds \\ &= \inf_{K>0}\inf_{T\geq 0}\sup_{t\geq T}\int_t^{t+1}\int_\Omega |\varphi(x,s)|\chi_{\{|\varphi(x,s)|\geq K\}}dxds \\ &= \inf_{T\geq 0}\inf_{K>0}\sup_{t\geq T}\int_t^{t+1}\int_\Omega |\varphi(x,s)|\chi_{\{|\varphi(x,s)|\geq K\}}dxds. \end{aligned} \tag{21.5}$$

Indeed, the first equality follows from a.t.u.i. of $\varphi(\cdot)$, the second equality holds because the mapping

$$K \mapsto \overline{\lim_{t\to+\infty}} \int_t^{t+1}\int_\Omega |\varphi(x,s)|\chi_{\{|\varphi(x,s)|\geq K\}}dxds$$

is nonincreasing and for each $a : [0, +\infty) \mapsto \overline{\mathbb{R}}$ the equality

$$\overline{\lim_{t\to+\infty}} a(t) = \inf_{T\geq 0}\sup_{t\geq T} a(t)$$

holds, and the last equality follows from the basic properties of infimum.

Step 2 We set

$$\delta(T) := \inf_{K>0} \sup_{t \geq T} \int_t^{t+1} \int_{\Omega} |\varphi(x,s)| \chi_{\{|\varphi(x,s)| \geq K\}} dx ds, \tag{21.6}$$

$T \geq 0$, and notice that (21.5) directly implies the existence of $\tilde{T} \geq 0$ such that

$$\delta(T) < +\infty \text{ for each } T \geq \tilde{T} \text{ and } \delta(T) \searrow 0 \text{ as } T \to \infty. \tag{21.7}$$

Step 3 According to (21.6) and (21.7), for each $T \geq \tilde{T}$ there exists $K_T > 0$ such that

$$\sup_{t \geq T} \int_t^{t+1} \int_{\Omega} |\varphi(x,s)| \chi_{\{|\varphi(x,s)| \geq K\}} dx ds < \delta(T) + \frac{1}{T} < +\infty, \tag{21.8}$$

for each $K \geq K_T$.

Step 4 Since for each $n = 0, 1, \ldots$

$$\begin{aligned} \int_{\tilde{T}+n}^{\tilde{T}+n+1} \int_{\Omega} |\varphi(x,s)| dx ds &= \int_{\tilde{T}+n}^{\tilde{T}+n+1} \int_{\Omega} |\varphi(x,s)| \chi_{\{|\varphi(x,s)| \leq K_{\tilde{T}}\}} dx ds \\ &+ \int_{\tilde{T}+n}^{\tilde{T}+n+1} \int_{\Omega} |\varphi(x,s)| \chi_{\{|\varphi(x,s)| \geq K_{\tilde{T}}\}} dx ds \\ &\leq K_{\tilde{T}} \text{meas}(\Omega) + \delta(\tilde{T}) + \frac{1}{\tilde{T}} < +\infty, \end{aligned}$$

where the first inequality follows from (21.8), and the second inequality holds because $\text{meas}(\Omega) < +\infty$, then absolute continuity of the Lebesgue integral implies that for each $T > \tilde{T}$ and $t \in [\tilde{T}, T]$ there exists $K(\tilde{T}, T) > 0$ such that

$$\int_t^{t+1} \int_{\Omega} |\varphi(x,s)| \chi_{\{|\varphi(x,s)| \geq K\}} dx ds \leq \int_{\tilde{T}}^{T+1} \int_{\Omega} |\varphi(x,s)| \chi_{\{|\varphi(x,s)| \geq K\}} dx ds < \frac{1}{T}$$

for each $K \geq K(\tilde{T}, T)$, that is,

$$\sup_{t\in[\tilde{T},T]} \int_t^{t+1} \int_\Omega |\varphi(x,s)| \chi_{\{|\varphi(x,s)|\geq K\}} dxds \leq \frac{1}{N}, \tag{21.9}$$

for each $T > \tilde{T}$ and $K \geq \tilde{K}_T^{\tilde{T}} := \sup_{t\in[\tilde{T},T]} \{K_T; K(\tilde{T}, T)\}$.

Step 5 Inequalities (21.8) and (21.9) imply that

$$\sup_{t\geq\tilde{T}} \int_t^{t+1} \int_\Omega |\varphi(x,s)| \chi_{\{|\varphi(x,s)|\geq K\}} dxds < \delta(T) + \frac{1}{T},$$

for each $T > \tilde{T}$ and $K \geq \tilde{K}_T^{\tilde{T}}$. Thus, according to (21.6),

$$\delta(\tilde{T}) = \inf_{K>0} \sup_{t\geq\tilde{T}} \int_t^{t+1} \int_\Omega |\varphi(x,s)| \chi_{\{|\varphi(x,s)|\geq K\}} dxds < \delta(T) + \frac{1}{T}, \tag{21.10}$$

for each $T > \tilde{T}$.

Step 6 Since the function

$$K \mapsto \sup_{t\geq\tilde{T}} \int_t^{t+1} \int_\Omega |\varphi(x,s)| \chi_{\{|\varphi(x,s)|\geq K\}} dxds$$

is nonincreasing, we have that

$$\lim_{K\to+\infty} \sup_{t\geq\tilde{T}} \int_t^{t+1} \int_\Omega |\varphi(x,s)| \chi_{\{|\varphi(x,s)|\geq K\}} dxds = \delta(\tilde{T}) < \delta(T) + \frac{1}{T}. \tag{21.11}$$

for each $T > \tilde{T}$, where the inequality follows from (21.10). According to (21.7), $\delta(T) + \frac{1}{T} \searrow 0$ as $T \to +\infty$. Therefore, (21.11) implies that

$$\lim_{K\to+\infty} \sup_{t\geq\tilde{T}} \int_t^{t+1} \int_\Omega |\varphi(x,s)| \chi_{\{|\varphi(x,s)|\geq K\}} dxds = 0,$$

that is, $\varphi(\cdot)$ is t.u.i.

21.3 Examples of Applications

Let $N, M = 1, 2, \ldots, \Omega \subset \mathbb{R}^N$ be a bounded domain with sufficiently smooth boundary $\partial\Omega$. We consider a problem of long-time behavior of all globally defined weak solutions for the non-autonomous parabolic problem (named RD-system)

$$\begin{cases} y_t = a\Delta y - f(x,t,y), \quad x \in \Omega,\ t > 0, \\ y|_{\partial\Omega} = 0, \end{cases} \tag{21.12}$$

as $t \to +\infty$, where $y = y(x,t) = (y^{(1)}(x,t), \ldots, y^{(M)}(x,t))$ is unknown vector-function, $f = f(x,t,y) = (f^{(1)}(x,t,y), \ldots, f^{(M)}(x,t,y))$ is given function, a is real $M \times M$ matrix with positive symmetric part.

We suppose that the listed below assumptions hold.

Assumption I Let $p_i \geq 2$ and $q_i > 1$ are such that $\frac{1}{p_i} + \frac{1}{q_i} = 1$, for any $i = 1, 2, \ldots, M$. Moreover, there exists a positive constant d such that $\frac{1}{2}(a + a^*) \geq dI$, where I is unit $M \times M$ matrix, a^* is a transposed matrix for a.

Assumption II The interaction function $f = (f^{(1)}, \ldots, f^{(M)}) : \Omega \times \mathbb{R}_+ \times \mathbb{R}^M \to \mathbb{R}^M$ satisfies the standard Carathéodory's conditions, i.e. the mapping $(x,t,u) \to f(x,t,u)$ is continuous in $u \in \mathbb{R}^M$ for a.e. $(x,t) \in \Omega \times \mathbb{R}_+$, and it is measurable in $(x,t) \in \Omega \times \mathbb{R}_+$ for any $u \in \mathbb{R}^M$.

Assumption III (Growth Condition) There exist an a.t.u.i. in $L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$ function $c_1 : \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$ and a constant $c_2 > 0$ such that

$$\sum_{i=1}^{M} \left| f^{(i)}(x,t,u) \right|^{q_i} \leq c_1(x,t) + c_2 \sum_{i=1}^{M} \left| u^{(i)} \right|^{p_i}$$

for any $u = (u^{(1)}, \ldots, u^{(M)}) \in \mathbb{R}^M$, and a.e. $(x,t) \in \Omega \times \mathbb{R}_+$.

Assumption IV (Sign Condition) There exists a constant $\alpha > 0$ and an a.t.u.i. in $L_1^{\mathrm{loc}}(\mathbb{R}_+; L_1(\Omega))$ function $\beta : \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\sum_{i=1}^{M} f^{(i)}(x,t,u) u^{(i)} \geq \alpha \sum_{i=1}^{M} \left| u^{(i)} \right|^{p_i} - \beta(x,t)$$

for any $u = (u^{(1)}, \ldots, u^{(M)}) \in \mathbb{R}^M$, and a.e. $(x,t) \in \Omega \times \mathbb{R}_+$.

In further arguments we will use standard functional Hilbert spaces $H = (L_2(\Omega))^M$, $V = (H_0^1(\Omega))^M$, and $V^* = (H^{-1}(\Omega))^M$ with standard respective inner products and norms $(\cdot,\cdot)_H$ and $\|\cdot\|_H$, $(\cdot,\cdot)_V$ and $\|\cdot\|_V$, and $(\cdot,\cdot)_{V^*}$ and $\|\cdot\|_{V^*}$,

vector notations $\mathbf{p} = (p_1, p_2, \ldots, p_M)$ and $\mathbf{q} = (q_1, q_2, \ldots, q_M)$, and the spaces

$$\mathbf{L_p}(\Omega) := L_{p_1}(\Omega) \times \ldots \times L_{p_M}(\Omega), \quad \mathbf{L_q}(\Omega) := L_{q_1}(\Omega) \times \ldots \times L_{q_M}(\Omega),$$
$$\mathbf{L_p}(\tau, T; \mathbf{L_p}(\Omega)) := L_{p_1}(\tau, T; L_{p_1}(\Omega)) \times \ldots \times L_{p_M}(\tau, T; L_{p_M}(\Omega)),$$
$$\mathbf{L_q}(\tau, T; \mathbf{L_q}(\Omega)) := L_{q_1}(\tau, T; L_{q_1}(\Omega)) \times \ldots \times L_{q_M}(\tau, T; L_{q_M}(\Omega)),\ 0 \le \tau < T < +\infty.$$

Let $0 \le \tau < T < +\infty$. A function $y = y(x,t) \in \mathbf{L}_2(\tau, T; V) \cap \mathbf{L_p}(\tau, T; \mathbf{L_p}(\Omega))$ is called a *weak solution* of Problem (21.12) on $[\tau, T]$, if for any function $\varphi = \varphi(x) \in (C_0^\infty(\Omega))^M$, the following identity holds

$$\frac{d}{dt}\int_\Omega y(x,t)\cdot\varphi(x)dx + \int_\Omega \{a\nabla y(x,t)\cdot\nabla\varphi(x) + f(x,t,y(x,t))\cdot\varphi(x)\}dx = 0 \tag{21.13}$$

in the sense of scalar distributions on (τ, T).

In the general case Problem (21.12) on $[\tau, T]$ with initial condition $y(x,\tau) = y_\tau(x)$ in Ω has more than one weak solution with $y_\tau \in H$ (cf. Balibrea et al. [2] and references therein).

Assumptions I–IV and Chepyzhov and Vishik [4, pp. 283–284] (see also Zgurovsky et al. [11, Chapter 2] and references therein) provide the existence of a weak solution of Cauchy problem (21.12) with initial data $y(\tau) = y^{(\tau)}$ on the interval $[\tau, T]$, for any $y^{(\tau)} \in H$. The proof is provided by standard Faedo–Galerkin approximations and using local existence Carathéodory's theorem instead of classical Peano results. A priori estimates are similar. Formula (21.13) and definition of the derivative for an element from $\mathscr{D}([\tau, T]; V^* + \mathbf{L_q}(\Omega))$ yield that each weak solution $y \in X_{\tau,T}$ of Problem (21.12) on $[\tau, T]$ belongs to the space $W_{\tau,T}$. Moreover, each weak solution of Problem (21.12) on $[\tau, T]$ satisfies the equality:

$$\int_\tau^T \int_\Omega \left[\frac{\partial y(x,t)}{\partial t}\cdot\psi(x,t) + a\nabla y(x,t)\cdot\nabla\psi(x,t) + f(x,t,y(x,t))\cdot\psi(x,t)\right]dxdt = 0, \tag{21.14}$$

for any $\psi \in X_{\tau,T}$. For fixed τ and T, such that $0 \le \tau < T < +\infty$, we denote

$$\mathscr{D}_{\tau,T}(y^{(\tau)}) = \{y(\cdot) \mid y \text{ is a weak solution of (21.12) on } [\tau, T],\ y(\tau) = y^{(\tau)}\}, \quad y^{(\tau)} \in H.$$

We remark that $\mathscr{D}_{\tau,T}(y^{(\tau)}) \ne \emptyset$ and $\mathscr{D}_{\tau,T}(y^{(\tau)}) \subset W_{\tau,T}$, if $0 \le \tau < T < +\infty$ and $y^{(\tau)} \in H$. Moreover, the concatenation of Problem (21.12) weak solutions is a weak solutions too, i.e. if $0 \le \tau < t < T$, $y^{(\tau)} \in H$, $y(\cdot) \in \mathscr{D}_{\tau,t}(y^{(\tau)})$, and $v(\cdot) \in \mathscr{D}_{t,T}(y(t))$, then

$$z(s) = \begin{cases} y(s),\ s \in [\tau, t], \\ v(s),\ s \in [t, T], \end{cases}$$

belongs to $\mathscr{D}_{\tau,T}(y^{(\tau)})$; cf. Zgurovsky et al. [12, pp. 55–56].

Each weak solution y of Problem (21.12) on a finite time interval $[\tau, T] \subset \mathbb{R}_+$ can be extended to a global one, defined on $[\tau, +\infty)$. For arbitrary $\tau \geq 0$ and $y^{(\tau)} \in H$ let $\mathscr{D}_\tau(y^{(\tau)})$ be the set of all weak solutions (defined on $[\tau, +\infty)$) of Problem (21.12) with initial data $y(\tau) = y^{(\tau)}$. Let us consider the family $\mathscr{K}_\tau^+ = \cup_{y^{(\tau)} \in H} \mathscr{D}_\tau(y^{(\tau)})$ of all weak solutions of Problem (21.12) defined on the semi-infinite time interval $[\tau, +\infty)$.

Consider the Fréchet space

$$C^{\mathrm{loc}}(\mathbb{R}_+; H) := \{y : \mathbb{R}_+ \to H \ : \ \Pi_{t_1,t_2} y \in C([t_1, t_2]; H) \text{ for any } [t_1, t_2] \subset \mathbb{R}_+\},$$

where Π_{t_1,t_2} is the restriction operator to the interval $[t_1, t_2]$; Chepyzhov and Vishik [3, p. 918]. We remark that the sequence $\{f_n\}_{n\geq 1}$ converges (converges weakly respectively) in $C^{\mathrm{loc}}(\mathbb{R}_+; H)$ towards $f \in C^{\mathrm{loc}}(\mathbb{R}_+; H)$ as $n \to +\infty$ if and only if the sequence $\{\Pi_{t_1,t_2} f_n\}_{n\geq 1}$ converges (converges weakly respectively) in $C([t_1, t_2]; H)$ towards $\Pi_{t_1,t_2} f$ as $n \to +\infty$ for any finite interval $[t_1, t_2] \subset \mathbb{R}_+$.

We denote $T(h)y(\cdot) = y_h(\cdot)$, where $y_h(t) = y(t + h)$ for any $y \in C^{\mathrm{loc}}(\mathbb{R}_+; H)$ and $t, h \geq 0$.

In the non-autonomous case we notice that $T(h)\mathscr{K}_0^+ \not\subseteq \mathscr{K}_0^+$. Therefore (see Gorban et al. [8]), we need to consider *united trajectory space* that includes all globally defined on any $[\tau, +\infty) \subseteq \mathbb{R}_+$ weak solutions of Problem (21.12) shifted to $\tau = 0$:

$$\mathscr{K}_\cup^+ := \bigcup_{\tau\geq 0} \left\{ y(\cdot + \tau) \in W^{\mathrm{loc}}(\mathbb{R}_+) \ : \ y(\cdot) \in \mathscr{K}_\tau^+ \right\}. \tag{21.15}$$

Note that $T(h)\{y(\cdot + \tau) \ : \ y \in \mathscr{K}_\tau^+\} \subseteq \{y(\cdot + \tau + h) \ : \ y \in \mathscr{K}_{\tau+h}^+\}$ for any $\tau, h \geq 0$. Therefore,

$$T(h)\mathscr{K}_\cup^+ \subseteq \mathscr{K}_\cup^+$$

for any $h \geq 0$. Further we consider *extended united trajectory space* for Problem (21.12):

$$\mathscr{K}_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}^+ = \mathrm{cl}_{C^{\mathrm{loc}}(\mathbb{R}_+;H)} \left[\mathscr{K}_\cup^+\right], \tag{21.16}$$

where $\mathrm{cl}_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}[\,\cdot\,]$ is the closure in $C^{\mathrm{loc}}(\mathbb{R}_+; H)$. We note that

$$T(h)\mathscr{K}_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}^+ \subseteq \mathscr{K}_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}^+$$

for each $h \geq 0$, because

$$\rho_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}(T(h)u, T(h)v) \leq \rho_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}(u, v) \text{ for any } u, v \in C^{\mathrm{loc}}(\mathbb{R}_+; H),$$

where $\rho_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$ is a standard metric on Fréchet space $C^{\mathrm{loc}}(\mathbb{R}_+; H)$.

Let us provide the result characterizing the compactness properties of shifted solutions of Problem (21.12) in the induced topology from $C^{\mathrm{loc}}(\mathbb{R}_+; H)$.

Theorem 21.2 *Let Assumptions I–IV hold. If* $\{y_n\}_{n\geq 1} \subset \mathcal{K}^+_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$ *is an arbitrary sequence, which is bounded in* $L_\infty(\mathbb{R}_+; H)$*, then there exist a subsequence* $\{y_{n_k}\}_{k\geq 1} \subseteq \{y_n\}_{n\geq 1}$ *and an element* $y \in \mathcal{K}^+_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$ *such that*

$$\|\Pi_{\tau,T} y_{n_k} - \Pi_{\tau,T} y\|_{C([\tau,T];H)} \to 0, \quad k \to +\infty, \tag{21.17}$$

for any finite time interval $[\tau, T] \subset (0, +\infty)$*. Moreover, for any* $y \in \mathcal{K}^+_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$ *the estimate holds*

$$\|y(t)\|^2_H \leq \|y(0)\|^2_H e^{-c_3 t} + c_4, \tag{21.18}$$

for any $t \geq 0$*, where positive constants* c_3 *and* c_4 *do not depend on* $y \in \mathcal{K}^+_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$ *and* $t \geq 0$*.*

Proof This statement directly follows from Gorban et al. [8, Theorem 4.1] and Theorem 21.1.

A set $\mathscr{P} \subset \mathscr{F}^{\mathrm{loc}}(\mathbb{R}_+) \cap L_\infty(\mathbb{R}_+; H)$ is said to be a *uniformly attracting set* (cf. Chepyzhov and Vishik [3, p. 921]) for the extended united trajectory space $\mathcal{K}^+_{\mathscr{F}^{\mathrm{loc}}(\mathbb{R}_+)}$ of Problem (21.12) in the topology of $\mathscr{F}^{\mathrm{loc}}(\mathbb{R}_+)$, if for any bounded in $L_\infty(\mathbb{R}_+; H)$ set $\mathscr{B} \subseteq \mathcal{K}^+_{\mathscr{F}^{\mathrm{loc}}(\mathbb{R}_+)}$ and any segment $[t_1, t_2] \subset \mathbb{R}_+$ the following relation holds:

$$\mathrm{dist}_{\mathscr{F}_{t_1,t_2}}(\Pi_{t_1,t_2} T(t)\mathscr{B}, \Pi_{t_1,t_2}\mathscr{P}) \to 0, \quad t \to +\infty, \tag{21.19}$$

where $\mathrm{dist}_{\mathscr{F}_{t_1,t_2}}$ is the Hausdorff semi-metric.

A set $\mathscr{U} \subset \mathcal{K}^+_{\mathscr{F}^{\mathrm{loc}}(\mathbb{R}_+)}$ is said to be a *uniform trajectory attractor* of the translation semigroup $\{T(t)\}_{t\geq 0}$ on $\mathcal{K}^+_{\mathscr{F}^{\mathrm{loc}}(\mathbb{R}_+)}$ in the induced topology from $C^{\mathrm{loc}}(\mathbb{R}_+; H)$, if

1. $\mathscr{U}$ is a compact set in $C^{\mathrm{loc}}(\mathbb{R}_+; H)$ and bounded in $L_\infty(\mathbb{R}_+; H)$;
2. $\mathscr{U}$ is strictly invariant with respect to $\{T(h)\}_{h\geq 0}$, i.e. $T(h)\mathscr{U} = \mathscr{U}$ $\forall h \geq 0$;
3. $\mathscr{U}$ is a minimal uniformly attracting set for $\mathcal{K}^+_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$ in the topology of $C^{\mathrm{loc}}(\mathbb{R}_+; H)$, i.e. $\mathscr{U}$ belongs to any compact uniformly attracting set $\mathscr{P}$ of $\mathcal{K}^+_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$: $\mathscr{U} \subseteq \mathscr{P}$.

Note that uniform trajectory attractor of the translation semigroup $\{T(t)\}_{t\geq 0}$ on $\mathcal{K}^+_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$ in the induced topology from $C^{\mathrm{loc}}(\mathbb{R}_+; H)$ coincides with the classical global attractor for the continuous semi-group $\{T(t)\}_{t\geq 0}$ defined on $\mathcal{K}^+_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$.

Assumptions I–IV are sufficient conditions for the existence of uniform trajectory attractor for weak solutions of Problem (21.12) in the topology of $C^{\mathrm{loc}}(\mathbb{R}_+; H)$.

Theorem 21.3 *Let Assumptions I–IV hold. Then there exists an uniform trajectory attractor* $\mathscr{U} \subset \mathscr{K}^{+}_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$ *of the translation semigroup* $\{T(t)\}_{t\geq 0}$ *on* $\mathscr{K}^{+}_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$ *in the induced topology from* $C^{\mathrm{loc}}(\mathbb{R}_+;H)$. *Moreover, there exists a compact in* $C^{\mathrm{loc}}(\mathbb{R}_+;H)$ *uniformly attracting set* $\mathscr{P} \subset C^{\mathrm{loc}}(\mathbb{R}_+;H) \cap L_\infty(\mathbb{R}_+;H)$ *for the extended united trajectory space* $\mathscr{K}^{+}_{C^{\mathrm{loc}}(\mathbb{R}_+;H)}$ *of Problem (21.12) in the topology of* $C^{\mathrm{loc}}(\mathbb{R}_+;H)$ *such that* $\mathscr{U}$ *coincides with* ω*-limit set of* $\mathscr{P}$*:*

$$\mathscr{U} = \bigcap_{t\geq 0} \mathrm{cl}_{C^{\mathrm{loc}}(\mathbb{R}_+;H)} \left[\bigcup_{h\geq t} T(h)\mathscr{P} \right]. \tag{21.20}$$

Proof This statement directly follows from Gorban et al. [8, Theorem 3.1] and Theorem 21.1.

21.4 Conclusions

Asymptotic translation uniform integrability condition for a function acting from positive semi-axe of time line to a Banach space is equivalent to uniform integrability condition. As a result, we claim only asymptotic (as time $t \to +\infty$) assumptions of translation compactness for parameters of non-autonomous reaction-diffusion equations.

References

1. Babin, A.V., Vishik, M.I.: Attractors of Evolution Equations (in Russian). Nauka, Moscow (1989)
2. Balibrea, F., Caraballo, T., Kloeden, P.E., Valero, J.: Recent developments in dynamical systems: three perspectives. Int. J. Bifurcation Chaos (2010). https://doi.org/10.1142/S0218127410027246
3. Chepyzhov, V.V., Vishik, M.I.: Evolution equations and their trajectory attractors. J. Math. Pures Appl. **76**, 913–964 (1997)
4. Chepyzhov, V.V., Vishik, M.I.: Attractors for Equations of Mathematical Physics. American Mathematical Society, Providence (2002)
5. Gajewski, H., Gröger, K., Zacharias, K.: Nichtlineare operatorgleichungen und operatordifferentialgleichungen. Akademie-Verlag, Berlin (1978)
6. Gluzman, M.O., Gorban, N.V., Kasyanov, P.O.: Lyapunov type functions for classes of autonomous parabolic feedback control problems and applications. Appl. Math. Lett. (2014). https://doi.org/10.1016/j.aml.2014.08.006
7. Gorban, N.V., Kasyanov, P.O.: On regularity of all weak solutions and their attractors for reaction-diffusion inclusion in unbounded domain. Contin. Distrib. Syst. Theory Appl. Solid Mech. Appl. **211**, (2014). https://doi.org/10.1007/978-3-319-03146-0$_$15
8. Gorban, N.V., Kapustyan, O.V., Kasyanov, P.O.: Uniform trajectory attractor for non-autonomous reaction-diffusion equations with Caratheodory's nonlinearity. Nonlinear Anal. Theory Methods Appl. **98**, 13–26 (2014). https://doi.org/10.1016/j.na.2013.12.004

9. Gorban, N.V., Kapustyan, O.V., Kasyanov, P.O., Paliichuk, L.S.: On global attractors for autonomous damped wave equation with discontinuous nonlinearity. Contin. Distrib. Syst. Theory Appl. Solid Mech. Appl. **211** (2014). https://doi.org/10.1007/978-3-319-03146-0$_$16
10. Temam, R.: Infinite-Dimensional Dynamical Systems in Mechanics and Physics. Applied Mathematical Sciences, vol. 68. Springer, New York (1988)
11. Zgurovsky, M.Z., Mel'nik, V.S., Kasyanov, P.O.: Evolution Inclusions and Variation Inequalities for Earth Data Processing II. Springer, Berlin (2011)
12. Zgurovsky, M.Z., Kasyanov, P.O., Kapustyan, O.V., Valero, J., Zadoianchuk, N.V.: Evolution Inclusions and Variation Inequalities for Earth Data Processing III. Springer, Berlin (2012)

Chapter 22
Automation of Impulse Processes Control in Cognitive Maps with Multirate Sampling Based on Weights Varying

Victor D. Romanenko and Yuriy L. Milyavsky

Abstract Automated control systems for multirate impulse processes in cognitive maps are considered. Some coordinates of the cognitive map can be measured and changed with small sampling period while others need longer sampling period. Thus, the impulse process is decomposed into two subsystems described as first-order difference equations systems with different sampling periods. Effects of the fast subsystem on the slow subsystem and vice versa are considered as disturbances which should be suppressed. Control for both subsystems is implemented not via external inputs (i.e. varying resources of the cognitive map) but via map's edges varying, which means that decision-maker modifies degree of influence of one cognitive map node on another one. Two approaches for control system design are proposed. The first approach is based on invariant ellipsoids method which allows robust stabilization of the system. The second approach is based on generalized variance minimization which allows setting some of the coordinates at predefined levels. Both approaches are verified on the real cognitive map of IT company HR management process.

22.1 Introduction

Cognitive map (CM) is a weighted directed graph with nodes representing coordinates of complex systems and edges representing relations between these coordinates. In [1] dynamic process in CM is discussed, which appears as a result of impulse disturbance at some CM node. This dynamic transient is named "CM

V. D. Romanenko · Y. L. Milyavsky (✉)
Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
e-mail: ipsa@kpi.ua

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_22

impulse process" in [1]. Impulse propagation rule in CM free motion is formalized in [1] in the form of difference equation

$$\Delta Y_i(k+1) = \sum_{j=1}^{n} a_{ij} \Delta Y_j(k), \tag{22.1}$$

where $\Delta Y_i(k) = Y_i(k) - Y_i(k-1), i = 1, 2, \ldots, n$; a_{ij}—weight of digraph edge connecting the j-th node with the i-th one. In vector form the expression (22.1) is written as

$$\Delta \overline{Y}(k+1) = A \Delta \overline{Y}(k), \tag{22.2}$$

where A—weighted incidence matrix $(n \times n)$, $\Delta \overline{Y}$—vector of CM nodes coordinates' increments Y_i .

In [2–6] control of CM impulse processes in complex systems is automated by means of external control vector design based on varying of resources associated with CM nodes in closed-loop control systems, using known methods of discrete controllers design. For this purpose forced motion equation under CM impulse process is formulated:

$$\Delta Y_i(k+1) = \sum_{j=1}^{n} a_{ij} \Delta Y_j(k) + b_i \Delta u_i(k), \tag{22.3}$$

where $\Delta u_i(k) = u_i(k) - u_i(k-1)$—controls increments implemented via varying available resources of coordinates $Y_i(k)$ and affecting directly CM nodes. In vector form Eq. (22.3) can be written as

$$\Delta \overline{Y}(k+1) = A \Delta \overline{Y}(k) + B \Delta \overline{u}(k). \tag{22.4}$$

Here elements of matrix B which correspond to controls in vector $\Delta \overline{u}$ are equal to ones, and others are zeros.

In [7] new principle of CM impulse process control design is discussed. It is based on edges' weights varying as control inputs in closed-loop system. This is possible when we can change the degree of influence $\Delta a_{ij}(k)$ of one CM node on another one at the k-th sampling period. For this the following model of forced motion of CM impulse process with unirate sampling is proposed:

$$\Delta \overline{Y}(k+1) = A \Delta \overline{Y}(k) + L(k) \Delta \overline{a}(k), \tag{22.5}$$

where matrix $L(k)$ is composed of measured coordinates $Y_\mu(k)$, $\mu \neq i$ which affect coordinates $\Delta Y_i(k+1)$ via edges with varying coefficients $\Delta a_{i\mu}(k)$. To design control vector $\Delta \overline{a}(k)$ quadratic criterion in the form of generalized variance is used as optimality criterion.

In [5, 8] control automation for complex systems' CM impulse processes with multirate sampling is proposed. Some of coordinates $Y_i (i = 1, 2, \ldots, p)$ are measured in discrete time with small sampling period T_0 and other coordinates $Y_\eta (\eta = p + 1, \ldots, n)$ are measured with period

$$h = mT_0, \tag{22.6}$$

where m is integer greater than 1. To describe dynamics of this system impulse processes models are proposed for sampling periods T_0 and h respectively. Based on quadratic optimality criteria fast and slow controllers are designed which generate controls that vary resources of nodes coordinates Y_i and Y_η.

22.2 Problem Statement

Implementation of control system of CM impulse process based on control inputs $\Delta\overline{u}(k)$ according to model (22.4) by means of varying resources of CM nodes [2–6, 8] is fraught with difficulties because it presupposes forced varying of CM nodes of complex system according to control law. It raises additional tensions in complex system's operating because of personal, antagonistic human factors.

Thus, present paper considers development of the new principal of CM impulse processes control in closed-loop system based on weights varying (started in [7]). We suppose that the first group of CM nodes coordinates $\overline{Y}_f (p \times 1)$ is measured (fixed) with sampling period T_0. These coordinates can be controlled by fast controller via edges weights $\Delta\overline{a}_f$ varying with period T_0. The second group of coordinates $\overline{Y}_s$ is measured with sampling period h. For these coordinates control vector $\Delta\overline{a}_s$ is generated by slow controller with period h.

To describe dynamics of forced motion of CM impulse process with multirate sampling unirate model (22.5) is split into two parts. The first part of the model is written as

$$\Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h + (l+1)T_0\right] = A_{11}\Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h + lT_0\right] + A_{12}\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right] + G_f\left[\left[\frac{k}{m}\right]h + lT_0\right]\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h + lT_0\right], \tag{22.7}$$

where $\left[\frac{k}{m}\right]$ is integer part of $\frac{k}{m}$, $l = 0, 1, \ldots, m-1$, $\Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h + lT_0\right] = \overline{Y}_f\left[\left[\frac{k}{m}\right]h + lT_0\right] - \overline{Y}_f\left[\left[\frac{k}{m}\right]h + (l-1)T_0\right]$ is the vector $(p \times 1)$ of increments of nodes coordinates measured with sampling period T_0. Vector $\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right] = \Delta\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right]$ for $l = 0$ and is zero otherwise. Coefficients increments vector $\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h + lT_0\right] = \overline{a}_f\left[\left[\frac{k}{m}\right]h + lT_0\right] - \overline{a}_f\left[\left[\frac{k}{m}\right]h + (l-1)T_0\right]$ has dimension q_1. Then matrices have the following dimensions: $A_{11}(p \times p)$, $A_{12}(p \times (n-p))$, $G_f(p \times q_1)$.

The second part of the mathematical model of CM impulse process with period $h = mT_0$ is written as

$$\Delta\overline{Y}_s\left[\left(\left[\frac{k}{m}\right]+1\right)h\right] = A_{22}\Delta\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right] + A_{21}\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h + lT_0\right] + G_s\left[\left[\frac{k}{m}\right]h\right]\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right], l = 0, 1, 2, \ldots, m-1, \quad (22.8)$$

where $\Delta\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right] = \overline{Y}_s\left[\left[\frac{k}{m}\right]h\right] - \overline{Y}_s\left[\left(\left[\frac{k}{m}\right]-1\right)h\right]$ is the vector $((n-p)\times 1)$ of increments of CM coordinates measured with sampling period h, $\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right] = \overline{a}_s\left[\left[\frac{k}{m}\right]h\right] - \overline{a}_s\left[\left(\left[\frac{k}{m}\right]-1\right)h\right]$ is the vector $(\eta\times 1)$ of increments of edges' weights. Vector $\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$ is defined as

$$\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h + lT_0\right] = \Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h\right] + \Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h + T_0\right] + \ldots + \Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h + (m-1)T_0\right].$$

It is easy to prove that [8]

$$\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h + lT_0\right] = \overline{Y}_f\left[\left[\frac{k}{m}\right]h + (m-1)T_0\right] - \overline{Y}_f\left[\left[\frac{k}{m}\right]h - T_0\right].$$

Rule for creating matrices G_f, G_s and vectors of weights increments $\Delta\overline{a}_f, \Delta\overline{a}_s$ in models (22.7), (22.8) are described in details in [7]. When decomposing initial model (22.5) into these two parts components $\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right]$ in model (22.7) and $\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$ in model (22.8) are considered as disturbances (with multirate sampling) affecting main dynamic coordinates (state variables) of CM.

The first problem in the present work is to design multirate state controllers to suppress constrained disturbances in impulse processes—$\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right]$ in model (22.7) and $\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$ in model (22.8) respectively. The second problem is to design closed-loop control system for impulse processes (22.7), (22.8) to shift CM nodes to different levels under effect of disturbances with multirate sampling. Thus, to solve both problems control vectors $\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h + lT_0\right], \Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right]$ in the form of CM weights increments with multirate sampling should be designed.

22.3 Suppression of Constrained Disturbances in Impulse Processes with Multirate Sampling Based on Invariant Ellipsoids Method

Probabilistic characteristics of disturbances $\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right]$ in the first model (22.7) and $\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]$ in the second model (22.8) are unknown. So, for solving the first problem we suppose that these disturbances are neither random nor harmonic. Arbitrary constrained external disturbances can be suppressed in terms of invariant ellipsoids; this technique can be found in [9, 10].

Invariant ellipsoid with respect to state variables of the model (22.7) is written as following:

$$E_{\Delta\overline{Y}_f}=\left\{\Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\in\Re^p:\right.$$
$$\left.\Delta\overline{Y}_f^T\left[\left[\frac{k}{m}\right]h+lT_0\right]P_f^{-1}\Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\le 1\right\}, \tag{22.9}$$

where $P_f>0$ is ellipsoid matrix. This ellipsoid is invariant w.r.t. vector $\Delta\overline{Y}_f$ if from $\Delta\overline{Y}_f(0)\in E_{\Delta\overline{Y}_f}$ it follows that $\Delta\overline{Y}_f\left(\left[\frac{k}{m}\right]h+lT_0\right)\in E_{\Delta\overline{Y}_f}$ for all time samples. Disturbances $\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right]$ in model (22.7) should be constrained in L_∞-norm:

$$\|\Delta\widetilde{\overline{Y}}_s\|_\infty=\sup\left[\Delta\widetilde{\overline{Y}}_s^T\left[\left[\frac{k}{m}\right]h\right]\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right]\right]^{1/2}\le 1. \tag{22.10}$$

Invariant ellipsoid with respect to state variables of the model (22.8) is written as following:

$$E_{\Delta\overline{Y}_s}=\left\{\Delta\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right]\in\Re(n-p):\Delta\overline{Y}_s^T\left[\left[\frac{k}{m}\right]h\right]P_s^{-1}\Delta\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right]\le 1\right\}, \tag{22.11}$$

where $P_s>0$ is ellipsoid matrix. Disturbances $\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]$ in model (22.7) should be constrained in L_∞-norm:

$$\|\Delta\widetilde{\overline{Y}}_f\|_\infty=\sup\left[\Delta\widetilde{\overline{Y}}_f^T\left[\left[\frac{k}{m}\right]h+lT_0\right]\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\right]^{1/2}$$
$$\le 1, l=0,1,\ldots,m-1 \tag{22.12}$$

This ellipsoid is invariant w.r.t. vector $\Delta\overline{Y}_s$ if from $\Delta\overline{Y}_s(0)\in E_{\Delta\overline{Y}_s}$ it follows that $\Delta\overline{Y}_s\left(\left[\frac{k}{m}\right]h\right)\in E_{\Delta\overline{Y}_s}$ for all time samples.

We will suppress constrained slow disturbances $\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right]$ in Eq. (22.7) with measured coordinates $\Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]$ using vector $\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]$ generated by fast state controller

$$\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]=-K_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]. \tag{22.13}$$

This controller should ensure optimal speed of change of CM weights coefficients' vector $\Delta\overline{a}_f$ (as control vector) depending on speed of change of measured nodes' coordinates $\Delta\overline{Y}_f$.

We will also suppress constrained fast disturbances $\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]$ in Eq. (22.8) with measured coordinates $\Delta\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right]$ using vector $\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right]$ generated by slow state controller

$$\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right]=-K_s\left[\left[\frac{k}{m}\right]h\right]\Delta\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right]. \tag{22.14}$$

If coordinates $\Delta\overline{Y}_f$, $\Delta\overline{Y}_s$ in (22.13), (22.14) respectively remain constant then increments $\Delta\overline{a}_f$, $\Delta\overline{a}_s$ will be zero. Thus, control vectors $\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]$, $\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right]$ are intended to stabilize transit process in complex system's dynamics described by CM impulse process models (22.7), (22.8) under the effect of disturbances $\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]$, $\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right]$ constrained by (22.10), (22.12).

Algorithms for control laws (22.13), (22.14) for impulse processes (22.7), (22.8) are based on minimization of the optimality criteria:

$$\mathrm{tr}P_f\to\min, \tag{22.15}$$

$$\mathrm{tr}P_s\to\min, \tag{22.16}$$

guaranteeing minimal invariant ellipsoids (22.9), (22.11) with maximal suppression of any disturbances constrained only by their range (22.10), (22.12).

Based on models (22.7), (22.8) and control laws (22.13), (22.14), equations of fast and slow closed-loop CM impulse process control subsystems respectively can be written as:

$$\Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h+(l+1)T_0\right]=\left(A_{11}-G_f\left[\left[\frac{k}{m}\right]h+lT_0\right]K_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\right)\times\Delta\overline{Y}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]+A_{12}\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right], \tag{22.17}$$

$$\Delta\overline{Y}_s\left[\left(\left[\frac{k}{m}\right]+1\right)h\right]=\left(A_{22}-G_s\left[\left[\frac{k}{m}\right]h\right]K_s\left[\left[\frac{k}{m}\right]h\right]\right)\times\Delta\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right]+A_{21}\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h+lT_0\right],\quad(22.18)$$

where $\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]=\overline{Y}_f\left[\left[\frac{k}{m}\right]h+(m-1)T_0\right]-\overline{Y}_f\left[\left[\frac{k}{m}\right]h-T_0\right]$. We suppose that both subsystems are controllable.

In [9] invariant ellipsoid method is utilized for unirate state-space system with L_∞-constrained disturbances, using linear matrix inequalities. In the present case for closed-loop systems (22.17), (22.18) linear matrix inequalities will be written in the following way:

$$\frac{1}{\alpha_1}\left(A_{11}-G_f\left[\left[\frac{k}{m}\right]h+lT_0\right]K_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\right)P_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\times\left(A_{11}-G_f\left[\left[\frac{k}{m}\right]h+lT_0\right]K_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\right)^T-P_f\left[\left[\frac{k}{m}\right]h+lT_0\right]+\frac{A_{12}A_{12}^T}{1-\alpha_1}\le 0;\quad(22.19)$$

$$\frac{1}{\alpha_2}\left(A_{22}-G_s\left[\left[\frac{k}{m}\right]h\right]K_s\left[\left[\frac{k}{m}\right]h\right]\right)P_s\left[\left[\frac{k}{m}\right]h\right]\times\left(A_{22}-G_s\left[\left[\frac{k}{m}\right]h\right]K_s\left[\left[\frac{k}{m}\right]h\right]\right)^T-P_s\left[\left[\frac{k}{m}\right]h\right]+\frac{A_{21}A_{21}^T}{1-\alpha_2}\le 0,\quad(22.20)$$

where $0<\alpha_1<1, 0<\alpha_2<1$.

After multiplication we obtain:

$$\frac{1}{\alpha_1}\left(A_{11}P_f\left[\left[\frac{k}{m}\right]h+lT_0\right]A_{11}^T-G_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\times K_f\left[\left[\frac{k}{m}\right]h+lT_0\right]P_f\left[\left[\frac{k}{m}\right]h+lT_0\right]A_{11}^T\right.$$
$$-A_{11}P_f\left[\left[\frac{k}{m}\right]h+lT_0\right]K_f\left[\left[\frac{k}{m}\right]h+lT_0\right]^T G_f^T\left[\left[\frac{k}{m}\right]h+lT_0\right]+G_f\left[\left[\frac{k}{m}\right]h+lT_0\right]K_f\left[\left[\frac{k}{m}\right]h+lT_0\right]$$
$$\left.\times P_f\left[\left[\frac{k}{m}\right]h+lT_0\right]K_f^T\left[\left[\frac{k}{m}\right]h+lT_0\right]G_f^T\left[\left[\frac{k}{m}\right]h+lT_0\right]\right)-P_f\left[\left[\frac{k}{m}\right]h+lT_0\right]+\frac{A_{12}A_{12}^T}{1-\alpha_1}\le 0;\quad(22.21)$$

$$\frac{1}{\alpha_2}\left(A_{22}P_s\left[\left[\frac{k}{m}\right]h\right]A_{22}^T - G_s\left[\left[\frac{k}{m}\right]h\right]K_s\left[\left[\frac{k}{m}\right]h\right]P_s\left[\left[\frac{k}{m}\right]h\right]A_{22}^T\right.$$
$$- A_{22}P_s\left[\left[\frac{k}{m}\right]h\right]K_s^T\left[\left[\frac{k}{m}\right]h\right]G_s^T\left[\left[\frac{k}{m}\right]h\right]$$
$$\left. + G_s\left[\left[\frac{k}{m}\right]h\right]K_s\left[\left[\frac{k}{m}\right]h\right]P_s\left[\left[\frac{k}{m}\right]h\right]K_s^T\left[\left[\frac{k}{m}\right]h\right]G_s^T\left[\left[\frac{k}{m}\right]h\right]\right)$$
$$-P_s\left[\left[\frac{k}{m}\right]h\right] + \frac{A_{21}A_{21}^T}{1-\alpha_2} \le 0. \tag{22.22}$$

These inequalities are nonlinear w.r.t. $P_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$ and $K_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$ in (22.21) and $P_s\left[\left[\frac{k}{m}\right]h\right]$, $K_s\left[\left[\frac{k}{m}\right]h\right]$ in (22.22). To linearize them the following substitutions are proposed:

$$M_f = K_f P_f;\ M_s = K_s P_s. \tag{22.23}$$

Auxiliary matrices $D_f = D_f^T$, $D_s = D_s^T$ are also introduced such that

$$\begin{pmatrix} D_f & M_f \\ M_f^T & P_f \end{pmatrix} \ge 0, \tag{22.24}$$

$$\begin{pmatrix} D_s & M_s \\ M_s^T & P_s \end{pmatrix} \ge 0. \tag{22.25}$$

From Schur formula it follows that if $P_f > 0$, $P_s > 0$ inequalities (22.24), (22.25) are equivalent to $D_f \ge M_f P_f^{-1} M_f^T = K_f P_f K_f^T$, $D_s \ge M_s P_s^{-1} M_s^T = K_s P_s K_s^T$. Then for inequalities (22.21), (22.22) to be fulfilled it is sufficient that

$$\frac{1}{\alpha_1}\left(A_{11}P_f\left[\left[\frac{k}{m}\right]h + lT_0\right]A_{11}^T - G_f\left[\left[\frac{k}{m}\right]h + lT_0\right]M_f\left[\left[\frac{k}{m}\right]h + lT_0\right]A_{11}^T\right.$$
$$- A_{11}M_f^T\left[\left[\frac{k}{m}\right]h + lT_0\right]G_f^T\left[\left[\frac{k}{m}\right]h + lT_0\right]$$
$$\left. + G_f\left[\left[\frac{k}{m}\right]h + lT_0\right]D_f\left[\left[\frac{k}{m}\right]h + lT_0\right]G_f^T\left[\left[\frac{k}{m}\right]h + lT_0\right]\right)$$
$$-P_f\left[\left[\frac{k}{m}\right]h + lT_0\right] + \frac{A_{12}A_{12}^T}{1-\alpha_1} \le 0; \tag{22.26}$$

$$\frac{1}{\alpha_2}\Big(A_{22}P_s\left[\left[\frac{k}{m}\right]h\right]A_{22}^T - G_s\left[\left[\frac{k}{m}\right]h\right]M_s\left[\left[\frac{k}{m}\right]h\right]A_{22}^T - A_{22}M_s^T\left[\left[\frac{k}{m}\right]h\right]G_s^T\left[\left[\frac{k}{m}\right]h\right] + G_s\left[\left[\frac{k}{m}\right]h\right]D_s\left[\left[\frac{k}{m}\right]h\right]G_s^T\left[\left[\frac{k}{m}\right]h\right]\Big) - P_s\left[\left[\frac{k}{m}\right]h\right] + \frac{A_{21}A_{21}^T}{1-\alpha_2} \le 0. \quad (22.27)$$

We minimize criterion (22.15) for fast subsystem (22.17) under constrains (22.24), (22.26) w.r.t. to variables $P_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$, $M_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$, $D_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$ using semidefinite programming. If $\hat{P}_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$, $\hat{M}_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$, $\hat{D}_f\left[\left[\frac{k}{m}\right]h + lT_0\right]$ are obtained estimates ensuring minimum of (22.15) under (22.24), (22.26) then optimal gain matrix of fast controller (22.13) is written based on (22.23) as

$$\hat{K_f}\left[\left[\frac{k}{m}\right]h + lT_0\right] = \hat{M}_f\left[\left[\frac{k}{m}\right]h + lT_0\right]\hat{P}_f{}^{-1}\left[\left[\frac{k}{m}\right]h + lT_0\right]. \quad (22.28)$$

In the same manner, we minimize criterion (22.16) for slow subsystem (22.18) under constrains (22.25), (22.27) w.r.t. to variables $P_s\left[\left[\frac{k}{m}\right]h\right]$, $M_s\left[\left[\frac{k}{m}\right]h\right]$, $D_s\left[\left[\frac{k}{m}\right]h\right]$. Let $\hat{P}_s\left[\left[\frac{k}{m}\right]h\right]$, $\hat{M}_s\left[\left[\frac{k}{m}\right]h\right]$, $\hat{D}_s\left[\left[\frac{k}{m}\right]h\right]$ ensure minimum of (22.16) under (22.25), (22.27). Then optimal gain matrix of slow controller (22.14) is written as

$$\hat{K}_s\left[\left[\frac{k}{m}\right]h\right] = \hat{M}_s\left[\left[\frac{k}{m}\right]h\right]\hat{P}_s{}^{-1}\left[\left[\frac{k}{m}\right]h\right].$$

22.4 Design of Multirate Impulse Processes Control Systems for Stabilization of CM Nodes

To design automated impulse processes control system for stabilization of CM nodes coordinates it is necessary to write vectors $\overline{Y}_f$ and $\overline{Y}_s$ in models (22.7), (22.8) in full coordinates of CM nodes:

$$\overline{Y}_f\left[\left[\frac{k}{m}\right]h + (l+1)T_0\right] = (I_{11} + A_{11} - A_{11}q_1^{-1})\overline{Y}_f\left[\left[\frac{k}{m}\right]h + lT_0\right] + A_{12}\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right] + G_f\left[\left[\frac{k}{m}\right]h + lT_0\right]\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h + lT_0\right], \quad (22.29)$$

$l = 0, 1, 2, \ldots, m-1,$

$$\overline{Y}_s\left[\left(\left[\frac{k}{m}\right]+1\right)h\right] = (I_{22}+A_{22}-A_{22}q_2^{-1})\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right]$$
$$+A_{21}\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]+G_s\left[\left[\frac{k}{m}\right]h\right]\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right], \qquad (22.30)$$

where q_1^{-1}, q_2^{-1} are inverse shift operators for sampling periods T_0 and $h = mT_0$ respectively; I_{11}, I_{22} are identity matrices with dimensions $p \times p$ and $(n-p) \times (n-p)$ respectively.

To design control law for fast controller quadratic optimality criterion is utilized:

$$J_f\left[\left[\frac{k}{m}\right]h+(l+1)T_0\right] = E\left\{\left[\overline{Y}_f\left[\left[\frac{k}{m}\right]h+(l+1)T_0\right]-\overline{W}_f\right]^T\right.$$
$$\left.\times\left[\overline{Y}_f\left[\left[\frac{k}{m}\right]h+(l+1)T_0\right]-\overline{W}_f\right]+\Delta\overline{a}_f^T\left[\left[\frac{k}{m}\right]h+lT_0\right]R_f\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\right\},$$
$$l = 0, 1, \ldots, m-1, \quad (22.31)$$

where $\overline{W}_f$ is a set-point vector of predefined levels for CM nodes $\overline{Y}_f$, R_f is positive-definite weight matrix. Minimizing this criterion w.r.t. vector $\Delta\overline{a}_f$, having considered (22.29), we obtain the fast controller equation:

$$\frac{\partial J_f\left[\left[\frac{k}{m}\right]h+(l+1)T_0\right]}{\partial\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]} = 2G_f^T\left[\left[\frac{k}{m}\right]h+lT_0\right]\left\{\left(I_{11}+A_{11}-A_{11}q_1^{-1}\right)\right.$$
$$\times\overline{Y}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]+G_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]$$
$$\left.+A_{12}\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right]-\overline{W}_f\right\}+2R_f\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]=0.$$

Hence we get control law of the fast controller:

$$\Delta\overline{a}_f\left[\left[\frac{k}{m}\right]h+lT_0\right] = -\left(G_f^T\left[\left[\frac{k}{m}\right]h+lT_0\right]G_f\left[\left[\frac{k}{m}\right]h+lT_0\right]+R_f\right)^{-1}$$
$$\times G_f^T\left[\left[\frac{k}{m}\right]h+lT_0\right]\left\{(I_{11}+A_{11}-A_{11}q_1^{-1})\overline{Y}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]\right.$$
$$\left.+A_{12}\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{m}\right]h\right]-\overline{W}_f\right\}. \qquad (22.32)$$

To design control law for slow controller with sampling period h the following optimality criterion is proposed:

$$J_s\left[\left[\frac{k}{m}\right]h\right] = E\left\{\left[\overline{Y}_s\left[\left(\left[\frac{k}{m}\right]+1\right)h\right]-\overline{W}_s\right]^T \times\left[\overline{Y}_s\left[\left(\left[\frac{k}{m}\right]+1\right)h\right]-\overline{W}_s\right]+\Delta\overline{a}_s^T\left[\left[\frac{k}{m}\right]h\right]R_s\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right]\right\}, \quad (22.33)$$

where $\overline{W}_s$ is a set-point vector of predefined levels for CM nodes $\overline{Y}_s$, R_s is positive-definite weight matrix. Minimizing this criterion w.r.t. vector $\Delta\overline{a}_s$, having considered (22.30), we obtain the fast controller equation:

$$\frac{\partial J_s\left[\left(\left[\frac{k}{m}\right]+1\right)h\right]}{\partial\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right]} = 2G_s^T\left[\left[\frac{k}{m}\right]h\right]\left\{\left(I_{22}+A_{22}-A_{22}q_2^{-1}\right)\times\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right]+G_s\left[\left[\frac{k}{m}\right]h\right]\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right] +A_{21}\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]-\overline{W}_s\right\}+2R_s\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right]=0. \quad (22.34)$$

Hence we get control law of the slow controller:

$$\Delta\overline{a}_s\left[\left[\frac{k}{m}\right]h\right] = -\left(G_s^T\left[\left[\frac{k}{m}\right]h\right]G_s\left[\left[\frac{k}{m}\right]h\right]+R_s\right)^{-1}G_s^T\left[\left[\frac{k}{m}\right]h\right] \times\left\{(I_{22}+A_{22}-A_{22}q_2^{-1})\overline{Y}_s\left[\left[\frac{k}{m}\right]h\right]+A_{21}\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{m}\right]h+lT_0\right]-\overline{W}_s\right\}. \quad (22.35)$$

22.5 Example of Human Resources Management in IT Company Based on CM Weights Increments with Multirate Sampling

CM of IT company's human resources (HR) development process was built (Fig. 22.1). The nodes of this CM can be split into two groups:

1 Nodes measured with sampling period $T_0 = 1$ month, included into vector $\overline{Y}_f$. These are: career and staff management y_1, bonuses for early completion of tasks y_3, bonuses for new skills development y_4, level of quality monitoring y_5, planning of staff training process y_6, average salary y_7, company finance per employee y_8, employees satisfaction y_9, promotion perspectives y_{10}, spending on employees' sports y_{11}, staff

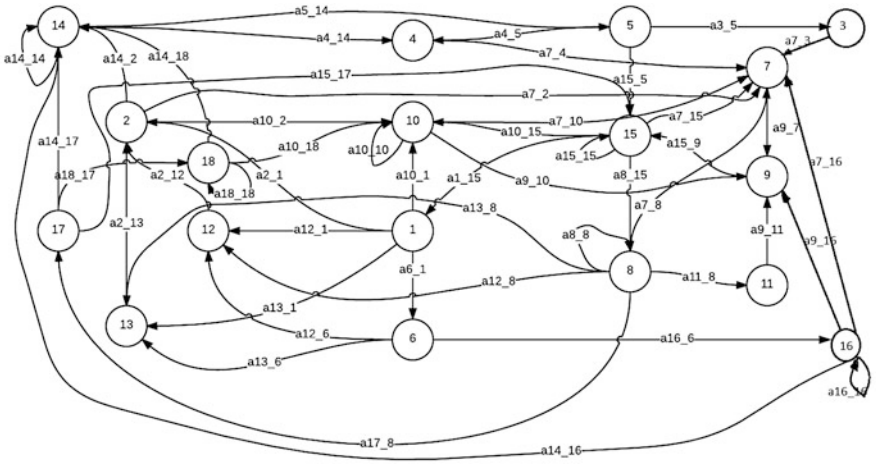


Fig. 22.1 CM of HR management

training without specialization change y_{12}, products innovativeness y_{15}, supporting staff training y_{16}.

2 CM nodes measured with sampling period $h = 6T_0 = 6$ months, included into vector $\overline{Y}_s$. These are staff certification y_2, staff retraining with change of main specialization y_{13}, professional skills y_{14}, spending on research and development y_{17}, graduate school effectiveness y_{18}.

Initial weights of this CM edges are the following: $a_{1,15} = 0.5$; $a_{2,1} = 0.4$; $a_{2,12} = 0.2$; $a_{2,13} = 0.5$; $a_{3,5} = -0.8$; $a_{4,5} = -0.8$; $a_{4,8} = 0.7$; $a_{4,14} = 0.5$; $a_{5,14} = 0.3$; $a_{6,1} = 0.5$; $a_{7,2} = 0.4$; $a_{7,3} = 0.2$; $a_{7,4} = 0.2$; $a_{7,8} = 0.5$; $a_{7,10} = 0.05$; $a_{7,15} = 0.2$; $a_{7,16} = 0.1$; $a_{8,8} = 0.4$; $a_{8,15} = 0.4$; $a_{9,7} = 0.4$; $a_{9,10} = 0.2$; $a_{9,11} = 0.3$; $a_{9,16} = 0.4$; $a_{10,1} = 0.5$; $a_{10,2} = 0.4$; $a_{10,10} = 0.5$; $a_{10,15} = 0.4$; $a_{10,18} = 0.4$; $a_{11,8} = 0.4$; $a_{12,1} = 0.4$; $a_{12,6} = 0.4$; $a_{12,8} = 0.5$; $a_{13,1} = 0.4$; $a_{13,6} = 0.4$; $a_{13,8} = 0.5$; $a_{14,2} = 0.7$; $a_{14,14} = 0.4$; $a_{14,16} = 0.1$; $a_{14,17} = 0.4$; $a_{14,18} = 0.4$; $a_{15,5} = 0.3$; $a_{15,9} = 0.15$; $a_{15,14} = 0.4$; $a_{15,15} = 0.3$; $a_{15,17} = 0.25$; $a_{16,6} = 0.8$; $a_{16,16} = 0.3$; $a_{17,8} = 0.3$; $a_{18,17} = 0.3$; $a_{18,18} = 0.4$.

Fast impulse process model (22.29) is written as

$$\overline{Y}_f\left[\left[\frac{k}{6}\right]h + (l+1)T_0\right] = (I_{11} + A_{11} - A_{11}q_1^{-1})\overline{Y}_f\left[\left[\frac{k}{6}\right]h + lT_0\right] + A_{12}\Delta\widetilde{\overline{Y}}_s\left[\left[\frac{k}{6}\right]h\right] +$$

$$+G_f\left[\left[\frac{k}{6}\right]h + lT_0\right]\Delta\overline{a}_f\left[\left[\frac{k}{6}\right]h + lT_0\right], l = 0, 1, \ldots, 5,$$

where matrices are

$$A_{11} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{1,15} & 0 \\ 0 & 0 & 0 & a_{3,5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{4,5} & 0 & 0 & a_{4,8} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{6,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{7,3} & a_{7,4} & 0 & 0 & 0 & a_{7,8} & 0 & a_{7,10} & 0 & 0 & a_{7,15} & a_{7,15} \\ 0 & 0 & 0 & 0 & 0 & 0 & a_{8,8} & 0 & 0 & 0 & 0 & a_{8,15} & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{9,7} & 0 & 0 & a_{9,10} & a_{9,11} & 0 & 0 & a_{9,16} \\ a_{10,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{10,10} & 0 & 0 & a_{10,15} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a_{11,8} & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{12,1} & 0 & 0 & 0 & a_{12,6} & 0 & a_{12,8} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{15,5} & 0 & 0 & 0 & a_{15,9} & 0 & 0 & 0 & a_{15,15} & 0 \\ 0 & 0 & 0 & 0 & a_{16,6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{16,16} \end{pmatrix},$$

$$A_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{4,14} & 0 & 0 \\ 0 & 0 & a_{5,14} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a_{7,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a_{10,2} & 0 & 0 & 0 & a_{10,18} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{15,14} & a_{15,17} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Control vector of weights increments for fast subsystem is the following:

$$\Delta\overline{a}_f = \left(\Delta a_{1,15}\ \Delta a_{3,5}\ \Delta a_{4,5}\ \Delta a_{6,1}\ \Delta a_{7,8}\ \Delta a_{10,1}\ \Delta a_{11,8}\ \Delta a_{12,6}\ \Delta a_{16,6} \right)^T.$$

Then matrix G_f is the following:

$$G_f\left[\left[\frac{k}{6}\right]h + lT_0\right] = G_f(kT_0) =$$

$$\begin{pmatrix} Y_{15}(kT_0) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & Y_5(kT_0) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & Y_5(kT_0) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & Y_1(kT_0) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & Y_8(kT_0) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & Y_1(kT_0) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & Y_8(kT_0) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & Y_6(kT_0) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & Y_6(kT_0) \end{pmatrix}.$$

Slow impulse process model (22.30) of this CM looks like

$$\overline{Y}_s\left[\left(\left[\frac{k}{m}\right]+1\right)h\right] = (I_{22}+A_{22}-A_{22}q_2^{-1})\overline{Y}_s\left[\left[\frac{k}{6}\right]h\right] + A_{21}\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{6}\right]h+lT_0\right] + \\ +G_s\left[\left[\frac{k}{6}\right]h\right]\Delta\overline{a}_s\left[\left[\frac{k}{6}\right]h\right],$$

where $\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{6}\right]h+lT_0\right] = \Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{6}\right]h+5T_0\right]-\Delta\widetilde{\overline{Y}}_f\left[\left[\frac{k}{6}\right]h-lT_0\right]$ and matrices are the following:

$$A_{21} = \begin{pmatrix} a_{2,1} & 0\,0\,0 & 0 & 0 & 0 & 0\,0\,0 & a_{2,12} & 0 & 0 \\ a_{13,1} & 0\,0\,0 & a_{13,6} & 0 & a_{13,8} & 0\,0\,0 & 0 & 0 & 0 \\ 0 & 0\,0\,0 & 0 & 0 & 0 & 0\,0\,0 & 0 & 0 & a_{14,16} \\ 0 & 0\,0\,0 & 0 & 0 & a_{17,8} & 0\,0\,0 & 0 & 0 & 0 \\ 0 & 0\,0\,0 & 0 & 0 & 0 & 0\,0\,0 & 0 & 0 & 0 \end{pmatrix},$$

$$A_{22} = \begin{pmatrix} 0 & a_{2,13} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a_{14,2} & 0 & a_{14,14} & a_{14,17} & a_{14,18} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{18,17} & a_{18,18} \end{pmatrix}.$$

Matrix G_s and increments vector $\Delta\overline{a}_s$ are

$$G_s\left[\left[\frac{k}{6}\right]h\right] = \begin{pmatrix} y_{13}\left[\left[\frac{k}{6}\right]h\right] & 0 \\ 0 & 0 \\ 0 & y_2\left[\left[\frac{k}{6}\right]h\right] \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \Delta\overline{a}_s\left[\left[\frac{k}{6}\right]h\right] = \begin{pmatrix} \Delta a_{2,13}\left[\left[\frac{k}{6}\right]h\right] \\ \Delta a_{14,2}\left[\left[\frac{k}{6}\right]h\right] \end{pmatrix}.$$

We simulated dynamics of both closed-loop subsystems using proposed methods of control. MatLab environment was used for simulation, specifically, SeDuMi package was useful to solve linear matrix inequalities [11]. Suppose that initial impulse affects all CM nodes negatively and aim of controller is to stabilize the system. Using the first method developed in the present work, stabilizing the system is ensured by minimizing invariant ellipsoids in which the increments of CM nodes' coordinates are placed. The coordinates increments converge to zero during the impulse process as it is shown on Fig. 22.2. Designed control inputs dynamics (in the form of CM weights' increments) are shown in Fig. 22.3.

Using the second method developed in the present work, stabilizing the system can be understood as getting the nodes coordinates back to level before initial impulse disturbance. Dynamics of CM nodes full coordinates and edges weights are shown on Figs. 22.4 and 22.5 respectively.

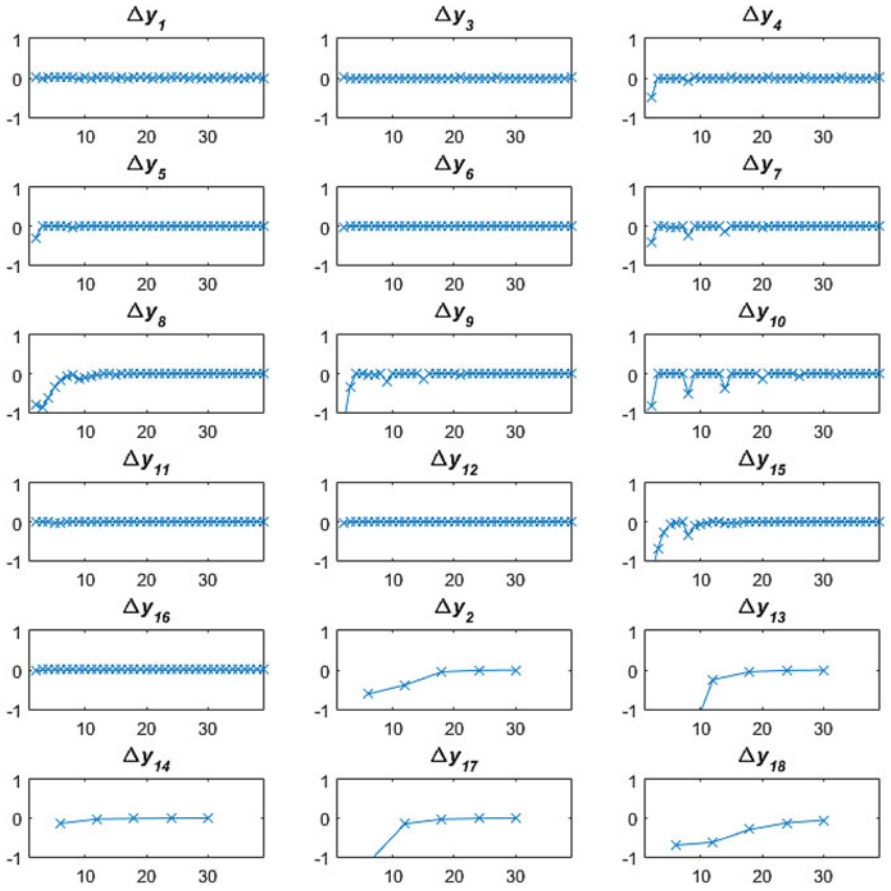


Fig. 22.2 CM nodes increments dynamics in the closed-loop system based on invariant ellipsoids method

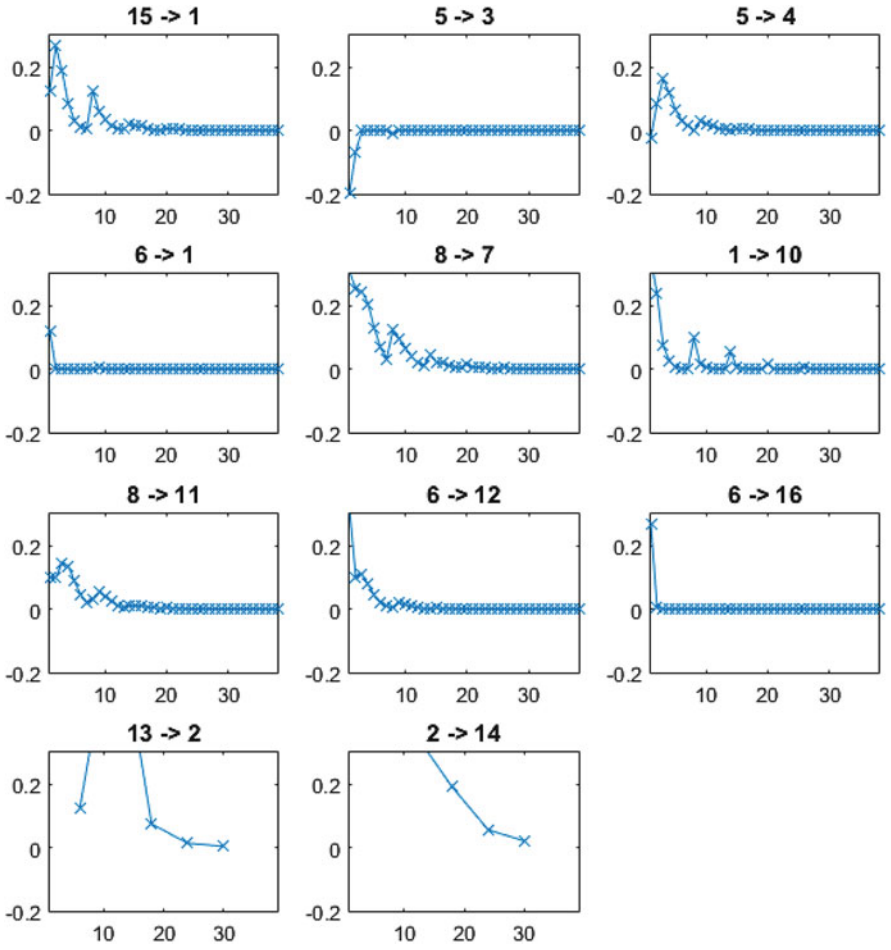


Fig. 22.3 CM weights increments dynamics in the closed-loop system based on invariant ellipsoids method

22.6 Conclusion

The present work develops the new principal of complex systems' dynamics control based on mathematical models of impulse processes in CM. Methods of automated control of CM impulse process with multirate sampling are discussed, where some of the coordinates are measured with small sampling period T_0 and other coordinates are measured with big sampling period $h = mT_0$. Basic impulse process model is written in the form of two interacting subsystems. The first subsystem describes dynamics of the first part of CM nodes $\overline{Y}_f$ in impulse process mode with sampling period T_0, and the second subsystem describes dynamics of other part of CM nodes $\overline{Y}_s$ with sampling period h. In the first subsystem changes of coordinates

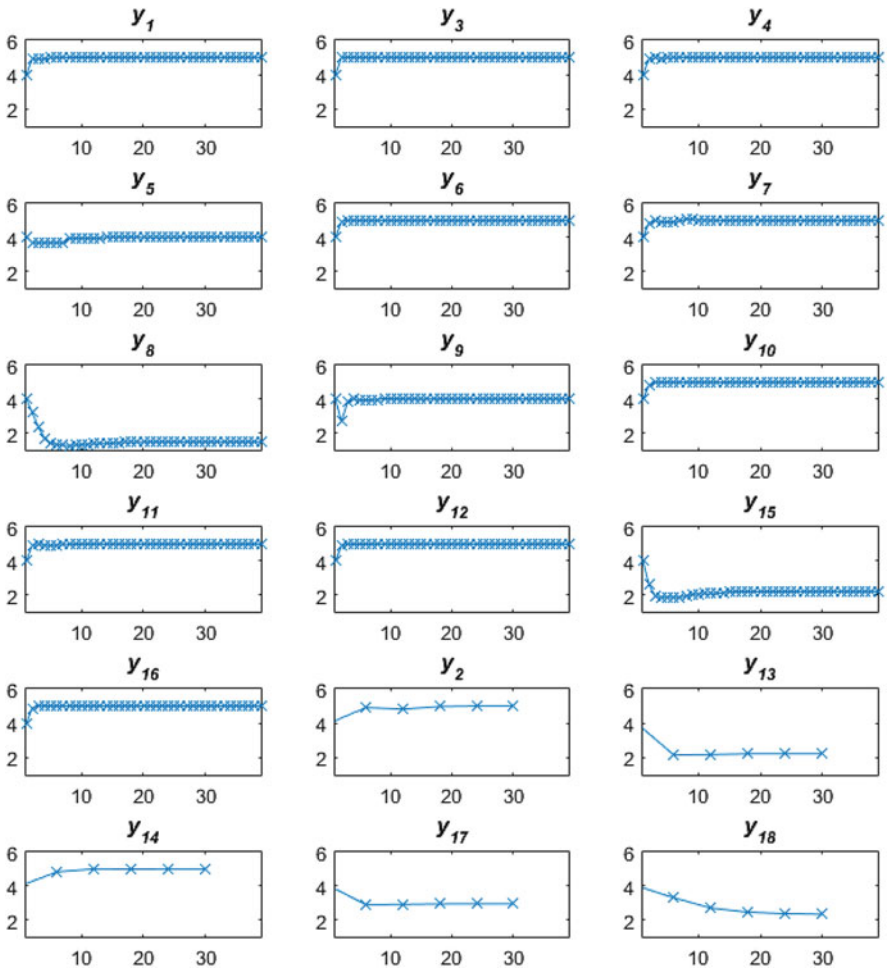


Fig. 22.4 CM nodes coordinates dynamics in the closed-loop system based on quadratic optimality criterion

$\overline{Y}_s$ are considered as disturbances, and in the second subsystem changes of $\overline{Y}_f$ are disturbances.

Unlike the report [8], here designed controls are implemented via CM edges' weights varying. They are generated in the closed-loop system with multirate sampling based on automated control theory methods. The following problems are solved in the present work:

- multirate state controllers are designed for suppression of constrained disturbances in the first and the second subsystems of CM impulse processes with multirate sampling based on invariant ellipsoids method;

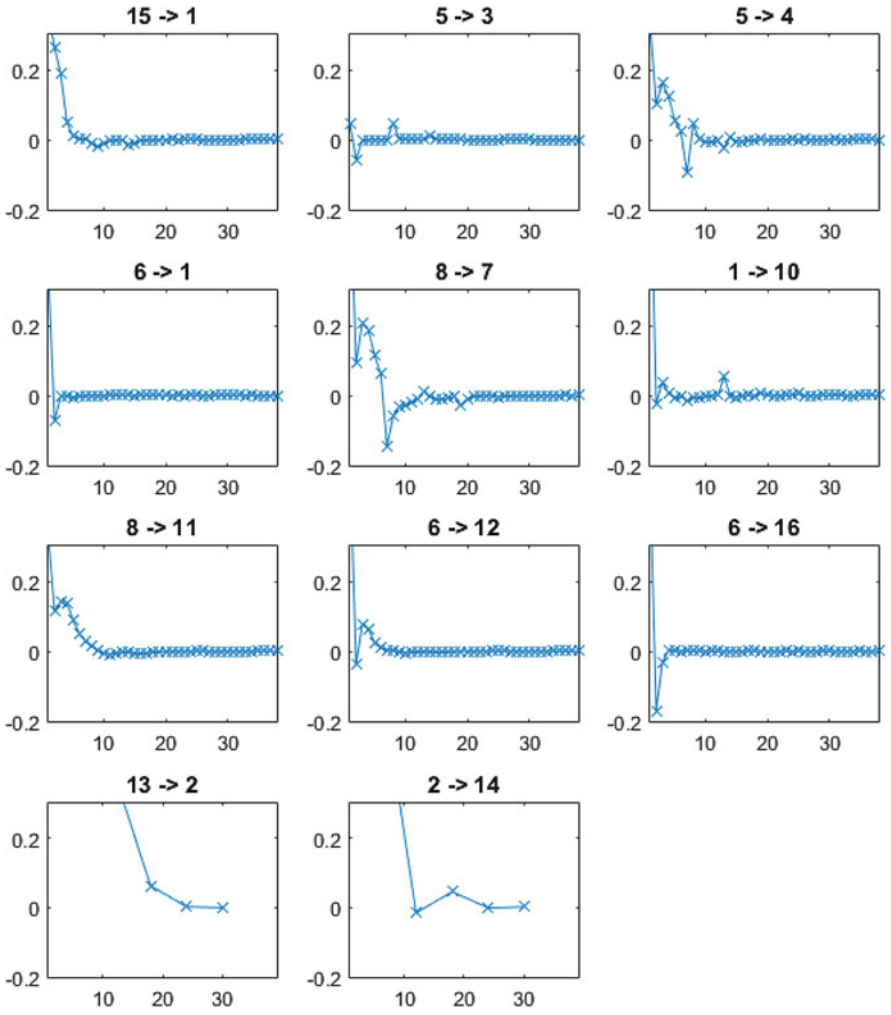


Fig. 22.5 CM weights increments dynamics in the closed-loop system based on quadratic optimality criterion

- control vectors are designed based on varying weights of the first and the second subsystems of CM impulse processes with multirate sampling ensuring transition of CM nodes coordinates from one level to another with further stabilization;
- impulse process in CM of human resources management in IT company is simulated with weights varying according to control laws generated by both proposed methods.

References

1. Roberts, F.: Discrete Mathematical Models with Applications to Social, Biological, and Environmental Problems. Englewood Cliffs, Prentice-Hall (1976)
2. Romanenko, V., Milyavsky, Y.: Stabilizing of impulse processes in cognitive maps based on state-space models. Syst. Res. Inf. Technol. **1**, 26–42 (2014) (in Russian)
3. Romanenko, V., Milyavsky, Y., Reutov, A.: Adaptive control method for unstable impulse processes in cognitive maps based on reference model. J. Autom. Inf. Sci. **47**(3), 11–23 (2015)
4. Romanenko, V., Milyavsky, Y.: Impulse processes stabilization in cognitive maps of complex systems based on modal regulators. Cybern. Comput. Eng. **179**, 43–55 (2015) (in Russian)
5. Zgurovsky, M., Romanenko, V., Milyavsky, Y.: Principles and methods of impulse processes control in cognitive maps of complex systems. Part I. J. Autom. Inf. Sci. **48**(1), 36–45 (2016)
6. Romanenko, V., Milyavsky, Y.: Adaptive coordinating control of interacting cognitive maps' vertices relations in impulse mode. Syst. Res. Inf. Technol. **3**, 109–120 (2015) (in Russian)
7. Romanenko, V., Milyavsky, Y.: Control method in cognitive maps based on weights increments. Cybern. Comput. Eng. **184**, 44–55 (2016)
8. Zgurovsky, M., Romanenko, V., Milyavsky, Y.: Adaptive control of impulse processes in complex systems cognitive maps with multirate coordinates sampling. In: Sadovnichiy, V., Zgurovsky, M. (eds.) Advances in Dynamical Systems and Control, vol. 69, pp. 363–374. Springer International Publishing, Basel (2016)
9. Polyak, B., Scherbakov, P.: Robust Stability and Control. Nauka, Moscow (2002) (in Russian)
10. Nazin, S., Polyak, B., Topunov, M.: Rejection of bounded exogenous disturbances by the method of invariant ellipsoids. Autom. Remote. Control. **68**(3), 467–486 (2007)
11. SeDuMi. Optimization over symmetric cones. Available via http://sedumi.ie.lehigh.edu/

Chapter 23
On Approximation of an Optimal Control Problem for Ill-Posed Strongly Nonlinear Elliptic Equation with p-Laplace Operator

Peter I. Kogut and Olha P. Kupenko

Abstract We study an optimal control problem for one class of non-linear elliptic equations with p-Laplace operator and L^1-nonlinearity in their right-hand side. We deal with such case of nonlinearity when we cannot expect to have a solution of the state equation for any given control. After defining a suitable functional class in which we look for solutions and a special cost functional, we prove the existence of optimal pairs. In order to handle the inherent degeneracy of the p-Laplacian and strong non-linearity in the right-hand side of elliptic equation, we use a two-parametric (ε, k)-regularization of p-Laplace operator, where we approximate it by a bounded monotone operator, and involve a special fictitious optimization problem. We derive existence of optimal solutions to the parametrized optimization problems at each (ε, k)-level of approximation. We also deduce the differentiability of the state for approximating problem with respect to the controls and obtain an optimality system based on the Lagrange principle. Further we discuss the asymptotic behaviour of the optimal solutions to regularized problems as the parameters ε and k tend to zero and infinity, respectively.

23.1 Introduction

Let Ω be a bounded open subset of $\mathbb{R}^N$ $(N \geq 3)$. We assume that its boundary $\partial\Omega$ is of the class $C^{1,1}$ and there exists a point $x_0 \in \operatorname{int}\Omega$ such that Ω is star-shaped with respect to x_0, i.e. $(\sigma - x_0, \nu(\sigma)) \geq 0$ for $\mathscr{H}^{N-1}$-a.a. $\sigma \in \partial\Omega$. Let

P. I. Kogut
Differential Equations Department, Oles Honchar National Dnipro University, Dnipro, Ukraine
e-mail: p.kogut@i.ua

O. P. Kupenko (✉)
System Analysis and Control Department, Dnipro National Technical University "Dnipro Polytechnics", Dnipro, Ukraine

Institute for Applied and System Analysis of National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_23

$F : \mathbb{R} \to [0, +\infty)$ be a mapping satisfying conditions $F \in C^1_{loc}(\mathbb{R})$, F is a convex function, and $F(z) \geq \exp(C_F z)$ for all $z \in \mathbb{R}$ with some constant $C_F > 0$. Let $f(z) = F'(z)$ and we assume that

$$f(z) \geq F(z), \quad \forall\, z \in \mathbb{R}, \quad \text{and} \quad \left| \int_{-\infty}^{0} z f(z)\, dz \right| < +\infty. \tag{23.1}$$

We are concerned with the following optimal control problem

$$\text{Minimize } J(u, y) = \frac{1}{2} \int_\Omega |y - y_d|^2\, dx + \frac{1}{q} \int_\Omega |u|^q\, dx + \frac{\alpha}{p'} \int_\Omega |f(y)|^{p'}\, dx, \tag{23.2}$$

subject to constrains

$$-\Delta_p\, y = f(y) + u \quad \text{in } \Omega, \tag{23.3}$$

$$y = 0 \quad \text{on } \partial\Omega, \tag{23.4}$$

$$u \in L^q(\Omega),\ y \in W^{1,p}_0(\Omega), \tag{23.5}$$

where $\alpha > 0$ is a given weight which is assumed to be small enough, $2 \leq p < N$, $q > p'$, $p' = p/(p-1) \in (1, 2]$ stands for the conjugate exponent, $\Delta_p\, y = \operatorname{div}\left(|\nabla y|^{p-2} \nabla y\right)$ is the p-Laplacian, and $y_d \in L^2(\Omega)$ is a given distribution.

The state $y \in W^{1,p}_0(\Omega)$ is called a weak solution of (23.3) and (23.4) if y belongs to the set

$$Y = \left\{ y \in W^{1,p}_0(\Omega) \;\middle|\; f(y) \in L^1(\Omega) \right\}, \tag{23.6}$$

and the integral identity

$$\int_\Omega |\nabla y|^{p-2} (\nabla y, \nabla \varphi)\, dx = \int_\Omega f(y)\varphi\, dx + \int_\Omega u\varphi\, dx \tag{23.7}$$

holds for every test function $\varphi \in C^\infty_0(\Omega)$.

A physical motivation to the study of optimal control problem (23.2)–(23.5), various applications of this type of boundary value problems (BVPs), and their main characteristic features are described in details in our recent paper [9] (see also [11, 12, 14]). We just mention here that the problem (23.3) and (23.4) can be seen as the stationary counterpart of evolution equations with nonlinear diffusion and the indicated BVP is ill-posed in general. Since the exponential growth of the term $f(y)$ can lead to the blow-up of the solutions of the corresponding evolution problems, it means that there is no reason to suppose that a weak solution to (23.3) and (23.4) for a given $u \in L^q(\Omega)$, even if it exists, is unique. As for the last term in the cost functional, it plays rather special role and it is unknown whether this OCP is consistent without this stabilizing term. In Sect. 23.2 we clarify this point in more details.

It is worth to note here that the optimal control problem (23.2)–(23.5) in the case of $p = 2$ and $f(y) = e^y$ was first discussed in detail by Casas et al. [5]. The problem of existence and uniqueness of the underlying boundary value problem and the corresponding optimal control problem was treated and an optimality system has been derived and analyzed. Analogous results for the case of general nonlinear elliptic equations of the type $\mathrm{div}(a(\nabla\cdot)) + f(\cdot)$ remained open. In this article we treat the case of the p-Laplacian, where $a(\nabla y) = |\nabla y|^{p-2}\nabla y$ and $p \geq 2$. The corresponding strongly nonlinear differential operator $-\mathrm{div}(|\nabla y|^{p-2}\nabla y) - f(y)$ is not monotone and, in principle, has degeneracies as ∇y tends to zero. Moreover, when the term $|\nabla y|^{p-2}$ is regarded as the coefficient of the Laplace operator, we also have the case of unbounded coefficients (see, for example, [8]). Because of this and L^1-boundedness of the function $f(y)$ there are serious hurdles to deduce an a priori estimate for the weak solutions of BVP (23.3) and (23.4) in the standard Sobolev space $W_0^{1,p}(\Omega)$. As a result, we focus on the case when for each admissible control $u \in L^q(\Omega)$, the original BVP (23.3) and (23.4) possesses a special type of weak solutions satisfying some extra state constraint (see the control-state inequality (23.10)). However, in this case it is not an easy matter to touch directly on this special set because its structure and the main topological properties are unknown in general (for the details we refer to [5, 9]). To lighten this problem and make the corresponding optimization procedure more feasible, some regularization and approximation of the optimal control problem (23.2)–(23.5) are necessary.

Using a monotone and bounded approximation $\mathscr{F}_k(|\nabla y|^2)$ of $|\nabla y|^2$ and following in many aspect our recent paper [6] (see also [7, 13]), we introduce a special family of optimization problems with fictitious controls and show that an optimal pair to the original optimal control problem can be attained by optimal solutions to the approximating ones provided the parameters $k \in \mathbb{N}$ and $\varepsilon > 0$ possess some special asymptotic properties. With that in mind we consequently provide the well-posedness analysis for the perturbed partial differential equations as well as for the corresponding fictitious optimal control problems. After that we pass to the limits as $k \to \infty$ and $\varepsilon \to 0$. Since the fictitious optimization problems are stated for the quasi-linear elliptic equations with coercive and monotone operators without any state and control constraints, the approximation and regularization approach is not only considered to be useful for the mathematical analysis, but also for the purpose of numerical simulations.

The plan of the paper is as follows. In Sect. 23.2 we study the existence of a solution for the original problem (23.2)–(23.5). A two-parametric family of approximating optimal control problems with a fictitious control is introduced in Sect. 23.3. We show here that each of these problems is consistent and admits at least one solution at each (ε, k)-level of approximation. Section 23.4 contains the proof of the main results of this paper (Theorems 23.5 and 23.6) and deals with the asymptotic analysis of the sequences of optimal solutions to the approximating problems (23.26)–(23.29), provided the parameter ε varies within a strictly decreasing sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ of positive real numbers satisfying some special condition. Finally, in Sect. 23.5 we deduce the differentiability of the state for approximating

problem with respect to the controls u and v and obtain the optimality system for each (ε, k)-level of approximation based on the Lagrange principle.

23.2 On Consistency of Optimal Control Problem (23.2)–(23.5)

Before proceeding further with qualitative analysis of the optimal control problem (23.2)–(23.5), we make use of the following results (see [9, Lemma 2.4 and Proposition 3.1]).

Lemma 23.1 *Let $y = y(u) \in Y$ be a weak solution to BVP (23.3) and (23.4) for a given $u \in L^q(\Omega)$. Then $f(y) \in W^{-1,p'}(\Omega)$, where $W^{-1,p'}(\Omega)$ stands for the dual space to $W_0^{1,p}(\Omega)$ with $p' = p/(p-1) \in (1; 2]$,*

$$\langle f(y), z\rangle_{W^{-1,p'}(\Omega);W^{1,p}(\Omega)} = \int_\Omega z\, f(y)\, dx, \quad \forall\, z \in W_0^{1,p}(\Omega), \tag{23.8}$$

and y satisfies the energy equality

$$\int_\Omega |\nabla y|^p\, dx = \int_\Omega y\, f(y)\, dx + \int_\Omega yu\, dx. \tag{23.9}$$

Proposition 23.1 *Let $u \in L^q(\Omega)$ and let $y = y(u) \in W_0^{1,p}(\Omega)$ be a weak solution to BVP (23.3) and (23.4). Assume that $f(y) \in L^{p'}(\Omega)$ and f satisfies properties (23.1). Then*

$$\left(\frac{N}{p} - 1\right) \int_\Omega |\nabla y|^p\, dx \le N \int_\Omega F(y)\, dx - \int_\Omega u\, (x - x_0, \nabla y)\, dx, \tag{23.10}$$

where $x_0 \in \operatorname{int} \Omega$ is an arbitrary point.

Theorem 23.1 *Let $u \in L^q(\Omega)$ and let $y = y(u) \in Y$ be a weak solution to BVP (23.3) and (23.4) such that y satisfies the inequality (23.10) and properties (23.1) hold true. Then*

$$\int_\Omega y\, f(y)\, dx \le C_1 \|u\|^{p'}_{L^q(\Omega)} + C_2 \|u\|^{p'-1}_{L^q(\Omega)} + C_3, \tag{23.11}$$

$$\|y\|_{W_0^{1,p}(\Omega)} \le C_4 \|u\|^{p'-1}_{L^q(\Omega)} + C_5, \tag{23.12}$$

for some positive constants C_i, $1 \le i \le 5$, independent of u and y.

It spite of the fact that inequality (23.10) makes sense even if we do not assume fulfillment of the inclusion $f(y) \in L^{p'}(\Omega)$ but have only that $y \in Y$ and $u \in$

$L^q(\Omega)$, it is unknown whether this inequality holds for an arbitrary weak solution to BVP (23.3) and (23.4). Since the existence and uniqueness of the weak solutions to the original BVP is an open question for arbitrary given control $u \in L^q(\Omega)$ with $q > p'$, the following result reflects some interesting properties of the Dirichlet boundary value problem (23.3) and (23.4) (see Theorem 7.2 in [9]).

Theorem 23.2 *Let $p \in [2, N)$ and let $u \in L^q(\Omega)$ with $q > p'$ be an arbitrary admissible control such that the boundary value problem (23.3) and (23.4) is solvable. Then the Dirichlet boundary value problem (23.3) and (23.4) admits a weak solution $y \in Y \subset W_0^{1,p}(\Omega)$ satisfying the inequality (23.10).*

Mainly inspired by this theorem, it has been proposed in [9] to reformulate the original control problem (23.2)–(23.5) to the following setting:

$$\text{Minimize } J(u, y) = \frac{1}{2}\int_\Omega |y - y_d|^2\, dx + \frac{1}{q}\int_\Omega |u|^q\, dx, \tag{23.13}$$

subject to constrains

$$-\Delta_p y = f(y) + u \quad \text{in } \Omega, \tag{23.14}$$

$$y = 0 \quad \text{on } \partial\Omega, \tag{23.15}$$

$$\int_\Omega |\nabla y|^p\, dx \le \frac{Np}{N-p}\int_\Omega F(y)\, dx - \frac{p}{N-p}\int_\Omega u\,(x - x_0, \nabla y)\, dx, \tag{23.16}$$

$$u \in L^q(\Omega),\ y \in W_0^{1,p}(\Omega), \tag{23.17}$$

where, from the formal point of view, the inequality (23.16) plays the role of an extra control-state constraint.

As follows from Theorem 23.2, the reformulated version (23.13)–(23.17) becomes a consistent optimization problem with a nonempty set of feasible solutions. Moreover, this problem has at least one solution for each $y_d \in L^2(\Omega)$ (see Theorem 4.1 in [9]). However, because of the inequality (23.16), there are serious hurdles to derive the corresponding optimality conditions for the problem (23.13)–(23.17) and provide its numerical simulations.

On the other hand, the validity of inequality (23.16) is a direct consequence of the condition $f(y) \in L^{p'}(\Omega)$. Hence, it is reasonable to consider instead of the problem (23.13)–(23.17), its regularized version in the form of the optimal control problem (23.2)–(23.5). As a result, its consistency immediately follows from Proposition 23.1, and moreover, the set of feasible solutions to the problem (23.2)–(23.5)

$$\Xi = \left\{ (u, y) \;\middle|\; \begin{array}{c} u \in L^q(\Omega), \quad y \in Y, \quad f(y) \in L^{p'}(\Omega), \\ \displaystyle\int_\Omega |\nabla y|^{p-2}\,(\nabla y, \nabla\varphi)\, dx = \int_\Omega f(y)\varphi\, dx \\ \displaystyle + \int_\Omega u\varphi\, dx,\ \forall\, \varphi \in W_0^{1,p}(\Omega) \end{array} \right\} \tag{23.18}$$

is always nonempty. Indeed, if we take an arbitrary function $\widetilde{y} \in C_0^\infty(\Omega)$ and put $\widetilde{u} := -\Delta_p \widetilde{y} - f(\widetilde{y})$, then $\widetilde{u} \in L^q(\Omega)$, $\widetilde{y} \in W_0^{1,p}(\Omega)$, and $f(\widetilde{y}) \in L^{p'}(\Omega)$. Hence, $\widetilde{y} \in Y \subset W_0^{1,p}(\Omega)$ is a weak solution to the boundary value problem (23.3) and (23.24) for given $\widetilde{u}$ and $(\widetilde{u}, \widetilde{y}) \in \Xi$. Let us show that the optimal control problem (23.2)–(23.5) is solvable.

Theorem 23.3 ([9]) *Let $p \in [2, N)$ and $q > p'$. Then, for a given $y_d \in L^2(\Omega)$, the optimal control problem (23.2)–(23.5) has at least one solution.*

Proof Since $J(u, y) \geq 0$ for all $(u, y) \in \Xi$, it follows that there exists a nonnegative value $\mu \geq 0$ such that $\mu = \inf_{(u,y)\in\Xi} J(u, y)$. Let $\{(u_k, y_k)\}_{k\in\mathbb{N}}$ be a minimizing sequence to the problem (23.2)–(23.5), i.e.

$$(u_k, y_k) \in \Xi \ \forall k \in \mathbb{N} \quad \text{and} \quad \lim_{k\to\infty} J(u_k, y_k) = \mu.$$

So, we can suppose that $J(u_k, y_k) \leq \mu + 1$ for all $k \in \mathbb{N}$. Taking into account the definition of the set Ξ, it follows from Proposition 23.1 and Theorem 23.1 that

$$\|y_k\|_{W_0^{1,p}(\Omega)} \leq C_4 \|u_k\|_{L^q(\Omega)}^{p'-1} + C_5$$

for some constants C_4 and C_5 independent of k and u_k. Then

$$\begin{aligned}
\|y_k\|_{W_0^{1,p}(\Omega)} + \|u_k\|_{L^q(\Omega)} &\leq C_5 + C_4 \left(q J(u_k, y_k)\right)^{\frac{p'-1}{q}} + \left(q J(u_k, y_k)\right)^{\frac{1}{q}} \\
&\leq C_5 + C_4 \left(q(\mu+1)\right)^{\frac{p'-1}{q}} + \left(q(\mu+1)\right)^{\frac{1}{q}}, \quad \forall k \in \mathbb{N}, \\
\|f(y_k)\|_{L^{p'}(\Omega)} &\leq \left(\frac{p'}{\alpha}(\mu+1)\right)^{1/p'}.
\end{aligned}$$

Thus, without loss of generality, we can suppose that there exists a subsequence of $\{(u_k, y_k)\}_{k\in\mathbb{N}}$ (still denoted by the same index) and a pair $(u^0, y^0) \in L^q(\Omega) \times W_0^{1,p}(\Omega)$ such that

$$(u_k, y_k) \rightharpoonup (u^0, y^0) \ \text{ weakly in } \ L^q(\Omega) \times W_0^{1,p}(\Omega) \ \text{ as } k \to \infty,$$

Let us show now that the limit pair (u^0, y^0) is related by the integral identity (23.7). By the Sobolev Embedding Theorem, the injection $W_0^{1,p}(\Omega) \hookrightarrow L^p(\Omega)$ is compact. Hence, the weak convergence $y_k \rightharpoonup y$ in $W_0^{1,p}(\Omega)$ implies the strong convergence in $L^p(\Omega)$. Therefore, up to a subsequence, we can suppose that $y_k(x) \to y(x)$ for almost every point $x \in \Omega$. As a result, we have the pointwise convergence: $f(y_k) \to f(y)$ almost everywhere in Ω. Since the sequence $\{f(y_k)\}_{k\in\mathbb{N}}$ is bounded in $L^{p'}(\Omega)$, it follows that

$$f(y_k) \rightharpoonup f(y^0) \quad \text{weakly in } \ L^{p'}(\Omega). \tag{23.19}$$

As a result, the limit passage in the right-hand side of the equality

$$\int_\Omega |\nabla y_k|^{p-2}\,(\nabla y_k, \nabla\varphi)\,dx = \int_\Omega f(y_k)\varphi\,dx + \int_\Omega u_k\varphi\,dx,\ \forall\,\varphi \in C_0^\infty(\mathbb{R}^N) \tag{23.20}$$

becomes trivial.

As for the limit passage as $k \to \infty$ in the left-hand side of (23.20), we make use of the following result (see Boccardo and Murat [2, Theorem 2.1]): if

(i) $y_k \to y^0$ weakly in $W_0^{1,p}(\Omega)$, strongly in $L^p(\Omega)$ and a.e. in Ω;
(ii) $u_k \to u^0$ strongly in $W^{-1,p'}(\Omega)$;
(iii) the sequence $\{f_k\}_{k\in\mathbb{N}}$ is bounded in $L^1(\Omega)$;
(iv) $-\mathrm{div}(|\nabla y_k|^{p-2}\nabla y_k) = f_k + u_k$ in $\mathscr{D}'(\Omega)$ for all $k \in \mathbb{N}$,

then, within a subsequence,

$$\nabla y_k \to \nabla y^0 \ \text{ strongly in } L^r(\Omega)^N \text{ for any } 1 \le r < p \text{ and a.e. in } \Omega. \tag{23.21}$$

In our case, the fulfilment of condition (iii) is guaranteed by the property (23.19). However, instead of (ii), we have the weak convergence $u_k \rightharpoonup u$ in $L^q(\Omega$ for the given $q > p'$. Since, by Sobolev embedding theorem, $W_0^{1,p}(\Omega)$ is continuously embedded in $L^{p^*}(\Omega)$ with $p^* = \frac{pN}{N-p}$, we have by duality arguments that $\left(L^{p^*}(\Omega)\right)^*$ is continuously embedded in $W^{-1,p'}(\Omega)$. So, if we define

$$p_* = (p^*)' = \frac{pN}{pN - N + p},$$

then we have

$$L^r(\Omega) \subset L^{p_*}(\Omega) \subset W^{-1,p'}(\Omega) \quad \forall\, r > \frac{pN}{pN - N + p}.$$

It is easy to check that $\frac{pN}{pN-N+p} < p' = \frac{p}{p-1}$ for all $p \ge 2$. Hence, the weak convergence $u_k \rightharpoonup u$ in $L^q(\Omega)$ with $q > p'$, implies the strong convergence in $W^{-1,p'}(\Omega)$. As for the rest assumptions of Boccardo–Murat Theorem they are obviously satisfied in our case. Hence, the pointwise convergence property (23.21), continuity of the mapping $\xi \mapsto |\xi|^{p-2}\xi$, and Vitali's theorem imply that

$$|\nabla y_k|^{p-2}\nabla y_k \to |\nabla y^0|^{p-2}\nabla y^0 \quad \text{strongly in } L^r(\Omega)^N \text{ for all } 1 \le r < p'. \tag{23.22}$$

Thus, taking these facts into account and passing to the limit in the integral identity (23.20) as $k \to \infty$, we see that y^0 is a weak solution to BVP (23.3) and (23.4) for the given $u^0 \in L^q(\Omega)$. Hence, (u^0, y^0) is a feasible pair to the problem (23.2)–(23.5). To conclude the proof, it remains to take into account the lower

semi-continuity of the cost functional $J : L^q(\Omega) \times W_0^{1,p}(\Omega) \to \mathbb{R}$ with respect to the weak convergence in $L^q(\Omega) \times W_0^{1,p}(\Omega)$ and property (23.19). This yields

$$\mu = \inf_{(u,y)\in\Xi} J(u, y) = \lim_{k\to\infty} J(u_k, y_k) \geq J(u^0, y^0).$$

Thus, $(u^0, y^0) \in \Xi$ is an optimal pair to the problem (23.2)–(23.5).

23.3 On Approximating Optimal Control Problems and Their Previous Analysis

We introduce, as in [6], the following two-parameter family of perturbed operators

$$\Delta_{\varepsilon,k,p}(y) = \operatorname{div}\left(\left(\varepsilon + \mathscr{F}_k\left(|\nabla y|^2\right)\right)^{\frac{p-2}{2}} \nabla y\right), \tag{23.23}$$

where $\mathscr{F}_k : \mathbb{R}_+ \to \mathbb{R}_+$ is a non-decreasing $C^1(\mathbb{R}_+)$-function such that

$$\begin{gathered}\mathscr{F}_k(t) = t, \ \text{ if } \ t \in \left[0, k^2\right], \quad \mathscr{F}_k(t) = k^2 + 1, \ \text{ if } \ t > k^2 + 1, \quad \text{and} \\ t \leq \mathscr{F}_k(t) \leq t + \delta, \ \text{ if } \ k^2 \leq t < k^2 + 1 \quad \text{for some } \ \delta \in (0, 1), \\ \mathscr{F}_k'(t) \leq \delta^*, \ \text{ if } \ k^2 \leq t < k^2 + 1 \quad \text{for some } \ \delta^* > 1,\end{gathered} \tag{23.24}$$

and the constants δ and δ^* are independent of $k \in \mathbb{N}$. In particular, if

$$\mathscr{F}_k(t) = \begin{cases} t, & \text{if } \ 0 \leq t \leq k^2, \\ (k^2 - t)^3 + (k^2 - t)^2 + t, & \text{if } \ k^2 \leq t \leq k^2 + 1, \\ k^2 + 1, & \text{if } \ t \geq k^2 + 1. \end{cases} \tag{23.25}$$

then $\delta = 4/27$ and $\delta^* = 4/3$ satisfy (23.24). Hereinafter, we assume that the parameter ε varies within a strictly decreasing sequence of positive real numbers which converge to zero.

We now introduce the following perturbed optimal control problem (see, for comparison, [3, 4]).

$$\begin{aligned}\text{Minimize } \Big\{ I_{\varepsilon,k}(u, v, y) = \frac{1}{2}\int_\Omega |y - y_d|^2\, dx + \frac{k}{p'}\int_\Omega |v - T_\varepsilon(f(y))|^{p'}\, dx \\ + \frac{1}{q}\int_\Omega |u|^q\, dx + \frac{\alpha}{p'}\int_\Omega |v|^{p'}\, dx \Big\}\end{aligned} \tag{23.26}$$

subject to the constraints

$$-\Delta_{\varepsilon,k,p}(y) = v + u \quad \text{in} \quad \Omega, \tag{23.27}$$

$$y = 0 \quad \text{on } \partial\Omega, \tag{23.28}$$

$$v \in L^{p'}(\Omega), \quad u \in L^q(\Omega), \quad y \in H_0^1(\Omega). \tag{23.29}$$

Here, $T_\varepsilon : \mathbb{R} \to \mathbb{R}$ is the truncation operator defined by

$$T_\varepsilon(s) = \max\left\{\min\left\{s, \varepsilon^{-1}\right\}, -\varepsilon^{-1}\right\}. \tag{23.30}$$

We consider the function $v \in L^{p'}(\Omega)$ as a fictitious control.

For our further analysis, we make use of the following the notation

$$\|\varphi\|_{\varepsilon,k} = \left(\int_\Omega \left(\varepsilon + \mathscr{F}_k(|\nabla\varphi|^2)\right)^{\frac{p-2}{2}} |\nabla\varphi|^2\, dx\right)^{1/p} \quad \forall \varphi \in H_0^1(\Omega).$$

It is clear that the effect of such perturbation of $\Delta_p(y)$ is to provide its regularization around points where $|\nabla y(x)|$ is equal to zero and becomes unbounded. Indeed, for an arbitrary element $y^* \in H_0^1(\Omega)$ let us consider the level set $\Omega_k(y^*) := \left\{x \in \Omega \ : \ |\nabla y^*(x)| > \sqrt{k^2+1}\right\}$. Then

$$\begin{aligned}
|\Omega_k(y^*)| &:= \int_{\Omega_k(y^*)} 1\, dx \le \frac{1}{\sqrt{k^2+1}} \int_{\Omega_k(y^*)} |\nabla y^*(x)|\, dx \\
&\le \frac{1}{k} |\Omega_k(y^*)|^{\frac{1}{2}} \left(\int_{\Omega_k(y^*)} |\nabla y^*|^2\, dx\right)^{\frac{1}{2}} \\
&= \frac{1}{k}\left(\frac{1}{\varepsilon + k^2 + 1}\right)^{\frac{p-2}{4}} \left(\int_{\Omega_k(y^*)} \left(\varepsilon + \mathscr{F}_k(|\nabla y^*|^2)\right)^{\frac{p-2}{2}} |\nabla y^*|^2 dx\right)^{\frac{1}{2}} |\Omega_k(y^*)|^{\frac{1}{2}} \\
&\le \frac{1}{k^{\frac{p}{2}}} |\Omega_k(y^*)|^{\frac{1}{2}} \|y^*\|_{\varepsilon,k}^{\frac{p}{2}}.
\end{aligned}$$

Hence, the Lebesgue measure of the set $\Omega_k(y^*)$ satisfies the estimate

$$|\Omega_k(y^*)| \le \frac{1}{k^p} \|y^*\|_{\varepsilon,k}^p, \quad \forall\, y^* \in H_0^1(\Omega). \tag{23.31}$$

In what follows, we say that for given $\varepsilon > 0$, $k \in \mathbb{N}$, $u \in L^q(\Omega)$, and $v \in L^{p'}(\Omega)$, a distribution $y_{\varepsilon,k} \in H_0^1(\Omega)$ is the weak solution to boundary value problem (23.27) and (23.28) if

$$\int_\Omega (\varepsilon + \mathscr{F}_k(|\nabla y_{\varepsilon,k}|^2))^{\frac{p-2}{2}} \left(\nabla y_{\varepsilon,k}, \nabla\varphi\right)_{\mathbb{R}^N} dx = \int_\Omega (u+v)\varphi\, dx, \quad \forall\, \varphi \in C_0^\infty(\Omega), \tag{23.32}$$

or equivalently

$$\int_\Omega (\varepsilon + \mathscr{F}_k(|\nabla\varphi|^2))^{\frac{p-2}{2}} \left(\nabla\varphi, \nabla\varphi - \nabla y_{\varepsilon,k}\right)_{\mathbb{R}^N} dx$$
$$\geq \int_\Omega (u+v)(\varphi - y_{\varepsilon,k})\, dx, \quad \forall\, \varphi \in C_0^\infty(\Omega). \tag{23.33}$$

To prove the well-posedness of the boundary value problem (23.27) and (23.28), it is enough to show that $L^{p'}(\Omega)$ is continuously embedded in $H^{-1}(\Omega)$. Indeed, by Sobolev embedding theorem, we have: $H_0^1(\Omega) \hookrightarrow L^r(\Omega)$ compactly for all $r \in [1, 2N/(N-2))$. Hence, the compactness of injection $L^{r'}(\Omega) \hookrightarrow H^{-1}(\Omega)$ holds true if only $r' > 2N/(N+2)$. Since $p' > 2N/(N+2)$ for each $p \in \left[2, 2N/(N-1)\right)$, it follows by duality arguments that $L^{p'}(\Omega) \hookrightarrow H^{-1}(\Omega)$ with a compact embedding. So, there exists a constant $C_p > 0$ such that $\|v\|_{H^{-1}(\Omega)} \leq C_p \|v\|_{L^{p'}(\Omega)}$ for all $v \in L^{p'}(\Omega)$. Moreover, taking into account the estimate

$$\int_\Omega v\varphi\, dx = \langle v, \varphi\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} \leq \|v\|_{H^{-1}(\Omega)} \|\varphi\|_{H_0^1(\Omega)}$$
$$\leq C_p \|v\|_{L^{p'}(\Omega)} \|\varphi\|_{H_0^1(\Omega)}, \quad \forall\, \varphi \in C_0^\infty(\Omega)$$

and the inclusion $u + v \in L^{p'}(\Omega)$, we see that the right-hand side of (23.32) can be extended to a linear continuous functional on $H_0^1(\Omega)$,

$$L(\varphi) := \langle v, \varphi\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} + \int_\Omega u\varphi\, dx, \quad \forall\, \varphi \in H_0^1(\Omega).$$

Since the operator $-\Delta_{\varepsilon,k,p}(\cdot) : H_0^1(\Omega) \to H^{-1}(\Omega)$ is bounded, strictly monotone, semi-continuous, and coercive (see [6]), it follows from the general theory of monotone operators that for each $\varepsilon > 0$, $k \in \mathbb{N}$, $p \in \left[2, 2N/(N-1)\right)$, $u \in L^q(\Omega)$, and $v \in L^{p'}(\Omega)$, the boundary value problem (23.27) and (23.28) admits a unique weak solution $y_{\varepsilon,k} \in H_0^1(\Omega)$ satisfying the energy equality (see Theorem 4.5 in [6])

$$\|y_{\varepsilon,k}\|_{\varepsilon,k}^p = \left\langle v, y_{\varepsilon,k}\right\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} + \int_\Omega u y_{\varepsilon,k}\, dx. \tag{23.34}$$

From this it is easy to deduce that, for every positive value $\varepsilon > 0$ and integer $k \in \mathbb{N}$, the set of feasible solutions to the problem (23.26)–(23.29)

$$\Xi_{\varepsilon,k} = \left\{ (u, v, y) \;\middle|\; \begin{array}{c} u \in L^q(\Omega),\ v \in L^{p'}(\Omega),\ y \in H_0^1(\Omega),\ I_{\varepsilon,k}(u, v, y) < +\infty, \\ (u, v, y) \text{ are related by identity (23.32)} \end{array} \right\} \tag{23.35}$$

is nonempty.

For our further analysis, we assume that $p \in \big[2, 2N/(N-1)\big)$. We also need to obtain some appropriate a priory estimates for the weak solutions to problem (23.27) and (23.28). With that in mind, we make use of the following auxiliary results.

Proposition 23.2 *Let $k \in \mathbb{N}$ and $\varepsilon > 0$ be given. Then, for arbitrary $u \in L^q(\Omega)$, $v \in L^{p'}(\Omega)$, and $y \in H_0^1(\Omega)$, we have*

$$\left|\langle v, y\rangle_{H^{-1}(\Omega);H_0^1(\Omega)}\right| \le C_p \|v\|_{L^{p'}(\Omega)} \left[|\Omega|^{\frac{p-2}{2p}} \|y\|_{\varepsilon,k} + \|y\|_{\varepsilon,k}^{\frac{p}{2}}\right], \tag{23.36}$$

$$\left|\int_\Omega uy\,dx\right| \le C_p |\Omega|^{\frac{q-p'}{qp'}} \|u\|_{L^q(\Omega)} \left[|\Omega|^{\frac{p-2}{2p}} \|y\|_{\varepsilon,k} + \|y\|_{\varepsilon,k}^{\frac{p}{2}}\right]. \tag{23.37}$$

Proof Let us fix an arbitrary element y of $H_0^1(\Omega)$. We associate with this element the set $\Omega^k(y)$, where $\Omega^k(y) := \{x \in \Omega \ : \ |\nabla y(x)| > k\}$. Then

$$\begin{aligned}\int_\Omega uy\,dx &\le C_p \|u\|_{L^{p'}(\Omega)} \|y\|_{H_0^1(\Omega)} \\ &\le C_p |\Omega|^{\frac{q-p'}{qp'}} \|u\|_{L^q(\Omega)} \left(\|\nabla y\|_{L^2(\Omega\setminus\Omega^k(y))^N} + \|\nabla y\|_{L^2(\Omega^k(y))^N}\right),\end{aligned} \tag{23.38}$$

$$\begin{aligned}\langle v, y\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} &\le C_p \|v\|_{L^{p'}(\Omega)} \|y\|_{H_0^1(\Omega)} \\ &= C_p \|v\|_{L^{p'}(\Omega)} \left(\|\nabla y\|_{L^2(\Omega\setminus\Omega^k(y))^N} + \|\nabla y\|_{L^2(\Omega^k(y))^N}\right).\end{aligned} \tag{23.39}$$

Using the fact that

$$\begin{aligned}\|\nabla y\|_{L^2(\Omega\setminus\Omega^k(y))^N} &\le |\Omega|^{\frac{p-2}{2p}} \|\nabla y\|_{L^p(\Omega\setminus\Omega^k(y))^N} \\ &\le |\Omega|^{\frac{p-2}{2p}} \left(\int_{\Omega\setminus\Omega^k(y)} (\varepsilon + |\nabla y|^2)^{\frac{p-2}{2}} |\nabla y|^2\,dx\right)^{\frac{1}{p}}\end{aligned}$$

and

$$\begin{aligned}&\mathscr{F}_k(|\nabla y|^2) = |\nabla y|^2 \ \text{ a.e. in } \Omega \setminus \Omega^k(y), \text{ and} \\ &k^2 \le \mathscr{F}_k(|\nabla y|^2) \le k^2 + 1 \ \text{ a.e. in } \Omega^k(y), \quad \forall\, k \in \mathbb{N},\end{aligned}$$

we obtain

$$\|\nabla y\|_{L^2(\Omega\setminus\Omega^k(y))^N} \le |\Omega|^{\frac{p-2}{2p}} \left(\int_{\Omega\setminus\Omega^k(y)} (\varepsilon + \mathscr{F}_k(|\nabla y|^2))^{\frac{p-2}{2}} |\nabla y|^2\, dx\right)^{\frac{1}{p}}$$
$$= |\Omega|^{\frac{p-2}{2p}} \|y\|_{\varepsilon,k}, \tag{23.40}$$

$$\|\nabla y\|_{L^2(\Omega^k(y))^N} \le \left(\int_{\Omega^k(y)} (\varepsilon + \mathscr{F}_k(|\nabla y|^2))^{\frac{p-2}{2}} |\nabla y|^2\, dx\right)^{\frac{1}{2}} = \|y\|_{\varepsilon,k}^{\frac{p}{2}}. \tag{23.41}$$

As a result, inequalities (23.36) and (23.37) immediately follows from (23.38)–(23.41).

Definition 23.1 Let $\{u_{\varepsilon,k}, v_{\varepsilon,k}\}_{\substack{\varepsilon>0\\ k\in\mathbb{N}}} \subset L^q(\Omega)\times L^{p'}(\Omega)$ be an arbitrary sequence of admissible controls. We say that a two-parametric sequence $\{y_{\varepsilon,k}\}_{\substack{\varepsilon>0\\ k\in\mathbb{N}}} \subset H_0^1(\Omega)$ is bounded with respect to the $\|\cdot\|_{\varepsilon,k}$-quasi-seminorm if $\sup_{\substack{\varepsilon>0\\ k\in\mathbb{N}}} \|y_{\varepsilon,k}\|_{\varepsilon,k} < +\infty$.

Let us show that for every $(u, v) \in L^q(\Omega)\times L^{p'}(\Omega)$, the sequence of weak solutions to the boundary value problem (23.27) and (23.28) $\{y_{\varepsilon,k} = y_{\varepsilon,k}(u, v)\}_{\substack{\varepsilon>0\\ k\in\mathbb{N}}}$ is bounded with respect to the $\|\cdot\|_{\varepsilon,k}$-quasi-seminorm in the sense of Definition 23.1.

Indeed, the energy equality (23.34) together with estimates (23.36) and (23.37) immediately lead us to the relation

$$\|y_{\varepsilon,k}\|_{\varepsilon,k}^p := \int_\Omega \left(\varepsilon + \mathscr{F}_k(|\nabla y_{\varepsilon,k}|^2)\right)^{\frac{p-2}{2}} |\nabla y_{\varepsilon,k}|^2\, dx$$
$$= \langle v, y_{\varepsilon,k}\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} + \int_\Omega u y_{\varepsilon,k}\, dx$$
$$\le C_p \left(\|v\|_{L^{p'}(\Omega)} + |\Omega|^{\frac{q-p'}{qp'}} \|u\|_{L^q(\Omega)}\right) \left[|\Omega|^{\frac{p-2}{2p}} \|y_{\varepsilon,k}\|_{\varepsilon,k} + \|y_{\varepsilon,k}\|_{\varepsilon,k}^{\frac{p}{2}}\right]. \tag{23.42}$$

As a result, it follows from (23.42) that

$$\|y_{\varepsilon,k}\|_{\varepsilon,k} \le \max\left\{C_{u,v}^{\frac{2}{p}}, C_{u,v}^{\frac{1}{p-1}}\right\}, \quad \forall\, \varepsilon > 0,\ \forall\, k \in \mathbb{N},\ \forall\, u \in L^q(\Omega),\ \forall\, v \in L^{p'}(\Omega), \tag{23.43}$$

where

$$C_{u,v} := C_p \left(\|v\|_{L^{p'}(\Omega)} + |\Omega|^{\frac{q-p'}{qp'}} \|u\|_{L^q(\Omega)}\right) \left(|\Omega|^{\frac{p-2}{2p}} + 1\right). \tag{23.44}$$

To conclude this section, we give the following result.

Theorem 23.4 *Let $p \in [2, 2N/(N-1))$ and $q > p'$. Then, for every positive value $\varepsilon > 0$ and integer $k \in \mathbb{N}$, the approximating optimal control problem (23.26)–(23.29) has at least one solution.*

Proof Since $I_{\varepsilon,k}(u, v, y) \geq 0$ for all $(u, v, y) \in L^q(\Omega) \times L^{p'}(\Omega) \times H_0^1(\Omega)$, it follows that there exists a non-negative value $\mu_{\varepsilon,k}$ such that $\mu_{\varepsilon,k} = \inf_{(u,v,y)\in\Xi_{\varepsilon,k}} I_{\varepsilon,k}(u, v, y)$. Let $\{(u_{\varepsilon,k,m}, v_{\varepsilon,k,m}, y_{\varepsilon,k,m})\}_{m\in\mathbb{N}}$ be a minimizing sequence, i.e.

$$(u_{\varepsilon,k,m}, v_{\varepsilon,k,m}, y_{\varepsilon,k,m}) \in \Xi_{\varepsilon,k} \ \forall\, m \in \mathbb{N} \quad \text{and} \quad \lim_{m\to\infty} I_{\varepsilon,k}(u_{\varepsilon,k,m}, v_{\varepsilon,k,m}, y_{\varepsilon,k,m}) = \mu_{\varepsilon,k}.$$

So, without loss of generality, we can suppose that

$$\begin{aligned} I_{\varepsilon,k}(u_{\varepsilon,k,m}, v_{\varepsilon,k,m}, y_{\varepsilon,k,m}) &= \frac{1}{2}\int_\Omega |y_{\varepsilon,k,m} - y_d|^2\,dx + \frac{k}{p'}\int_\Omega |v_{\varepsilon,k,m} - T_\varepsilon(f(y_{\varepsilon,k,m}))|^{p'}\,dx \\ &\quad + \frac{1}{q}\int_\Omega |u_{\varepsilon,k,m}|^q\,dx + \frac{\alpha}{p'}\int_\Omega |v_{\varepsilon,k,m}|^{p'}\,dx \\ &\leq \mu_{\varepsilon,k} + 1 \quad \text{for all } m \in \mathbb{N}. \end{aligned} \tag{23.45}$$

Since $T_\varepsilon(f(y_{\varepsilon,k,m})) \in L^\infty(\Omega)$, it follows from (23.45) that the sequence of fictitious controls $\{v_{\varepsilon,k,m}\}_{k\in\mathbb{N}}$ is uniformly bounded in $L^{p'}(\Omega)$. The similar conclusion can be made for the sequence $\{u_{\varepsilon,k,m}\}_{k\in\mathbb{N}}$. So, we can admit the existence of elements $v^0_{\varepsilon,k} \in L^{p'}(\Omega)$ and $u^0_{\varepsilon,k} \in L^q(\Omega)$ such that (up to a subsequence)

$$v_{\varepsilon,k,m} \rightharpoonup v^0_{\varepsilon,k} \ \text{ in } \ L^{p'}(\Omega) \ \text{ and } \ u_{\varepsilon,k,m} \rightharpoonup u^0_{\varepsilon,k} \ \text{ in } \ L^q(\Omega) \ \text{ as } m \to \infty. \tag{23.46}$$

Moreover, in view of estimate (23.43), we see that the sequence $\{y_{\varepsilon,k,m}\}_{k\in\mathbb{N}}$ is bounded in $H_0^1(\Omega)$. Indeed, setting $\Omega^k(y_{\varepsilon,k,m}) := \{x \in \Omega : |\nabla y_{\varepsilon,k,m}(x)| > k\}$ for each $k \in \mathbb{N}$, we have

$$\begin{aligned} \|y_{\varepsilon,k,m}\|_{H_0^1(\Omega)} &\leq \|\nabla y_{\varepsilon,k,m}\|_{L^2(\Omega\setminus\Omega^k(y_{\varepsilon,k,m}))^N} + \|\nabla y_{\varepsilon,k,m}\|_{L^2(\Omega^k(y_{\varepsilon,k,m}))^N} \\ &\overset{\text{by (23.40) and (23.41)}}{\leq} \left[|\Omega|^{\frac{p-2}{2p}}\|y_{\varepsilon,k,m}\|_{\varepsilon,k} + \|y_{\varepsilon,k,m}\|^{\frac{p}{2}}_{\varepsilon,k}\right] \overset{\text{by (23.43)}}{<} +\infty. \end{aligned} \tag{23.47}$$

As a result, we deduce the existence of a subsequence of $\{y_{\varepsilon,k,m}\}_{k\in\mathbb{N}}$, denoted in the same way, and an element $y^0_{\varepsilon,k} \in H_0^1(\Omega)$ such that $y_{\varepsilon,k,m} \rightharpoonup y^0_{\varepsilon,k}$ in $H_0^1(\Omega)$ as $m \to \infty$. Let us prove that $y^0_{\varepsilon,k}$ is the solution of (23.27) and (23.28) with $v = v^0_{\varepsilon,k}$ and $u = u^0_{\varepsilon,k}$. Let us fix an arbitrary test function $\varphi \in C_0^\infty(\Omega)$ and pass to the limit

in the Minty inequality

$$\int_\Omega (\varepsilon + \mathscr{F}_k(|\nabla\varphi|^2))^{\frac{p-2}{2}} \left(\nabla\varphi, \nabla\varphi - \nabla y_{\varepsilon,k,m}\right)_{\mathbb{R}^N} dx$$
$$\geq \left\langle v_{\varepsilon,k,m}, (\varphi - y_{\varepsilon,k,m})\right\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} + \int_\Omega u_{\varepsilon,k,m}(\varphi - y_{\varepsilon,k,m})\, dx, \qquad (23.48)$$

as $m \to \infty$. In view of the convergences $v_{\varepsilon,k,m} \rightharpoonup v^0_{\varepsilon,k}$ in $L^{p'}(\Omega)$, $u_{\varepsilon,k,m} \rightharpoonup u^0_{\varepsilon,k}$ in $L^q(\Omega)$, and $y_{\varepsilon,k,m} \to y^0_{\varepsilon,k}$ strongly in $L^2(\Omega)$, we obtain

$$\lim_{m\to\infty} \int_\Omega (\varepsilon + \mathscr{F}_k(|\nabla\varphi|^2))^{\frac{p-2}{2}} \left(\nabla\varphi, \nabla y_{\varepsilon,k,m}\right)_{\mathbb{R}^N} dx$$
$$= \int_\Omega (\varepsilon + \mathscr{F}_k(|\nabla\varphi|^2))^{\frac{p-2}{2}} \left(\nabla\varphi, \nabla y^0_{\varepsilon,k}\right)_{\mathbb{R}^N} dx,$$

$$\lim_{m\to\infty} \left\langle v_{\varepsilon,k,m}, (\varphi - y_{\varepsilon,k,m})\right\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} = \left\langle v^0_{\varepsilon,k}, (\varphi - y^0_{\varepsilon,k})\right\rangle_{H^{-1}(\Omega);H_0^1(\Omega)},$$

$$\lim_{m\to\infty} \int_\Omega u_{\varepsilon,k,m}(\varphi - y_{\varepsilon,k,m})\, dx = \int_\Omega u^0_{\varepsilon,k}(\varphi - y^0_{\varepsilon,k})\, dx.$$

Thus, passing to the limit in relation (23.48) as $m \to \infty$, we arrive at the inequality

$$\int_\Omega (\varepsilon + \mathscr{F}_k(|\nabla\varphi|^2))^{\frac{p-2}{2}} \left(\nabla\varphi, \nabla\varphi - \nabla y^0_{\varepsilon,k}\right)_{\mathbb{R}^N} dx \geq \left\langle v^0_{\varepsilon,k}, (\varphi - y^0_{\varepsilon,k})\right\rangle_{H^{-1}(\Omega);H_0^1(\Omega)}$$
$$+ \int_\Omega u^0_{\varepsilon,k}(\varphi - y^0_{\varepsilon,k})\, dx, \quad \forall\, \varphi \in C_0^\infty(\Omega).$$

Finally, from the density of $C_0^\infty(\Omega)$ in $H_0^1(\Omega)$, we infer that the integral identity

$$\int_\Omega (\varepsilon + \mathscr{F}_k(|\nabla y^0_{\varepsilon,k}|^2))^{\frac{p-2}{2}} \left(\nabla y^0_{\varepsilon,k}, \nabla\varphi\right)_{\mathbb{R}^N} dx = \left\langle v^0_{\varepsilon,k}, \varphi\right\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} + \int_\Omega u^0_{\varepsilon,k}\varphi\, dx$$

holds for every $\varphi \in H_0^1(\Omega)$, and hence $y \in H_0^1(\Omega)$ is the solution to the boundary value problem (23.27) and (23.28) for $v = v^0_{\varepsilon,k}$ and $u = u^0_{\varepsilon,k}$. Since the solution of (23.27) and (23.28) is unique, the whole sequence $\{y_{\varepsilon,k,m}\}_{m\in\mathbb{N}}$ converges weakly to $y^0_{\varepsilon,k}$ in $H_0^1(\Omega)$. Thus, $\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k}\right) \in \Xi_{\varepsilon,k}$.

The fact that $\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k}\right)$ is an optimal solution to the problem (23.26)–(23.29) immediately follows from such observations: in accordance to the strong

convergence $y_{\varepsilon,k,m} \to y^0_{\varepsilon,k}$ in $L^2(\Omega)$, we have

$$T_\varepsilon(f(y_{\varepsilon,k,m})) \to T_\varepsilon(f(y^0_{\varepsilon,k})) \text{ a.e. in } \Omega \text{ and } \sup_{m\in\mathbb{N}} \|T_\varepsilon(f(y_{\varepsilon,k,m}))\|_{L^q(\Omega)} \le \varepsilon^{-\frac{1}{q}} |\Omega|^{\frac{1}{q}}.$$

Since $q > 2$, it follows that the sequence $T_\varepsilon(f(y_{\varepsilon,k,m})) \to T_\varepsilon(f(y^0_{\varepsilon,k}))$ strongly in $L^{p'}(\Omega)$. Combining this fact with the weak convergence $v_{\varepsilon,k,m} \rightharpoonup v^0_{\varepsilon,k}$ in $L^{p'}(\Omega)$ and the lower semi-continuity of the norm $\|\cdot\|_{L^{p'}(\Omega)}$ with respect to the weak convergence in $L^{p'}(\Omega)$, we finally obtain

$$\inf_{(u,v,y)\in\Xi_{\varepsilon,k}} I_{\varepsilon,k}(u,v,y) = \lim_{m\to\infty} I_{\varepsilon,k}(u_{\varepsilon,k,m}, v_{\varepsilon,k,m}, y_{\varepsilon,k,m}) \ge I_{\varepsilon,k}\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k}\right).$$

Thus, $\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k}\right)$ is an optimal solution to the problem (23.26)–(23.29).

23.4 Asymptotic Analysis of Approximating OCP (23.26)–(23.29)

Our main intention in this section is to show that optimal solutions to the original OCP (23.2)–(23.5) can be attained (in some sense) by optimal solutions to the approximating problems (23.26)–(23.29). With that in mind, we make use of the concept of variational convergence of constrained minimization problems (see [10]) and study the asymptotic behaviour of a family of OCPs (23.26)–(23.29) as $\varepsilon \to 0$ and $k \to \infty$. We begin with some auxiliary results concerning the weak compactness in $H^1_0(\Omega)$ of $\|\cdot\|_{\varepsilon,k}$-bounded sequences.

Lemma 23.2 *Let $\{u_{\varepsilon,k}\}_{\substack{\varepsilon>0\\k\in\mathbb{N}}} \subset L^q(\Omega)$ and $\{v_{\varepsilon,k}\}_{\substack{\varepsilon>0\\k\in\mathbb{N}}} \subset L^{p'}(\Omega)$ be arbitrary bounded sequences of admissible controls with associated states $\left\{y_{\varepsilon,k}\right\}_{\substack{\varepsilon>0\\k\in\mathbb{N}}} \subset H^1_0(\Omega)$, i.e. $y_{\varepsilon,k} = y_{\varepsilon,k}(u_{\varepsilon,k}, v_{\varepsilon,k})$ is a weak solution of (23.27) and (23.28). Then the sequence $\left\{y_{\varepsilon,k}\right\}_{\substack{\varepsilon>0\\k\in\mathbb{N}}}$ is bounded in $H^1_0(\Omega)$. Moreover, each cluster point y of the sequence $\left\{y_{\varepsilon,k}\right\}_{\substack{\varepsilon>0\\k\in\mathbb{N}}}$ with respect to the weak convergence in $H^1_0(\Omega)$, satisfies: $y \in W^{1,p}_0(\Omega)$.*

For the proof of this Lemma we refer to [6, Lemma 5.1].

Lemma 23.3 *Let $\{\varepsilon_i\}_{i\in\mathbb{N}}$, $\{k_i\}_{i\in\mathbb{N}}$, $\{u_i\}_{i\in\mathbb{N}} \subset L^q(\Omega)$, and $\{v_i\}_{i\in\mathbb{N}} \subset L^{p'}(\Omega)$ be sequences such that*

$$\varepsilon_i \to 0, \quad k_i \to \infty, \quad u_i \rightharpoonup u \text{ in } L^q(\Omega), \quad v_i \rightharpoonup v \text{ in } L^{p'}(\Omega), \tag{23.49}$$

where $p' = p/(p-1)$ *and* $2 \le p < 2N/(N-1)$. *Let* $y = y(u,v)$ *and* $y_i = y_{\varepsilon_i,k_i}(u_i, v_i)$ *be the solutions of*

$$-\operatorname{div}\left(|\nabla y|^{p-2}\nabla y\right) = v + u \quad \text{in} \quad \Omega, \tag{23.50}$$

$$y = 0 \quad \text{on} \ \partial\Omega \tag{23.51}$$

and

$$-\operatorname{div}\left(\left(\varepsilon_i + \mathscr{F}_{k_i}\left(|\nabla y|^2\right)\right)^{\frac{p-2}{2}}\nabla y\right) = v_i + u_i \quad \text{in} \quad \Omega, \tag{23.52}$$

$$y = 0 \quad \text{on} \ \partial\Omega, \tag{23.53}$$

respectively. Then

$$y_i \to y \ \text{ in } \ H_0^1(\Omega) \ \text{ as } \ i \to \infty, \tag{23.54}$$

$$\chi_{\Omega\setminus\Omega_k(y_i)}\nabla y_i \to \nabla y \ \text{strongly in } \ L^p(\Omega)^N, \tag{23.55}$$

$$\lim_{i\to\infty}\int_\Omega \left(\varepsilon_i + \mathscr{F}_{k_i}(|\nabla y_i|^2)\right)^{\frac{p-2}{2}}|\nabla y_i|^2\,dx = \int_\Omega |\nabla y|^p\,dx, \tag{23.56}$$

where $\Omega_{k_i}(y_i)$ *is defined as*

$$\Omega_{k_i}(y_i) := \left\{x \in \Omega \ : \ |\nabla y_i(x)| > \sqrt{k_i^2+1}\right\}. \tag{23.57}$$

Proof For the reader's convenience, we divide the proof into five steps.

Step 1 $y_i \rightharpoonup y$ *in* $H_0^1(\Omega)$. Taking into account the a priori estimate (23.43), we have

$$\|y_i\|^p_{\varepsilon_i,k_i} \le C_p\left[\|v_i\|_{L^{p'}(\Omega)} + |\Omega|^{\frac{q-p'}{qp'}}\|u_i\|_{L^q(\Omega)}\right]\left[|\Omega|^{\frac{p-2}{2p}}\|y_i\|_{\varepsilon_i,k_i} + \|y_i\|^{\frac{p}{2}}_{\varepsilon_i,k_i}\right]. \tag{23.58}$$

It follows from (23.58) that

$$\|y_i\|_{\varepsilon_i,k_i} \le \max\left\{C_i^{\frac{2}{p}}, C_i^{\frac{1}{p-1}}\right\}, \quad \forall\, i \in \mathbb{N}, \tag{23.59}$$

where $C_i := C_p\left(\|v_i\|_{L^{p'}(\Omega)} + |\Omega|^{\frac{q-p'}{qp'}}\|u_i\|_{L^q(\Omega)}\right)\left(|\Omega|^{\frac{p-2}{2p}} + 1\right)$.

Then from Lemma 23.2 we deduce the existence of a subsequence, denoted in the same way $\{y_i\}_{i\in\mathbb{N}} \subset H_0^1(\Omega)$ and an element $y \in W_0^{1,p}(\Omega)$ such that $y_i \rightharpoonup y$ in $H_0^1(\Omega)$. Let us prove that y is the solution of (23.50) and (23.51). With that in mind we fix an arbitrary test function $\varphi \in C_0^\infty(\Omega)$ and pass to the limit in the Minty inequality

$$\int_\Omega (\varepsilon_i + \mathscr{F}_{k_i}(|\nabla\varphi|^2))^{\frac{p-2}{2}} (\nabla\varphi, \nabla\varphi - \nabla y_i)_{\mathbb{R}^N}\, dx \geq \langle v_i, \varphi - y_i\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} + \int_\Omega u_i(\varphi - y_i)\, dx, \tag{23.60}$$

as $i \to \infty$. Taking into account that

$$(\varepsilon_i + \mathscr{F}_{k_i}(|\nabla\varphi|^2))^{\frac{p-2}{2}} \nabla\varphi \to |\nabla\varphi|^{p-2}\nabla\varphi \ \text{ strongly in } L^{p'}(\Omega)^N, \ \text{ with } \ p' = p/(p-1),$$

in view of the convergences $\nabla y_i \rightharpoonup \nabla y$ in $L^2(\Omega)^N$, $u_i \rightharpoonup u$ in $L^q(\Omega)$, $v_i \to v$ in $H^{-1}(\Omega)$ (by compactness of the embedding $L^{p'}(\Omega) \hookrightarrow H^{-1}(\Omega)$), and $y_i \to y$ in $L^2(\Omega)$, we obtain

$$\lim_{i\to\infty} \int_\Omega (\varepsilon_i + \mathscr{F}_{k_i}(|\nabla\varphi|^2))^{\frac{p-2}{2}} (\nabla\varphi, \nabla\varphi)_{\mathbb{R}^N}\, dx = \int_\Omega |\nabla\varphi|^{p-2} (\nabla\varphi, \nabla\varphi)_{\mathbb{R}^N}\, dx,$$

$$\lim_{i\to\infty} \int_\Omega (\varepsilon_i + \mathscr{F}_{k_i}(|\nabla\varphi|^2))^{\frac{p-2}{2}} (\nabla\varphi, \nabla y_i)_{\mathbb{R}^N}\, dx = \int_\Omega |\nabla\varphi|^{p-2} (\nabla\varphi, \nabla y)_{\mathbb{R}^N}\, dx,$$

$$\lim_{i\to\infty} \int_\Omega u_i(\varphi - y_i)\, dx = \int_\Omega u(\varphi - y)\, dx,$$

$$\lim_{i\to\infty} \langle v_i, \varphi - y_i\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} = \langle v, \varphi - y\rangle_{W^{-1,p'}(\Omega);W_0^{1,p}(\Omega)}.$$

Thus, passing to the limit in relation (23.60) as $i \to \infty$, we arrive at the inequality

$$\int_\Omega |\nabla\varphi|^{p-2} (\nabla\varphi, \nabla\varphi - \nabla y)_{\mathbb{R}^N}\, dx \geq \langle v, \varphi - y_i\rangle_{W^{-1,p'}(\Omega);W_0^{1,p}(\Omega)} + \int_\Omega u(\varphi - y)\, dx$$

for every $\varphi \in C_0^\infty(\Omega)$. Finally, from the density of $C_0^\infty(\Omega)$ in $W_0^{1,p}(\Omega)$, we infer that this inequality holds for every $\varphi \in W_0^{1,p}(\Omega)$, and hence $y \in W_0^{1,p}(\Omega)$ is the solution to the boundary value problem (23.50) and (23.51) in the sense of distributions. Since the solution of (23.50) and (23.51) is unique, the whole sequence $\{y_i\}_{i\in\mathbb{N}}$ converges weakly to $y = y(u, v)$ in $H_0^1(\Omega)$.

Step 2 $\chi_{\Omega\setminus\Omega_k(y_i)}\nabla y_i \rightharpoonup \nabla y$ *in* $L^p(\Omega)^N$. Following the definition of the sets $\Omega_{k_i}(y_i)$, we obtain

$$\int_\Omega |\chi_{\Omega\setminus\Omega_{k_i}(y_i)}\nabla y_i|^p\,dx = \int_{\Omega\setminus\Omega_{k_i}(y_i)} |\nabla y_i|^p\,dx$$

$$\le \int_{\Omega\setminus\Omega_{k_i}(y_i)} \Big(\varepsilon_i + \mathscr{F}_{k_i}(|\nabla y_i|^2)\Big)^{\frac{p-2}{2}} |\nabla y_i|^2\,dx, \le \|y_i\|^p_{\varepsilon_i,k_i} \overset{\text{by (23.43)}}{\le} C < +\infty, \quad \forall\, i \in \mathbb{N}.$$

Hence, taking a new subsequence if necessary, we infer the existence of a vector-valued function $g \in L^p(\Omega)^N$ such that $\chi_{\Omega\setminus\Omega_{k_i}(y_i)}\nabla y_i \rightharpoonup g$ in $L^p(\Omega)^N$ as $i\to\infty$, i.e.

$$\lim_{i\to\infty} \int_{\Omega\setminus\Omega_{k_i}(y_i)} (\nabla y_i, \nabla\varphi)\,dx = \int_\Omega (g, \nabla\varphi)\,dx, \quad \forall\varphi \in C_0^\infty(\Omega). \tag{23.61}$$

On the other hand, in view of the weak convergence $\nabla y_i \rightharpoonup \nabla y$ in $L^2(\Omega)^N$,

$$\int_\Omega (\nabla y, \nabla\varphi)\,dx = \lim_{i\to\infty}\int_\Omega (\nabla y_i, \nabla\varphi)\,dx$$

$$= \lim_{i\to\infty}\int_{\Omega\setminus\Omega_{k_i}(y_i)} (\nabla y_i, \nabla\varphi)\,dx + \lim_{i\to\infty}\int_{\Omega_{k_i}(y_i)} (\nabla y_i, \nabla\varphi)\,dx. \tag{23.62}$$

Since

$$\left|\int_{\Omega_{k_i}(y_i)} (\nabla y_i, \nabla\varphi)\,dx\right| \le \|\varphi\|_{C^1(\overline{\Omega})}\sqrt{|\Omega_{k_i}(y_i)|}\left(\int_{\Omega_{k_i}(y_i)} |\nabla y_i|^2\,dx\right)^{1/2}$$

$$\le \frac{\|\varphi\|_{C^1(\overline{\Omega})}}{\big(\varepsilon_i + k_i^2 + 1\big)^{\frac{p-2}{4}}}\sqrt{|\Omega_{k_i}(y_i)|}\,\|y_i\|^{\frac{p}{2}}_{\varepsilon_i,k_i} \overset{\text{by (23.31),(23.43)}}{\le} \|\varphi\|_{C^1(\overline{\Omega})}\frac{C}{k_i^{p-1}} \to 0 \text{ as } i\to\infty,$$

it follows from (23.61) and (23.62) that

$$\int_\Omega (g, \nabla\varphi)\,dx = \int_\Omega (\nabla y, \nabla\varphi)\,dx, \quad \forall\varphi \in C_0^\infty(\Omega).$$

Hence, $g = \nabla y$ almost everywhere in Ω and $\chi_{\Omega\setminus\Omega_k(y_i)}\nabla y_i \rightharpoonup \nabla y$ in $L^p(\Omega)^N$ holds.

Step 3 $\chi_{\Omega\setminus\Omega_k(y_i)}\nabla y_i \to \nabla y$ *in* $L^p(\Omega)^N$. For each $i \in \mathbb{N}$, we have the energy equalities

$$\int_\Omega (\varepsilon_i + \mathscr{F}_{k_i}(|\nabla y_i|^2))^{\frac{p-2}{2}}|\nabla y_i|^2\,dx = \langle v_i, y_i\rangle_{H^{-1}(\Omega);H_0^1(\Omega)} + \int_\Omega u_i y_i\,dx,$$

$$\int_\Omega |\nabla y|^p\,dx = \langle v, y\rangle_{W^{-1,p'}(\Omega);W_0^{1,p}(\Omega)} + \int_\Omega uy\,dx. \tag{23.63}$$

From (23.63) and the fact that $y_i \rightharpoonup y$ in $H_0^1(\Omega)$ and $v_i \to v$ in $H^{-1}(\Omega)$, we deduce

$$\lim_{i\to\infty}\int_\Omega (\varepsilon_i+\mathscr{F}_{k_i}(|\nabla y_i|^2))^{\frac{p-2}{2}}|\nabla y_i|^2dx=\lim_{i\to\infty}\left[\langle v_i, y_i\rangle_{H^{-1}(\Omega);H_0^1(\Omega)}+\int_\Omega u_i y_i dx\right]$$
$$=\langle v, y\rangle_{W^{-1,p'}(\Omega);W_0^{1,p}(\Omega)}+\int_\Omega uy\,dx \overset{\text{by (23.63)}_2}{=} \int_\Omega |\nabla y|^p\,dx. \tag{23.64}$$

Moreover, we have

$$\begin{aligned}\int_\Omega |\nabla y|^p\,dx &= \lim_{i\to\infty}\int_\Omega \left(\varepsilon_i+\mathscr{F}_{k_i}(|\nabla y_i|^2)\right)^{\frac{p-2}{2}}|\nabla y_i|^2\,dx\\ &\geq \limsup_{i\to\infty}\int_{\Omega\setminus\Omega_{k_i}(y_i)}\left(\varepsilon_i+\mathscr{F}_{k_i}(|\nabla y_i|^2)\right)^{\frac{p-2}{2}}|\nabla y_i|^2\,dx\\ &\overset{\text{by (23.57)}}{\geq} \limsup_{i\to\infty}\int_{\Omega\setminus\Omega_{k_i}(y_i)}\left(\varepsilon_i+|\nabla y_i|^2\right)^{\frac{p-2}{2}}|\nabla y_i|^2\,dx\\ &\geq \limsup_{i\to\infty}\int_\Omega \chi_{\Omega\setminus\Omega_{k_i}(y_i)}|\nabla y_i|^p\,dx \geq \liminf_{i\to\infty}\int_\Omega \chi_{\Omega\setminus\Omega_{k_i}(y_i)}|\nabla y_i|^p\,dx.\end{aligned} \tag{23.65}$$

Since $\chi_{\Omega\setminus\Omega_{k_i}(y_i)}\nabla y_i \rightharpoonup \nabla y$ in $L^p(\Omega)^N$, it follows from (23.65) that

$$\begin{aligned}\int_\Omega |\nabla y|^p\,dx &\geq \limsup_{i\to\infty}\int_\Omega \chi_{\Omega\setminus\Omega_{k_i}(y_i)}|\nabla y_i|^p\,dx \geq \liminf_{i\to\infty}\int_\Omega \chi_{\Omega\setminus\Omega_{k_i}(y_i)}|\nabla y_i|^p\,dx\\ &= \liminf_{i\to\infty}\|\chi_{\Omega\setminus\Omega_{k_i}(y_i)}\nabla y_i\|^p_{L^p(\Omega)^N} \geq \|\nabla y\|^p_{L^p(\Omega)^N}=\int_\Omega |\nabla y|^p\,dx.\end{aligned}$$

It remains to note that the weak convergence $\chi_{\Omega\setminus\Omega_{k_i}(y_i)}\nabla y_i \rightharpoonup \nabla y$ in $L^p(\Omega)^N$ and the convergence of their norms $\|\chi_{\Omega\setminus\Omega_{k_i}(y_i)}\nabla y_i\|_{L^p(\Omega)^N} \to \|\nabla y\|_{L^p(\Omega)^N}$ imply the strong convergence $\chi_{\Omega\setminus\Omega_{k_i}(y_i)}\nabla y_i \to \nabla y$ in $L^p(\Omega)^N$.

Step 4 $y_i \to y$ *in* $H_0^1(\Omega)$. From (23.55) and (23.65) we obtain

$$\lim_{i\to\infty}\int_{\Omega_{k_i}(y_i)}\left(\varepsilon_i+\mathscr{F}_{k_i}(|\nabla y_i|^2)\right)^{\frac{p-2}{2}}|\nabla y_i|^2\,dx=0. \tag{23.66}$$

We apply (23.66) to deduce

$$\lim_{i\to\infty}\int_{\Omega_k(y_i)}|\nabla y_i|^2\,dx \leq \lim_{i\to\infty}\int_{\Omega_k(y_i)}(\varepsilon_i+\mathscr{F}_k(|\nabla y_i|^2))^{\frac{p-2}{2}}|\nabla y_i|^2\,dx=0.$$

Now, combining this estimate and (23.55), we conclude that

$$\nabla y_i = \chi_{\Omega_k(y_i)} \nabla y_i + \chi_{\Omega \setminus \Omega_k(y_i)} \nabla y_i \to \nabla y \ \text{ strongly in } \ L^2(\Omega)^N.$$

The following noteworthy property is crucial for our further analysis.

Theorem 23.5 *Let* $\left\{(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k})\right\}_{\substack{\varepsilon>0\\ k\in\mathbb{N}}}$ *be an arbitrary sequence of optimal solutions to the approximating problems* (23.26)–(23.29)*. Assume that* $2 \le p < 2N/(N-1)$ *and the parameter* ε *varies within a strictly decreasing sequence* $\{\varepsilon_k\}_{k\in\mathbb{N}}$ *of positive real numbers such that*

$$\lim_{k\to\infty} \left(k\, \varepsilon_k^{p'}\right) = 0. \tag{23.67}$$

Then the sequence $\left\{(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k})\right\}_{\substack{\varepsilon>0\\ k\in\mathbb{N}}}$ *is bounded in* $L^q(\Omega) \times L^{p'}(\Omega) \times H^1_0(\Omega)$ *and any its cluster triple* (u^0, v^0, y^0) *with respect to the weak topology of* $L^q(\Omega) \times L^{p'}(\Omega) \times H^1_0(\Omega)$ *is such that* $v^0 = f(y^0)$ *and* (u^0, y^0) *is a feasible solution of the OCP* (23.2)–(23.5)*.*

Proof Let us take an arbitrary function $\widetilde{y} \in C^\infty_0(\Omega)$ and put $\widetilde{u} := -\Delta_p \widetilde{y} - f(\widetilde{y})$. Then $\widetilde{u} \in L^q(\Omega)$, $\widetilde{y} \in H^1_0(\Omega)$, and $f(\widetilde{y}) \in L^2(\Omega)$. Hence, $\widetilde{y} \in Y \subset H^1_0(\Omega)$ is a weak solution to the boundary value problem (23.3) and (23.4) for given $\widetilde{u}$, and, therefore, $(\widetilde{u}, \widetilde{y}) \in \Xi$. Having set

$$\widetilde{v}_{\varepsilon,k} := -\Delta_{\varepsilon,k,p}(\widetilde{y}) + \Delta_p\, \widetilde{y} + f(\widetilde{y}), \tag{23.68}$$

it is easy to check that $\widetilde{v}_{\varepsilon,k} \in L^2(\Omega)$ and $(\widetilde{u}, \widetilde{v}_{\varepsilon,k}, \widetilde{y}) \in \Xi_{\varepsilon,k}$.

Let us show that, for sufficiently small $\varepsilon > 0$ and $k \in \mathbb{N}$ large enough, there exists a constant $C > 0$ such that

$$\|\widetilde{v}_{\varepsilon,k} - T_\varepsilon\,(f(\widetilde{y}))\,\|^{p'}_{L^{p'}(\Omega)} \le C\varepsilon^{p'}. \tag{23.69}$$

Indeed, using the fact that $\widetilde{y} \in C^\infty_0(\Omega)$, we see that

$$f(\widetilde{y}) = T_\varepsilon\,(f(\widetilde{y})) \quad \text{for sufficiently small } \varepsilon > 0 \text{ and}$$
$$\mathscr{F}_k(|\nabla\widetilde{y}|^2) = |\nabla\widetilde{y}|^2 \quad \text{for } k \in \mathbb{N} \text{ large enough.}$$

In what follows, we make use of the notation $\Delta_\infty(\widetilde{y}) = \left(\nabla\widetilde{y}, D^2(\widetilde{y})\nabla\widetilde{y}\right)$. Then, for indicated ε and k, by smoothness of the function $\widetilde{y}$, we have

$$\begin{aligned}\widetilde{v}_{\varepsilon,k} - T_\varepsilon\,(f(\widetilde{y})) &= \operatorname{div}\left(|\nabla\widetilde{y}|^{p-2}\nabla\widetilde{y} - \left(\varepsilon + |\nabla\widetilde{y}|^2\right)^{\frac{p-2}{2}}\nabla\widetilde{y}\right)\\ &= |\nabla\widetilde{y}|^{p-2}\Delta\widetilde{y} + (p-2)|\nabla\widetilde{y}|^{p-4}\Delta_\infty(\widetilde{y})\end{aligned}$$

$$
\begin{aligned}
&- \left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{p-2}{2}} \Delta \widetilde{y} - (p-2)\left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{p-4}{2}} \Delta_\infty(\widetilde{y}) \\
&= - \left[\left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{p-2}{2}} - |\nabla \widetilde{y}|^{p-2}\right] \Delta \widetilde{y} \\
&\quad - (p-2)\left[\left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{p-4}{2}} - |\nabla \widetilde{y}|^{p-4}\right] \Delta_\infty(\widetilde{y}). \qquad (23.70)
\end{aligned}
$$

Setting $\Phi(\varepsilon) := \left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{p-2}{2}}$, we see that $\Phi \in C([0,1])$ and $\Phi \in C^1(0,1)$ because $p \geq 2$. Hence, by the mean value theorem, we deduce that

$$
\left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{p-2}{2}} - |\nabla \widetilde{y}|^{p-2} = \Phi(\varepsilon) - \Phi(0) = \varepsilon \Phi'(\varepsilon_0) = \frac{p-2}{2}\left(\varepsilon_0 + |\nabla \widetilde{y}|^2\right)^{\frac{p-4}{2}} \varepsilon, \qquad (23.71)
$$

where $\varepsilon_0 \in (0,1)$. The similar inference can be done for the function $\Psi(\varepsilon) := \left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{p-4}{2}}$ provided $p \geq 4$. Indeed, in this case we have

$$
\left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{p-4}{2}} - |\nabla \widetilde{y}|^{p-4} = \Psi(\varepsilon) - \Psi(0) = \varepsilon \Psi'(\varepsilon_1) = \frac{p-4}{2}\left(\varepsilon_1 + |\nabla \widetilde{y}|^2\right)^{\frac{p-6}{2}} \varepsilon, \qquad (23.72)
$$

for some $\varepsilon_1 \in (0,1)$. Utilizing (23.71) and (23.72), and using the fact that

$$
|\nabla \widetilde{y}|,\ \Delta_\infty(\widetilde{y}),\ \Delta \widetilde{y} \in L^\infty(\Omega),
$$

we can deduce from (23.70) existence of a positive constant C_1, depending only on ε_0, ε_1, $\widetilde{y}$, and p, such that the following estimate

$$
\|\widetilde{v}_{\varepsilon,k} - T_\varepsilon\left(f(\widetilde{y})\right)\|^r_{L^r(\Omega)} = \int_\Omega \left(\widetilde{v}_{\varepsilon,k} - T_\varepsilon\left(f(\widetilde{y})\right)\right)^r dx \leq C_1 \varepsilon^r \qquad (23.73)
$$

holds true for all $r \in [1,\infty)$, ε small enough, and k large enough, provided $p \geq 4$.

In the case if $p \in [2,4)$, we have

$$
\begin{aligned}
\left|\left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{p-4}{2}} - |\nabla \widetilde{y}|^{p-4}\right| \Delta_\infty(\widetilde{y}) &= \left|\left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{4-p}{2}} - |\nabla \widetilde{y}|^{4-p}\right| \\
&\times \frac{|\nabla \widetilde{y}|^{p-2}}{\left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{4-p}{2}}} \left(\frac{\nabla \widetilde{y}}{|\nabla \widetilde{y}|}, D^2(\widetilde{y}) \frac{\nabla \widetilde{y}}{|\nabla \widetilde{y}|}\right)
\end{aligned}
$$

$$\leq \left| \left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{\frac{4-p}{2}} - |\nabla \widetilde{y}|^{4-p} \right| \left(\varepsilon + |\nabla \widetilde{y}|^2\right)^{p-2-\frac{4-p}{2}} \|D^2(\widetilde{y})\|$$

$$\overset{\text{by (23.72)}}{\leq} \frac{4-p}{2} \left(\varepsilon_2 + |\nabla \widetilde{y}|^2\right)^{\frac{2-p}{2}} \varepsilon \left(1 + |\nabla \widetilde{y}|^2\right)^{\frac{3}{2}(p-2)} \|D^2(\widetilde{y})\|$$

Combining this estimate with (23.71) and using the same arguments as we did it before, we also can deduce the existence of a constant C_2, depending only on ε_0, ε_2, $\widetilde{y}$, and $p \in [2, 4)$, such that

$$\|\widetilde{v}_{\varepsilon,k} - T_\varepsilon(f(\widetilde{y}))\|^r_{L^r(\Omega)} = \int_\Omega \left(\widetilde{v}_{\varepsilon,k} - T_\varepsilon(f(\widetilde{y}))\right)^r dx \leq C_2 \varepsilon^r, \quad \forall r \geq 1. \tag{23.74}$$

Hence, the estimate (23.69) is valid. Taking this fact into account, we see that

$$I_{\varepsilon,k}\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k}\right) = \inf_{(u,v,y)\in \Xi_{\varepsilon,k}} I_{\varepsilon,k}(u, v, y) \leq I_{\varepsilon,k}\left(\widetilde{u}, \widetilde{v}_{\varepsilon,k}, \widetilde{y}\right)$$
$$\leq \frac{1}{2}\int_\Omega |\widetilde{y} - y_d|^2 dx + \frac{k}{p'} C\varepsilon^{p'} + \frac{1}{q}\int_\Omega |\widetilde{u}|^q dx + \frac{\alpha}{p'}\int_\Omega |\widetilde{v}_{\varepsilon,k}|^{p'} dx, \tag{23.75}$$

where

$$\int_\Omega |\widetilde{v}_{\varepsilon,k}|^{p'} dx \overset{\text{by (23.68)}}{\leq} 2^{p'}\left[\int_\Omega \left|-\Delta_{\varepsilon,k,p}(\widetilde{y}) + \Delta_p \widetilde{y}\right|^{p'} dx + \int_\Omega |f(\widetilde{y})|^{p'} dx\right]$$
$$\leq C\varepsilon^{p'} + 2^{p'}\|f(\widetilde{y})\|^{p'}_{L^\infty(\Omega)}|\Omega|. \tag{23.76}$$

Utilizing the estimates (23.75) and (23.76), and assuming that the sequence $\{\varepsilon\} = \{\varepsilon_k\}_{k\in\mathbb{N}}$ satisfies the condition (23.67), we obtain

$$\sup_{\substack{\varepsilon>0\\ k\in\mathbb{N}}} I_{\varepsilon,k}\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k}\right) = \sup_{\substack{\varepsilon>0\\ k\in\mathbb{N}}}\Big[\frac{1}{2}\int_\Omega |y^0_{\varepsilon,k} - y_d|^2 dx + \frac{k}{p'}\int_\Omega |v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))|^{p'} dx$$
$$+ \frac{1}{q}\int_\Omega |u^0_{\varepsilon,k}|^q dx + \frac{\alpha}{p'}\int_\Omega |v^0_{\varepsilon,k}|^{p'} dx\Big]$$
$$\leq \sup_{\substack{\varepsilon>0\\ k\in\mathbb{N}}}\Big[\frac{1}{2}\int_\Omega |y^0_{\varepsilon,k} - y_d|^2 dx + \frac{k}{p'}\varepsilon^{p'} + \frac{1}{q}\int_\Omega |u^0_{\varepsilon,k}|^q dx + \frac{\alpha}{p'}\int_\Omega |v^0_{\varepsilon,k}|^{p'} dx\Big]$$
$$\leq \frac{1}{2}\int_\Omega |\widetilde{y} - y_d|^2 dx + \frac{1}{q}\int_\Omega |\widetilde{u}|^q dx + \frac{\alpha}{p'} 2^{p'}\|f(\widetilde{y})\|^{p'}_{L^\infty(\Omega)}|\Omega| < +\infty$$

and, as a consequence, we can deduce the existence of a constant $C^* > 0$ independent on ε and k such that

$$\begin{aligned} &\sup_{\substack{\varepsilon>0\\ k\in\mathbb{N}}} \|v^0_{\varepsilon,k}\|_{L^{p'}(\Omega)} < C^*, \ \sup_{\substack{\varepsilon>0\\ k\in\mathbb{N}}} \|u^0_{\varepsilon,k}\|^q_{L^q(\Omega)} < C^*, \\ &\text{and } \sup_{\substack{\varepsilon>0\\ k\in\mathbb{N}}} \|v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))\|^{p'}_{L^{p'}(\Omega)} < C^* k^{-1}. \end{aligned} \tag{23.77}$$

Then the estimate (23.59) implies that

$$\sup_{\substack{\varepsilon>0\\ k\in\mathbb{N}}} \|y^0_{\varepsilon,k}\|_{\varepsilon_i,k_i} \le \max\left\{C_*^{\frac{2}{p}}, C_*^{\frac{1}{p-1}}\right\} \tag{23.78}$$

$$\text{with } C_* = C_p\left(C^* + |\Omega|^{\frac{q-p'}{qp'}} C^*\right)\left(|\Omega|^{\frac{p-2}{2p}} + 1\right),$$

i.e., in view of (23.47), we can suppose that the sequence $\left\{y^0_{\varepsilon,k}\right\}_{\substack{\varepsilon>0\\ k\in\mathbb{N}}}$ is bounded in $H^1_0(\Omega)$ and, therefore, the sequence of solutions to the approximating problems $\left\{\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k}\right)\right\}_{\substack{\varepsilon>0\\ k\in\mathbb{N}}}$ is compact with respect to the weak convergence in $L^q(\Omega) \times L^{p'}(\Omega) \times H^1_0(\Omega)$.

Let (u^0, v^0, y^0) be its any cluster triplet, i.e. up to a subsequence, we have

$$y^0_{\varepsilon,k} \rightharpoonup y^0 \text{ in } H^1_0(\Omega), \ u^0_{\varepsilon,k} \rightharpoonup u^0 \text{ in } L^q(\Omega), \ v^0_{\varepsilon,k} \rightharpoonup v^0 \text{ in } L^{p'}(\Omega). \tag{23.79}$$

Then Lemma 23.3 implies that $y^0 = y(u^0, v^0)$ is a weak solution to the problem (23.50) and (23.51). Moreover, for any $\varphi \in C^\infty_0(\Omega)$ we have

$$\begin{aligned} \left|\int_\Omega v^0 \varphi\, dx - \int_\Omega f(y^0)\varphi\, dx\right| &\le \left|\int_\Omega v^0_{\varepsilon,k}\varphi\, dx - \int_\Omega v^0 \varphi\, dx\right| \\ &+ \left|\int_\Omega v^0_{\varepsilon,k}\varphi\, dx - \int_\Omega f(y^0)\varphi\, dx\right| \\ &\le \left|\int_\Omega \left(v^0_{\varepsilon,k} - v^0\right)\varphi\, dx\right| + \int_\Omega \left|v^0_{\varepsilon,k} - T_\varepsilon\left(f(y^0_{\varepsilon,k})\right)\right| |\varphi|\, dx \\ &+ \int_\Omega \left|T_\varepsilon\left(f(y^0_{\varepsilon,k})\right) - f(y^0_{\varepsilon,k})\right| |\varphi|\, dx + \int_\Omega \left|f(y^0_{\varepsilon,k}) - f(y^0)\right| |\varphi|\, dx \\ &= J_0 + J_1 + J_2 + J_3, \end{aligned} \tag{23.80}$$

where

$$J_0 = \left| \int_\Omega \left(v^0_{\varepsilon,k} - v^0 \right) \varphi \, dx \right| \overset{\text{by (23.79)}_3}{\to} 0 \text{ as } \varepsilon \to 0 \text{ and } k \to \infty, \tag{23.81}$$

$$J_1 \le \| v^0_{\varepsilon,k} - T_\varepsilon \left(f(y^0_{\varepsilon,k}) \right) \|_{L^{p'}(\Omega)} \|\varphi\|_{L^p(\Omega)} \overset{\text{by (23.77)}}{\to} 0 \text{ as } \varepsilon \to 0 \text{ and } k \to \infty, \tag{23.82}$$

$$J_2 \le \| T_\varepsilon \left(f(y^0_{\varepsilon,k}) \right) - f(y^0_{\varepsilon,k}) \|_{L^1(\Omega)} \|\varphi\|_{L^\infty(\Omega)} \overset{\text{by definition of } T_\varepsilon}{\to} 0 \text{as } \varepsilon \to 0 \text{ and } k \to \infty, \tag{23.83}$$

$$J_3 \le \| f(y^0_{\varepsilon,k}) - f(y^0) \|_{L^1(\Omega)} \|\varphi\|_{L^\infty(\Omega)}.$$

Let us show that

$$f(y^0_{\varepsilon,k}) \to f(y^0) \text{ strongly in } L^1(\Omega) \text{ as } \varepsilon \to 0 \text{ and } k \to \infty. \tag{23.84}$$

With that in mind, we note that by compactness of the injection $H^1_0(\Omega) \hookrightarrow L^2(\Omega)$, we can deduce the existence of a subsequence of $\left\{ y^0_{\varepsilon,k} \right\}_{\substack{\varepsilon>0 \\ k\in\mathbb{N}}}$, denoted in the same way, such that $f(y^0_{\varepsilon,k}) \to f(y^0)$ almost everywhere in Ω. So, in order to conclude (23.84), it remains to establish the equi-integrability on Ω of the sequence $\left\{ f\left(y^0_{\varepsilon,k} \right) \right\}_{\substack{\varepsilon>0 \\ k\in\mathbb{N}}}$. For this purpose, we make use of the following relation, coming from the energy identity (23.42),

$$\int_\Omega y^0_{\varepsilon,k} f\left(y^0_{\varepsilon,k} \right) dx = \| y^0_{\varepsilon,k} \|^p_{\varepsilon,k} - \left\langle v^0_{\varepsilon,k} - f\left(y^0_{\varepsilon,k} \right), y^0_{\varepsilon,k} \right\rangle_{H^{-1}(\Omega);H^1_0(\Omega)} - \int_\Omega u^0_{\varepsilon,k} y^0_{\varepsilon,k} \, dx. \tag{23.85}$$

Since

$$\| y^0_{\varepsilon,k} \|^p_{\varepsilon,k} \overset{\text{by (23.78)}}{\le} A_1^p := \max \left\{ C_*^{\frac{2}{p}}, C_*^{\frac{1}{p-1}} \right\}, \tag{23.86}$$

$$\int_\Omega u^0_{\varepsilon,k} y^0_{\varepsilon,k} \, dx \overset{\text{by (23.37)}}{\le} C_p |\Omega|^{\frac{q-p'}{qp'}} \| u^0_{\varepsilon,k} \|_{L^q(\Omega)} \left[|\Omega|^{\frac{p-2}{2p}} \| y^0_{\varepsilon,k} \|_{\varepsilon,k} + \| y^0_{\varepsilon,k} \|^{\frac{p}{2}}_{\varepsilon,k} \right]$$
$$\overset{\text{by (23.86), (23.77)}}{\le} A_2 := C_p |\Omega|^{\frac{q-p'}{qp'}} C^* \left[|\Omega|^{\frac{p-2}{2p}} A_1 + A_1^{\frac{p}{2}} \right] \tag{23.87}$$

and

$$\left| \left\langle v^0_{\varepsilon,k} - f\left(y^0_{\varepsilon,k} \right), y^0_{\varepsilon,k} \right\rangle_{H^{-1}(\Omega);H^1_0(\Omega)} \right| \le \left| \left\langle v^0_{\varepsilon,k} - T_\varepsilon \left(f\left(y^0_{\varepsilon,k} \right) \right), y^0_{\varepsilon,k} \right\rangle_{H^{-1}(\Omega);H^1_0(\Omega)} \right|$$
$$+ \left| \left\langle f\left(y^0_{\varepsilon,k} \right) - T_\varepsilon \left(f\left(y^0_{\varepsilon,k} \right) \right), y^0_{\varepsilon,k} \right\rangle_{H^{-1}(\Omega);H^1_0(\Omega)} \right|$$

$$\overset{\text{by (23.36)}}{\leq} C_p \|v^0_{\varepsilon,k} - T_\varepsilon\left(f\left(y^0_{\varepsilon,k}\right)\right)\|_{L^{p'}(\Omega)} \left[|\Omega|^{\frac{p-2}{2p}} \|y^0_{\varepsilon,k}\|_{\varepsilon,k} + \|y^0_{\varepsilon,k}\|^{\frac{p}{2}}_{\varepsilon,k}\right]$$

$$+ C_p \|f\left(y^0_{\varepsilon,k}\right) - T_\varepsilon\left(f\left(y^0_{\varepsilon,k}\right)\right)\|_{L^{p'}(\Omega)} \left[|\Omega|^{\frac{p-2}{2p}} \|y^0_{\varepsilon,k}\|_{\varepsilon,k} + \|y^0_{\varepsilon,k}\|^{\frac{p}{2}}_{\varepsilon,k}\right]$$

$$\overset{\text{by (23.77), (23.86), (23.30)}}{\leq} C_p \left[|\Omega|^{\frac{p-2}{2p}} A_1 + A_1^{\frac{p}{2}}\right] \left(\varepsilon C^* + \varepsilon \widetilde{C}\right). \tag{23.88}$$

Utilizing the estimates (23.88), (23.87), (23.86), it follows from (23.85) that there exists a constant $M > 0$ independent of ε and k such that

$$\sup_{\substack{\varepsilon>0 \\ k\in\mathbb{N}}} \left| \int_\Omega y^0_{\varepsilon,k} f\left(y^0_{\varepsilon,k}\right) dx \right| \leq M. \tag{23.89}$$

We recall that a sequence $\{f_k\}_{k\in\mathbb{N}}$ is called equi-integrable on Ω if for any $\delta > 0$, there is a $\tau = \tau(\delta)$ such that $\int_S |f_k|\, dx < \delta$ for every measurable subset $S \subset \Omega$ of Lebesgue measure $|S| < \tau$. So, in order to show that the sequence $\left\{f\left(y^0_{\varepsilon,k}\right)\right\}_{\substack{\varepsilon>0 \\ k\in\mathbb{N}}}$ is equi-integrable on Ω, we take $m > 0$ such that

$$m > 2M\delta^{-1}. \tag{23.90}$$

We also set $\tau = \delta/(2f(m))$. Then for every measurable set $S \subset \Omega$ with $|S| < \tau$, we have

$$\int_S f(y^0_{\varepsilon,k})\, dx = \int_{\left\{x\in S\,:\, y^0_{\varepsilon,k}(x)>m\right\}} f(y^0_{\varepsilon,k})\, dx + \int_{\left\{x\in S\,:\, y^0_{\varepsilon,k}(x)\leq m\right\}} f(y^0_{\varepsilon,k})\, dx$$

$$\leq \frac{1}{m} \int_{\left\{x\in S\,:\, y^0_{\varepsilon,k}(x)>m\right\}} y^0_{\varepsilon,k} f(y^0_{\varepsilon,k})\, dx + \int_{\left\{x\in S\,:\, y^0_{\varepsilon,k}(x)\leq m\right\}} f(m)\, dx$$

$$\overset{\text{by (23.89)}}{\leq} \frac{M}{m} + f(m)|S| \overset{\text{by (23.90)}}{\leq} \frac{\delta}{2} + \frac{\delta}{2}.$$

As a result, the assertion (23.84) is a direct consequence of Lebesgue's Convergence Theorem. Thus,

$$J_3 \leq \|f(y^0_{\varepsilon,k}) - f(y^0)\|_{L^1(\Omega)} \|\varphi\|_{L^\infty(\Omega)} \overset{\text{by (23.84)}}{\to} 0 \text{ as } \varepsilon \to 0 \text{ and } k \to \infty.$$

Combining this fact with properties (23.81)–(23.83), we deduce from (23.80) that $v^0 = f(y^0)$ almost everywhere on Ω. Hence, by (23.79), we have

$$f(y^0) \in L^{p'}(\Omega) \quad \text{and} \quad v^0_{\varepsilon,k} \rightharpoonup f(y^0) \text{ in } L^{p'}(\Omega).$$

Thus, (u^0, y^0) a feasible solution of the OCP (23.2)–(23.5). The proof is complete.

We are now in a position to show that optimal solutions to the approximating OCP (23.26)–(23.29) lead in the limit to optimal pairs of the original OCP (23.2)–(23.5).

Theorem 23.6 *Let $2 \le p < 2N/(N-1)$ and let $\left\{(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k})\right\}_{\substack{\varepsilon>0\\ k\in\mathbb{N}}}$ be an arbitrary sequence of optimal solutions to the approximating problems (23.26)–(23.29), where the parameter ε varies within a strictly decreasing sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ of positive real numbers satisfying condition (23.67). Then, this sequence is bounded in $L^q(\Omega) \times L^{p'}(\Omega) \times H^1_0(\Omega)$ and any its cluster point (u^0, v^0, y^0) with respect to the weak topology is such that $v^0 = f(y^0)$ and (u^0, y^0) is solution of the OCP (23.2)–(23.5). Moreover, if for one subsequence we have $y^0_{\varepsilon,k} \rightharpoonup y^0$ in $H^1_0(\Omega)$, $u^0_{\varepsilon,k} \rightharpoonup u^0$ in $L^q(\Omega)$, and $v^0_{\varepsilon,k} \rightharpoonup v^0$ in $L^{p'}(\Omega)$, then the following properties hold*

$$y^0_{\varepsilon,k} \to y^0 \ \text{ in } \ H^1_0(\Omega), \quad u^0_{\varepsilon,k} \to u^0 \ \text{ in } \ L^q(\Omega), \tag{23.91}$$

$$v^0_{\varepsilon,k} \to f(y^0) \ \text{ in } \ L^{p'}(\Omega), \quad \frac{k}{p'}\int_\Omega |v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))|^{p'}\,dx \to 0, \tag{23.92}$$

$$\chi_{\Omega\setminus\Omega_k(y^0_{\varepsilon,k})}\nabla y^0_{\varepsilon,k} \to \nabla y^0 \ \textit{strongly in} \ L^p(\Omega)^N, \tag{23.93}$$

$$\lim_{\substack{\varepsilon\to 0\\ k\to\infty}} \int_\Omega \left(\varepsilon + \mathscr{F}_k(|\nabla y^0_{\varepsilon,k}|^2)\right)^{\frac{p-2}{2}} |\nabla y^0_{\varepsilon,k}|^2\,dx = \int_\Omega |\nabla y^0|^p\,dx, \tag{23.94}$$

$$\lim_{\substack{\varepsilon\to 0\\ k\to\infty}} I_{\varepsilon,k}(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k}) = \lim_{\substack{\varepsilon\to 0\\ k\to\infty}} J(u^0_{\varepsilon,k}, y^0_{\varepsilon,k}) = J(u^0, y^0). \tag{23.95}$$

Proof The boundedness of the sequence $\left\{(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k})\right\}_{\substack{\varepsilon>0\\ k\in\mathbb{N}}}$ has been proved in Theorem 23.5. Let (u^0, v^0, y^0) be its any cluster point with respect to the weak topology of $L^q(\Omega) \times L^{p'}(\Omega) \times H^1_0(\Omega)$. Let us take a subsequence, denoted in the same way, satisfying the property (23.79). Then

$$y^0_{\varepsilon,k} \rightharpoonup y^0 \text{ in } H^1_0(\Omega),\ u^0_{\varepsilon,k} \rightharpoonup u^0 \text{ in } L^q(\Omega),\ v^0_{\varepsilon,k} \rightharpoonup v^0 \text{ in } L^{p'}(\Omega) \text{ as } \varepsilon \to 0 \text{ and } k \to \infty, \tag{23.96}$$

and from Lemma 23.3 we get that $y^0_{\varepsilon,k} \to y^0$ strongly in $H^1_0(\Omega)$. As for the convergences (23.93) and (23.94), they follow from (23.55) and (23.56), respectively. Moreover, Theorem 23.5 implies that $v^0 = f(y^0)$, $f(y^0) \in L^{p'}(\Omega)$, and y^0 is a weak solution of (23.3) and (23.4) corresponding to $u = u^0$.

Let us prove that (u^0, y^0) is an optimal pair to the problem (23.2)–(23.5). Given an arbitrary feasible $(u, y) \in \Xi$, we define $u_{\varepsilon,k} = u$, $v_{\varepsilon,k} = T_\varepsilon(f(y))$, and $y_{\varepsilon,k}$ as the solution of the boundary value problem (23.27) and (23.28). Since $v_{\varepsilon,k} \in$

$L^{p'}(\Omega)$, it follows that $(u_{\varepsilon,k}, v_{\varepsilon,k}, y_{\varepsilon,k}) \in \Xi_{\varepsilon,k}$. By definition of the cut-off operator T_ε, we have

$$v_{\varepsilon,k} \to f(y) \quad \text{strongly in } L^{p'}(\Omega) \text{ as } \varepsilon \to 0 \text{ and } k \to \infty. \tag{23.97}$$

Then Lemma 23.3 implies the existence of an element $y^* \in W_0^{1,p}(\Omega)$ such that $y_{\varepsilon,k} \to y^*$ in $H_0^1(\Omega)$ and y^* satisfies the equality (in the sense of distributions)

$$-\operatorname{div}\left(|\nabla y^*|^{p-2}\nabla y^*\right) = f(y) + u \quad \text{in } \Omega.$$

On the other hand, the condition $(u, y) \in \Xi$ leads to the relation

$$-\operatorname{div}\left(|\nabla y|^{p-2}\nabla y\right) = f(y) + u \quad \text{in } \Omega.$$

Hence,

$$-\operatorname{div}\left(|\nabla y^*|^{p-2}\nabla y^*\right) + \operatorname{div}\left(|\nabla y|^{p-2}\nabla y\right) = 0$$

and, therefore,

$$\left\langle -\operatorname{div}\left(|\nabla y^*|^{p-2}\nabla y^*\right) + \operatorname{div}\left(|\nabla y|^{p-2}\nabla y^0\right), y^* - y\right\rangle_{W^{-1,p'}(\Omega);\, W_0^{1,p}(\Omega)} = 0.$$

Since the p-Laplace operator is strictly monotone, it follows that $y^* = y$ as element of $W_0^{1,p}(\Omega)$. Thus, from (23.97) and Lemma 23.3 we get that

$$\Xi_{\varepsilon,k} \ni (u_{\varepsilon,k}, v_{\varepsilon,k}, y_{\varepsilon,k}) \longrightarrow (u, f(y), y) \quad \text{strongly in } L^q(\Omega) \times L^{p'}(\Omega) \times H_0^1(\Omega). \tag{23.98}$$

Further, we make use of the following observation. Since $f \in C_{loc}(\mathbb{R})$, it follows from (23.30) and (23.97) that

$$v_{\varepsilon,k} - T_\varepsilon(f(y_{\varepsilon,k})) \longrightarrow 0 \quad \text{strongly in } L^{p'}(\Omega) \text{ as } \varepsilon \to 0 \text{ and } k \to \infty.$$

Hence, there exists a mapping $\varepsilon \mapsto k(\varepsilon)$, increasing to $+\infty$ and arguably depending on y, such that (see Section 1.2.2 in [1])

$$\lim_{\substack{\varepsilon\to 0\\ k\to\infty}} \left[\int_\Omega |v_{\varepsilon,k} - T_\varepsilon(f(y_{\varepsilon,k}))|^{p'}\, dx\right] = \lim_{\varepsilon\to 0}\left[\int_\Omega |v_{\varepsilon,k(\varepsilon)} - T_\varepsilon(f(y_{\varepsilon,k(\varepsilon)}))|^{p'}\, dx\right] = 0. \tag{23.99}$$

Utilizing (23.98) and (23.99), we have

$$\begin{aligned}\lim_{\substack{\varepsilon\to 0\\ k\to\infty}}\Big[\frac{1}{2}\int_\Omega |y_{\varepsilon,k}-y_d|^2\,dx+\frac{1}{q}\int_\Omega |u_{\varepsilon,k}|^q\,dx+\frac{\alpha}{p'}\int_\Omega |v_{\varepsilon,k}|^{p'}\,dx\Big]\\ =\lim_{\varepsilon\to 0}\Big[\frac{1}{2}\int_\Omega |y_{\varepsilon,k(\varepsilon)}-y_d|^2\,dx+\frac{1}{q}\int_\Omega |u_{\varepsilon,k(\varepsilon)}|^q\,dx+\frac{\alpha}{p'}\int_\Omega |v_{\varepsilon,k(\varepsilon)}|^{p'}\,dx\Big]\\ =\frac{1}{2}\int_\Omega |y-y_d|^2\,dx++\frac{1}{q}\int_\Omega |u|^q\,dx+\frac{\alpha}{p'}\int_\Omega |f(y)|^{p'}\,dx. \qquad (23.100)\end{aligned}$$

Since

$$\begin{aligned}0\le\limsup_{\substack{\varepsilon\to 0\\ k\to\infty}}\Big[\frac{k}{p'}\int_\Omega |v_{\varepsilon,k(\varepsilon)}-T_\varepsilon(f(y_{\varepsilon,k(\varepsilon)}))|^{p'}\,dx\Big]\\ \le\limsup_{k\to\infty}\Big[\frac{k}{p'}\limsup_{\varepsilon\to 0}\int_\Omega |v_{\varepsilon,k(\varepsilon)}-T_\varepsilon(f(y_{\varepsilon,k(\varepsilon)}))|^{p'}\,dx\Big]=0, \qquad (23.101)\end{aligned}$$

it follows from (23.100) and (23.101) that

$$\begin{aligned}\lim_{\substack{\varepsilon\to 0\\ k\to\infty}} I_{\varepsilon,k}(u_{\varepsilon,k(\varepsilon)},v_{\varepsilon,k(\varepsilon)},y_{\varepsilon,k(\varepsilon)})&=\limsup_{k\to\infty}\Big[\limsup_{\varepsilon\to 0} I_{\varepsilon,k}(u_{\varepsilon,k(\varepsilon)},v_{\varepsilon,k(\varepsilon)},y_{\varepsilon,k(\varepsilon)})\Big]\\ &=J(u,y). \qquad (23.102)\end{aligned}$$

Now, using (23.96), (23.77), the above identity, and the fact that $(u^0_{\varepsilon,k},v^0_{\varepsilon,k},y^0_{\varepsilon,k})$ is a solution of (23.26)–(23.29), we get

$$\begin{aligned}J(u^0,y^0)&\le\liminf_{\substack{\varepsilon\to 0\\ k\to\infty}} I_{\varepsilon,k}(u^0_{\varepsilon,k},v^0_{\varepsilon,k},y^0_{\varepsilon,k})\le\liminf_{\substack{\varepsilon\to 0\\ k\to\infty}} I_{\varepsilon,k}(u^0_{\varepsilon,k(\varepsilon)},v^0_{\varepsilon,k(\varepsilon)},y^0_{\varepsilon,k(\varepsilon)})\\ &\le\liminf_{\substack{\varepsilon\to 0\\ k\to\infty}} I_{\varepsilon,k}(u_{\varepsilon,k(\varepsilon)},v_{\varepsilon,k(\varepsilon)},y_{\varepsilon,k(\varepsilon)})\\ &\le\limsup_{\substack{\varepsilon\to 0\\ k\to\infty}} I_{\varepsilon,k}(u_{\varepsilon,k(\varepsilon)},v_{\varepsilon,k(\varepsilon)},y_{\varepsilon,k(\varepsilon)})=J(u,y).\end{aligned}$$

Since (u,y) is an arbitrary pair in Ξ, this implies that (u^0,y^0) is a solution of the original optimal control problem (23.2)–(23.5). Moreover, taking $(u,y)=(u^0,y^0)$ in the above inequalities, the relations (23.95) is proved. Finally, $(23.91)_2$ and (23.92) are the direct consequences of (23.95) and the convergence properties (23.96) established before.

23.5 Optimality Conditions for Approximating OCP (23.26)–(23.29)

The aim of this section is to derive the optimality system for approximating optimal control problem (23.26)–(23.29). With that in mind, we assume that the mapping $F : \mathbb{R} \to [0, +\infty)$ satisfies condition $F \in C^2_{loc}(\mathbb{R})$ and begin with investigation of differentiability of the mapping $(u, v) \mapsto y_{\varepsilon,k}(u, v)$. It is well known that in the case $\varepsilon = 0$ and $k = \infty$ this mapping is not necessarily Gâteaux differentiable even if $f(y) \equiv 0$. Indeed, let us consider the following boundary value problem

$$-\mathrm{div}\left(|\nabla y|^{p-2}\nabla y\right) = u \quad \text{in } \Omega,$$
$$y = 0 \quad \text{on } \partial\Omega,$$

where $p > 2$ and Ω is the unite open ball in R^N centered at the origin, $\Omega = B(0, 1)$.

It is easy to check that the states associated to $u_0(x) = 0$, $u_1(x) = -N$, and $u_t(x) := u_0(x) + tu_1(x)(x) = -tN$, for each $t > 0$, are

$$y_0(x) = 0, \quad y_1(x) = \frac{p-1}{p}\left(|x|^{\frac{p}{p-1}} - 1\right), \quad \text{and } y_t(x) = t^{\frac{1}{p-1}} y_1(x),$$

respectively. Then the mapping $u \mapsto y(u)$ is not Gâteaux differentiable at $u = u_0$, because the sequence

$$\left\{\frac{y_t - y_0}{t} = t^{\alpha} y_1(x)\right\}_{t>0}, \quad \text{with } \alpha = (2-p)/(p-1)$$

does not converge as $t \to 0$, if $p > 2$.

So, in order to derive an optimality system to the original optimal control problem (23.2)–(23.5) and provide its rigour mathematical substantiation, a direct application of the implicit function theorem or Ioffe–Tikhomirov theorem looks rather questionable. On the other hand, Theorem 23.6 reveals another way to characterize the optimal pairs to the problem (23.2)–(23.5). Namely, we can do it deriving an optimality system for the approximating problem (23.26)–(23.29) and studying then its asymptotic behaviour as $\varepsilon \to 0$ and $k \to \infty$.

We know that the boundary value problem (23.27)–(23.28) has a unique solution $y_{\varepsilon,k} \in H^1_0(\Omega)$ for every $u \in L^q(\Omega)$ and $v \in L^{p'}(\Omega)$. Let $G_{\varepsilon,k} : L^q(\Omega) \times L^{p'}(\Omega) \longrightarrow H^1_0(\Omega)$ be the mapping defined by $G_{\varepsilon,k}(u, v) = y_{\varepsilon,k}(u, v)$, where $y_{\varepsilon,k}(u, v)$ solution of (23.27)–(23.28) associated to u and v.

Theorem 23.7 *The mapping $G_{\varepsilon,k}$ is of the class C^1 and for any $u \in L^q(\Omega)$, $v \in L^{p'}(\Omega)$, $h_u \in L^q(\Omega)$, and $h_v \in L^{p'}(\Omega)$ the element*

$$z(h_u, h_v) = D_u G_{\varepsilon,k}(u, v)\,[h_u] + D_v G_{\varepsilon,k}(u, v)\,[h_v]$$

is the unique solution in $H^1_0(\Omega)$ *of the equation*

$$-\operatorname{div}\left(\rho_{\varepsilon,k}(y_{\varepsilon,k})\mathscr{A}_{\varepsilon,k}(y_{\varepsilon,k})\nabla z\right) = h_u + h_v, \tag{23.103}$$

where $y_{\varepsilon,k} = G_{\varepsilon,k}(u, v)$ *and*

$$\mathscr{A}_{\varepsilon,k}(y) = \left(I_N + (p-2)\mathscr{F}'_k\left(|\nabla y|^2\right)\left[\frac{\nabla y}{\sqrt{\varepsilon + \mathscr{F}_k\left(|\nabla y|^2\right)}} \otimes \frac{\nabla y}{\sqrt{\varepsilon + \mathscr{F}_k\left(|\nabla y|^2\right)}}\right]\right), \tag{23.104}$$

$$\rho_{\varepsilon,k}(y) = (\varepsilon + \mathscr{F}_k\left(|\nabla y|^2\right))^{\frac{p-2}{2}}. \tag{23.105}$$

Proof We apply the implicit function theorem. To this end we define the function $F : H^1_0(\Omega) \times L^q(\Omega) \times L^{p'}(\Omega) \longrightarrow H^{-1}(\Omega)$ by

$$F(y, u, v) = -\Delta_{\varepsilon,k,p}(y) - u - v.$$

It is immediate that F is of class C^1. Moreover the partial derivative $\frac{\partial F}{\partial y}(y, u, v) : H^1_0(\Omega) \longrightarrow H^{-1}(\Omega)$ is an isomorphism. Indeed, it is easy to see that

$$\frac{\partial F}{\partial y}(y, u, v)\,[z] = -\operatorname{div}\left(\rho_{\varepsilon,k}(y_{\varepsilon,k})\mathscr{A}_{\varepsilon,k}(y_{\varepsilon,k})\nabla z\right),$$

where the matrix $\mathscr{A}_{\varepsilon,k}(y)$ and the scalar function $\rho_{\varepsilon,k}(y)$ are given by (23.104) and (23.105) and possess the following properties:

$$\mathscr{A}_{\varepsilon,k}(y) \in L^\infty(\Omega; \mathbb{S}^N_{sym}); \tag{23.106}$$

$$\|\mathscr{A}_{\varepsilon,k}(y)\|_{L^\infty(\Omega;\mathbb{S}^N_{sym})} \le 1 + (p-2)\delta^*, \quad \forall\, \varepsilon > 0 \text{ and } \forall\, k \in \mathbb{N}; \tag{23.107}$$

$$\varepsilon^{\frac{p-2}{2}} \le \rho_{\varepsilon,k}(y) \le (\varepsilon + k^2 + 1)^{\frac{p-2}{2}} \text{ a.e. in } \Omega; \tag{23.108}$$

$$|\eta|^2 \le \left(\eta, \mathscr{A}_{\varepsilon,k}(y)\eta\right)_{\mathbb{R}^N} \le \left(1 + (p-2)\delta^*\right)|\eta|^2 \text{ a.e. in } \Omega,\ \forall\, \eta \in \mathbb{R}^N, \tag{23.109}$$

where $\mathbb{S}^N_{sym}$ is the set of all $N \times N$ symmetric matrices.

We note that properties (23.106) and (23.108) immediately follow from (23.104) and (23.105) and definition of the $C^1(\mathbb{R}_+)$-function $\mathscr{F}_k : \mathbb{R}_+ \to \mathbb{R}_+$. To prove the property (23.109), it is enough to take into account the following chain of

estimates

$$|\eta|^2 \le (\eta, I\eta)_{\mathbb{R}^N} \le (\eta, I\eta)_{\mathbb{R}^N} + (p-2)\mathscr{F}_k'\left(|\nabla y|^2\right)\left(\frac{\nabla y}{\sqrt{\varepsilon + \mathscr{F}_k\left(|\nabla y|^2\right)}}, \eta\right)^2_{\mathbb{R}^N}$$

$$= \left(\eta, \mathscr{A}_{\varepsilon,k}(y)\eta\right)_{\mathbb{R}^N} \le |\eta|^2 + (p-2)\mathscr{F}_k'\left(|\nabla y|^2\right)\left|\frac{\nabla y}{\sqrt{\varepsilon + \mathscr{F}_k\left(|\nabla y|^2\right)}}\right|^2 |\eta|^2$$

$$\le \left(1 + (p-2)\delta^*\right)|\eta|^2 \text{ a.e. in } \Omega,$$

because $\mathscr{F}_k'\left(|\nabla y|^2\right) = 0$ a.e. on the set $\Omega_k(y) := \left\{x \in \Omega \ : \ |\nabla y(x)| > \sqrt{k^2+1}\right\}$.

As a result, the isomorphism of the mapping

$$\frac{\partial F}{\partial y}(y, u, v) : H_0^1(\Omega) \longrightarrow H^{-1}(\Omega)$$

is a direct consequence of estimates (23.108) and (23.109) and the Lax-Milgram theorem. In addition, for every $u \in L^q(\Omega)$, $v \in L^{p'}(\Omega)$, and $y_{\varepsilon,k} = G_{\varepsilon,k}(u, v)$, we have that $F(y_{\varepsilon,k}, u, v) = 0$. Hence, by application of the implicit function theorem, we deduce that for any $(u^0, v^0) \in L^q(\Omega) \times L^{p'}(\Omega)$ there exists a neighborhood $\mathscr{U} \times \mathscr{V}$ of (u^0, v^0) in $L^q(\Omega) \times L^{p'}(\Omega)$ and a mapping $g : \mathscr{U} \times \mathscr{V} \longrightarrow H_0^1(\Omega)$ of class C^1 such that

$$F(g(u, v), u, v) = 0 \quad \forall\, (u, v) \in \mathscr{U} \times \mathscr{V}.$$

The mapping g obviously coincides with $G_{\varepsilon,k}$, which proves that $G_{\varepsilon,k}$ is of class C^1 and the expression of the derivative follows from the equation

$$\frac{\partial F}{\partial y}(y_{\varepsilon,k}, u, v)\Big[D_u G_{\varepsilon,k}(u, v)\,[h_u] + D_v G_{\varepsilon,k}(u, v)\,[h_v]\Big] + \frac{\partial F}{\partial u}(y_{\varepsilon,k}, u, v)\,[h_u] + \frac{\partial F}{\partial v}(y_{\varepsilon,k}, u, v)\,[h_v] = 0.$$

Now we observe that the problem (23.26)–(23.29) can be written in the form

$$\underset{(u,v)\in L^q(\Omega)\times L^{p'}(\Omega)}{\text{Minimize}} \quad J_{\varepsilon,k}(u, v) = \frac{1}{2}\int_\Omega |y_{\varepsilon,k}(u, v) - y_d|^2\, dx + \frac{k}{p'}\int_\Omega |v - T_\varepsilon(f(y_{\varepsilon,k}(u, v)))|^{p'}\, dx + \frac{1}{q}\int_\Omega |u|^q\, dx + \frac{\alpha}{p'}\int_\Omega |v|^{p'}\, dx. \tag{23.110}$$

In the functional $J_{\varepsilon,k}$ we distinguish two terms $J_{\varepsilon,k}(u,v) = F_{\varepsilon,k}(u,v) + j(u,v)$ with

$$F_{\varepsilon,k}(u,v) = \frac{1}{2}\int_\Omega |y_{\varepsilon,k}(u,v) - y_d|^2\,dx + \frac{k}{p'}\int_\Omega |v - T_\varepsilon(f(y_{\varepsilon,k}(u,v)))|^{p'}\,dx$$

and

$$j(u,v) = \frac{1}{q}\int_\Omega |u|^q\,dx + \frac{\alpha}{p'}\int_\Omega |v|^{p'}\,dx.$$

Now we make use of the Stampacchia's Theorem (see, for instance, Theorem 1.19 in [15]) which says that if $\Phi : \mathbb{R} \to \mathbb{R}$ is a Lipschitz continuous function and $z \in W_0^{1,p}(\Omega)$, then $\nabla\Phi(z)$ belongs to $L^p(\Omega)^N$ and $\nabla\Phi(z) = \Phi'(z)\nabla z$ almost everywhere in Ω. Setting $\Phi(t) = T_\varepsilon(f(t))$ and using the fact that f is of class C^1, from the differentiability of $G_{\varepsilon,k}$ and the chain rule it immediately follows that

$$\begin{aligned}\left(F_{\varepsilon,k}\right)'_u(u,v)[h_u] + \left(F_{\varepsilon,k}\right)'_v(u,v)[h_v] &= \int_\Omega (y_{\varepsilon,k}(u,v) - y_d)z(h_u,h_v)\,dx\\ &+ k\int_\Omega |v - T_\varepsilon(f(y_{\varepsilon,k}(u,v)))|^{p'-2}\left(v - T_\varepsilon(f(y_{\varepsilon,k}(u,v)))\right)\\ &\times\left(h_v - f'(y_{\varepsilon,k}(u,v))\chi_{\{|f(y_{\varepsilon,k}(u,v))|\le\varepsilon^{-1}\}}z(h_u,h_v)\right)dx\end{aligned}$$

$$= \int_\Omega \Psi_1 z(h_u,h_v)\,dx + \int_\Omega \Psi_2 h_v\,dx$$

with

$$z(h_u,h_v) = D_u G_{\varepsilon,k}(u,v)\,[h_u] + D_v G_{\varepsilon,k}(u,v)\,[h_v]\,,$$

for any $(u,v) \in L^q(\Omega)\times L^{p'}(\Omega)$ and $(h_u,h_v) \in L^q(\Omega)\times L^{p'}(\Omega)$. Here,

$$\begin{aligned}\Psi_1 &= y_{\varepsilon,k}(u,v) - y_d - k|v - T_\varepsilon(f(y_{\varepsilon,k}(u,v)))|^{p'-2}\\ &\times\left(v - T_\varepsilon(f(y_{\varepsilon,k}(u,v)))\right)f'(y_{\varepsilon,k}(u,v))\chi_{\{|f(y_{\varepsilon,k}(u,v))|\le\varepsilon^{-1}\}}\\ \Psi_2 &= k|v - T_\varepsilon(f(y_{\varepsilon,k}(u,v)))|^{p'-2}\left(v - T_\varepsilon(f(y_{\varepsilon,k}(u,v)))\right).\end{aligned}$$

As for the functional $j(u,v)$, we have

$$\begin{aligned}j'_u(u,v)[h_u] + j'_v(u,v)[h_v] &= \int_\Omega |u|^{q-2}uh_u\,dx\\ &+ \alpha\int_\Omega |v|^{p'-2}vh_v\,dx,\ \forall\,(h_u,h_v) \in L^q(\Omega)\times L^{p'}(\Omega).\end{aligned}$$

Now we introduce the adjoint state as follows

$$\begin{cases} -\mathrm{div}\Big(\rho_{\varepsilon,k}(y_{\varepsilon,k}(u,v))\mathscr{A}_{\varepsilon,k}(y_{\varepsilon,k}(u,v))\nabla\mu_{\varepsilon,k}\Big) = \Psi_1 \ \text{ in } \ \Omega, \\ \mu_{\varepsilon,k} = 0 \ \text{ on } \ \partial\Omega. \end{cases} \tag{23.111}$$

Using (23.103), we obtain

$$\begin{aligned} \left(F_{\varepsilon,k}\right)'_u(u,v)[h_u] + \left(F_{\varepsilon,k}\right)'_v(u,v)[h_v] = \int_\Omega [h_u + h_v]\,\mu_{\varepsilon,k}\,dx \\ + k\int_\Omega \Big[|v - T_\varepsilon(f(y_{\varepsilon,k}(u,v)))|^{p'-2}\left(v - T_\varepsilon(f(y_{\varepsilon,k}(u,v)))\right)\Big]h_v\,dx. \end{aligned} \tag{23.112}$$

We are now in a position to establish the main result of this section.

Theorem 23.8 *For given $\varepsilon > 0$, $k \in \mathbb{N}$, $2 \le p < 2N/(N-1)$, $q > p'$, and $y_d \in L^2(\Omega)$, let $\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}\right)$ be a local solution of (23.110). Assume that $F \in C^2_{loc}(\mathbb{R})$. Then there exist elements $y^0_{\varepsilon,k}, \mu_{\varepsilon,k} \in H^1_0(\Omega)$ such that the tuple $(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}, y^0_{\varepsilon,k}, \mu_{\varepsilon,k})$ satisfies the following Euler-Lagrange system to the problem (23.26)–(23.29)*

$$\begin{cases} -\mathrm{div}\left(\rho_{\varepsilon,k}(y^0_{\varepsilon,k})\nabla y^0_{\varepsilon,k}\right) = v^0_{\varepsilon,k} + u^0_{\varepsilon,k} \ \text{ in } \ \Omega, \\ y^0_{\varepsilon,k} = 0 \ \text{ on } \ \partial\Omega, \end{cases} \tag{23.113}$$

$$\begin{cases} -\mathrm{div}\Big(\rho_{\varepsilon,k}(y^0_{\varepsilon,k})\mathscr{A}_{\varepsilon,k}(y^0_{\varepsilon,k})\nabla\mu_{\varepsilon,k}\Big) = y^0_{\varepsilon,k} - y_d - k|v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))|^{p'-2} \\ \qquad \times \left(v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))\right) f'(y^0_{\varepsilon,k})\chi_{\left\{|f(y^0_{\varepsilon,k})| \le \varepsilon^{-1}\right\}} \ \text{ in } \ \Omega, \\ \mu_{\varepsilon,k} = 0 \ \text{ on } \ \partial\Omega, \end{cases} \tag{23.114}$$

$$\begin{cases} u^0_{\varepsilon,k} = -\left|\mu_{\varepsilon,k}\right|^{\frac{1}{q-1}}\mathrm{sign}(\mu_{\varepsilon,k}) \quad \text{a.e. in } \ \Omega, \\ \mu_{\varepsilon,k} = -\alpha|v^0_{\varepsilon,k}|^{p'-2}v^0_{\varepsilon,k} \\ \qquad - k\Big[|v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))|^{p'-2}\left(v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))\right)\Big], \quad \text{a.e. in } \ \Omega. \end{cases} \tag{23.115}$$

Remark 23.1 Here, we say that $\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}\right)$ is a local solution of (23.110) if there is a closed neighborhood $\mathscr{U}(u^0_{\varepsilon,k}) \times \mathscr{V}(v^0_{\varepsilon,k})$ of $\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}\right)$ in the norm topology

of $L^q(\Omega) \times L^{p'}(\Omega)$ satisfying

$$J_{\varepsilon,k}\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}\right) < J_{\varepsilon,k}(u, v) \quad \forall\, u \in \mathscr{U}(u^0_{\varepsilon,k}) \text{ and } \forall\, v \in \mathscr{V}(v^0_{\varepsilon,k})$$

such that $(u, v, y_{\varepsilon,k}(u, v))$ is a feasible triplet for (23.27) and (23.28) and $(u, v) \neq \left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}\right)$.

Proof Given $u \in L^q(\Omega)$ and $v \in L^{p'}(\Omega)$, since $\left(u^0_{\varepsilon,k}, v^0_{\varepsilon,k}\right)$ is a local optimal solution of (23.110), we have that

$$J_{\varepsilon,k}(u^0_{\varepsilon,k} + \rho(u - u^0_{\varepsilon,k}), v^0_{\varepsilon,k} + \rho(v - v^0_{\varepsilon,k})) \geq J_{\varepsilon,k}(u^0_{\varepsilon,k}, v^0_{\varepsilon,k})$$

for all $\rho > 0$ small enough.

Hence,

$$\begin{aligned}
0 \leq \frac{1}{\rho}\left(J_{\varepsilon,k}(u^0_{\varepsilon,k} + \rho(u - u^0_{\varepsilon,k}), v^0_{\varepsilon,k} + \rho(v - v^0_{\varepsilon,k})) - J_{\varepsilon,k}(u^0_{\varepsilon,k}, v^0_{\varepsilon,k})\right) \\
= \frac{F_{\varepsilon,k}(u^0_{\varepsilon,k} + \rho(u - u^0_{\varepsilon,k}), v^0_{\varepsilon,k} + \rho(v - v^0_{\varepsilon,k})) - F_{\varepsilon,k}(u^0_{\varepsilon,k}, v^0_{\varepsilon,k} + \rho(v - v^0_{\varepsilon,k}))}{\rho} \\
+ \frac{F_{\varepsilon,k}(u^0_{\varepsilon,k}, v^0_{\varepsilon,k} + \rho(v - v^0_{\varepsilon,k})) - F_{\varepsilon,k}(u^0_{\varepsilon,k}, v^0_{\varepsilon,k})}{\rho} \\
+ \frac{j((u^0_{\varepsilon,k} + \rho(u - u^0_{\varepsilon,k})), v^0_{\varepsilon,k} + \rho(v - v^0_{\varepsilon,k})) - j(u^0_{\varepsilon,k}, v^0_{\varepsilon,k} + \rho(v - v^0_{\varepsilon,k}))}{\rho} \\
+ \frac{j(u^0_{\varepsilon,k}, v^0_{\varepsilon,k} + \rho(v - v^0_{\varepsilon,k})) - j(u^0_{\varepsilon,k}, v^0_{\varepsilon,k})}{\rho}.
\end{aligned}$$

Now, taking $\rho \to 0$, we get

$$\begin{aligned}
0 \leq \left(F_{\varepsilon,k}\right)'_u (u^0_{\varepsilon,k}, v^0_{\varepsilon,k})\left[u - u^0_{\varepsilon,k}\right] + \left(F_{\varepsilon,k}\right)'_v (u^0_{\varepsilon,k}, v^0_{\varepsilon,k})\left[v - v^0_{\varepsilon,k}\right] \\
+ j'_u(u^0_{\varepsilon,k}, v^0_{\varepsilon,k})\left[u - u^0_{\varepsilon,k}\right] + j'_v(u^0_{\varepsilon,k}, v^0_{\varepsilon,k})\left[v - v^0_{\varepsilon,k}\right].
\end{aligned}$$

Finally, using the expression of $F'_{\varepsilon,k}$ given by (23.112) we obtain

$$\begin{aligned}
0 \leq \int_\Omega \left[u - u^0_{\varepsilon,k}\right] \mu_{\varepsilon,k}\, dx + \int_\Omega \left[v - v^0_{\varepsilon,k}\right] \mu_{\varepsilon,k}\, dx + \int_\Omega |u^0_{\varepsilon,k}|^{q-2} u^0_{\varepsilon,k}\left[u - u^0_{\varepsilon,k}\right] dx \\
+ \alpha \int_\Omega |v^0_{\varepsilon,k}|^{p'-2} v^0_{\varepsilon,k}\left[v - v^0_{\varepsilon,k}\right] dx \\
+ k \int_\Omega \left[|v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))|^{p'-2}\left(v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))\right)\right]\left[v - v^0_{\varepsilon,k}\right] dx.
\end{aligned}$$

Since $u \in L^q(\Omega)$ and $v \in L^{p'}(\Omega)$ are independent and arbitrary functions, we deduce from this relation the following equalities

$$\mu_{\varepsilon,k} = -|u^0_{\varepsilon,k}|^{q-2}u^0_{\varepsilon,k}, \quad \text{a. e. in } \Omega \tag{23.116}$$

$$\mu_{\varepsilon,k} = -\alpha|v^0_{\varepsilon,k}|^{p'-2}v^0_{\varepsilon,k} \tag{23.117}$$

$$- k\left[|v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))|^{p'-2}\left(v^0_{\varepsilon,k} - T_\varepsilon(f(y^0_{\varepsilon,k}))\right)\right] \quad \text{a. e. in } \Omega. \tag{23.118}$$

Thus, the optimality system (23.113)–(23.115) immediately follows from (23.111) and (23.116).

Acknowledgements Research funded by the DFG-cluster CE315: Engineering of Advanced Materials

References

1. Attouch, H.: Variational Convergence for Functions and Operators. Pitman Advanced Pub. Program, Boston (1984)
2. Boccardo, L., Murat, F.: Almost everywhere convergence of the gradients of solutions to elliptic and parabolic equations. Nonlinear Anal. Theory Methods Appl. **19**, 581–597 (1992)
3. Casas, E., Fernandez, L.A.: Optimal control of quasilinear elliptic equations with non differentiable coefficients at the origin. Rev. Mat. Univ. Complut. Madrid **4**(2–3), 227–250 (1991)
4. Casas, E., Fernandez, L.A.: Distributed controls of systems governed by a general class of quasilinear elliptic systems. J. Differ. Equ. **104**, 20–47 (1993)
5. Casas, E., Kavian, O., Puel, J.P.: Optimal control of an ill-posed elliptic semilinear equation with an exponential nonlinearity. ESAIM Control Optim. Calc. Var. **3**, 361–380 (1998)
6. Casas, E., Kogut, P.I., Leugering, G.: Approximation of optimal control problems in the coefficient for the p-Laplace equation. I. Convergence result. SIAM J. Control. Optim. **54**(3), 1406–1422 (2016)
7. Durante, T., Kupenko, O.P., Manzo, R.: On attainability of optimal controls in coefficients for system of Hammerstein type with anisotropic p-Laplacian. Ricerche Mat. **66**(2), 259–292 (2017)
8. Kogut, P.I., Kupenko, O.P.: On attainability of optimal solutions for linear elliptic equations with unbounded coefficients. Visnyk DNU. Series: Mathematical Modelling, Dnipropetrovsk: DNU **20**(4), 63–82 (2012)
9. Kogut, P.I., Kupenko, O.P.: On optimal control problem for an ill-posed strongly nonlinear elliptic equation with p-Laplace operator and L^1-type of nonlinearity. Discrete Contin. Dynam. Syst. Ser. B (2018, to appear)
10. Kogut, P.I., Leugering, G.: Optimal Control Problems for Partial Differential Equations on Reticulated Domains. Approximation and Asymptotic Analysis, Series: Systems and Control. Birkhäuser, Boston (2011)
11. Kogut, P.I., Putchenko, A.O.: On approximate solutions to one class of non-linear Dirichlet elliptic boundary value problems. Visnyk DNU. Series: Mathematical Modelling, Dnipropetrovsk: DNU **24**(8), 27–55 (2016)

12. Kogut, P.I., Manzo, R., Putchenko, A.O.: On approximate solutions to the Neumann elliptic boundary value problem with non-linearity of exponential type. Bound. Value Probl. **2016**(1), 1–32 (2016)
13. Kupenko, O.P., Manzo, R.: Approximation of an optimal control problem in the coefficients for variational inequality with anisotropic p-Laplacian. Nonlinear Differ. Equ. Appl. **35**(23) (2016). https://doi.org/10.1007/s00030-016-0387-9
14. Kupenko, O.P., Manzo, R.: On optimal controls in coefficients for ill-posed non-linear elliptic Dirichlet boundary value problems. Discrete Contin. Dyn. Syst. Ser. B **23**(4), 1363–1393 (2018). https://doi.org/10.3934/dcdsb.2018155
15. Orsina, L.: Elliptic Equations with Measure Data. Preprint, Sapienza University of Rome (2011)

Chapter 24
Approximate Optimal Regulator for Distributed Control Problem with Superposition Functional and Rapidly Oscillating Coefficients

Olena A. Kapustian

Abstract In this paper, we consider the optimal stabilization problem on infinite time interval for a parabolic equation with rapidly oscillating coefficients and non-decomposable quadratic cost functional with superposition type operator. In general, to find the exact formula of optimal regulator is not possible for such a problem, because the Fourier method cannot be directly applied. But the transition to the homogenized parameters greatly simplifies the structure of the problem. Assuming that the problem with the homogenized coefficients already admits optimal regulator, we ground approximate optimal control in the feedback form for the initial problem. We give an example of superposition operator for specific conditions in this paper.

24.1 Introduction

In this work, we focus on the finding effective methods of control for complicated systems with distributed parameters on infinite time interval, initiated in the works [3, 11]. Finding control in the feedback form (regulator) plays important role here. In [6–10, 12, 14] it was proposed and substantiated a procedure for constructing approximate optimal feedback control for a wide class of infinite-dimensional processes in micro-inhomogeneous medium both on finite and infinite time interval. We use some known facts on G-convergence theory from [2, 4, 16]. In this paper from this point of view we consider the optimal control problem on infinite time interval for a parabolic equation with rapidly oscillating coefficients and non-decomposable quadratic cost functional with superposition type operator. The case of finite time interval was considered in [5]. In general, to find the exact formula of optimal regulator is not possible for problem with rapidly oscillating coefficients, because we cannot directly apply the Fourier decomposition method.

O. A. Kapustian (✉)
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_24

But the transition to the problem with homogenized parameters greatly simplifies the structure of the problem. Assuming that the problem with the homogenized coefficients already admits optimal feedback form, we ground approximate optimal regulator for the initial problem. For a deeper understanding of the problem, we give an example of superposition operator for specific conditions in this paper.

24.2 Setting of the Problem

Let $\Omega \subset R^n$ be a bounded domain, $\varepsilon \in (0, 1)$ is a small parameter. In cylinder $Q = (0, +\infty) \times \Omega$ controlled process $\{y, u\}$ is described by the problem

$$\begin{cases} \frac{\partial y}{\partial t} = A^\varepsilon y + u(t, x), \\ y|_{\partial\Omega} = 0, \\ y(0, x) = y_0^\varepsilon, \end{cases} \tag{24.1}$$

$$u \in U \subseteq L^2(Q), \tag{24.2}$$

$$J_\varepsilon(y, u) = \int\limits_Q q_\varepsilon(t, x, y(t, x))y(t, x)dtdx + \int\limits_Q u^2(t, x)dtdx \to \inf, \tag{24.3}$$

where U is convex and closed subset of $L^2(Q)$, $0 \in U$,

$$A^\varepsilon = div(a^\varepsilon \nabla), \quad a^\varepsilon(x) = a\Big(\frac{x}{\varepsilon}\Big),$$

a is measurable, symmetric, periodic matrix, satisfying the conditions of uniform ellipticity and boundedness: $\exists \nu_1 > 0$, $\nu_2 > 0$ $\forall \eta \in R^n$ $\forall\, x \in R^n$

$$\nu_1 \sum_{i=1}^n \eta_i^2 \le \sum_{i,j=1}^n a_{i,j}(x)\eta_i\eta_j \le \nu_2 \sum_{i=1}^n \eta_i^2, \tag{24.4}$$

$q_\varepsilon : (0, +\infty) \times \Omega \times R \mapsto R$ is a Caratheodory function, i.e. it is measurable with respect to the first two variables and continuous with respect to the third variable, and there exist functions $C_1 \in L^2(Q)$, $C_2 \in L^1(Q)$, and constant $C > 0$, independent of $\varepsilon \in (0, 1)$ such that for all $\xi \in R$ and almost all (a. a.) $(t, x) \in Q$ the following inequalities hold

$$\begin{gathered} |q_\varepsilon(t, x, \xi)| \le C|\xi| + C_1(t, x), \\ q_\varepsilon(t, x, \xi)\xi \ge -C_2(t, x). \end{gathered} \tag{24.5}$$

Then superposition operator $q_\varepsilon(t, x, \cdot) : L^2(Q) \mapsto L^2(Q)$ is continuous [1]. Hence, by conditions (24.4), (24.5) and the properties of solutions of problem (24.1) (see Lemma 24.1) we obtain [11] that problem (24.1)–(24.3) has solution $\{\bar{y}^\varepsilon, \bar{u}^\varepsilon\}$ (optimal process) in class $W(0, +\infty) \times L^2(Q)$, where $W(0, T)$ with $T \leq \infty$ is a class of functions $y \in L^2\left(0, T; H_0^1(\Omega)\right)$, which have distributed derivatives with respect to t from $L^2\left(0, T; H^{-1}(\Omega)\right)$. In general case, we are not able to find an exact optimal feedback law for problem (24.1)–(24.3). However, in many cases [4] a transition to homogenized parameters simplifies the structure of the problem. We will assume that the problem with homogenized coefficients already admits optimal feedback control of the form $u[t, x, y(t, x)]$.

The main goal of this paper is to prove the fact that the regulator $u[t, x, y(t, x)]$ realizes an approximate feedback control in initial problem (24.1)–(24.3), i.e. for any $\eta > 0$

$$\left| J_\varepsilon(\bar{y}^\varepsilon, \bar{u}^\varepsilon) - J_\varepsilon(y_\varepsilon, u[t, x, y_\varepsilon]) \right| < \eta \tag{24.6}$$

for $\varepsilon > 0$ small enough, where y_ε is a solution of problem (24.1)–(24.3) with control $u[t, x, y_\varepsilon]$.

24.3 Main Results

We shall use $\|\cdot\|$ to denote the norm and $(\cdot, \cdot)$ to denote the inner product in $L^2(\Omega)$. Let us assume that there exists a Caratheodory function $q : (0, +\infty) \times \Omega \times R \mapsto R$ such that

$$\begin{gathered} \forall\, r > 0 \ \forall\, T > 0 \quad q_\varepsilon(t, x, \xi) \to q(t, x, \xi) \text{ weakly in } L_2(Q_T) \\ \text{uniformly for } |\xi| \leq r, \end{gathered} \tag{24.7}$$

where $Q_T = (0, T) \times \Omega$.

We refer to the following problem

$$\begin{cases} \frac{\partial y}{\partial t} = A^0 y + u(t, x), \\ y|_{\partial\Omega} = 0, \\ y|_{t=0} = y_0, \end{cases} \tag{24.8}$$

$$u \in U \subseteq L^2(Q), \tag{24.9}$$

$$J(y, u) = \int_Q q(t, x, y(t, x)) y(t, x) dt dx + \int_Q u^2(t, x) dt dx \to \inf, \tag{24.10}$$

as an homogenized one for problem (24.1)–(24.3). Here a constant matrix a^0 is homogenized for a^ε [9], $A^0 = div(a^0\nabla)$, $y_0 \in L^2(\Omega)$ such that

$$y_0^\varepsilon \to y_0 \text{ weakly in } L^2(\Omega) \text{ as } \varepsilon \to 0. \tag{24.11}$$

In further arguments we will use the following result about convergence of parabolic operators which is the consequence of G-convergence of A^ε to A^0 [16].

Lemma 24.1 ([2, 16]) *Let $y_0^\varepsilon \to y_0$ weakly in $L^2(\Omega)$, $u_\varepsilon \to u$ weakly in $L^2(Q_T)$. Then $y^\varepsilon \to y$ in $L^2(Q_T)$ and in $C\left([\delta, T]; L^2(\Omega)\right)$ $\forall \delta > 0$, where y^ε is the solution of problem (24.1) on $(0, T)$ with control u_ε, y is the solution of problem (24.8) on $(0, T)$ with control u.*

Let us assume the following conditions hold:

$$\text{problem (24.8)–(24.10) has a unique solution } \{\overline{y}, \overline{u}\}; \tag{24.12}$$

$$\text{there exists a measurable map } u : [0, +\infty) \times \Omega \times L^2(\Omega) \mapsto L^2(\Omega) \text{ such that}$$

$$\overline{u}(t, x) \equiv u[t, x, \overline{y}(t)]; \tag{24.13}$$

$$\begin{gathered} \exists\, D > 0 \text{ such that } \forall\, t \geq 0,\ y, z \in L^2(\Omega) \\ u[t, x, 0] \in L^2(Q), \\ \|u[t, x, y] - u[t, x, z]\| \leq D\|y - z\|. \end{gathered} \tag{24.14}$$

Moreover,

$$\exists\, \hat{T} > 0\ \forall\, t \geq \hat{T}\ \forall\, y \in L^2(\Omega) \int_\Omega u[t, x, y(x)]y(x)dx < \nu_1\lambda\|y\|^2, \tag{24.15}$$

where $\lambda > 0$ is taken from Poincare inequality

$$\forall\, y \in H_0^1(\Omega)\ \|\nabla y\|^2 \geq \lambda\|y\|^2.$$

Before we formulate the main result, we give a typical example of problem (24.1)–(24.3), for which conditions (24.5), (24.7), (24.12)–(24.15) hold.

Example Let $\{X_i\}$, $\{\lambda_i\}$ be solutions of spectrum problem

$$\begin{cases} A^0 X_i = -\lambda_i X_i, \\ X_i|_{\partial\Omega} = 0, \end{cases}$$

$$U = \left\{ v \in L^2(Q) \,\middle|\, \forall i \in \overline{1, p} \ \left| \int_\Omega v(t, x)X_i(x)dx \right| \leq \xi_i \text{ for a.a. } t > 0 \right\}, \tag{24.16}$$

where $p \geq 1$, and $\xi_1 > 0, \ldots, \xi_p > 0$ are fixed numbers,

$$q_\varepsilon(x, \xi) = g\left(\frac{x}{\varepsilon}\right)\xi,$$

where g is measurable, bounded, non-negative, periodic function with mean value $\langle g \rangle$ [4]. Then conditions (24.5), (24.7) hold for $q(x, \xi) = \langle g \rangle \xi$. Moreover, problem (24.8) and (24.10) becomes a classical linear quadratic problem that has the unique solution [11]. Thus, condition (24.12) holds.

Suppose that

$$\forall\, i \in \overline{1, p} \quad \left|(y_0, X_i)\right| > \frac{\xi_i}{R_i},$$

where $R_i = -\lambda_i + \sqrt{\lambda_i^2 + 1}$.

In this case, using Fourier decomposition method, it's easy to obtain optimal control in feedback form for problem (24.8)–(24.10) [7]:

$$u[t, x, y(t)] = \sum_{i=1}^{p} \left(\alpha_i(t)\,(y(t), X_i) + \beta_i(t)\right) X_i(x) - \sum_{i=p+1}^{\infty} R_i\,(y(t), X_i)\, X_i(x), \tag{24.17}$$

where

$$\alpha_i(t) = \begin{cases} 0, & t \in [0, t_i], \\ -R_i & t > t_i, \end{cases} \qquad \beta_i(t) = \begin{cases} -\xi_i sign\,(y(t), X_i)\,, & t \in [0, t_i], \\ 0, & t > t_i, \end{cases}$$

and $t_i > 0$ is a unique solution of the equation

$$t_i = \frac{1}{\lambda_i} \ln\left(\frac{R_i}{\sqrt{\lambda_i^2 + 1}}\left(1 + \frac{\lambda_i}{\xi_i}\,|(y(t), X_i)|\right) e^{\lambda_i t}\right), t \in [0, t_i],$$

where y is a solution of problem (24.8) with control (24.17).

Using (24.17), we can define optimal regulator $u : [0, +\infty) \times \Omega \times L^2(\Omega) \mapsto L^2(\Omega)$ in the following form

$$u[t, x, y] = \sum_{i=1}^{p} \left(\alpha_i(t)\,(y, X_i) + \beta_i(t)\right) X_i(x) - \sum_{i=p+1}^{\infty} R_i\,(y, X_i)\, X_i(x), \tag{24.18}$$

where

$$\alpha_i(t) = \begin{cases} 0, & t \in [0, t_i], \\ -R_i & t > t_i, \end{cases} \quad \beta_i(t) = \begin{cases} -\xi_i sign\,(y_0, X_i), & t \in [0, t_i], \\ 0, & t > t_i, \end{cases}$$

$$t_i = \frac{1}{\lambda_i} \ln \left(\frac{R_i}{\sqrt{\lambda_i^2 + 1}} \left(1 + \frac{\lambda_i}{\xi_i} |(y_0, X_i)| \right) e^{\lambda_i t} \right).$$

Then conditions (24.13) and (24.14) hold. Moreover, for $t \geq \hat{T} := \max\limits_{1 \leq i \leq p} \{t_i\}$ we deduce the inequality

$$(u[t, y], y) \leq 0,$$

which implies (24.15).

Other examples of optimal regulator for infinite-dimensional control problems one can find in [7].

Let us return to problem (24.1)–(24.3). Using regulator (24.13), we consider the problem

$$\begin{cases} \frac{\partial y}{\partial t} = A^\varepsilon y + u[t, x, y], \\ y\,|_{\partial\Omega} = 0, \\ y\,\big|_{t=0} = y_0^\varepsilon. \end{cases} \tag{24.19}$$

Under conditions (24.14) problem (24.19) has a unique solution $y_\varepsilon \in C([0, +\infty); L^2(\Omega))$ which belongs to the class $W(0, T)$ for every $T > 0$ [5, 15]. Moreover, due to (24.15) for some $\gamma > 0$

$$\forall\, t \geq \hat{T} \quad \frac{d}{dt} \|y_\varepsilon(t)\|^2 + \gamma \|y_\varepsilon(t)\|^2 \leq 0. \tag{24.20}$$

In particular, $y_\varepsilon \in L^2(Q)$ and $J_\varepsilon(y_\varepsilon, u[t, x, y_\varepsilon]) < \infty$.

The main result of this article is the following theorem.

Theorem 24.1 *Let conditions (24.4), (24.5), (24.7), (24.12)–(24.15) hold and, moreover, there exists a positive function l, $l \in L^\infty(Q_T)$ $\forall\, T > 0$ such that for all $\varepsilon \in (0, 1)$*

$$|q_\varepsilon(t, x, \xi_1) - q_\varepsilon(t, x, \xi_2)| \leq l(t, x)|\xi_1 - \xi_2|. \tag{24.21}$$

Then for an arbitrary $\eta > 0$ *there exists* $\bar{\varepsilon} \in (0, 1)$ *such that* $\forall \varepsilon \in (0, \bar{\varepsilon})$

$$\left| J_\varepsilon \left(\bar{y}^\varepsilon, \bar{u}^\varepsilon \right) - J_\varepsilon \left(y_\varepsilon, u \left[t, x, y_\varepsilon (t, x) \right] \right) \right| < \eta,$$

where $\{\bar{y}^\varepsilon, \bar{u}^\varepsilon\}$ *is an optimal process for problem (24.1), (24.3),* y_ε *is the solution of problem (24.19), control* $u \left[t, x, y_\varepsilon (t, x) \right]$ *is defined from (24.13).*

Proof At the beginning we show that as $\varepsilon \to 0$ both the solution y_ε of problem (24.19) and the solution $\bar{y}^\varepsilon$ of problem (24.1) and (24.3) tend to $\overline{y}$ in some sense, where $\{\overline{y}, \overline{u}\}$ is the optimal process in problem (24.8) and (24.10). Firstly we consider problem (24.19). For almost all (a.a.) $t > 0$ and for some $D_1 > 0$ the following estimate holds for the solution y_ε

$$\frac{d}{dt} \left\| y_\varepsilon (t) \right\|^2 + 2v_1 \left\| y_\varepsilon (t) \right\|^2_{H_0^1} \le D_1 \left(\left\| y_\varepsilon (t) \right\|^2 + 1 \right). \tag{24.22}$$

Using Gronwall's Lemma, from (24.22) we obtain that for every $T > 0$ the sequence $\{y_\varepsilon\}$ is bounded in $W(0, T)$. Then, by Compactness Lemma [13] there exists a function z such that along subsequence for every $T > 0$

$$\begin{gathered} y_\varepsilon \to z \text{ in } L^2 (Q_T) \text{ and almost everywhere in } Q, \\ y_\varepsilon(t) \to z(t) \text{ in } L^2 (\Omega) \text{ for a. a. } t > 0, \text{ and weakly } \forall\, t \ge 0. \end{gathered} \tag{24.23}$$

Moreover, from (24.20)

$$y_\varepsilon \to z \text{ weakly in } L^2 (Q). \tag{24.24}$$

From (24.14) we derive that

$$u[t, x, y_\varepsilon] \to u[t, x, z] \text{ in } L^2 (Q_T) \ \forall\, T > 0, \tag{24.25}$$

$$u[t, x, y_\varepsilon] \to u[t, x, z] \text{ weakly in } L^2 (Q). \tag{24.26}$$

From Lemma 24.1 we obtain that z is the solution of problem (24.19) with operator A^0 and initial data y^0, and

$$y_\varepsilon \to z \text{ in } C \left([\delta, T]; L^2 (\Omega) \right) \ \forall 0 < \delta < T. \tag{24.27}$$

Since the optimal control problem (24.8) and (24.10) has a unique solution $\{\overline{y}, \overline{u}\}$ and formula $\overline{u}(t, x) = u[t, x, \overline{y}(t, x)]$ is valid for control $\overline{u}$, then $\overline{y}$ is the solution of problem (24.19) with operator A^0 and initial data y^0. However, this problem also has a unique solution, so $\overline{y} \equiv z$, and moreover, convergences (24.23)–(24.27) hold as $\varepsilon \to 0$ (not only along subsequence).

The following result is proved in [5]

Lemma 24.2 *Let functions q_ε satisfy conditions (24.5), (24.7), (24.21) and for some $T > 0$ $y_\varepsilon \to y$ in $L^2(Q_T)$. Then*

$$q_\varepsilon(t, x, y_\varepsilon) \to q(t, x, y) \text{ weakly in } L^2(Q_T).$$

In all further arguments we will denote by J_ε^T (or J^T) functional (24.3) (or (24.10)) over Q_T.

From Lemma 24.2 and (24.23), (24.25) we derive that for every $T > 0$

$$J_\varepsilon^T\left(y_\varepsilon, u\left[t, x, y_\varepsilon\right]\right) \to J^T\left(\overline{y}, \overline{u}\right), \ \varepsilon \to 0. \tag{24.28}$$

Moreover, from (24.20), (24.22) for $T > \hat{T}$

$$\int_T^\infty \|y_\varepsilon(t)\|^2 dt \le \frac{1}{\gamma}\|y_0^\varepsilon\|^2 e^{(D_1+\gamma)\hat{T}} e^{-\gamma T}. \tag{24.29}$$

Then (24.28) and (24.29) imply convergence

$$J_\varepsilon\left(y_\varepsilon, u\left[t, x, y_\varepsilon\right]\right) \to J\left(\overline{y}, \overline{u}\right), \ \varepsilon \to 0. \tag{24.30}$$

Now we consider the optimal process $\{\bar{y}^\varepsilon, \bar{u}^\varepsilon\}$ of problem (24.1)–(24.3). Let z_ε be a solution of problem (24.1) with control $u \equiv 0$. Then

$$2v_1\lambda \int_0^\infty \|z_\varepsilon(t)\|^2 dt \le \|y_0^\varepsilon\|^2.$$

The last inequality and optimality of $\{\bar{y}^\varepsilon, \bar{u}^\varepsilon\}$ imply an inequality

$$-\int_Q C_2(t,x)dtdx + \int_Q \left(\bar{u}^\varepsilon\right)^2(t,x)dtdx \le$$

$$\le J_\varepsilon\left(\bar{y}^\varepsilon, \bar{u}^\varepsilon\right) \le \frac{C}{2v_1\lambda}\|y_0^\varepsilon\|^2.$$

Therefore, the sequence $\{\bar{u}^\varepsilon\}$ is bounded in $L^2(Q)$. Then, there exists $v \in L^2(Q)$ such that along some subsequence

$$\bar{u}^\varepsilon \to v \text{ weakly in } L^2(Q), \ \varepsilon \to 0.$$

By the boundedness of $\{\bar{u}^\varepsilon\}$ in $L^2(Q)$ and estimate

$$\frac{d}{dt}\left\|\bar{y}^\varepsilon(t)\right\|^2 + 2v_1\left\|\bar{y}^\varepsilon(t)\right\|_{H_0^1}^2 \le 2\left|\left(\bar{y}^\varepsilon(t), \bar{u}^\varepsilon(t)\right|\right., \tag{24.31}$$

and by Gronwall's Lemma, we deduce the boundedness of the sequence $\{\bar{y}^\varepsilon\}$ in $W(0,T)$ for every $T > 0$. Then along subsequence it tends to some function y as $\varepsilon \to 0$ within the meaning of (24.23). Using Lemma 24.1, we obtain that y is the solution of problem (24.8) with control v, and $\bar{y}^\varepsilon$ tends to y in the sense of (24.27).

Let us show that the process $\{y, v\}$ is optimal for problem (24.8)–(24.10). From the optimality of $\{\bar{y}^\varepsilon, \bar{u}^\varepsilon\}$ for arbitrary $u \in L^2(Q)$ the following inequality holds

$$J_\varepsilon\left(\bar{y}^\varepsilon, \bar{u}^\varepsilon\right) \le J_\varepsilon\left(p_\varepsilon, u\right), \tag{24.32}$$

where p_ε is the solution of problem (24.1) with control u. Hence, replacing u^ε on u, estimate (24.31) holds for p_ε. Thus, $\{p_\varepsilon\}$ is bounded in $W(0,T)$ for every $T > 0$. With the above thinking we obtain that p_ε converges to some function p as $\varepsilon \to 0$ in the meaning of (24.23). Moreover, p is the solution of problem (24.8) with control u and p_ε converges to p in the meaning of (24.27). Analyzing problem (24.8), we deduce that for some $\delta > 0$, $C_\delta > 0$ for all $t \ge 0$

$$\frac{d}{dt}\|p(t)\|^2 + \delta\|p(t)\|^2 \le C_\delta\|u(t)\|^2. \tag{24.33}$$

In particular, inequality (24.33) implies that for $T > 0$

$$\|p(T)\|^2 \le \left(\|y_0\|^2 + C_\delta\|u\|_{L^2(Q)}\right)e^{-\frac{\delta T}{2}} + C_\delta\int_{\frac{T}{2}}^{\infty}\|u(t)\|^2dt. \tag{24.34}$$

Moreover, for some $D_2 > 0$ for every $T > 0$

$$\int_T^{+\infty}\|p_\varepsilon(t)\|^2dt \le D_2\left(\|p_\varepsilon(T)\|^2 + \int_T^{+\infty}|u(t)|^2dt\right). \tag{24.35}$$

Then from inequality (24.32) we obtain for all $T > 0$:

$$J_\varepsilon^T\left(\bar{y}^\varepsilon, \bar{u}^\varepsilon\right) \le \int_{Q_T} q_\varepsilon(t,x,p_\varepsilon(t,x))p_\varepsilon(t,x)dtdx + \int_0^{+\infty}|u(t)|^2dt +$$

$$\int_T^{\infty}\int_{\Omega} C_2(t,x)dtdx + D_2\,\|p_\varepsilon(T)\|^2 + D_2\int_T^{+\infty}\|u(t)\|^2dt. \tag{24.36}$$

Hence,

$$\liminf_{\varepsilon\to 0} J_\varepsilon^T\left(\bar{y}^\varepsilon, \bar{u}^\varepsilon\right) \ge \int_{Q_T} q(t,x,y(t,x))y(t,x)dtdx + \liminf_{\varepsilon\to 0}\int_0^T\|\bar{u}^\varepsilon(t)\|^2dt \ge J^T(y,v), \tag{24.37}$$

$$\limsup_{\varepsilon\to 0} J_\varepsilon^T\left(\bar{y}^\varepsilon,\bar{u}^\varepsilon\right)\le D_2\|p(T)\|^2+\int\limits_{Q_T} q(t,x,p(t,x)p(t,x)dtdx+$$

$$\int\limits_T^\infty\int\limits_\Omega C_2(t,x)dtdx+\int\limits_0^{+\infty}|u(t)|^2dt+D_2\int\limits_T^{+\infty}|u(t)|^2dt. \tag{24.38}$$

Thus from inequalities (24.37), (24.38) and (24.34) for $T\to\infty$ it follows that

$$J(y,v)\le J(p,u),$$

so $\{y,v\}=\{\bar{y},\bar{u}\}$ is optimal process in problem (24.8) and (24.10).

Now in previous arguments we put $u=\bar{u}$. Then for corresponding solutions $\bar{p}_\varepsilon$ of problem (24.1) due to (24.12) we have $\bar{p}_\varepsilon\to\bar{y}$ in the meaning of (24.23) and

$$\limsup_{\varepsilon\to 0}\int\limits_0^{+\infty}|\bar{u}^\varepsilon(t)|^2dt\le\int\limits_0^{+\infty}|\bar{u}(t)|^2dt+D_2\|\bar{y}(T)\|^2+D_2\int\limits_T^{+\infty}|\bar{u}(t)|^2dt+\int\limits_T^\infty\int\limits_\Omega C_2(t,x)dtdx,$$

and for $T\to\infty$ we get

$$\limsup_{\varepsilon\to 0}\int\limits_0^{+\infty}|\bar{u}^\varepsilon(t)|^2dt\le\int\limits_0^{+\infty}|\bar{u}(t)|^2dt, \tag{24.39}$$

which guarantees strong convergence of $\bar{u}^\varepsilon$ to $\bar{u}$ in $L_2(Q)$ together with weak convergence.

As inequality (24.33) takes place for $\{\bar{y}^\varepsilon,\bar{u}^\varepsilon\}$, then strong convergence of $\{\bar{y}^\varepsilon$ in $L^2(Q_T)$ and strong convergence of $\bar{u}^\varepsilon$ to $\bar{u}$ in $L_2(Q)$ allow us to pass to the limit for $\varepsilon\to 0$ and obtain

$$J_\varepsilon\left(\bar{y}^\varepsilon,\bar{u}^\varepsilon\right)\to J\left(\overline{y},\overline{u}\right),\ \varepsilon\to 0. \tag{24.40}$$

Theorem is proved.

Remark 24.1 Theorem guarantees convergence not only for quality criteria but also for controls and phase variables in the following way:

$$\bar{u}^\varepsilon-u[t,x,y_\varepsilon]\to 0 \text{ in } L^2(Q),\ \varepsilon\to 0,$$

$$\bar{y}^\varepsilon-y_\varepsilon\to 0 \text{ in } C\left([\delta,T];L^2(\Omega)\right)\ \forall 0<\delta<T.$$

24.4 Conclusion

In this paper, the optimal control problem for a parabolic equation with rapidly oscillating coefficients and non-decomposable quadratic cost functional with superposition type operator was considered on infinite time interval. Using some results for corresponding problem with homogenized parameters, the approximate optimal control in the feedback form (regulator) for the initial problem was grounded.

References

1. Appell, J., Zabreiko, P.: Nonlinear Superposition Operators. Cambridge University Press, Cambridge (1990)
2. Denkowski, Z., Mortola, S.: Asymptotic behavior of optimal solutions to control problems for systems described by differential inclusions corresponding to partial differential equations. J. Optim. Theory Appl. **78**, 365–391 (1993)
3. Fursikov, A.V.: Optimal Control of Distributed Systems. Theory and Applications. AMS, Providence (1999)
4. Jikov, V.V., Kozlov, S.M., Oleinik, O.A.: Homogenization of Differential Operators and Integral Functionals. Springer, Berlin (1994)
5. Kapustian, O.A., Sobchuk, V.V.: Approximate homogenized synthesis for distributed optimal control problem with superposition type cost functional. Stat. Optim. Inf. Comput. **6** 233–239 (2018)
6. Kapustian, O.A., Sukretna, A.V.: Approximate averaged synthesis of the problem of optimal control for a parabolic equation. Ukr. Math. J. **56**(10), 1653–1664 (2004)
7. Kapustyan, O.A.: Approximate synthesis of optimal bounded control for a parabolic boundary-value problem. Ukr. Math. J. **54**(12), 2067–2074 (2002)
8. Kapustyan, E.A., Nakonechnyi, A.G.: Optimal bounded control synthesis for a parabolic boundary-value problem with fast oscillatory coefficients. J. Autom. Inf. Sci. **31**(12), 33–44 (1999)
9. Kapustyan, O.V., Kapustyan, O.A., Sukretna, A.V.: Approximate bounded synthesis for one weakly nonlinear boundary-value problem. Nonlinear Oscil. **12**(3), 297–304 (2009)
10. Kapustyan, O.V., Kapustian, O.A., Sukretna, A.V.: Approximate stabilization for a nonlinear parabolic boundary-value problem. Ukr. Math. J. **63**(5), 759–767 (2011)
11. Lions, J.-L.: Optimal Control of Systems, Governed by Partial Differential Equations. Springer, Berlin (1971)
12. Mashchenko, S.O.: A mathematical programming problem with the fuzzy set of indices of constraints. Cybern. Syst. Anal. **49**(1), 62–68 (2013)
13. Mel'nik, V.S.: Minimization of superposition functionals. Cybern. Syst. Anal. **30**(1), 41–46 (1994)
14. Pichkur, V.V., Sasonkina, M.S.: Practical stabilization of discrete control systems. Int. J. Pure Appl. Math. **81**(6), 877–884 (2012)
15. Sell, G.R., You, Y.: Dynamics of Evolutionary Equations. Springer, New York (2002)
16. Zhikov, V.V., Kozlov, S.M., Oleinik, O.A.: G-convergence of parabolic operators. Russ. Math. Surv. **36**, 9–60 (1983)

Chapter 25
Divided Optimal Control for Parabolic-Hyperbolic Equation with Non-local Pointed Boundary Conditions and Quadratic Quality Criterion

Volodymyr O. Kapustyan and Ivan O. Pyshnograiev

Abstract We obtain necessary and sufficient conditions for finding the divided optimal control for parabolic-hyperbolic equations with non-local boundary conditions and general quadratic criterion in the special norm. The initial data, which guarantee the classical solvability of the problem, was drown out. The unique solvability of problem is established, systems kernels are estimated, and the convergence of solutions of the problem is proved.

25.1 Introduction

The investigation of complex systems behaviour is very important problem nowadays. It appears in different fields of human life. There is a big amount of the methods for solving these problems [1, 2]. One of them is the considering of mathematical models including mixed boundary value problems.

Such problems were studied by various scientists [3–9].

In the paper we found the conditions for the divided optimal control for parabolic-hyperbolic equations with non-local boundary conditions and general quadratic quality criterion in special norm. Its approximate solution was considered in [10].

V. O. Kapustyan
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
e-mail: kapustyanv@ukr.net

I. O. Pyshnograiev (✉)
World Data Center for Geoinformatics and Sustainable Development, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
e-mail: pyshnograiev@wdc.org.ua

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_25

25.2 The Problem Statement

Let the controlled process $y(x,t) \in C^1(\bar{D}) \cap C^2(D_-) \cap C^{2,1}(D_+)$ in D satisfy the equation

$$Ly(x,t) = g(x)\hat{u}(t) \tag{25.1}$$

with initial

$$y(x,-\alpha) = \varphi(x) \tag{25.2}$$

and boundary conditions

$$y(0,t) = 0,\ y'(0,t) = y'(1,t),\ -\alpha \le t \le T, \tag{25.3}$$

where $D = \{(x,t) : 0 < x < 1, -\alpha < t \le T, \alpha, T > 0\}$, $D_- = \{(x,t) : 0 < x < 1, -\alpha < t \le 0\}$, $D_+ = \{(x,t) : 0 < x < 1, 0 < t \le T\}$,

$$Ly = \begin{cases} y_t - y_{xx},\ \mathrm{t} \ge 0, \\ y_{tt} - y_{xx},\ \mathrm{t} < 0. \end{cases}$$

and

$$\hat{u}(t) = \begin{cases} u(t),\ \mathrm{t} \ge 0, \\ v(t),\ \mathrm{t} < 0. \end{cases}$$

This boundary value problem was solved in [11].

It is needed to find the control $v^*(t) \in C[-\alpha, 0) : |v^*(t)| \le 1; |u^*(0)| \le l_0; \xi^*(t) \in L_2[0,T] : |\xi^*(t)| \le l_1$ almost everywhere on $[0,T]$, which minimizes the functional

$$I(\hat{u}) = 0.5(\hat{\alpha}||y(.,T) - \psi(.)||_D^2 + \hat{\beta}_1 \int_{-\alpha}^{0} ||y(.,t)||_D^2 dt + \hat{\beta}_2 \int_0^T ||y(.,t)||_D^2 dt +$$

$$+\hat{\gamma}_1 \int_{-\alpha}^{0} v^2(t)dt + \hat{\gamma}_2(u^2(0) + \int_0^T \xi^2(t)dt)) =$$

$$= 0.5(\sum_{i=0}^{\infty}(\hat{\alpha}\ (y_i(T) - \psi_i)^2 + \hat{\beta}_1 \int_{-\alpha}^{0} y_i^2(t)dt + \hat{\beta}_2 \int_0^T y_i^2(t)dt +$$

$$+\hat{\gamma}_1 \int_{-\alpha}^{0} v^2(t)dt + \hat{\gamma}_2(u^2(0) + \int_0^T \xi^2(t)dt)),\ u(t) = u(0) + \int_0^T \xi(t)dt. \tag{25.4}$$

Because of strict convexity functional (25.4) by control it has a single point of minimum $(v^*(t), u^*(0), \xi^*(t))$ $C[-\alpha, 0) \times R^1 \times L_2(0, T)$, which is characterized by following optimality conditions

$$\int_{-\alpha}^{0} [\hat{\gamma}_1 v^*(t) + \int_{-\alpha}^{0} \mathscr{K}_1^{(1)}(t,\tau) v^*(\tau) d\tau + \mathscr{K}_2^{(1)}(t) u^*(0) + \int_0^T \mathscr{K}_3^{(1)}(t,\tau)\xi^*(\tau) d\tau -$$

$$- \mathscr{M}_1^{(1)}(t,\varphi) - \mathscr{M}_2^{(1)}(t,\psi)][v(t) - v^*(t)]dt \geq 0, \forall |v(t)| \leq 1,$$

$$(\hat{\gamma}_2 u^*(0) + \int_{-\alpha}^{0} \mathscr{K}_1^{(2)}(\tau) v^*(\tau) d\tau + \mathscr{K}_2^{(2)} u^*(0) + \int_0^T \mathscr{K}_3^{(2)}(\tau)\xi^*(\tau) d\tau -$$

$$- \mathscr{M}_1^{(2)}(\varphi) + \mathscr{M}_2^{(2)}(\psi))[u(0) - u^*(0)] \geq 0, \forall |u(0)| \leq l_0,$$

$$\int_0^T [\hat{\gamma}_2 \xi^*(t) + \int_{-\alpha}^{0} \mathscr{K}_1^{(3)}(t,\tau) v^*(\tau) d\tau + \mathscr{K}_2^{(3)}(t) u^*(0) + \int_0^T \mathscr{K}_3^{(3)}(t,\tau)\xi^*(\tau) d\tau -$$

$$- \mathscr{M}_1^{(3)}(t,\varphi) - \mathscr{M}_2^{(3)}(t,\psi)][\xi(t) - \xi^*(t)]dt \geq 0, |\xi(t)| \leq l_1, \tag{25.5}$$

where

$$\mathscr{K}_1^{(1)}(t,\tau) = g_0^2 \mathscr{K}_{0,1}^{(1)}(t,\tau) + \sum_{k=1}^{\infty} (g_{2k-1}^2 \mathscr{K}_{2k-1,1}^{(1,2k-1)}(t,\tau) +$$

$$+2\, g_{2k-1} g_{2k} \mathscr{K}_{2k,1}^{(1,2k-1)}(t,\tau) + g_{2k}^2 \mathscr{K}_{2k,1}^{(1,2k)}(t,\tau)),$$

$$\mathscr{K}_2^{(1)}(t) = g_0^2 \mathscr{K}_{0,2}^{(1)}(t) + \sum_{k=1}^{\infty} (g_{2k-1}^2 \mathscr{K}_{2k-1,2}^{(1,2k-1)}(t) + 2\, g_{2k-1} g_{2k} \mathscr{K}_{2k,2}^{(1,2k-1)}(t) +$$

$$+ g_{2k}^2 \mathscr{K}_{2k,2}^{(1,2k)}(t)),$$

$$\mathscr{K}_3^{(1)}(t,\tau) = g_0^2 \mathscr{K}_{0,3}^{(1)}(t,\tau) + \sum_{k=1}^{\infty} (g_{2k-1}^2 \mathscr{K}_{2k-1,3}^{(1,2k-1)}(t,\tau) +$$

$$+2\, g_{2k-1} g_{2k} \mathscr{K}_{2k,3}^{(1,2k-1)}(t,\tau) + g_{2k}^2 \mathscr{K}_{2k,3}^{(1,2k)}(t,\tau)),$$

$$\mathscr{M}_1^{(1)}(t,\varphi) = \mathscr{M}_{0,1}^{(1)}(t)\, \varphi_0 g_0 + \sum_{k=1}^{\infty} (g_{2k-1}\, (\mathscr{M}_{2k-1,1}^{(1,2k-1)}(t)\varphi_{2k-1} + \mathscr{M}_{2k,1}^{(1,2k-1)}(t)\varphi_{2k}) +$$

$$+ g_{2k} (\mathscr{M}_{2k-1,1}^{(1,2k)}(t)\varphi_{2k-1} + \mathscr{M}_{2k,1}^{(1,2k)}(t)\varphi_{2k})),$$

$$\mathscr{M}_2^{(1)}(t,\psi)=\mathscr{M}_{0,2}^{(1)}(t)\psi_0 g_0+\sum_{k=1}^{\infty}(g_{2k-1}(\mathscr{M}_{2k-1,2}^{(1,2k-1)}(t)\psi_{2k-1}+\mathscr{M}_{2k,2}^{(1,2k-1)}(t)\psi_{2k})+$$
$$+g_{2k}\mathscr{M}_{2k,2}^{(1,2k)}(t)\psi_{2k}),\ \mathscr{K}_1^{(2)}(t)=\mathscr{K}_2^{(1)}(t),$$
$$\mathscr{K}_2^{(2)}=\mathscr{K}_{0,2}^{(2)}g_0^2+\sum_{k=1}^{\infty}(g_{2k-1}^2\mathscr{K}_{2k-1,2}^{(2,2k-1)}+2g_{2k-1}g_{2k}\mathscr{K}_{2k,2}^{(2,2k-1)}+g_{2k}^2\mathscr{K}_{2k,2}^{(2,2k)}),$$
$$\mathscr{K}_3^{(2)}(t)=\mathscr{K}_{0,3}^{(2)}(t)g_0^2+\sum_{k=1}^{\infty}(g_{2k-1}^2\mathscr{K}_{2k-1,3}^{(2,2k-1)}(t)+2g_{2k-1}g_{2k}\mathscr{K}_{2k,3}^{(2,2k-1)}(t)+$$
$$+g_{2k}^2\mathscr{K}_{2k,3}^{(2,2k)}(t)),$$
$$\mathscr{M}_1^{(2)}(\varphi)=\mathscr{M}_{0,1}^{(2)}\,\varphi_0 g_0+\sum_{k=1}^{\infty}(g_{2k-1}\,(\mathscr{M}_{2k-1,1}^{(2,2k-1)}\varphi_{2k-1}+\mathscr{M}_{2k,1}^{(2,2k-1)}\varphi_{2k})+$$
$$+g_{2k}(\mathscr{M}_{2k-1,1}^{(2,2k)}\varphi_{2k-1}+\mathscr{M}_{2k,1}^{(2,2k)}\varphi_{2k})),$$

$$\mathscr{M}_2^{(2)}(\psi)=\mathscr{M}_{0,2}^{(2)}\,\psi_0 g_0+\sum_{k=1}^{\infty}(g_{2k-1}\,(\mathscr{M}_{2k-1,2}^{(2,2k-1)}\psi_{2k-1}+\mathscr{M}_{2k,2}^{(2,2k-1)}\psi_{2k})+$$
$$+g_{2k}\mathscr{M}_{2k,2}^{(2,2k)}\psi_{2k}),\ \mathscr{K}_1^{(3)}(t,\tau)=\mathscr{K}_3^{(1)}(\tau,t),\ \mathscr{K}_2^{(3)}(t)=\mathscr{K}_3^{(2)}(t),$$
$$\mathscr{K}_3^{(3)}(t,\tau)=g_0^2\mathscr{K}_{0,3}^{(3)}(t,\tau)+\sum_{k=1}^{\infty}(g_{2k-1}^2\mathscr{K}_{2k-1,3}^{(3,2k-1)}(t,\tau)+$$
$$+2\,g_{2k-1}g_{2k}\mathscr{K}_{2k,3}^{(3,2k-1)}(t,\tau)+g_{2k}^2\mathscr{K}_{2k,3}^{(3,2k)}(t,\tau)),$$
$$\mathscr{M}_1^{(3)}(t,\varphi)=\mathscr{M}_{0,1}^{(3)}(t)\,\varphi_0 g_0+\sum_{k=1}^{\infty}(g_{2k-1}\,(\mathscr{M}_{2k-1,1}^{(3,2k-1)}(t)\varphi_{2k-1}+$$
$$+\mathscr{M}_{2k,1}^{(3,2k-1)}(t)\varphi_{2k})+g_{2k}(\mathscr{M}_{2k-1,1}^{(3,2k)}(t)\varphi_{2k-1}+\mathscr{M}_{2k,1}^{(3,2k)}(t)\varphi_{2k})),$$

$$\mathscr{M}_2^{(3)}(t,\psi)=\mathscr{M}_{0,2}^{(3)}(t)\,\psi_0 g_0+\sum_{k=1}^{\infty}(g_{2k-1}\,(\mathscr{M}_{2k-1,2}^{(3,2k-1)}(t)\psi_{2k-1}+$$
$$+\mathscr{M}_{2k,2}^{(3,2k-1)}(t)\psi_{2k})+g_{2k}\mathscr{M}_{2k,2}^{(3,2k)}(t)\psi_{2k}).\quad(25.6)$$

25.3 The Problem Solving

25.3.1 Unbounded Control

Let's consider the unbounded control. Then the system of variational inequalities (25.5) takes the form

$$\hat{\gamma}_1 v^*(t) + \int_{-\alpha}^{0} \mathscr{K}_1^{(1)}(t,\tau)v^*(\tau)d\tau + \mathscr{K}_2^{(1)}(t)u^*(0) + \int_{0}^{T} \mathscr{K}_3^{(1)}(t,\tau)\xi^*(\tau)d\tau =$$

$$= \mathscr{M}_1^{(1)}(t,\varphi) + \mathscr{M}_2^{(1)}(t,\psi), t \in [-\alpha, 0),$$

$$\hat{\gamma}_2 u^*(0) + \int_{-\alpha}^{0} \mathscr{K}_1^{(2)}(\tau)v^*(\tau)d\tau + \mathscr{K}_2^{(2)}u^*(0) + \int_{0}^{T} \mathscr{K}_3^{(2)}(\tau)\xi^*(\tau)d\tau =$$

$$= \mathscr{M}_1^{(2)}(\varphi) + \mathscr{M}_2^{(2)}(\psi),$$

$$\hat{\gamma}_2 \xi^*(t) + \int_{-\alpha}^{0} \mathscr{K}_1^{(3)}(t,\tau)v^*(\tau)d\tau + \mathscr{K}_2^{(3)}(t)u^*(0) + \int_{0}^{T} \mathscr{K}_3^{(3)}(t,\tau)\xi^*(\tau)d\tau =$$

$$= \mathscr{M}_1^{(3)}(t,\varphi) + \mathscr{M}_2^{(3)}(t,\psi), t \in (0, T]. \tag{25.7}$$

Using the estimations of (25.6), we establish inequalities

$$||\mathscr{K}_1^{(1)}||_{C(-\alpha,0)\times C(-\alpha,0)} \le C\sum_{k=1}^{\infty}(g_{2k-1}^2||\mathscr{K}_{2k-1,1}^{(1,2k-1)}||_{C(-\alpha,0)\times C(-\alpha,0)} + 2|g_{2k-1}| \times$$

$$\times|g_{2k}|||\mathscr{K}_{2k,1}^{(1,2k-1)}||_{C(-\alpha,0)\times C(-\alpha,0)} + g_{2k}^2||\mathscr{K}_{2k,1}^{(1,2k)}||_{C(-\alpha,0)\times C(-\alpha,0)}) \le$$

$$\le C\sum_{k=1}^{\infty}(g_{2k-1}^2 + g_{2k}^2)(\frac{\hat{\alpha}}{\lambda_k^2\exp(2\lambda_k^2 T)} + \frac{\hat{\beta}_1}{\lambda_k^2} + \frac{\hat{\beta}_2}{\lambda_k^4}),$$

$$||\mathscr{K}_2^{(1)}||_{C(-\alpha,0)} = ||\mathscr{K}_1^{(2)}||_{C(-\alpha,0)} \le C\sum_{k=1}^{\infty}(g_{2k-1}^2 + g_{2k}^2)\,(\frac{\hat{\alpha}}{\lambda_k^2\exp(\lambda_k^2 T)} + \frac{\hat{\beta}_1}{\lambda_k^3} + \frac{\hat{\beta}_2}{\lambda_k^4}),$$

$$||\mathscr{K}_3^{(1)}||_{C(-\alpha,0)\times C(0,T)} = ||\mathscr{K}_1^{(3)}||_{C(-\alpha,0)\times C(0,T)} \le$$

$$\le C\sum_{k=1}^{\infty}(g_{2k-1}^2 + g_{2k}^2)(\frac{\hat{\alpha}}{\lambda_k^2\exp(\lambda_k^2 T)} + \frac{\hat{\beta}_2}{\lambda_k^4}),$$

$$|\mathscr{K}_2^{(2)}| \le C\sum_{k=1}^{\infty}(g_{2k-1}^2 + g_{2k}^2)\,(\frac{\hat{\alpha}}{\lambda_k^2} + \frac{\hat{\beta}_1}{\lambda_k^4} + \frac{\hat{\beta}_2}{\lambda_k^2}),$$

$$||\mathscr{K}_3^{(2)}||_{C(0,T)} = ||\mathscr{K}_2^{(3)}||_{C(0,T)} \le C\sum_{k=1}^{\infty}(g_{2k-1}^2 + g_{2k}^2)\,(\frac{\hat{\alpha}}{\lambda_k^2} + \frac{\hat{\beta}_2}{\lambda_k^2}),$$

$$||\mathscr{K}_3^{(3)}||_{C(0,T)\times C(0,T)} \le C\sum_{k=1}^{\infty}(g_{2k-1}^2 + g_{2k}^2)\,(\frac{\hat{\alpha}}{\lambda_k^2} + \frac{\hat{\beta}_2}{\lambda_k^2});$$

$$||\mathscr{M}_1^{(1)}(.,\varphi)||_{C(-\alpha,0)} \le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)\,(|\varphi_{2k-1}| + |\varphi_{2k}|)\times$$

$$\times\sum_{i,j=2k-1}^{2k}||\mathscr{M}_{i,1}^{(1,j)}||_{C(-\alpha,0)} \le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)\,(|\varphi_{2k-1}| + |\varphi_{2k}|)\times$$

$$\times(\frac{\hat{\alpha}}{\lambda_k^2\exp(2\lambda_k^2 T)} + \frac{\hat{\beta}_1}{\lambda_k} + \frac{\hat{\beta}_2}{\lambda_k^3}),$$

$$||\mathscr{M}_2^{(1)}(.,\psi)||_{C(-\alpha,0)} \le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)\,(|\psi_{2k-1}| + |\psi_{2k}|)\times$$

$$\times\sum_{i,j=2k-1}^{2k}||\mathscr{M}_{i,2}^{(1,j)}||_{C(-\alpha,0)} \le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)\,(|\psi_{2k-1}| + |\psi_{2k}|)\frac{\hat{\alpha}}{\lambda_k\exp\lambda_k^2 T},$$

$$|\mathscr{M}_1^{(2)}(\varphi)| \le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)\,(|\varphi_{2k-1}| + |\varphi_{2k}|)\sum_{i,j=2k-1}^{2k}|\mathscr{M}_{i,1}^{(2,j)}| \le$$

$$\le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)(|\varphi_{2k-1}| + |\varphi_{2k}|)(\frac{\hat{\alpha}}{\lambda_k\exp(\lambda_k^2 T)} + \frac{\hat{\beta}_1}{\lambda_k^2} + \frac{\hat{\beta}_2}{\lambda_k^3}),$$

$$|\mathscr{M}_2^{(2)}(\psi)| \le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)\,(|\psi_{2k-1}| + |\psi_{2k}|)\sum_{i,j=2k-1}^{2k}|\mathscr{M}_{i,2}^{(2,j)}| \le$$

$$\le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)\,(|\psi_{2k-1}| + |\psi_{2k}|)\frac{\hat{\alpha}}{\lambda_k},$$

$$||\mathscr{M}_1^{(3)}(.,\varphi)||_{C(0,T)} \le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)(|\varphi_{2k-1}| + |\varphi_{2k}|)\times$$

$$\times\sum_{i,j=2k-1}^{2k}||\mathscr{M}_{i,1}^{(3,j)}||_{C(0,T)} \le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)(|\varphi_{2k-1}| + |\varphi_{2k}|)\times$$

$$\times(\frac{\hat{\alpha}}{\lambda_k\exp(\lambda_k^2 T)} + \frac{\hat{\beta}_2}{\lambda_k^3}),$$

$$||\mathscr{M}_2^{(3)}(.,\psi)||_{C(0,T)} \le C\sum_{k=1}^{\infty}(|g_{2k-1}|+|g_{2k}|)\,(|\psi_{2k-1}|+|\psi_{2k}|)\times$$

$$\times\sum_{i,j=2k-1}^{2k}||\mathscr{M}_{i,2}^{(3,j)}||_{C(0,T)} \le C\sum_{k=1}^{\infty}(|g_{2k-1}|+|g_{2k}|)\,(|\psi_{2k-1}|+|\psi_{2k}|)\,\frac{\hat{\alpha}}{\lambda_k}. \tag{25.8}$$

From the estimations (25.8), the theorem on symmetric operator [12] and conditions of initial boundary value problem solvability [11] it follows that the convergence of the series

$$\sum_{k=1}^{\infty}\frac{|\psi_{2k-1}|+|\psi_{2k}|}{\lambda_k} \tag{25.9}$$

ensure the convergence of the series in the right-hand sides of the inequalities (25.8), i.e. kernels and the right part of the system (25.7) are continuous functions.

Let's establish the uniquely solvability of the system (25.7). For this purpose we will determine the operator

$$\hat{\mathscr{A}}\theta(.) = \Gamma_{3\times3}\theta(t) + \mathscr{A}\theta(.),$$

where $(\theta(t))' = (v(t), u(0), \xi(t)) \in L_2(-\alpha,0)\times R^1\times L_2(0,T)$, $\Gamma_{3\times3} = diag\{\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_2\}$, operator $\mathscr{A}$ is determined by the remaining members of the left-hand sides of the system of equations (25.7).

Because of the convergence of the series (25.9) operator $\hat{\mathscr{A}}$ operates out of the space $L_2(-\alpha,0)\times R^1\times L_2(0,T)$ to the $L_2(-\alpha,0)\times R^1\times L_2(0,T)$. Also it is linear and continuous. Let's prove the following theorem.

Theorem 25.1 *If functions $\psi(x)$, $\varphi(x)$, $g(x)$ are such that the series (25.9) coincides with the convergent series*

$$\sum_{k=1}^{\infty}\lambda_k^2(|\varphi_{2k-1}|+|\varphi_{2k}|),$$

$$\sum_{k=1}^{\infty}\lambda_k(|g_{2k-1}|+|g_{2k}|),$$

then the system (25.7) has a single solution in space $C(-\alpha,0)\times R^1\times L_2(0,T)$.

Proof The space $L_2(-\alpha,0)\times R^1\times L_2(0,T)$ is a Hilbert space with a natural inner product

$$\langle\theta,\tilde{\theta}\rangle_3=\int\limits_{-\alpha}^{0}v(t)\tilde{v}(t)dt+u(0)\,\tilde{u}(0)+\int\limits_{0}^{T}\xi(t)\tilde{\xi}(t)\,dt,$$

where $(\theta(t))'=(v_(t),u(0),\xi(t)),\ (\tilde{\theta}(t))'=\ (\tilde{v}(t),\tilde{u}(0),\tilde{\xi}(t))$.

Using the solution of initial boundary value problem [11] and (25.6) we select from the functional (25.4) quadratic by control $v(t),\ t\in[-\alpha,0);\ u(0),\xi(t)\in[0,T]$ part and subtract the value

$$0.5(\hat{\gamma}_1\int\limits_{-\alpha}^{0}v^2(t)dt+\hat{\gamma}_2u^2(0)+\int\limits_{0}^{T}\xi^2(t)dt))$$

from it. That is, we get functional

$$0\le\tilde{I}=0.5\langle\mathscr{A}\theta(.),\theta(.)\rangle_3.$$

This implies a positive definition of the operator $\hat{\mathscr{A}}$ and unique solvability of the system (25.7) in the space $L_2(-\alpha,0)\times R^1\times L_2(0,T)$. In addition, an estimation

$$||\theta||_3=(\int\limits_{-\alpha}^{0}v^2(t)dt\ +u^2(0)+\int\limits_{0}^{T}\xi^2(t)dt)^{1/2}\ \le C\sum_{i=1}^{2}(||\mathscr{M}_i^{(1)}||_{L_2(-\alpha,0)}+$$

$$+|\mathscr{M}_i^{(2)}|+||\mathscr{M}_i^{(3)}||_{L_2(0,T)})\le C\sum_{k=1}^{\infty}(|g_{2k-1}|+\ |g_{2k}|)((|\varphi_{2k-1}|+|\varphi_{2k}|)\times$$

$$\times(\frac{\hat{\alpha}}{\lambda_k\exp(\lambda_k^2T)}+\frac{\hat{\beta}_1}{\lambda_k}\ +\frac{\hat{\beta}_2}{\lambda_k^3})+(|\psi_{2k-1}|\ +|\psi_{2k}|)\frac{\hat{\alpha}}{\lambda_k})<\infty \tag{25.10}$$

is made for the solutions of this system.

From the first equation of this system, considering it as an identity on the solution $\theta(t)$, we find an estimate

$$||\frac{dv(.)}{dt}||_{L_2(0,T)}\le C\ (||\frac{d\mathscr{K}_1^{(1)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(-\alpha,0)}||v||_{L_2(0,T)}+$$

$$+||\frac{d\mathscr{K}_2^{(1)}(.)}{dt}||_{L_2(0,T)}|u(0)|+||\frac{d\mathscr{K}_3^{(1)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(0,T)}||\xi||_{L_2(0,T)}\ +$$

$$+||\frac{d\mathscr{M}_1^{(1)}(.,\varphi)}{dt}||_{L_2(-\alpha,0)}+||\frac{d\mathscr{M}_2^{(1)}(.,\psi)}{dt}||_{L_2(-\alpha,0)}). \tag{25.11}$$

Let's estimate the norms of functions derivatives standing on the right side of the inequality (25.10).

$$||\frac{d\mathscr{K}_1^{(1)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(-\alpha,0)} \le C\sum_{k=1}^{\infty}[g_{2k-1}^2 \times$$

$$\times(||\frac{d\mathscr{K}_{2k-1,1}^{(1,2k-1)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(-\alpha,0)} + ||\frac{d\mathscr{K}_{2k,1}^{(1,2k-1)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(-\alpha,0)}) +$$

$$+g_{2k}^2(||\frac{d\mathscr{K}_{2k,1}^{(1,2k)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(-\alpha,0)}) +$$

$$+||\frac{d\mathscr{K}_{2k,1}^{(1,2k-1)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(-\alpha,0))}] \le C\sum_{k=1}^{\infty}(g_{2k-1}^2 + g_{2k}^2)(\frac{\hat{\alpha}}{\lambda_k\exp(2\lambda_k^2T)} + \frac{\hat{\beta}_1}{\lambda_k} + \frac{\hat{\beta}_2}{\lambda_k^3}),$$

$$||\frac{d\mathscr{K}_2^{(1)}(.)}{dt}||_{L_2(-\alpha,0)} \le C\sum_{k=1}^{\infty}[g_{2k-1}^2(||\frac{d\mathscr{K}_{2k-1,2}^{(1,2k-1)}(.)}{dt}||_{L_2(-\alpha,0)} +$$

$$+||\frac{d\mathscr{K}_{2k,2}^{(1,2k-1)}(.)}{dt}||_{L_2(-\alpha,0))} + g_{2k}^2(||\frac{d\mathscr{K}_{2k,2}^{(1,2k)}(.)}{dt}||_{L_2(-\alpha,0)} + ||\frac{d\mathscr{K}_{2k,2}^{(1,2k-1)}(.)}{dt}||_{L_2(-\alpha,0))}] \le$$

$$\le C\sum_{k=1}^{\infty}(g_{2k-1}^2 + g_{2k}^2)(\frac{\hat{\alpha}}{\lambda_k\exp(\lambda_k^2T)} + \frac{\hat{\beta}_1}{\lambda_k^2} + \frac{\hat{\beta}_2}{\lambda_k^3}),$$

$$||\frac{d\mathscr{K}_3^{(1)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(0,T)} \le C\sum_{k=1}^{\infty}[g_{2k-1}^2||\frac{d\mathscr{K}_{2k-1,3}^{(1,2k-1)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(0,T)} +$$

$$+||\frac{d\mathscr{K}_{2k,3}^{(1,2k-1)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(0,T)}) + g_{2k}^2(||\frac{d\mathscr{K}_{2k,3}^{(1,2k)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(0,T)} +$$

$$+||\frac{d\mathscr{K}_{2k,3}^{(1,2k-1)}(.,.)}{dt}||_{L_2(-\alpha,0)\times L_2(0,T))}] \le C\sum_{k=1}^{\infty}(g_{2k-1}^2 + g_{2k}^2)\ (\frac{\hat{\alpha}}{\lambda_k\exp(\lambda_k^2T)} + \frac{\hat{\beta}_2}{\lambda_k^3}),$$

$$||\frac{d\mathscr{M}_1^{(1)}(.,\varphi)}{dt}||_{L_2(-\alpha,0)} \le C\sum_{k=1}^{\infty}[|g_{2k-1}|\ (||\frac{d\mathscr{M}_{2k-1,1}^{(1,2k-1)}(.)}{dt}||_{L_2(-\alpha,0)}|\varphi_{2k-1}| +$$

$$+||\frac{d\mathscr{M}_{2k,1}^{(1,2k-1)}(.)}{dt}||_{L_2(-\alpha,0)}|\varphi_{2k}|) + |g_{2k}|\ (||\frac{d\mathscr{M}_{2k-1,1}^{(1,2k)}(.)}{dt}||_{L_2(-\alpha,0)}|\varphi_{2k-1}| +$$

$$+||\frac{d\mathscr{M}_{2k,1}^{(1,2k)}(.)}{dt}||_{L_2(-\alpha,0)}|\varphi_{2k}|)] \le C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)(|\varphi_{2k-1}| + |\varphi_{2k}|) \times$$

$$\times(\frac{\hat{\alpha}}{\lambda_k\exp(2\lambda_k^2T)} + \hat{\beta}_1 + \frac{\hat{\beta}_2}{\lambda_k^2}),$$

$$||\frac{d\mathscr{M}_2^{(1)}(.,\psi)}{dt}||_{L_2(-\alpha,0)} \le C\sum_{k=1}^{\infty}[|g_{2k-1}|\ (||\frac{d\mathscr{M}_{2k-1,2}^{(1,2k-1)}(.)}{dt}||_{L_2(-\alpha,0)}|\psi_{2k-1}| +$$

$$+||\frac{d\mathscr{M}_{2k,2}^{(1,2k-1)}(.)}{dt}||_{L_2(-\alpha,0)}|\psi_{2k}|) + |g_{2k}|||\frac{d\mathscr{M}_{2k,2}^{(1,2k)}(.)}{dt}||_{L_2(-\alpha,0)}\ |\psi_{2k}|] \leq$$

$$\leq C\sum_{k=1}^{\infty}(|g_{2k-1}| + |g_{2k}|)(|\psi_{2k-1}| + |\psi_{2k}|)\ \frac{\hat{\alpha}}{\exp\lambda_k^2 T}. \tag{25.12}$$

When the conditions of the theorem are fulfilled the estimations (25.10)–(25.12) guarantee the absolute continuity of the control $v(t)$ •

Let's consider the following case.

Let $v(t) = 0, t < 0; \xi(t) = 0, t > 0, \hat{\beta}_1 = \hat{\beta}_2 = 0$. Then from the second equation of the system (25.7) we find

$$u(0) = \frac{\mathscr{M}_1^{(2)}(\varphi) + \mathscr{M}_2^{(2)}(\psi)}{\hat{\gamma}_2 + \mathscr{K}_2^{(2)}},$$

where

$$\mathscr{K}_2^{(2)} = (\hat{U}_{0,+}^0(T)g_0)^2 + \sum_{k=1}^{\infty}[(\hat{U}_{2k-1,+}^{2k-1}(T)g_{2k-1})^2 + (\hat{U}_{2k,+}^{2k-1}(T)g_{2k-1} + \hat{U}_{2k,+}^{2k}(T)g_{2k})^2],$$

$$\mathscr{M}_1^{(2)}(\varphi) = -\Phi_{0,+}^0(T)\ \hat{U}_{0,+}^0(T)\varphi_0 g_0 - \sum_{k=1}^{\infty}[\Phi_{2k-1,+}^{2k-1}(T)\ \hat{U}_{2k-1,+}^{2k-1}(T)\varphi_{2k-1}\ g_{2k-1}) +$$

$$+(\Phi_{2k,+}^{2k-1}(T)\varphi_{2k-1} + \Phi_{2k,+}^{2k}(T)\varphi_{2k})(\hat{U}_{2k,+}^{2k-1}(T)g_{2k-1} + \hat{U}_{2k,+}^{2k}(T)g_{2k})],$$

$$\mathscr{M}_2^{(2)}(\varphi) = \hat{U}_{0,+}^0(T)\psi_0\ g_0 + \sum_{k=1}^{\infty}[\hat{U}_{2k-1,+}^{2k-1}(T)\psi_{2k-1}\ g_{2k-1}) +$$

$$+\psi_{2k}(\hat{U}_{2k,+}^{2k-1}(T)g_{2k-1} + \hat{U}_{2k,+}^{2k}(T)g_{2k})].$$

The values of the Fourier coefficients of the optimal trajectories also correspond to this control at the point $t = T$:

$$y_0(T) = \Phi_{0,+}^0(T)\varphi_0 + g_0\ U_{0,+}^0(T)u(0),$$

$$y_{2k-1}(T) = \Phi_{2k-1,+}^{2k-1}(T)\varphi_{2k-1} + \ g_{2k-1}\hat{U}_{2k-1,+}^{2k-1}(T)u(0),$$

$$y_{2k}(T) = \Phi_{2k,+}^{2k-1}(T)\varphi_{2k-1} + \ \Phi_{2k,+}^{2k}(T)\varphi_{2k} + (g_{2k-1}\ \hat{U}_{2k,+}^{2k-1}(T) + g_{2k}\hat{U}_{2k,+}^{2k}(T))\ u(0).$$

And the value of the optimality criterion is

$$I = 0.5[\sum_{k=0}^{\infty}\ \hat{\alpha}(y_i(T) - \psi_i)^2 + \ \hat{\gamma}_2 u^2(0)].$$

The above formulas remain valid for $\hat{\gamma}_2 = 0$, but in this case they do not solve the problem with minimal energy: the control

$$\lim_{\hat{\gamma}_2 \to 0} u(0) = \frac{\mathscr{M}_1^{(2)}(\varphi) + \mathscr{M}_2^{(2)}(\psi)}{\mathscr{K}_2^{(2)}}$$

does not enforce conditions

$$\lim_{\hat{\gamma}_2 \to 0} y_i(0) = \psi_i, i \geq 0 \tag{25.13}$$

since then they would have to satisfy equality

$$\frac{\psi_0 - \Phi_{0,+}^0(T)\varphi_0}{g_0\, U_{0,+}^0(T)} = \frac{\psi_{2k-1} - \Phi_{2k-1,+}^{2k-1}(T)\varphi_{2k-1}}{g_{2k-1}\, \hat{U}_{2k-1,+}^{2k-1}(T)} =$$

$$= \frac{\psi_{2k} - \Phi_{2k,+}^{2k-1}(T)\varphi_{2k-1} - \Phi_{2k,+}^{2k}(T)\varphi_{2k}}{(g_{2k-1}\, \hat{U}_{2k,+}^{2k-1}(T) + g_{2k}\hat{U}_{2k,+}^{2k}(T))} = \lim_{\hat{\gamma}_2 \to 0} u(0), \tag{25.14}$$

which is impossible in the case of arbitrary functions $\varphi(x), \psi(x), g(x)$ from Theorem 25.1.

25.3.2 Bounded Control

Let $u^*(0) = 0, \xi^*(t) = 0, t > 0$, and optimal control $v^*(t)$ has the following structure: $v^*(t) = -1, t \in [-\alpha, \bar{\xi}_1]; |v^*(t)| < 1, t \in (\bar{\xi}_1, 0]$. Then this control satisfies the expressions

$$v^*(t) = -1, -1 + \int_{\bar{\xi}_1}^{0} \mathscr{K}_1^{(1)}(t,\tau)v^*(\tau)d\tau > \mathscr{M}_1^{(1)}(t,\varphi) + \mathscr{M}_2^{(1)}(t,\psi) +$$

$$+ \int_{-\alpha}^{\bar{\xi}_1} \mathscr{K}_1^{(1)}(t,\tau)d\tau, t \in [-\alpha, \bar{\xi}_1);$$

$$v^*(t) + \int_{\bar{\xi}_1}^{0} \mathscr{K}_1^{(1)}(t,\tau)\, v^*(\tau)d\tau = \mathscr{M}_1^{(1)}(t,\varphi) + \mathscr{M}_2^{(1)}(t,\psi) +$$

$$+ \int_{-\alpha}^{\bar{\xi}_1} \mathscr{K}_1^{(1)}(t,\tau)d\tau, |v^*(t)| < 1, t \in [\bar{\xi}_1, 0). \tag{25.15}$$

The number $\bar{\xi}_1$ here defined as the only solution of the equation

$$v^*(\bar{\xi}_1) = -1, \tag{25.16}$$

where $v^*(t)$ is the solution of the equation from (25.15).

It should be noted that from the results of the paragraph with the unbounded control it follows that the Eq. (25.15) has a single absolutely continuous solution.

References

1. Kseniia, I., Ivan, P.: A composite indicator of K-society measurement. In: Proceedings of the 11th International Conference on ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer, pp. 161–171 (2015)
2. Zgurovsky, M., Boldak, A., Yefremov, K., Pyshnograiev, I.: Modeling and investigating the behavior of complex socio-economic systems. In: Conference Proceedings of 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 1113–1116 (2017)
3. Kapustyan, V.O., Pyshnograiev, I.O.: Problem of optimal control for parabolic-hyperbolic equations with nonlocal point boundary conditions and semidefinite quality criterion. Ukrainskyi Matematychnyi Zhurnal. **67**(8), 1068–1081 (2016)
4. Kapustyan, V.O., Kapustian, O.A., Mazur, O.K.: Problem of optimal control for the Poisson equation with nonlocal boundary conditions. J. Math. Sci. **201**, 325–334 (2014)
5. Kapustyan, V.O., Kapustyan, O.V., Kapustian, O.A., Mazur, O.K.: The optimal control problem for parabolic equation with nonlocal boundary conditions in circular sector. In: Continuous and Distributed Systems II, pp. 297–314. Springer, Cham (2015)
6. Infante, G.: Positive solutions of nonlocal boundary value problems with singularities. Discrete Contin. Dyn. Syst. **3**, 377384 (2009)
7. Martin-Vaquero, J., Wade, B.A.: On efficient numerical methods for an initial-boundary value problem with nonlocal boundary conditions. Appl. Numer. Math. **36**, 3411–3418 (2010)
8. Nakhusheva, Z.A.: Nonlocal problem for the Lavrentev-Bitsadze equation and its analogs in the theory of equations of mixed parabolic-hyperbolic type. Differ. Equ. **49**(10), 1299–1306 (2013)
9. Repin, O.A., Kumykova, S.K.: A nonlocal problem for a mixed-type equation whose order degenerates along the line of change of type. Russ. Math. **57**(9), 49–56 (2013)
10. Kapustyan, V.O., Pyshnograiev, I.O.: Approximate Optimal Control for Parabolic Hyperbolic Equations with Nonlocal Boundary Conditions and General Quadratic Quality Criterion. Advances in Dynamical Systems and Control, pp. 387–401. Springer, Cham (2016)
11. Kapustyan, V.O., Pyshnograiev, I.O.: The conditions of existence and uniqueness of the solution of a parabolic-hyperbolic equation with nonlocal boundary conditions [Ukrainian]. Science News NTUU "KPI" **4**, 72–86 (2012)
12. Egorov, A.I.: Optimal Control for Heating and Diffusing Processes. Nauka, Moscow (1978)

Chapter 26
Quasi-Linear Differential-Deference Game of Approach

Lesia V. Baranovska

Abstract The paper is devoted to the games of approach. We consider a controlled object whose dynamics is described by the linear differential system with pure time delay or the differential-difference system with commutative matrices in Euclidean space. The approaches to the solutions of these problems are proposed which based on the Method of Resolving Functions and the First Direct Method of L.S. Pontryagin. The guaranteed times of the game termination are found, and corresponding control laws are constructed. The results are illustrated by a model example.

26.1 Introduction

We consider the game problems of approach, which are central to the theory of conflict-controlled processes. They were the basis of the emergence of the theory, are the most informative and of considerable interest to researchers. The impetus for their development was given by real applications in economics, space technology, military affairs, biology, medicine, etc.

Conflict-controlled processes is a section of the mathematical control theory which is studying the manipulation of moving objects operated under in conditions of conflict and uncertainty. The evolution of an object can be described by systems of difference, ordinary differential, differential-difference, integral, integro-differential equations, systems of equations with distributed parameters, systems of equations with fractional derivatives, impulse influences and their various combinations (hybrid systems).

The term differential game is used for games in which the dynamics of an object is described by a system of ordinary differential equations. If the process is described by more complicated equations, possessing the semigroup property, then

L. V. Baranovska (✉)
Institute for Applied System Analysis, National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine
e-mail: lesia@baranovsky.org

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*, Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_26

the term dynamic games is used. Finally, conflict-controlled processes are the most common term for determining the range of issues relating to game problems.

There are two types of dynamic games: games of degree and games of kind (see [1]). On the trajectory of the dynamical system, there is a function that depends on the initial state and on the player's control. In games of the first type, the goal of the first player is to minimize this function, set on the system trajectories, the purpose of the other one is to maximize it. In games of the second type, this functionality is the time of the exit of the trajectory of an object to a given terminal set, and the problem is to analyze the possibility of the pursuit of a trajectory of a system to a terminal set (the game of approach) or the deviation of the trap escape from this set (the deviation game).

The well-known pursuit strategies were mostly designed for military purposes. In practice, the rule of positional pursuit (see Fig. 26.1) and the rule of parallel pursuit (see Fig. 26.2) are widely used.

In the theory of differential games, along with the Pontryagin-Pshenichny's backward procedures (see [2, 3]), Krasovskii rule of extreme aiming (see [4]) and Isaacs's ideology (see [1]), there exist effective methods that constitutes share a separate direction.

Fig. 26.1 Positional pursuit

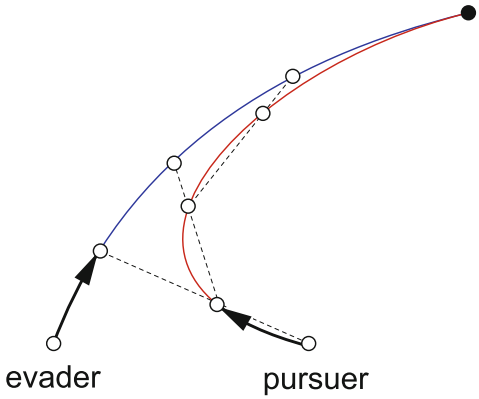
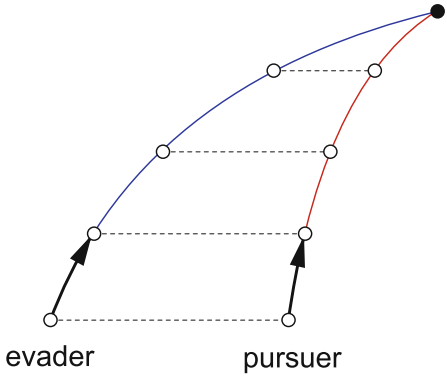


Fig. 26.2 Parallel pursuit



These are the First Direct Method of L.S. Pontryagin and the Method of Resolving Functions (see [5]). They are combined by the general principle of constructing controls of the pursuer on the basis of the Filippov-Castain multidimensional choice theorem (see [6]) and they provide a theoretical justification for the rule of parallel pursuit (see Fig. 26.2).

In this paper, the Method of Resolving Functions is chosen as the main tool for research, widely used to study conflict-controlled processes of various nature (see [5, 7]). The processes with fractional derivatives are studied in (see [8]), game problems of successive convergence are discussed in (see [9]), a general scheme of the method of resolving functions is given in (see [7]), the applied problem of soft meeting is solved in (see [10]), the nonstationary problems are considered in (see [11–14]), a variant of the matrix resolving functions are proposed in (see [15]), an approach games problem under the failure of controlling devices are considered in (see [16, 17]), and in (see [18, 19]) the cases of integral constraints on control are examined.

The future of many processes depends not only on the present state, but is also significantly determined by the entire prehistory. Numerous problems in the theory of automatic control, engineering, mechanics, radiophysics, biology, economics are described by differential equations with delay. For example, transport delay usually occurs in systems in which matter, energy or signals are transmitted over a distance (see [20]). In control systems, where one of the links is a person, the delay in the reaction of a person is important in constructing a mathematical model of the entire system. Distributed time delay occurs in the modeling of feeding systems and combustion chambers in a liquid monopropellant rocket motor with pressure feeding (see [21]). Great contribution to the development of these directions is made by Bellman R., Cooke K., Lunel S.M.V., Mitropolskii U.A., Myshkis A.D., Norkin S.B., Hale J.C., Azbelev N.V., Maksimov V.P., Rakhmatulina L.F. and others.

In (see [22–25]) the modification of the Method of Resolving Function for the differential-difference pursuit games is described, pursuit differential-difference games of approach with non-fixed time are considered in (see [26, 27]), system with time-varying delay is considered in (see [28]), in (see [29, 30]) the pursuit games with differential-difference equations of a neutral type are studied, an analytic approach based on the Method of Resolving Functions to study the differential-difference games of approach with commutative matrices is suggested in (see [31]), and the differential-difference games of approach for objects with different inertial are proposed in (see [32, 33]).

An attractive side of the Method of Resolving Functions is the fact that it allows us to effectively use modern technology of set-valued mappings and their selectors in the substantiation of game constructions and to obtain meaningful results on their basis (see [5]).

For dynamical systems whose evolution is described by differential-difference system with a cylindrical terminal set under the condition of L.S. Pontryagin introduces a resolving function, through which the game's end time is determined. The peculiarity of the basic scheme of the method is the fact that the time of the

end of the game depends on a selector, the choice of which is in the power of the pursuer.

The resolving function characterizes the course of the game. When, at some point in time, the integral from it becomes a unit, this means that the trajectory falls onto the terminal set. Sufficient conditions for solvability of the problem of approach with a terminal set are provided. The pursuit process is divided into two stages.

On the first one $[0,\ t_*)$, where t_* is the moment of switching, the Method of Resolving Functions with using by the pursuer at the time t of the entire run-time control prehistory $v_t\,(\cdot)$ work. When at the instant t_* the integral of the resolving function turns into unity, the process of pursuit is switched to the First Direct Method of L.S. Pontryagin which is realized within the class of countercontrols in quasistrategy. In other words, from the moment of switching to the calculated moment, the ending of the game "stretches" time, and, in this area, the resolving function is considered to be zero, since it does not make any sense to accumulate it.

26.2 Differential-Difference Games of Approach with Commutative Matrices

Let $\mathbb{R}^n$ be an Euclidean space of points $z = (z_1, \ldots, z_n)$ and $K\,(\mathbb{R}^n)$ be a set of nonempty compacts in $\mathbb{R}^n$.

We consider the problem of approach for the system of differential-difference equations of retarded type (see [34–36]):

$$\dot{z}\,(t) = Az\,(t) + Bz\,(t-\tau) + \phi\,(u\,,\ v)\,,\quad z \in \mathbb{R}^n\,,\quad u \in U\,,\quad v \in V\,, \tag{26.1}$$

where A and B are square constant matrices of order n; $U,\ V \in K\,(\mathbb{R}^n)$; $\phi\colon U \times V \to \mathbb{R}^n$, is jointly continuous in its variables; $\tau = \text{const} > 0\,.$

The phase vector consists of geometric coordinates, velocities and accelerations of the pursuer and the evader.

Let $z\,(t)$ be a solution of Eq. (26.1) under the initial condition

$$z\,(t) = z^0\,(t)\,,\quad -\tau \le t \le 0\,, \tag{26.2}$$

where function $z^0\,(t)$ is absolutely continuous on $[-\tau\,,\ 0]\,.$

The piece of the trajectory $z^t\,(\cdot\,)$, where

$$z^t\,(\cdot\,) = \{z\,(t+s)\,,\quad -\tau \le s \le 0\}$$

will be referred to as the state of system (26.1) at the moment t.

Definition 26.1 (See [37, 38]) For each $k = 1, 2, \ldots$, the time-delay exponential is defined as follows

$$exp_\tau\{B,t\} = \begin{cases} \Theta, \ -\infty < t < -\tau; \\ I, \ -\tau \le t < 0; \\ I + B\frac{t}{1!} + B^2\frac{(t-\tau)^2}{2!} + \cdots + B^k\frac{(t-(k-1)\tau)^k}{k!}, \ (k-1)\tau \le t \le k\tau, \end{cases}$$

where Θ is a zero matrix.

Lemma 26.1 (See [37, 38]) *Let $z(t)$ be a continuous solution to the system (26.1) with commutative matrices A and B under the initial condition in (26.2). Then,*

$$\begin{aligned} z(t) &= exp\{A(t+\tau)\}exp_\tau\{B_1, t-\tau\}z^0(-\tau) \\ &+ \int_{-\tau}^{0} exp\{A(t-\tau)\}exp_\tau\{B_1, t-\tau-s\}[\dot{z}^0(s) - Az^0(s)]ds \\ &+ \int_0^t exp\{A(t-\tau-s)\}exp_\tau\{B_1, t-\tau-s\}\phi(u(s), v(s))\,ds, \end{aligned}$$

or, in another form,

$$\begin{aligned} z(t) &= F(t)\,a + \int_{-\tau}^{0} F(t-\tau-s)\,b(s)\,ds \\ &+ \int_0^t F(t-\tau-s)\,\phi(u(s), v(s))\,ds, \end{aligned}$$

where we denote

$$a = exp\{A\tau\}z^0(-\tau), \quad b(t) = exp\{A\tau\}[\dot{z}^0(t) - Az^0(t)],$$

and matrix

$$F(t) = exp\{At\}exp_\tau\{B_1, t\}, \ t \ge 0, \quad B_1 = exp\{-A\tau\}B,$$

is a solution to the similar system

$$\dot{z}(t) = Az(t) + Bz(t-\tau)$$

under the initial condition

$$F(t) \equiv exp\{At\}, \quad -\tau \leq t \leq 0.$$

Let us examine the differential-difference system (see [31]) as an example:

$$\dot{z}(t) = Az(t) + Bz(t-\tau) + u(t) - v(t), \quad z \in \mathbb{R}^{2n},$$

where

$$A = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix},$$

0 is a zero matrix, I is a unit matrix of order n,

$$U = \left\{ \begin{pmatrix} -u(t) \\ 0 \end{pmatrix} : u \in \mathbb{R}^n, \ ||u|| \leq 2 \right\}, \quad V = \left\{ \begin{pmatrix} 0 \\ -v(t) \end{pmatrix} : v \in \mathbb{R}^n, \ ||v|| \leq 1 \right\}.$$

The initial condition is equal to

$$z^0(t) = \left(z_1^0(t), z_2^0(t) \right), \quad -1 \leq t \leq 0.$$

We observe that matrices A and B are commutative, and $AB = BA = \Theta$, $A^n = A$, $B^n = B$.

From Lemma 26.1, we see that the functional matrix $F(t)$ is a solution to the similar system

$$\begin{pmatrix} F_{11}(t) & F_{12}(t) \\ F_{21}(t) & F_{22}(t) \end{pmatrix} \otimes I =$$

$$\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} F_{11}(t) & F_{12}(t) \\ F_{21}(t) & F_{22}(t) \end{pmatrix} \otimes I + \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} \cdot \begin{pmatrix} F_{11}(t-1) & F_{12}(t-1) \\ F_{21}(t-1) & F_{22}(t-1) \end{pmatrix} \otimes I =$$

$$\begin{pmatrix} F_{11}(t) & F_{12}(t) \\ F_{21}(t-1) & F_{22}(t-1) \end{pmatrix} \otimes I$$

and it satisfies the initial condition $F(t) \equiv exp\{At\}, \quad -\tau \leq t \leq 0$. Since

$$B_1 = exp\{-A\} \cdot B = \left(I_n - A + \frac{A^2}{2!} - \frac{A^3}{3!} + \cdots + (-1)^n \frac{A^n}{n!} + \cdots \right) \cdot B = B,$$

we obtain

$$F(t) = exp\{At\} \cdot exp_\tau\{B, t\}$$

$$= \left(I_n + At + A^2\frac{t^2}{2!} + A^3\frac{t^3}{3!} + \cdots + A^n\frac{t^n}{n!} + \cdots\right)$$

$$\cdot \left(I_n + Bt + B^2\frac{(t-1)^2}{2!} + B^3\frac{(t-2)^3}{3!} + \cdots + B^n\frac{(t-(n-1))^n}{n!} + \cdots\right)$$

$$= I_n + Bt + B^2\frac{(t-1)^2}{2!} + B^3\frac{(t-2)^3}{3!} + \cdots + B^n\frac{(t-(n-1))^n}{n!} + \cdots$$

$$+At + A^2\frac{t^2}{2!} + A^3\frac{t^3}{3!} + \cdots + A^n\frac{t^n}{n!} + \cdots = \begin{pmatrix} e^t & 0 \\ 0 & F_{22}(t) \end{pmatrix} \otimes I,$$

where

$$F_{22}(t) = exp_1\{I, t\} = 1 + \frac{t}{1!} + \frac{(t-1)^2}{2!} + \frac{(t-2)^3}{3!} + \cdots + \frac{(t-(k-1))^k}{k!},$$

$$(k-1) \le t \le k,\ k = 0, 1, 2, \ldots.$$

The terminal set has cylindrical form, i.e.

$$M^* = M_0 + M, \tag{26.3}$$

where M_0 is a linear subspace in $\mathbb{R}^n$and M is a compact set from the orthogonal complement of M_0 in $\mathbb{R}^n$.

The players choose their controls in the form of certain functions. Thus, the pursuer and the evader affect the process (26.1), pursuing their own goals. The goal of the pursuer (u) is in the shortest time to bring a trajectory of the process to a certain closed set M^*; the goal of the evader (v) is to avoid a trajectory of the process from meeting with the terminal set (26.3) on a whole semi-infinite interval of time or if is impossible to maximally postpone the moment of meeting.

Now we describe what kind of information is available to the pursuer in the course of the game.

Denote by Ω_U, Ω_V the sets of Lebesgue measurable functions $u(t), v(t)$, $u(t) \in U$, $v(t) \in V$, $t \ge 0$, respectively. A mapping that puts into correspondence to a state $z^0(\cdot)$ some element in Ω_V is called an open-loop strategy of the evader, specific realization of this strategy for a given initial state $z^0(\cdot)$ of process (26.1) is called an open-loop control. In the process of the game (26.1), (26.3), the evader applies open-loop controls $v(\cdot) \in \Omega_V$.

Function

$$u(t) = u\left(z^0(\cdot),\ t,\ v(t)\right),$$

such that $v(\cdot) \in \Omega_V$ implies $u(\cdot) \in \Omega_U$ is called countercontrol (stroboscopic strategy of Hajek (see [39])) of pursuer corresponding to initial state $z^0(\cdot)$. The game is evolving on the closed time interval $[0,\ T]$. We assume that the pursuer chooses his control in the form

$$u(t) = u\left(z^0(\cdot),\ t,\ v_t(\cdot)\right), \quad t \geq 0,$$

where $v_t(\cdot) = \{v(s) : s \in [0,\ t],\ v(\cdot) \in \Omega_V\}$, and $u(\cdot) \in \Omega_U$.

Under these hypotheses, we will play the role of the pursuer and find sufficient conditions on the parameters of the problem (26.1), (26.3), insuring the game termination for certain guaranteed time.

Let π be the orthogonal projector from $\mathbb{R}^n$ onto the subspace L. Consider the set-valued mapping

$$W(t,\ v) = \pi F(t)\phi(U,\ v), \quad W(t) = \bigcap_{v \in V} W(t,\ v),$$

where $F(t)$ is defined in Lemma 26.1.

Condition 1 (Pontryagin's Condition) The mapping $W(t) \neq \emptyset$ for all $t \geq 0$.

Remark 26.1 For the linear process ($\phi(u, v) = u - v$)

$$W(t) = \pi K(t) U \overset{*}{-} \pi K(t) V,$$

where $\overset{*}{-}$ is a geometric subtraction of the sets (Minkowski' difference) (see [43]).

By virtue of the assumptions on the process parameters, the set-valued mapping $W(t,\ v)$ is continuous on the set $[0,\ +\infty) \times V$ in Hausdorff metric. Consequently, as follows from Condition 1, the mapping $W(t)$ is upper semi-continuous and therefore Borel measurable function (see [44]). Hence, there exists at least one Borelian selection $g(t)$, $g(t) \in W(t)$, $t \geq 0$ (see [45]). Let us denote by $G = \{g(\cdot) : g(t) \in W(t),\ t \geq 0\}$ the set of all Borelian selections of the set-valued mapping $W(t)$. For fixed $g(\cdot) \in G$ we put

$$\xi\left(t,\ z^0(\cdot),\ g(\cdot)\right) =$$

$$= \pi F(t) a + \int_{-\tau}^{0} \pi F(t - \tau - s) b(s) ds + \int_0^t g(s) ds,$$

and consider the resolving function

$$\alpha\left(t, s, z^0(\cdot), m, v, g(\cdot)\right) = \alpha_{W(t-\tau-s,v)-g(t-\tau-s)}\left(m - \xi\left(t,\ z^0(\cdot),\ g(\cdot)\right)\right)$$

for $t \geq s \geq 0$, $v \in V$, $m \in M$, $x \in \mathbb{R}^n$.

By virtue of the properties of the superposition of set-valued mappings and functions, it is Borel measurable function in s, v (see [5]). Finally, denote

$$\alpha\left(t, s, z^0(\cdot), v, g(\cdot)\right) = \max_{m \in M} \alpha\left(t, s, z^0(\cdot), m, v, g(\cdot)\right),$$

and then we obtain the resolving function

$$\begin{aligned} \alpha\left(t, s, z^0(\cdot), v, g(\cdot)\right) &= \sup\{\alpha \geq 0 : \\ [W(t-\tau-s, v) - g(t-\tau-s)] \cap \alpha\left[M - \xi\left(t, z^0(\cdot), g(\cdot)\right)\right] &\neq \emptyset\}. \end{aligned} \tag{26.4}$$

Moreover, we also observe that function $\alpha\left(t,\ s,\ z^0(\cdot),\ v,\ g(\cdot)\right) = +\infty$ for all $s \in [0,\ t]$, $v \in V$, if and only if $\xi\left(t,\ z^0(\cdot),\ g(\cdot)\right) \in M$. If for some $t \geq 0$ $\xi\left(t,\ z^0(\cdot),\ \gamma(\cdot)\right) \notin M$, then function (26.4) assumes finite values.

Define the function T by

$$\begin{aligned} T &= T\left(z^0(\cdot),\ g(\cdot)\right) \\ &= \inf\left\{t \geq 0 : \int_0^t \inf_{v \in V} \alpha\left(t,\ s,\ z^0(\cdot),\ v,\ g(\cdot)\right) ds \geq 1\right\}, \quad g(\cdot) \in G. \end{aligned} \tag{26.5}$$

If the inequality in the curly brackets is not satisfied for all $t \geq 0$, we set $T\left(z^0(\cdot), g(\cdot)\right) = +\infty$.

Theorem 26.1 *Let the conflict controlled process (26.1), (26.3)) with the initial condition (26.2) and commutative matrices A and B satisfy Condition 1, and let the set M be convex, for the given initial state $z^0(\cdot)$ and some selection $g^0(\cdot) \in G$ $T = T\left(z^0(\cdot),\ g^0(\cdot)\right) < +\infty$.*

Then a trajectory of the process (26.1), (26.3) can be brought by the pursuer from $z^0(\cdot)$ to the terminal set M^* at the moment T under arbitrary admissible controls of the evader.

Proof Let $v(\cdot) \in \Omega_V$. First consider the case when $\xi\left(T,\ z^0(\cdot),\ g^0(\cdot)\right) \notin M$. We introduce the controlling function

$$\begin{aligned} h(t) &= h\left(T,\ t,\ s,\ z^0(\cdot), v(\cdot),\ g^0(\cdot)\right) \\ &= 1 - \int_0^t \alpha\left(T,\ s,\ z^0(\cdot),\ v(s),\ g^0(\cdot)\right) ds, \quad t \geq 0. \end{aligned}$$

From the definition of time T, there exists a switching time $t_* = t_*(v(\cdot))$, $0 < t_* \leq T$, such that $h(t_*) = 0$.

Let us describe the rules by which the pursuer constructs his control on the so-called active and the passive parts, $[0\,,\ t_*)$ and $[t_*\,,\ T]$, respectively.

Consider the set-valued mapping

$$U_1(s,v)=\Big\{u\in U\ :\ \pi F\left(T-\tau-s\right)\phi\left(u,\ v\right)-g^0\left(T-\tau-s\right)$$
$$\in\alpha\left(T,\ s,\ z^0\left(\cdot\right),v\left(s\right),g^0\left(\cdot\right)\right)\left[M-\xi\left(T,z^0\left(\cdot\right),g^0\left(\cdot\right)\right)\right]\Big\}.$$

From assumptions concerning the process (26.1), (26.3) parameters, with account of properties of the resolving function, it follows that the mapping $U_1\left(s,\ v\right)$ is a Borel measurable function on the set $[0,\ T]\times V$. Then selection

$$u_1\left(s,\ v\right)=\operatorname{lex}\min U_1\left(s,\ v\right)$$

appears as a jointly Borel measurable function in its variables (see [44]). The pursuer's control on the interval $[0,\ t_*)$ is constructed in the following form

$$u\left(s\right)=u_1\left(s,\ v\left(s\right)\right),$$

being superposition of Borel measurable functions it is also Borel measurable function (see [44]).

The pursuer's control on the interval $[0,\ t_*)$ is constructed in the following form

$$u\left(s\right)=u_1\left(s,\ v\left(s\right)\right),$$

being superposition of Borel measurable functions it is also Borel measurable function (see [44]).

Set

$$\alpha\left(T,\ s,\ z^0\left(\ \cdot\ \right),\ v\left(s\right),\ g^0\left(\ \cdot\ \right)\right)\equiv 0,\quad s\in[t_*,\ T].$$

Then the mapping

$$U_2\left(s,\ v\right)$$
$$=\left\{u\in U:\pi F\left(T-\tau-s\right)\phi\left(u,\ v\right)-g^0\left(T-\tau-s\right)=0\right\},\quad s\in[t_*,\ T],\ v\in V$$

is Borel measurable function in its variables, and its selection

$$u_2\left(s,\ v\right)=\operatorname{lex}\min U_2\left(s,\ v\right)$$

is Borel measurable function also.

On the interval $[t_*,\ T]$ we set the pursuer's control equal to

$$u(s) = u_2(s,\ v(s)) \ . \tag{26.6}$$

It is measurable function too (see [5, 9]).

Let $\xi\left(T,\ z^0(\cdot),\ g^0(\cdot)\right) \in M$. In this case, we choose the pursuer's control on the interval $[0,\ T]$ in the form (26.6).

Thus, the rules are defined, to which the pursuer should follow in constructing his control. We will now show that if the pursuer follows these rules in the course of the game, a trajectory of process (26.1) hits the terminal set at the time T under arbitrary admissible controls of the evader.

By virtue of Lemma 26.1, the Cauchy formula for the system (26.1) implies the representation

$$\begin{aligned} \pi z(T) = \pi F(T)\, a + \int_{-\tau}^{0} \pi F(T-\tau-s)\, b(s)\, ds \\ + \int_0^T \pi F(T-\tau-s)\, \phi(u(s),\ v(s))\, ds. \end{aligned} \tag{26.7}$$

First we examine the case when $\xi\left(T,\ z^0(\cdot),\ g^0(\cdot)\right) \notin M$.

By adding and subtracting from the right-hand side of Eq. (26.7) the value $\int_0^T g^0(T-\tau-s)\,\mathrm{d}s$, one can deduce

$$\begin{aligned} &\pi z(T) \\ &= \left[\pi F(T)\, a + \int_{-\tau}^{0} \pi F(T-\tau-s)\, b(s)\,\mathrm{d}s + \int_0^T g^0(T-\tau-s)\,\mathrm{d}s\right] \\ &\quad + \int_0^T \left[\pi F(T-\tau-s)\, \phi(u(s),\ v(s)) - g^0(T-\tau-s)\right]\mathrm{d}s \\ &\quad \in \xi\left(T,\ z^0(\cdot),\ g^0(\cdot)\right) + \\ &\int_0^T \alpha\left(T,\ s,\ z^0(\cdot),\ v,\ g^0(\cdot)\right)[M - \xi\left(T,\ z^0(\cdot),\ g^0(\cdot)\right)]\mathrm{d}s \\ &= \xi\left(T,\ z^0(\cdot),\ g^0(\cdot)\right) + \int_0^T \alpha\left(T,\ s,\ z^0(\cdot),\ v,\ g^0(\cdot)\right) M\,\mathrm{d}s \\ &\quad - \int_0^T \alpha\left(T,\ s,\ z^0(\cdot),\ v,\ g^0(\cdot)\right) \xi\left(T,\ z^0(\cdot),\ g^0(\cdot)\right)\mathrm{d}s. \end{aligned} \tag{26.8}$$

By virtue (26.8) and $\alpha\left(T,\ s,\ z^0(\cdot),\ v(s),\ g^0(\cdot)\right)=0, s\in[t_*,\ T]$ we have the inclusion

$$\pi z(T)\in\xi\left(T,\ z^0(\cdot),\ g^0(\cdot)\right)\left[1-\int_0^{t_*}\alpha\left(T,\ s,\ z^0(\cdot),\ v(s),\ g^0(\cdot)\right)\mathrm{d}s\right]$$
$$+\int_0^{t_*}\alpha\left(T,\ s,\ z^0(\cdot),\ v(s),\ g^0(\cdot)\right)M\mathrm{d}s.$$

Since $\int_0^{t_*}\alpha\left(T,\ s,\ z^0(\cdot),\ v(s),\ g^0(\cdot)\right)\mathrm{d}s=1$ and the set M is convex then $\pi z(T)\in M$. Then, applying the rule of the pursuer control for the case when $\xi\left(T,\ z^0(\cdot),\ g^0(\cdot)\right)\in M$, we obtain the inclusion $\pi z(T)\in M$. The proof is therefore complete.

Corollary 26.1 *Assume that the differential-difference game of approach (26.1), (26.3) is linear* $(\phi(u,\ v)=u-v)$, *matrices A and B are commutative, Condition 1 holds, there exists a continuous positive function* $r(t),\ r:\mathbb{R}\to\mathbb{R}$, *and a number* $l\geq 0$ *such that* $\pi F(t)U=r(t)S,\ M=lS$, *where S is the unit ball centered at zero in the subspace L.*

Then when $\xi\left(t,\ z^0(\cdot),\ g(\cdot)\right)\notin lS$, *the resolving function (26.4) is the largest root of the quadratic equation for* $\alpha>0$

$$\left\|\pi F(t-\tau-s)v+g(t-\tau-s)-\alpha\xi\left(t,\ z^0(\cdot),\ g(\cdot)\right)\right\|=$$
$$=r(t-\tau-s)+\alpha l. \tag{26.9}$$

Proof By virtue of the assumptions of Corollary 26.1, we conclude from expression (26.4) that the resolving function $\alpha\left(T,\ s,\ z^0(\cdot),\ v,\ g(\cdot)\right)$ for fixed values of its arguments is the maximal number α such that

$$[r(t-\tau-s)S-\pi F(t-\tau-s)v-g(t-\tau-s)]\cap$$
$$\alpha\left[lS-\xi\left(t,\ z^0(\cdot),\ g(\cdot)\right)\right]\neq\emptyset.$$

The last expression is equivalent to the inclusion

$$\pi F(t-\tau-s)v+g(t-\tau-s)-\alpha\xi\left(t,\ z^0(\cdot),\ g(\cdot)\right)\in$$
$$[r(t-\tau-s)+\alpha l)]S.$$

Due to the linearity of the left-hand side of this inclusion in α, the vector $\pi F(t-\tau-s)v+g(t-\tau-s)-\alpha\xi\left(t,\ z^0(\cdot),\ g(\cdot)\right)$ lies on the boundary of the ball $[r(t-\tau-s)+\alpha ql]S$ for the maximal value of α. In other words, the length of this vector is equal to the radius of this ball that is demonstrated by (26.9). The proof is complete.

26.3 Differential-Difference Games of Approach with Pure Time Delay

We consider the problem of approach, which is described by the system of differential-difference equations with pure time delay (see [38, 43, 44])

$$\dot{z}(t) = Bz(t-\tau) + \phi(u,\ v), \quad z \in \mathbb{R}^n,\ \ u \in U,\ \ v \in V,\ \ t \geq 0, \tag{26.10}$$

with the initial condition (26.2).

Lemma 26.2 (See [41]) *Let $z(t)$ be a continuous solution to the system (26.10) under the initial condition (26.2). Then,*

$$z(t) = exp_\tau\{B,t\}z^0(-\tau) + \int_{-\tau}^{0} exp_\tau\{B,t-\tau-s\}\dot{z}^0(s)\,ds$$
$$+ \int_0^t exp_\tau\{B,t-\tau-s\}\phi(u(s),v(s))\,ds.$$

The terminal set has the cylindrical form (26.3). Function

$$u(t) = u\left(z^0(\cdot),\ t,\ v(t)\right),$$

such that $v(\cdot) \in \Omega_V$ implies $u(\cdot) \in \Omega_U$ is called countercontrol stroboscopic strategy of Hajek (see [39]) of pursuer corresponding to initial state $z^0(\cdot)$. The game is evolving on the closed time interval $[0,\ T]$. We assume that the pursuer chooses his control in the form

$$u(t) = \begin{cases} u_1\left(z^0(\cdot),\ t,\ v(t)\right), & t \in [0,\ t_*); \\ u_2\left(z^0(\cdot),\ t,\ v(t)\right), & t \in [t_*,\ T], \end{cases}$$

where $[0,\ t_*)$ is the active interval time, $[t_*,\ T]$ is the passive one, and $t_* = t_*(v(\cdot))$ is the moment of switching from the Method of Resolving Functions in first interval time to the First Direct Method of L.S. Pontryagin in the second one.

We introduce set-valued mappings

$$\bar{W}(t,\ v) = \pi exp_\tau\{B,t\}\phi(U,\ v),$$
$$\bar{W}(t) = \bigcap_{v \in V} \bar{W}(t,\ v),$$

Condition 2 The mapping $\bar{W}(t) \neq \emptyset$ for all $t \geq 0$.

The mapping $\bar{W}$ is upper semi-continuous and therefore Borel measurable function (see [40, 41]). Hence, there exists at least one Borelian selection $g(t)$, $g(t) \in \bar{W}(t)$ (see [42]). Denote by $G = \{g(t) : g(t) \in \bar{W}(t), \quad t \geq 0\}$ the set of all Borelian selections of the set-valued mapping $\bar{W}(t)$. For fixed $g(\cdot) \in G$ we put

$$\xi\left(t, z^0(\cdot), g(\cdot)\right) =$$

$$= \pi exp_\tau\{B, t\}z^0(-\tau) + \int_{-\tau}^{0} \pi exp_\tau\{B, t-\tau-s\}\dot{z}^0(s)\mathrm{d}s + \int_0^t g(s)\mathrm{d}s,$$

and consider the resolving function

$$\alpha\left(t, s, z^0(\cdot), v, g(\cdot)\right) = \sup\{\alpha \geq 0 :$$

$$\left[\bar{W}(t-\tau-s, v) - g(t-\tau-s)\right] \cap \alpha\left[M - \xi\left(t, z^0(\cdot), g(\cdot)\right)\right] \neq \emptyset\}. \tag{26.11}$$

The function $\alpha\left(t, s, z^0(\cdot), v, g(\cdot)\right)$ is summable for $s \in [0, t]$ (see [5]).

We introduce the function (26.5). The value $T = T\left(z^0(\cdot), g(\cdot)\right)$ for the initial state $z^0(\cdot)$ of the system (26.10) and some selector $g^0(\cdot) \in G$ is the guaranteed moment of capture by the pursuer of the evader according to the Method of Resolving Functions.

On the other hand, we set

$$P\left(z^0(\cdot), g(\cdot)\right)$$

$$= \min\left\{t \geq 0 : \pi exp_\tau\{B, t\}z^0(-\tau) + \int_{-\tau}^{0} \pi exp_\tau\{B, t-\tau-s\}\dot{z}^0(s)\mathrm{d}s\right.$$

$$\left. \in M - \int_0^t \bar{W}(t-\tau-s)\mathrm{d}s\right\}. \tag{26.12}$$

Let us show that the quantity (26.3) is the guaranteed moment of the end of the game of approach according to the First Direct Method of L.S. Pontryagin (see [45]).

Theorem 26.2 *Let the conflict controlled process (26.10), (26.3) with the initial condition (26.2) satisfy Condition 2, the set M be convex, $P\left(z^0(\cdot)\right) < +\infty$, when $P\left(z^0(\cdot)\right)$ is defined by formula (26.3).*

Then a trajectory of the process (26.10), (26.3) can be brought by the pursuer from $z^0(\cdot)$ to the terminal set M^* at the moment $P\left(z^0(\cdot)\right)$.

Proof For simplicity of presentation, denote $P_0 = P\left(z^0(\cdot)\right)$. We have the following inclusion

$$\pi exp_\tau\{B, P_0\}z^0(-\tau) + \int_{-\tau}^{0} \pi exp_\tau\{B, P_0 - \tau - s\}\dot{z}^0(s)\mathrm{d}s$$
$$\in M - \int_0^{P_0} \bar{W}(P_0 - \tau - s)\mathrm{d}s.$$

Since, there exist point $m \in M$ and selection $g(\cdot) \in G$ such that

$$\pi exp_\tau\{B, P_0\}z^0(-\tau) + \int_{-\tau}^{0} \pi exp_\tau\{B, P_0 - \tau - s\}\dot{z}^0(s)\mathrm{d}s$$
$$= m - \int_0^{P_0} g(P_0 - \tau - s)\mathrm{d}s.$$

Consider the set-valued mapping

$$\begin{aligned} U(s, v) = \{u \in U &: \pi exp_\tau\{B, P_0 - \tau - s\}\phi(u, v) \\ &- g(P_0 - \tau - s) = 0\}, \quad s \in [0, P_0], \quad v \in V. \end{aligned} \tag{26.13}$$

The mapping $U(s, v)$ and selection $u(s, v) = \operatorname{lex}\min U(s, v)$ are Borel measurable functions in its variables.

We set the pursuers control equal to

$$u(s) = u(s, v(s)), \quad s \in [0, P_0],$$

where $v(s), v(s) \in V$, is an arbitrary admissible control of the evader, and it will be a Borel measurable function of time.

From the relation (26.13) with (26.3) we obtain

$$\pi z(P_0) = \pi exp_\tau\{B, P_0\}z^0(-\tau) + \int_{-\tau}^{0} \pi exp_\tau\{B, P_0 - \tau - s\}\dot{z}^0(s)\mathrm{d}s$$
$$+ \int_0^{P_0} \pi exp_\tau\{B, P_0 - \tau - s\}\phi(u(s), v(s))\mathrm{d}s = m \in M.$$

This means that $z(P_0) \in M^*$. The proof is therefore complete.

Theorem 26.3 *Let the conflict controlled process (26.10), (26.3) with the initial condition (26.2) satisfy Condition 2.*

Then the inclusion

$$\pi exp_\tau\{B,t\}z^0(-\tau)+\int_{-\tau}^{0}\pi exp_\tau\{B,t-\tau-s\}\dot{z}^0(s)ds$$
$$\in M-\int_0^t \bar{W}(t-\tau-s)ds,\quad t\geq 0,$$

holds if and only if a selection $g(\cdot)\in G$ *exists, such that* $\xi\left(t,\ z^0(\cdot),\ g(\cdot)\right)\in M$.

Proof Letting

$$\pi exp_\tau\{B,t\}z^0(-\tau)+\int_{-\tau}^{0}\pi exp_\tau\{B,t-\tau-s\}\dot{z}^0(s)\mathrm{d}s$$
$$\in M-\int_0^t \bar{W}(t-\tau-s)\mathrm{d}s.$$

There exist point $m\in M$ and selection $g(\cdot)\in G$ such that

$$\pi exp_\tau\{B,t\}z^0(-\tau)+\int_{-\tau}^{0}\pi exp_\tau\{B,t-\tau-s\}\dot{z}^0(s)\mathrm{d}s$$
$$=m-\int_0^t g(t-\tau-s)\mathrm{d}s,$$

which is equivalent to $\xi\left(t,\ z^0(\cdot)\ \ g(\cdot)\right)=m\in M$.

Using the reverse line of reasoning we come to the required result. The proof is therefore complete.

Theorem 26.4 *Let the conflict controlled process (26.10), (26.3) with the initial condition (26.2) satisfy Condition 2, and let the set M be convex, for the given initial state* $z^0(\cdot)$ *and some selection* $g^0(\cdot)\in G$ $T=T\left(z^0(\cdot),\ g^0(\cdot)\right)<+\infty$.

Then a trajectory of the process (26.10), (26.3) can be brought by the pursuer from $z^0(\cdot)$ *to the terminal set* M^* *at the moment* T.

Proof Let $v(s),\ v(s)\in V,\ s\in[0,\ T]$ be an arbitrary Borel measurable function. First, consider the case when $\xi\left(T,\ z^0(\cdot),\ g^0(\cdot)\right)\notin M$. We introduce the controlling function

$$h(t)==1-\int_0^t \alpha\left(T,\ s,\ z^0(\cdot),\ v(s),\ g^0(\cdot)\right)ds,\quad t\geq 0.$$

From the definition of time T, there exists a switching time $t_*=t_*(v(\cdot)),\quad 0<t_*\leq T$, such that $h(t_*)=0$.

Let us describe the rules by which the pursuer constructs his control on the so-called active and the passive parts, $[0,\ t_*)$ and $[t_*,\ T]$, respectively.

Consider the set-valued mapping

$$U_1(s,v)=\Big\{u\in U\,:\,\pi exp_\tau\{B,T-\tau-s\}\phi\,(u,\,v)-g^0\,(T-\tau-s)$$
$$\in\alpha\left(T,\,s,\,z^0\,(\cdot\,),v\,(s),g^0\,(\cdot\,)\right)\left[M-\xi\left(T,z^0\,(\cdot\,),g^0\,(\cdot\,)\right)\right]\Big\}.$$

It follows from assumptions concerning the process (26.10), (26.3) parameters, with account of properties of the resolving function, that the mapping $U_1\,(s,\,v)$ is a Borel measurable function on the set $[0,\,T]\times V$. Then selection

$$u_1\,(s,\,v)=\operatorname{lex}\min U_1\,(s,\,v)$$

appears as a jointly Borel measurable function in its variables (see [41]).

The pursuer's control on the interval $[0,\,t_*)$ is constructed in the following form

$$u\,(s)=u_1\,(s,\,v\,(s))\,,$$

being a superposition of Borel measurable functions it is also Borel measurable function (see [42]).

Set

$$\alpha\left(T,\,s,\,z^0\,(\,\cdot\,),\,v\,(s),\,g^0\,(\,\cdot\,)\right)\equiv 0,\quad s\in[t_*,\,T]\,.$$

Then the mapping

$$U_2\,(s,\,v)$$
$$=\left\{u\in U:\pi exp_\tau\{B,T-\tau-s\}\phi\,(u,\,v)-g^0\,(T-\tau-s)=0\right\},\quad s\in[t_*,\,T]\,,\,v\in V$$

is Borel measurable function in its variables, and its selection

$$u_2\,(s,\,v)=\operatorname{lex}\min U_2\,(s,\,v)$$

is Borel measurable function as well.

On the interval $[t_*\,,\,T]$ we set the pursuer's control equal to

$$u\,(s)=u_2\,(s\,,\,v\,(s))\,. \tag{26.14}$$

It is measurable function too.

Let $\xi\left(T,\,z^0\,(\,\cdot\,),\,g^0\,(\,\cdot\,)\right)\in M$. In this case, we choose the pursuer's control on the interval $[0,\,T]$ in the form (26.14).

Thus, the rules are defined, to which the pursuer should follow in constructing his control. We will now show that if the pursuer follows these rules in the course

of the game, a trajectory of process (26.10) hits the terminal set at the time T under arbitrary admissible controls of the evader.

By virtue of Lemma 26.2, the Cauchy formula for the system (26.10) implies the representation

$$\pi z(T) = \pi exp_\tau\{B, T\}z^0(-\tau) + \int_{-\tau}^{0} \pi exp_\tau\{B, T-\tau-s\}\dot{z}^0(s)\mathrm{d}s + \int_0^T \pi exp_\tau\{B, T-\tau-s\}\phi(u(s), v(s))\mathrm{d}s. \tag{26.15}$$

First, we examine the case when $\xi\left(T, z^0(\cdot), g^0(\cdot)\right) \notin M$.

By adding and subtracting from the right-hand side of Eq. (26.15) the value $\int_0^T g^0(T-\tau-s)\,\mathrm{d}s$, one can deduce

$$\pi z(T) \in \xi\left(T, z^0(\cdot), g^0(\cdot)\right)\left[1 - \int_0^{t_*} \alpha\left(T, s, z^0(\cdot), v(s), g^0(\cdot)\right)\mathrm{d}s\right] + \int_0^{t_*} \alpha\left(T, s, z^0(\cdot), v(s), g^0(\cdot)\right) M\mathrm{d}s.$$

Since $\int_0^{t_*} \alpha\left(T, s, z^0(\cdot), v(s), g^0(\cdot)\right)\mathrm{d}s = 1$ and the set M is convex then $\pi z(T) \in M$. Then, applying the rule of the pursuer control for the case when $\xi\left(T, z^0(\cdot), g^0(\cdot)\right) \in M$, we obtain the inclusion $\pi z(T) \in M$. The proof is therefore complete.

Corollary 26.2 *Let the conflict-controlled process (26.10), (26.3) with the initial condition (26.2) satisfy Condition 2.*

Then for any initial state $z^0(\cdot)$ there exists a selection $g^0(\cdot) \in G$ such that

$$T\left(z^0(\cdot), g^0(\cdot)\right) \le P\left(z^0(\cdot)\right).$$

The effectiveness of the Method of Resolving Functions, sufficient conditions that are easily verified, the ability to quickly build the resolution function, using the modern techniques of set-valued mappings and their selections, prove the relevance of this method for solving differential-difference games that are of great practical importance.

Acknowledgements The author is grateful to Academician Zgurovsky M.Z. for the possibility of the publication and to professor Kasyanov P.O. for assistance in publication this article.

References

1. Isaacs, R.: Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization. Wiley, New York (1965)
2. Pontryagin, L.S.: Selected Scientific Works, vol. 2. Nauka, Moscow (1988)
3. Pshenitchnyi, B.N., Ostapenko, V.V.: Differential Games. Naukova Dumka, Kyiv (1992)
4. Krasovskii, N.N.: Game-Theoretical Control Problems. Springer, New York (1988). https://doi.org/10.1007/978-1-4612-3716-7
5. Chikrii, A.A.: Conflict-Controlled Processes. Springer, Dordrecht (2013)
6. Filippov, A.F.: Differential Equations with Discontinuous Right-Hand Sides. Nauka, Moscow (1985)
7. Chikrii, A.A.: An analytical method in dynamic pursuit games. Proc. Steklov Inst. Math. **271**(1), 69–85 (2010)
8. Chikrii, A.A., Eidelman, S.D.: Generalized Mittag-Leffler matrix functions in game problems for evolutionary equations of fractional order. Cybern. Syst. Anal. **36**(3), 315–338 (2000)
9. Chikrii, A.A., Kalashnikova, S.F.: Pursuit of a group of evaders by a single controlled object. Cybernetics **23**(4), 437–445 (1987)
10. Albus, J., Meystel, A., Chikrii, A.A., Belousov, A.A., Kozlov, A.J.: Analytical method for solution of the game problem of softlanding for moving objects. Cybern. Syst. Anal. **37**(1), 75–91 (2001)
11. Baranovskaya, L.V., Chikrii, A.A., Chikrii, Al.A.: Inverse Minkowski functional in a nonstationary problem of group pursuit. J. Comput. Syst. Sci. Int. **36**(1), 101–106 (1997)
12. Chikrii, Al.A.: On nonstationary game problem of motion control. J. Autom. Inf. Sci. **47**(11), 74–83 (2015). https://doi.org/10.1615/JAutomatInfScien.v47.i11.60
13. Pepelyaev, V.A., Chikrii, Al.A.: On the game dynamics problems for nonstationary controlled processes. J. Autom. Inf. Sci. **49**(3), 13–23 (2017). https://doi.org/10.1615/JAutomatInfScien.v49.i3.30
14. Kryvonos, I.Iu., Chikrii, Al.A., Chikrii, K.A.: On an approach scheme in nonstationary game problems. J. Autom. Inf. Sci. **45**(8), 41–58 (2013). https://doi.org/10.1615/JAutomatInfScien.v45.i8.40
15. Chikrii, A.A., Chikrii, G.Ts.: Matrix resolving functions in game problems of dynamics. Proc. Steklov Inst. Math. **291**(1), 56–65 (2015)
16. Chikriy, A.A., Baranovskaya, L.V., Chikriy, Al.A.: An approach game problem under the failure of controlling devices. J. Autom. Inf. Sci. **32**(5), 1–8 (2000). https://doi.org/10.1615/JAutomatInfScien.v32.i5.10
17. Baranovskaya, L.V., Chikriy, A.A.: Game problems for a class of Hereditaly systems. J. Autom. Inf. Sci. **29**(2–3), 87–97 (1997). https://doi.org/10.1615/JAutomatInfScien.v29.i2-3.120
18. Chikrii, A.A., Belousov, A.A.: On linear differential games with integral constrains. Trudy Instituta Matematiki i Mekhaniki UrO RAN **15**(4), 290–301 (2009)
19. Bigun, Ya.I., Kryvonos, I.Iu., Chikrii, Al.A., Chikrii, K.A.: Group approach under phase constraints. J. Autom. Inf. Sci. **46**(4), 1–8 (2014). https://doi.org/10.1615/JAutomatInfScien.v46.i4.10
20. Elsgolts, L.E., Norkin, S.B.: Differential Equations with Deviating Argument. Nauka, Moscow (1971)
21. Kolmanovskii, V.B., Richard, J.P.: Stability of some linear systems with delays. J. IEEE Trans. Autom. Control. **44**(5), 984–989 (1999)
22. Baranovskaya, L.V., Baranovskaya, G.G.: On differential-difference group pursuit game. Dopov. Akad. Nauk Ukr. **3**, 12–15 (1997)
23. Baranovskaya, G.G., Baranovskaya, L.V.: Group pursuit in quasilinear differential-difference games. J. Autom. Inf. Sci. **29**(1), 55–62 (1997). https://doi.org/10.1615/JAutomatInfScien.v29.i1.70

24. Baranovskaya, L.V., Chikrii, Al.A.: On one class of difference-differential group pursuit games. In: Multiple Criteria and Game Problems under Uncertainty. Proceedings of the Fourth International Workshop, September 1996, Moscow, vol. 11, pp. 814 (1996)
25. Chikrii, A.A., Baranovskaya, L.V.: A type of controlled system with delay. Cybern. Comput. Technol. **107**, 1–8 (1998)
26. Baranovskaya, L.V.: About one class of difference games of group rapprochement with unfixed time. Sci. World **1**(2(18)), 10–12 (2015)
27. Baranovska, L.V.: The group pursuit differential-difference games of approach with non-fixed time. Naukovi Visti NTUU KPI **4**, 18–22 (2011)
28. Liubarshchuk, I.A., Bihun, Ya.I., Cherevko, I.M.: Game problem for systems with time-varying delay. Problemy Upravleniya I Informatiki **2**, 79–90 (2016)
29. Baranovskaya, L.V.: A method of resolving functions for one class of pursuit problems. East.-Eur. J. Enterp. Technol. **74**(4), 4–8 (2015). https://doi.org/10.15587/1729-4061.2015.39355
30. Kyrychenko, N.F., Baranovskaya, L.V., Chyckrij, Al.A.: On the class of linear differential-difference games of pursuit. Dopov. Akad. Nauk Ukr. **6**, 24–26 (1997)
31. Baranovska, L.V.: On quasilinear differential-difference games of approach. Problemy upravleniya i informatiki **4**, 5–18 (2017)
32. Baranovska, L.V.: The modification of the method of resolving functions for the difference-differential pursuit's games. Naukovi Visti NTUU KPI **4**, 1420 (2012)
33. Baranovskaya, L.V.: Method of resolving functions for the differential-difference pursuit game for different-inertia objects. Adv. Dyn. Syst. Control. **69**, 159–176 (2016). https://doi.org/10.1007/978-3-319-40673-2
34. Bellman, R., Cooke, K.L.: Differential-Difference Equations. Academic, Cambridge (1963)
35. Osipov, Yu.S.: Differential games of systems with delay. Dokl. Akad. Nauk **196**(4), 779–782 (1971)
36. Khusainov, D.Ya., Benditkis, D.D., Diblik, J.: Weak delay in systems with an aftereffect. Funct. Differ. Equ. **9**(3–4), 385–404 (2002)
37. Diblk, J., Morvkov, B., Khusainov, D., Kukharenko, A.: Delayed exponential functions and their application to representations of solutions of linear equations with constant coefficients and with single delay. In: Proceedings of the 2nd International Conference on Mathematical Models for Engineering Science, and Proceedings of the 2nd International Conference on Development, Energy, Environment, Economics, and Proceedings of the 2nd International Conference on Communication and Management in Technological Innovation and Academic Globalization, 10–12 December 2011, pp. 82–87 (2011)
38. Khusainov, D.Ya., Diblik, J., Ruzhichkova, M.: Linear dynamical systems with aftereffect. In: Representation of Decisions, Stability, Control, Stabilization. GP Inform-Analytics Agency, Kiev (2015)
39. Hajek, O.: Pursuit Games: An Introduction to the Theory and Applications of Differential Games of Pursuit and Evasion. Dover Publications, New York (2008)
40. Pshenichnyi, B.N.: Convex Analysis and Extreme Challenges. Nauka, Moscow (1980)
41. Joffe, A.D., Tikhomirov, V.: Theory of Extremal Problems. North Holland, Amsterdam (1979)
42. Aubin, J-P., Frankovska, He.: Set-Valued Analysis. Birkhause, Boston (1990)
43. Diblk, J., Khusainov, D.Y.: Representation of solutions of linear discrete systems with constant coefficients and pure delay. Adv. Differ. Equ., 1687–1847 (2006). https://doi.org/10.1155/ADE/2006/80825
44. Baranovska, L.V.: Method of resolving functions for the pursuit game with a pure time-lag. In: System Analysis and Information Technology: 19th International Conference SAIT 2017, Kyiv, Ukraine, 22–25 May 2017. Proceedings ESC IASA NTUU Igor Sikorsky Kyiv Polytechnic Institute, p. 18 (2017)
45. Nikolskii, M.S.: L.S. Pontryagins First Direct Method in Differential Games. Izdat. Lomonosov Moscow State University (Izdat. Gos. Univ.), Moscow (1984)

Chapter 27
The Problem of a Function Maximization on a Type-2 Fuzzy Set

S. O. Mashchenko and D. O. Kapustian

Abstract The article focuses on generalizing the concept of the maximizing alternative in the case of the objective function maximization problem on the type-2 fuzzy set (T2FS) of feasible alternatives. An extension of the natural order relation to the class of fuzzy sets is used for comparison of fuzzy sets of alternatives membership degrees. It is shown that such a fuzzy preference relation provides fuzzy sets of membership degrees of T2FSs of feasible alternatives to be normal. With the help of this preference relation a fuzzy set of non-dominated alternatives is constructed. The notion of α-level non-dominated alternative is introduced. It is shown that this is a solution to the optimization problem. In this problem the objective function is maximized with a bounded secondary membership degree of the T2FS of feasible alternatives. The problem of choosing alternatives according to the two criteria (the objective function and secondary degrees of membership to the T2FS of feasible alternatives) is formulated. Its Pareto optimal solutions are called the effective maximizing alternatives. Their properties are investigated.

27.1 Introduction

The problem of maximizing a given non-fuzzy function on a given fuzzy set (type-1) of feasible elements has been considered by Negoita and Ralescu [1]. These authors interpreted the function to be maximized as a membership function of some fuzzy goal set. A significant contribution to solve this problem was made by Orlovsky [2]. He proposed two concepts for solving this problem and investigated their equivalence. The first solution concept makes use of levels sets of a fuzzy set. The second one is based on the idea of representing the support of a fuzzy set of solutions as a set of Pareto optimal solutions to a two-criteria problem. It maximizes both the objective function and the membership function of the fuzzy set

S. O. Mashchenko (✉) · D. O. Kapustian
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*,
Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_27

of feasible alternatives. Models and methods of fuzzy mathematical programming are sufficiently well-developed [3–5].

Although fuzzy sets are the main tool for formalizing of uncertainty in describing the decision-making problem under the Bellman-Zadeh approach, they can also be vague. Mendel and John [6] emphasize that there are such sources of uncertainty in fuzzy sets. These are: the meanings of words that are used in the description of fuzzy sets; an ambiguity of the opinions of various experts; a measurement noise; a data noise. All these problems lead to uncertainty of the fuzzy set membership function.

The main instrument of description such uncertainty is fuzzy set type-2 (T2FS) theory, which introduced by Zadeh in 1971 [7]. The degree of membership of elements in a usual fuzzy set is given by a value on the interval [0, 1], whereas the degree of membership of elements in a T2FS is a fuzzy set on [0, 1]. It can be seen that, mathematically, a T2FS A is a mapping $A : X \to [0,1]^{[0,1]}$ (see [7]). Mendel and John [6] provide the following definition which is based on the ideas invented by Karnik and Mendel [8].

A T2FS A is characterized by a type-2 membership function (T2MF) $\tilde{\mu}_A(x, y)$, where $x \in X$ and $y \in Y(x) \subseteq [0, 1]$, that is,

$$A = \{(x, \tilde{\mu}_A(x, y)) : \ x \in X, \ y \in Y(x) \subseteq [0, 1]\}.$$

Sometimes it is convenient to use this definition in combination with remarks of Harding et al. [9] and Aisbett et al. [10] in which the notion of T2FS is characterized by a T2MF

$$\mu_A(x, y) = \begin{cases} \tilde{\mu}_A(x, y), & y \in Y(x); \\ 0, & \text{elsewhere}, \end{cases}$$

$x \in X$ which is expanded on $y \in Y(x) \subseteq [0, 1]$. Consequently,

$$A = \{(x, \tilde{\mu}_A(x, y)) : \ x \in X, \ y \in Y \subseteq [0, 1]\}.$$

The main lack of T2FSs is some difficulties of their understanding and using. Despite these problems, T2FSs are widely used in decision making theory [11–14].

27.2 Formulation of the Problem

When solving the function maximizing problem on a fuzzy set F of feasible alternatives of universal set X of alternatives, the main idea of the approach of Negoita and Ralescu [1] consisted of its formulation as a problem of achieving a fuzzy goal set G, membership function of which was specified by the maximizing function. Further, to solve this, the Bellman-Zade concept was used [15]. According to this approach the goal of decision making G and the set of feasible alternatives

F are considered as equitable fuzzy sets. This makes it possible to fairly simply determine the solution set D. The "optimal" solution is defined as any alternative, which maximizes membership function of the fuzzy set D and is named as maximizing alternative. We apply this idea when set of solutions is T2FS.

Let X be a universal set of alternatives. Denote by the F T2FS of feasible alternatives in X which is characterized by a T2MF $\mu_F(x, y)$, $x \in X$, $y \in Y \subseteq [0, 1]$. Let $g(x)$ is function $X \to R$. We suppose $g(x) \in [0, 1]$ for $\forall x \in X$ and maximize it on the T2FS F of feasible alternatives. We formulate the problem to achieve fuzzy goal. Let this goal is the fuzzy set G_0 in X with membership function $\mu_{G_0}(x) = g(x)$, $x \in X$. This implies that better membership degrees $\mu_{G_0}(x)$ to the fuzzy set G_0 correspond to better values of the function $g(x)$, which we want to maximize. The intersection of F and G_0 we define as the fuzzy solution set D. For this end we according to [6] represent fuzzy goal set G_0 in the form of T2FS G with T2MF

$$\mu_G(x, y) = \begin{cases} 1, & y = g(x), \\ 0, & \text{elsewhere}, \end{cases} \tag{27.1}$$

$x \in X$, $y \in [0, 1]$. We define the fuzzy set D of solution as the T2FS resulting from the intersection of F and G. According to [6]

$$\mu_D(x, y) = \max_{u,v \in [0,1],\ \min\{u,v\}=y} \min\{\mu_F(x, u), \mu_G(x, v)\},\ x \in X,\ y \in Y \subseteq [0, 1] \tag{27.2}$$

is its T2MF. Which alternative in X can be chosen as maximizing?

Consider an example.

Example 1 Suppose that T2FS of feasible alternatives F is given on the universal set $X = [0, 12]$. Its T2MF $\mu_F(x, y)$, $x \in X$, $y \in Y \subseteq [0, 1]$ takes values in the set $\{0, 5; 1\}$. Let $g(x) = x/12$ is objective function, which is maximizing. The goal T2FS G with T2MF $\mu_G(x, y)$ in the form (1) corresponds to the objective function $g(x)$. Figure 27.1 shows the 0, 5-level lines (thin lines) and 1-level lines (thick lines) of the T2MFs of feasible alternatives F (solid lines) and goal G (dashed lines). A fuzzy solution to this problem is the T2FS $D = F \bigcap G$ whose T2MF has 0, 5-level lines (thin lines) and 1-level lines (thick lines). They are shown on Fig. 27.2. On this figure we see that the alternative $x = 6$ has the maximal primary degree of membership to the T2FS of solutions which is equal to 0, 5 with secondary grade 0, 5; but the alternative $x = 5, 25$ has the maximal primary degree of membership which is equal to $7/16$ with secondary grade 1. Which alternative should be chosen? First, second or . . . ?

The sets $FY_D(x) = \bigcup_{y \in Y(x)}(y, \mu_D(x, y))$ of the membership degrees of the alternatives $x \in X$ to the T2FS D of solutions are fuzzy. For example, we have single point fuzzy sets $FY_D(6) = \{(0, 5; 0, 5)\}$ for the alternative $x = 6$ and $FY_D(5, 25) = \{(7/16; 1)\}$ for the alternative $x = 5, 25$. Therefore, to solve the problem one needs to learn how to compare the alternatives. Then the alternative

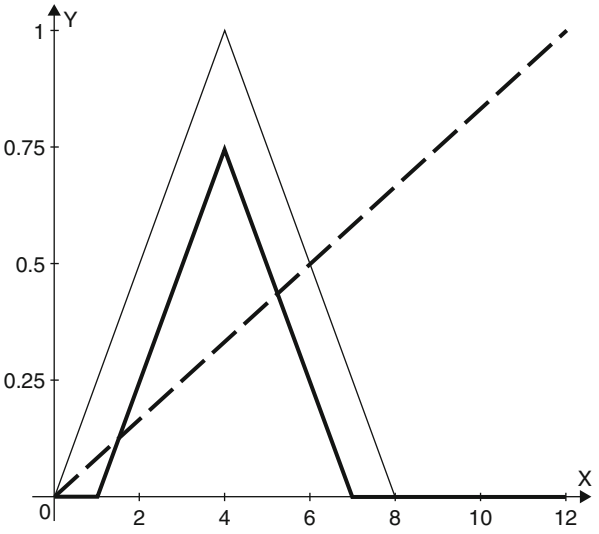


Fig. 27.1 The levels lines of the F and G T2MFs

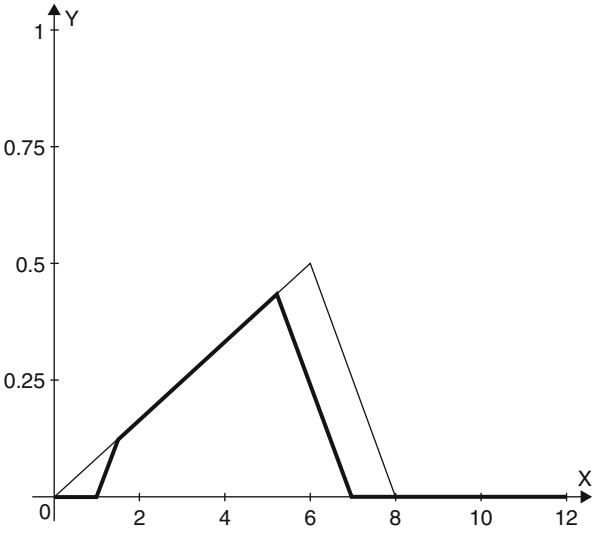


Fig. 27.2 The levels lines of the D T2MF

which corresponds to the 'best' fuzzy set of membership degrees can be thought of as a solution to the formulated problem of choosing the maximizing alternative.

In the case, where fuzzy sets of membership degrees of T2FS can be represented as fuzzy numbers, it is advisable to use known ranking methods that have been extensively investigated in the last few years because of their applicability in classical fuzzy optimization. There are many approaches to compare fuzzy numbers [15–20]. An excellent review is presented in [21]. A large number of methods for comparing fuzzy numbers can be explained by the wide scope of their applications.

Nevertheless, sometimes fuzzy sets of the T2FS membership degrees cannot be represented as fuzzy numbers. This is, for instance, the case when membership degrees have discrete values as in Example 1. Also, for example, in [22, 23] such T2FSs are obtained as the result of union and intersection operations of fuzzy sets with fuzzy number of operands. In view of this in the present article we focus on the more general case of T2FSs.

27.3 Preliminaries

27.3.1 A Fuzzy Preference Relation

In this section, we give some definitions from [24] which we will use further.

A fuzzy preference relation (FPR) R on the set X is a fuzzy subset of the product set $X \times X$ with the membership function $\eta_R : X \times X \to [0, 1]$. It is assumed that the FPR is reflexive, that is $\eta_R(x, x) = 1 \; \forall x \in X$.

For a given FPR R we define two fuzzy relations:

– fuzzy indifference relation $I = R \bigcap R^{-1}$ with the membership function

$$\eta_I(x, y) = \min\{\eta_R(x, y), \eta_R(y, x)\}\,, \; x, y \in X;$$

– fuzzy strict preference relation $S = R \backslash R^{-1}$ with the membership function

$$\eta_S(x, y) = \max\{0, \eta_R(x, y) - \eta_R(y, x)\}\,, \; x, y \in X.$$

For FPR R on a given set X we introduce the fuzzy set ND of non-dominated alternatives with the membership function

$$\begin{gathered}\eta_{ND}(x) = \min_{y \in X}\{1 - \eta_S(x, y)\} = 1 - \max_{y \in X} \eta_S(x, y) = \\ 1 - \max_{y \in X}\{\eta_R(x, y) - \eta_R(y, x)\}.\end{gathered} \tag{27.3}$$

By the value $\eta_{ND}(x)$ we mean the degree with which alternative x is not dominated by any one of the elements of the set X.

27.3.2 Extension of a Fuzzy Preference Relation to the Class of Fuzzy Sets

Formally this problem can be stated as follows. Let $\mu_F : Y \times Y \to [0, 1]$ be a membership function of FPR F on a given set Y and Φ is a class of fuzzy sets of Y. What is the form of the FPR F on the class Φ?

To obtain the desired FPR we argue as follows [25]. If a fuzzy set A belongs to Φ and has the membership function $\lambda_A(y)$, $y \in Y$, then the value

$$\eta(A, y') = \max_{y \in Y} \min\{\lambda_A(y), \mu_F(y, y')\}$$

can be taken as a degree to which the fuzzy set A is preferred to the element $y' \in Y$. Similarly, the value

$$\eta(y', A) = \max_{y \in Y} \min\{\lambda_A(y), \mu_F(y', y)\}$$

is a degree to which the reverse preference is true. Now if A' and A'' are two arbitrary fuzzy sets on Y, then the value

$$\eta(y', A) = \max_{y \in Y} \min\{\lambda_{A'}(y), \max_{z \in Y} \min\{\lambda_{A''}(z), \mu_F(y, z)\}\} =$$

$$\max_{y,z \in Y} \min\{\lambda_{A'}(y), \lambda_{A''}(z), \mu_F(y, z)\}$$

is a degree of the preference $A' \tilde{F} A''$, where we write $\tilde{F}$ for the FPR on Φ with the membership function η. Such a FPR $\tilde{F}$ defined on Φ is induced by the FPR F defined on the universal set Y.

When Y is the set of real numbers and F is the natural order "$\geq$" on Y, the induced FPR $\tilde{F}$ has the form

$$\eta(A', A'') = \max_{y,z \in Y,\ y \geq z} \min\{\lambda_{A'}(y), \lambda_{A''}(z)\}.$$

In a more particular case, when the class Φ is smaller this formula can be simplified. According to [25] for any two normal convex fuzzy sets A' and A'' on Y one of the equations holds

$$\eta(A', A'') = 1 \text{ or } \eta(A', A'') = \max_{y \in Y} \min\{\lambda_{A'}(y), \lambda_{A''}(y)\}.$$

It should be noted that this coincides with the result obtained in [26] for fuzzy numbers which are also normal and convex fuzzy sets. This indicates a certain universality of such approach.

27.4 Fuzzy Set of Non-dominated Alternatives

Let T2FS D be a solution to the maximization problem of function $g(x)$ on the T2FS F. Its T2MF $\mu_D(x, y)$, $x \in X$, $y \in Y \subseteq [0, 1]$ is given by (27.2). Firstly we

simplify formula (27.2). Formulas (27.1) and (27.2) imply that

$$\mu_D(x,y) = \max_{u,v\in[0,1],\ \min\{u,v\}=y} \min\{\mu_F(x,u), \mu_G(x,v)\} =$$

$$\max\{\max_{u,v\in[0,1],\ u\le v,\ u=y} \min\{\mu_F(x,u), \mu_G(x,v)\},$$

$$\max_{u,v\in[0,1],\ u>v,\ v=y} \min\{\mu_F(x,u), \mu_G(x,v)\}\} =$$

$$\max\{\max_{v\in[0,1],\ y\le v} \min\{\mu_F(x,y), \mu_G(x,v)\}, \max_{u\in[0,1],\ u>y} \min\{\mu_F(x,u), \mu_G(x,y)\}\} =$$

$$\begin{cases} \max\{\min\{\mu_F(x,y),0\}, \max_{u\in[0,1],\ u>y} \min\{\mu_F(x,u),0\}\}, & y > g(x), \\ \max\{\min\{\mu_F(x,y),1\}, \max_{u\in[0,1],\ u>y} \min\{\mu_F(x,u),1\}\}, & y = g(x), \\ \max\{\min\{\mu_F(x,y),1\}, \max_{u\in[0,1],\ u>y} \min\{\mu_F(x,u),0\}\}, & y < g(x), \end{cases}$$

$$= \begin{cases} 0, & y > g(x), \\ \max\{\mu_F(x,y), \max_{u\in[0,1],\ u>y} \mu_F(x,u)\}, & y = g(x), \\ \mu_F(x,y), & y < g(x). \end{cases}$$

Whence we finally get

$$\mu_D(x,y) = \begin{cases} 0, & y > g(x), \\ \max_{u\in[0,1],\ u\ge y} \mu_F(x,u)\}, & y = g(x), \\ \mu_F(x,y), & y < g(x). \end{cases} \tag{27.4}$$

The sets $FY_D(x) = \bigcup_{y\in Y}(y, \mu_D(x,y))$ of membership degrees of the corresponding alternatives $x \in X$ to the T2FS D of solutions are fuzzy. In view of this, to compare alternatives we use a FPR $\tilde{F}$ with the membership function $\eta(x', x'') = \eta(FY_D(x'), FY_D(x''))$. The FPR $\tilde{F}$ is an extension of the natural order "$\ge$" on the set of real numbers to the class of fuzzy sets which are defined for $y \in Y \subseteq [0,1]$. Recall from Sect. 27.3.2 that the membership function of the FPR $\tilde{F}$ takes the form

$$\eta(x', x'') = \max_{u,v\in Y,\ u\ge v} \min\{\mu_D(x',u), \mu_D(x'',v)\},\ x', x'' \in X. \tag{27.5}$$

We call a T2FS on X with T2MF $\varphi(x,y)$, $x \in X$, $y \in Y \subseteq [0,1]$ normal with respect to (w.r.t.) secondary grades if

$$\max_{y\in Y} \varphi(x,y) = 1,\ \forall x \in X. \tag{27.6}$$

Formula (27.1) implies obvious proposition.

Proposition 27.1 *T2FS G of goal is normal w.r.t. secondary grades.*

Lemma 27.1 *The fuzzy relation $\tilde{F}$ on X is FPR if, and only if, the T2FS of feasible alternatives F is normal w.r.t. secondary grades.*

Proof Note that the membership function of the fuzzy set $\bigcup_{y\in Y}(y, \mu_F(x^*, y))$ of membership degrees to the T2FS F for the alternative $x^* \in X$ has the form $\mu_F(x^*, y)$. Similarly, $\mu_G(x^*, y)$ and $\mu_D(x^*, y)$ are the membership functions of the fuzzy sets of membership degrees to the T2FSs G and D for the alternative $x^* \in X$, respectively.

Sufficiency Assume that the T2FSs of feasible alternatives F and goal G (according to Proposition 27.1) are normal w.r.t. secondary grades, that is,

$$\max_{y\in Y} \mu_F(x, y) = 1, \ \max_{y\in Y} \mu_G(x, y) = 1, \ \forall x \in X. \tag{27.7}$$

We intend to show that

$$\max_{y\in Y} \mu_D(x, y) = 1, \ \forall x \in X. \tag{27.8}$$

Assume on the contrary that $\exists x' \in X$ for which

$$\mu_D(x', y) < 1, \ \forall y \in Y. \tag{27.9}$$

Set $u' = \arg\max_{y\in Y} \mu_F(x', y)$, $v' = \arg\max_{y\in Y} \mu_F(x', y)$ and $y' = \min\{u', v'\}$. Formulas (27.7) imply that $\mu_F(x', u') = 1$ and $\mu_G(x', v') = 1$. According to (27.2)

$$\mu_D(x', y') = \max_{u,v\in Y, \ \min\{u,v\}=y'} \min\{\mu_F(x', u), \mu_G(x', v)\} \geq$$

$$\min\{\mu_F(x', u'), \mu_G(x', v')\} = \min\{1, 1\} = 1$$

A contradiction to (27.9). Therefore, equalities (27.8) hold and the T2FS D of solutions is normal w.r.t. secondary grades.

Now it is necessary to show that the fuzzy relation $\tilde{F}$ is reflexive, that is, $\eta(x, x) = 1, \forall x \in X$. Assume on the contrary that $\exists x' \in X$ for which

$$\eta(x', x') < 1. \tag{27.10}$$

Denote by u' and v' arbitrary elements of the set $\mathrm{Argmax}_{y\in Y} \mu_D(x', y) = 1$. Without loss of generality we assume that $u' \geq v'$. Formula (27.8) entails $\mu_D(x', u') = 1$ and $\mu_D(x', v') = 1$. Therefore, according to formula (27.5)

$$\eta(x', x') = \max_{u,v\in Y, \ u\geq v} \min\{\mu_F(x', u), \mu_G(x', v)\} \geq$$

$$\min\{\mu_F(x', u'), \mu_G(x', v')\} = \min\{1, 1\} = 1,$$

we have a contradiction to (27.10). Thus, the fuzzy relation $\tilde{F}$ is reflexive.

Necessity Assume that the fuzzy relation $\tilde{F}$ is reflexive, that is,

$$\eta(x, x) = 1, \ \forall x \in X. \tag{27.11}$$

We first show that equalities (27.8) hold.

Assume on the contrary that $\exists x' \in X$ for which $\mu_D(x', y) < 1$ for any $y \in Y$, hence, particularly, for $y = u'$ and $y = v'$ satisfying

$$\min\{\mu_F(x', u'), \mu_G(x', v')\} = \max_{u,v \in Y, \ u \geq v} \min\{\mu_F(x', u), \mu_G(x', v)\}.$$

In view of $\min\{\mu_F(x', u'), \mu_G(x', v')\} < 1$ formula (27.5) implies that $\eta(x', x') < 1$, this is a contradiction to (27.11). Thus, equalities (27.8) do hold which implies that the T2FS D of solutions is normal w.r.t. secondary grades.

We are going to check that equalities (27.7) hold. Assume on the contrary that $\exists x' \in X$ for which at least one of the equalities does not hold. Assume, without loss of generality, that $\mu_F(x', u) < 1 \ \forall u \in Y$. This inequality particularly holds for $u = u'$ satisfying

$$\min\{\mu_F(x', u'), \mu_G(x', v')\} = \max_{u,v \in Y} \min\{\mu_F(x', u), \mu_G(x', v)\} \tag{27.12}$$

Setting $y' = \min\{u', v'\}$ we can represent (27.12) in the form

$$\min\{\mu_F(x', u'), \mu_G(x', v')\} = \max_{u,v \in Y, \ \min\{u,v\} = y'} \min\{\mu_F(x', u), \mu_G(x', v)\}.$$

In view of $\min\{\mu_F(x', u'), \mu_G(x', v')\} < 1$ formula (27.2) entails $\mu_D(x', v') < 1$, which is a contradiction to (27.8). Thus, equalities (27.7) do hold and the T2FS of feasible alternatives F is normal w.r.t. secondary grades.

The proof of Lemma 27.1 is complete.

Remark 27.1 The assumptions of Lemma 27.1 restrict applicability of this approach to problems of maximizing a given function on a T2FS of feasible alternatives.

Since FPR is introduced on the set X of alternatives, the original problem is reduced to the problem of finding a fuzzy set of non-dominated alternatives and then choosing an alternative which is 'the best' in some sense.

Now we intend to construct on the set X a fuzzy set ND of non-dominated alternatives. According to formulas (27.3) and (27.5), its membership function has the form

$$\begin{gathered} \eta_{ND}(x) = 1 - \max_{y \in X}\{\eta(x, y) - \eta(y, x)\} = \\ 1 - \max_{y \in X}\{\max_{u,v \in Y, \ u \geq v} \min\{\mu_D(y, u), \mu_D(x, v)\} - \\ \max_{u,v \in Y, \ u \geq v} \min\{\mu_D(x, u), \mu_D(y, v)\}\}. \end{gathered} \tag{27.13}$$

According to the Bellman-Zadeh approach [27] the maximizing alternative (with the maximal degree of membership) is usually chosen as the "best" alternative to the fuzzy set. In the situations when such an alternative is not easily found, one selects an alternative with a degree of membership not less than a given value $\alpha \in [0, 1]$. We apply this idea to solve the original problem.

An alternative $x^* \in X$ satisfying $\eta_{ND}(x) \geq \alpha$ will be called an α-level non-dominated alternative to the problem of the function $g(x)$ maximization on the F.

Observe that $\eta_{ND}(x)$ is the degree of non-dominance of the alternative $x \in X$. Hence, $\eta_{ND}(x) \geq \alpha$ implies that there is no alternative dominating the alternative x with a membership degree greater than $1 - \alpha$, $\alpha \in [0, 1]$ on the set X. The membership function $\eta_{ND}(x)$ of the fuzzy set ND of non-dominated alternatives is complicated enough. Therefore, a method is needed which enables one to select α-level non-dominated alternatives by making only implicit use of this membership function. Consider the problem:

$$g(x) \to \max \tag{27.14}$$

subject to

$$\begin{gathered} \mu_F(x, y) \geq \alpha, \\ u \geq g(x), \\ x \in X,\ y \in Y \subseteq [0, 1]. \end{gathered} \tag{27.15}$$

Theorem 27.1 *Assume that the T2FS F of feasible alternatives is normal w.r.t. secondary grades and (x^*, y^*) is an optimal solution to problem (27.14) and (27.15). Then x^* is an α-level non-dominated alternative.*

Proof Suppose that (x^*, y^*) is the optimal solution to problem (27.14) and (27.15). Denote $y^* = g(x^*)$. Then (x^*, u^*, y^*) is optimal solution to the problem:

$$y \to \max \tag{27.16}$$

subject to

$$\begin{gathered} \mu_F(x, y) \geq \alpha, \\ u \geq g(x), \\ y = g(x), \\ x \in X,\ y, u \in Y \subseteq [0, 1]. \end{gathered} \tag{27.17}$$

We show the vector (x^*, u^*, y^*) is optimal solution to the problem:

$$y \to \max \tag{27.18}$$

subject to

$$\begin{gathered} \max_{u\in[0,1],\ u\geq g(x)} \mu_F(x,u) \geq \alpha, \\ y = g(x), \\ x \in X,\ y,u \in Y \subseteq [0,1]. \end{gathered} \tag{27.19}$$

Assume on the contrary that the vector (x^*, u^*, y^*) is not optimal solution to problem (27.18) and (27.19). In view of constraints (27.19) hold, then $\exists \bar{x} \in X$, $\exists \bar{y}, \bar{u} \in [0,1]$ for which $\bar{y} > y^*$, $\mu_F(x,\bar{u}) \geq \alpha$, $\bar{y} = g(\bar{x})$, $\bar{u} \geq g(\bar{x})$. This implies the vector $(\bar{x}, \bar{u}, \bar{x})$ satisfies to constraints (27.17) and has better value of the objective function, than (x^*, u^*, y^*). This is a contradiction. Formula (27.4) implies obvious inequality

$$\mu_D(x,y) \geq \max_{u\in[0,1],\ u\geq y=g(x)} \mu_F(x,u).$$

Whence optimal solution to the problem (27.18) and (27.19) (x^*, y^*) is optimal solution to the problem:

$$y \to \max \tag{27.20}$$

subject to

$$\begin{gathered} \mu_D(x,u) \geq \alpha, \\ x \in X,\ y \in Y \subseteq [0,1]. \end{gathered} \tag{27.21}$$

We consider fuzzy relation $\tilde{F}$ on X. According to the Lemma 27.1 $\tilde{F}$ be FPR with membership function $\eta_{ND}(x^*)$ in the form (27.13). We intend to show that

$$\eta_{ND}(x^*) = 1 - \max_{x\in X}\{\max_{u,y\in Y,\ u\geq y} \min\{\mu_D(x,u), \mu_D(x^*,y)\} -$$

$$\max_{u,y\in Y,\ u\geq y} \min\{\mu_D(x^*,u), \mu_D(x,y)\}\} \geq \alpha.$$

Assume on the contrary that there is $\tilde{x} \in X$ such that

$$\begin{gathered} \max_{u,y\in Y,\ u\geq y} \min\{\mu_D(\tilde{x},u), \mu_D(x^*,y)\} - \\ \max_{u,y\in Y,\ u\geq y} \min\{\mu_D(x^*,u), \mu_D(\tilde{x},y)\} \geq 1-\alpha. \end{gathered} \tag{27.22}$$

Claim 1 there exists a $\tilde{y} \in Y$ satisfying

$$\mu_D(\tilde{x},\tilde{y})\} \geq \alpha. \tag{27.23}$$

Indeed, if this is not the case, then $\mu_D(x, y)\} < \alpha, \forall y \in Y$ and thereupon

$$\max_{y \in Y} \mu_D(x, y)\} < \alpha.$$

Claim 2 $y^* \geq \tilde{y}$. Assume on the contrary that $y^* < \tilde{y}$. This inequality together with (27.23) implies that the vector (x^*, y^*) is not the optimal solution to problem (27.20) and (27.21), and we get a contradiction.

Combining pieces together yields

$$\max_{u, y \in Y,\ u \geq y} \min\{\mu_D(x^*, u), \mu_D(\tilde{x}, y)\} \geq$$

$$\min\{\mu_D(x^*, y^*), \mu_D(\tilde{x}, \tilde{y})\} \geq \min\{\alpha, \alpha\} = \alpha.$$

Now formula (27.22) entails

$$\max_{u, y \in Y,\ u \geq y} \min\{\mu_D(x^*, u), \mu_D(\tilde{x}, y)\} >$$

$$1 - \alpha + \max_{u, y \in Y,\ u \geq y} \min\{\mu_D(x^*, u), \mu_D(\tilde{x}, y)\} \geq 1 - \alpha + \alpha = 1$$

which is a contradiction in view of $\mu_D(x, y)\} \leq 1$ for $\forall x \in X, \forall y \in Y$.

The proof of Theorem 27.1 is complete.

27.5 Effective Maximizing Alternatives

It is easy to construct an example with nonlinear objective function in which problem (27.14) and (27.15) has the set of optimal solutions (x^*, y^*) with the same values of objective function $g(x)$, but with different secondary grades of membership $\mu_F(x^*, y^*) \geq \alpha$. In this case, it is sensible to select an alternative which corresponds to the optimal solution to the following problem:

$$\mu_F(x, y) \to \max \tag{27.24}$$

subject to

$$\begin{gathered} g(x) \geq g(x^*) \\ y \geq g(x) \\ x \in X,\ y \in Y \subseteq [0, 1]. \end{gathered} \tag{27.25}$$

This observation leads to the following conclusion. When selecting a maximizing alternative one should try to take as large as possible both the objective function (problem (27.14) and (27.15)) and the secondary grade (problem (27.24)

and (27.25)). In other words, one should choose only those alternatives that are called Pareto optimal in the two-criteria optimization problem:

$$\begin{gathered} g(x) \to \max, \\ \mu_D(x, y) \to \max \end{gathered} \tag{27.26}$$

subject to

$$\begin{gathered} y \geq g(x) \\ x \in X,\ y \in Y \subseteq [0, 1]. \end{gathered} \tag{27.27}$$

Maximized in this problem are the objective function and the secondary grade of membership to the T2FS F of feasible alternatives.

Recall that the vector $(x, y), x \in X, y \in Y$ dominates the vector $(x', y'), x' \in X, y' \in Y$ in two-criteria problem (27.26) and (27.27) (notation $(x, y) \succ (x', y')$) if the inequalities $g(x) \geq g(x')$ and $\mu_F(x, y) \geq \mu_F(x', y')$ hold and at least one of these is strict. This concept allows us to define the set of Pareto optimal solutions to two-criteria problem. This observation leads to the following conclusion. When selecting a maximizing alternative one should try to take as large as possible both the objective function (27.26) and (27.27).

The vector (x^*, y^*) is called Pareto optimal solution to two-criteria problem (27.26) and (27.27) if there is no such vector $(x, y), x \in X, y \in Y$ which dominates (x^*, y^*). We denote the set of Pareto optimal solutions to problem (27.26) and (27.27) by $P = \{(x^*, y^*) : \nexists x \in X,\ \nexists y \in Y\ \ (x, y) \succ (x^*, y^*)\}$.

Remark 27.2 Assume that the Pareto optimal solutions to problem (27.26) and (27.27) exist. Sufficient conditions for this to hold are widely known. For instance, $\mu_F(x, y)$ is continuous and X, Y are bicompacts, or $\mu_F(x, y)$ is arbitrary and X, Y are finite sets.

We call $x^* \in X$ the effective maximizing alternative if there exists $y^* \in Y$ such that the vector (x^*, y^*) is a Pareto optimal solution to problem (27.26) and (27.27), that is, $(x^*, y^*) \in P$.

The properties of effective maximizing are partially investigated in the following theorem.

Theorem 27.2 *Assume that the T2FS F of feasible alternatives is normal w.r.t. secondary grades and that the vector (x^*, y^*) is a Pareto optimal solution to two-criteria optimization problem (27.26) and (27.27). Then x^* is the effective maximizing alternative which is also $\mu_F(x^*, y^*)$-level non-dominated.*

Proof Assume that $(x^*, y^*) \in P$. By definition, x^* is then the effective maximizing alternative. Setting $\alpha = \mu_F(x^*, y^*)$ we observe that $\alpha \in [0, 1]$. Further, we intend to show that the vector (x^*, y^*) is the optimal solution to problem (27.14) and (27.15). It is obvious that system of constraints (27.15) is compatible. Assume that system (27.15) has a feasible solution $(\tilde{x}, \tilde{y})$ for which the inequalities $g(\tilde{x}) \geq$

$g(x^*)$ and $\mu_F(\tilde{x}, \tilde{y}) > \mu_D(x^*, y^*)$ hold. Then $(\tilde{x}, \tilde{y}) \succ (x^*, y^*)$ and thereupon $(x^*, y^*) \notin P$ by the definition of Pareto optimal solutions to problem (27.26) and (27.27). We get a contradiction. Thus, the vector (x^*, y^*) is the optimal solution to problem (27.14) and (27.15). It remains to note that according to Theorem 27.1 x^* is the α-level non-dominated alternative for $\alpha = \mu_F(x^*, y^*)$.

The proof of Theorem 27.2 is complete.

Recall that Pareto optimal solutions are incomparable. Therefore, to select an effective maximizing alternative $x^* \in X$ a decision-maker (DM) chooses a compromise between the desire to obtain the largest possible values of both the objective function $g(x)$ and secondary membership degree $\mu_F(x^*, y)$. In this situation the best values for one criterion lead to worse values for the other.

Let us now get back to Example 1. It is easy to verify that the set P of Pareto optimal solutions to problem (27.26) and (27.27) contains only two vectors $(x', y') = (6; 0, 5)$ and $(x'', y'') = (5, 25; 7/16)$ with the values $\mu_F(x', y') = 0, 5$ and $\mu_F(x'', y'') = 1$, respectively. They are incomparable because $0, 5 = g(x') > g(x'') = 7/16$ and $0, 5 = \mu_F(x', y') < \mu_F(x'', y'') = 1$. Which effective maximizing alternative does a DM select? There are two options: either $x' = 6$ with the values of the objective function $g(x') = 0, 5$ and secondary membership degree $\mu_F(x', y') = 0, 5$ or $x'' = 5, 25$ with $g(x'') = 7/16$ and $\mu_F(x'', y'') = 1$. The DM particular choice depends on what is more important for him/her: the value of objective function or secondary degree of membership to the T2FS of feasible alternatives.

The methods of multi-criteria (particularly, two-criteria) optimization are rather effective, well understood, have been sufficiently well studied (two-criteria in particular). These methods provide a variety of opportunities to extract information from the DM about his/her preferences on the set of criteria (in this case this is the preference between the objective function value and secondary degrees of membership to the T2FS of feasible alternatives) and use this information to obtain Pareto optimal solutions. Surveys of multi-objective optimization methods can be found in Branke et al. [28], Sawaragi et al. [29] and Steuer [30].

27.6 Conclusion

The T2FSs are an extension of the classical fuzzy sets. The former can model greater uncertainty than the latter. However, complexity of the theory of T2FSs and related computations precludes the wide use of T2FS in practical applications. The present article demonstrates that the Bellman-Zadeh approach can be successfully applied to decision-making problems which are defined on T2FSs. There are known areas of applications of mathematical programming problems in the case when different types of uncertainty are present in the membership function of the fuzzy set feasible alternatives. It is expected that the method developed in the present article will be useful in these areas.

References

1. Negoita, C.V., Ralescu, D.A.: Applications of Fuzzy Sets to Systems Analysis. Birkhauser/Wiley, Basel/New York (1975)
2. Orlovsky, S.A.: On programming with fuzzy constraint sets. Kybernetes **6**, 197–201 (1977). https://doi.org/10.1108/eb005453
3. Carlsson, C., Fuller, R.: Fuzzy Reasoning in Decision Making and Optimization. Studies in Fuzziness and Soft Computing. Physica-Verlag, Heidelberg (2002)
4. Lodwik, W.A.: Fuzzy Optimization, Recent Advances and Applications. Springer, Berlin (2010)
5. Ivokhin, E.V., Almodars, B.S.K.: To approaches for solving transportation problem with fuzzy resources. J. Autom. Inf. Sci. **46**(10), 45–57 (2014). https://doi.org/10.1615/JAutomatInfScien.v46
6. Mendel, J.M., John, R.I.: Type-2 fuzzy sets made simple. IEEE Trans. Fuzzy Syst. **10**(2), 117–124 (2002)
7. Zadeh, L.A.: Quantitative fuzzy semantics. Inform. Sci. **3**(2), 159–176 (1971)
8. Karnik, N.N., Mendel, J.M.: An introduction to type-2 fuzzy logic systems. Univ. of Southern Calif., Los Angeles (online). http://sipi.usc.edu/~mendel/report
9. Harding, J., Walker, C., Walker, E.: The variety generated by the truth value algebra of T2FSs. Fuzzy Sets Syst. **161**, 735–749 (2010)
10. Aisbett, J., Rickard, J.T., Morgenthaler, D.G.: Type-2 fuzzy sets as functions on spaces. IEEE Trans. Fuzzy Syst. **18**(6), 841–844 (2010)
11. Yager, R.R.: Fuzzy subsets of type II in decisions. J. Cybern. **10**, 37–159 (1980)
12. Chaneau, J.L., Gunaratne, M., Altschaeffl, A.G.: An application of type-2 sets to decision making in engineering. In: Bezdek, J. (ed.) Analysis of Fuzzy Information, vol. II: Artificial Intelligence and Decision Systems. CRC Press, Boca Raton (1987)
13. John, R.I.: Type-2 inferencing and community transport scheduling. In: Proceedings of 4th European Congress Intelligent Techniques Soft Computing, Aachen, Germany, pp. 1369–1372 (1996)
14. Kapustyan, E.A., Nakonechnyi, O.G.: The minimax problems of pointwise observation for a parabolic boundary-value problem. J. Autom. Inf. Sci. **34**, 52–63 (2002)
15. Baas, S.M., Kwakernaak, H.: Rating and ranking of multiple-aspect alternative using fuzzy sets. Automatica **13**(1), 47–58 (1977)
16. Chanas, S., Kuchta, D.: Multiobjective programming in optimization of interval objective functions a generalized approach. Eur. J. Oper. Res. **94**(3), 594–598 (1996)
17. Detyniecki, M., Yager, R.R.: Ranking fuzzy numbers using α-weighted valuations. Int. J. Uncertainty Fuzziness Knowledge Based Syst. **8**(3), 573–591 (2000)
18. Dubois, D., Prade, H.: Ranking of fuzzy numbers in the setting of possibility theory. Inform. Sci. **30**(3), 183–224 (1983)
19. Sengupta, A., Kumar, T.P.: Fuzzy preference ordering of intervals. Fuzzy Preference Ordering of Interval Numbers in Decision Problems. Studies in Fuzziness and Soft Computing, pp. 59–89. Springer, Berlin (2009)
20. Yager, R.R., Detyniecki, M., Bouchon-Meunier, B.: A context-dependent method for ordering fuzzy numbers using probabilities. Inf. Sci. **30**(3), 237–255 (2001)
21. Skalna et al.: Ordering of Fuzzy Numbers. Advances in Fuzzy Decision Making. Studies in Fuzziness and Soft Computing, vol. 333. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26494-3-2
22. Mashchenko, S.O.: Generalization of Germeyer's criterion in the problem of decision making under the uncertainty conditions with the fuzzy set of the states of nature. J. Autom. Inf. Sci. **44**(12), 26–34 (2012). https://doi.org/10.1615/JautomatInfScien.v44.i10.20
23. Mashchenko, S.O.: A mathematical programming problem with the fuzzy set of indices of constraints. Cybern. Syst. Anal. **49**(1), 62–68 (2013). https://doi.org/10.1007/s10559-013-9485-4

24. Odovsky, S.A.: Decision-making with a fuzzy preference relation. Int. J. Fuzzy Sets Syst. **1**(3), 155–167 (1978)
25. Orlovsky, S.A.: On formalization of a general fuzzy mathematical problem. Fuzzy Sets Syst., **3**(1), 311–321 (1980)
26. Chang, D.-Y.: Applications of the extent analysis method on fuzzy AHP. Eur. J. Oper. Res. **95**(3), 649–655 (1996)
27. Bellman, R.E., Zadeh, L.A.: Decision-making in a fuzzy environment. Manag. Sci. **17**, 8141–8164 (1970)
28. Branke, J. et al.: Multiobjective Optimization, Interactive and Evolutionary Approaches. A State-of-the-Art Survey. Springer, Berlin (2008)
29. Sawaragi, Y., Nakayama, H., Tanino, T.: Theory of Multiobjective Optimization. Academic, Orlando (1985)
30. Steuer, R.E.: Multiple Criteria Optimization: Theory, Computation, and Applications. Wiley, Chichester (1986)

Chapter 28
Using Wavelet Techniques to Approximate the Subjacent Risk of Death

Generating Alternative Scenarios via Bootstrap

F. G. Morillas Jurado and I. Baeza Sampere

Abstract In Actuarial science, graduation techniques have been used extensively: the large number of scientific papers and technical documentation published evidences this fact (see Ayuso M et al. (Estadística Actuarial Vida. UBe, Barcelona (2007)), Baeza Sampere and Morillas Jurado (Rev. Anales del Instituto de Actuarios Españoles 135–164 (2011)), London (Graduation: The Revision of Estimates. ACTEX Publications, Connecticut (1985)), Cairns, et al. (Scand Actuar J 2(3):79–113 (2008)) and the references therein). Graduation techniques are defined by Haberman and Renshaw J Inst Actuar 110:135–156 (1983) as *a set of principles and techniques for use that are used on raw data so that a more appropriate basis is obtained to make inferences and calculations of premiums, reserves and other variables of interest in the financial and insurance sector*. Solvency II (Directive 2009/138/EC) is the regulatory framework used for risk management and for the supervision of insurance companies. This normative is effective from 1/2016 and it establishes the technical exigencies to be applied on some procedures such as mathematical provisioning or pricing. Related to these procedures, this normative introduces the concepts of best-estimate and margin-risk (see also CEIOPS: QIS5 Technical specifications. Technical Report. European Commissions-Internal market and services DG (2010)). The purpose of the former is to approximate the expected loss, and the latter to control the deviation from the best-estimate. In life actuarial methodologies, the probabilities of death are used explicitly and so the estimation of q_x has a great impact on *best-estimate*. A widely recognized technique is that of wavelet techniques which have been applied in several fields such as engineering, digital processing of images, medicine, economy, finances, etc. (see Mallat (A wavelet tour of signal processing. Elsevier, Oxford (2009)). In this paper, we use wavelet techniques to achieve a good approximation of q_x with the aim of

F. G. Morillas Jurado (✉) · I. Baeza Sampere
Department of Applied Economy, University of Valencia, Valencia, Spain
e-mail: Francisco.Morillas@uv.es; Ismael.Baeza@uv.es

V. A. Sadovnichiy, M. Z. Zgurovsky (eds.), *Modern Mathematics and Mechanics*,
Understanding Complex Systems, https://doi.org/10.1007/978-3-319-96755-4_28

obtaining a good estimation of the *best-estimate*. Therefore, as we can deduce from Standards IFRS 17 (view [16]), the knowledge of future scenarios is an important key point. In this research, the *best-estimate* is obtained using a wavelet-based graduation technique similar to Meneu et al. El Factor de Sostenibilidad: Diseños alternativos y valoración financiero-actuarial de sus efectos sobre los parámetros del sistema. Economía Española y Protección Social, V, 63–96 (2013). Then, using resampling techniques over the deviations between the graduated values and the observed values, we can obtain an approximation to the variability of the estimation and go on to construct several series of alternative scenarios of death. Complementarily, we note that the wavelet techniques present some problems when the amount of data is small; for this reason, we countercheck the cases where we have low-frequency series (as is usually the case in the context of Life Insurers).

28.1 Introduction

In recent times, the study of longevity has experienced a boom due to its application in different fields of insurance and finance. Issues such as pensions, dependence, longevity/survival in the public sphere or pricing/provisioning, design of new products, insurance or finance (Solvency Directive) and impact on retirement plans (public or private) in the insurance ambit have led to these studies being considered socially relevant.

The instruments that summarize the study of survival or death are the life tables or mortality tables. These tables collect information, such as the age of death, risk of dying at age x, the number of persons who survive or die at each age or the probability of dying between ages.

It is important to note that the value of the risk of death at a given age is generally unknown and this fact has prompted a large number of studies in order to estimate the underlying value at each age x. These death risks, or their estimates, are necessary, for example, to calculate different policy premiums or to estimate the Sustainability Factor used in public pension systems [21]. To do these estimations, the technical specifications [4] and the standards IFRS [16] are used.

A way of simplifying the study of mortality is to consider that:

1. The biometric functions are continuous.
2. The underlying probabilities cannot be observed directly: we only can perceive the true values plus random fluctuations and they are indistinguishable.
3. The rates have a structural behavior [1, 14].

These assumptions justify the extensive development of graduation techniques and enable us to articulate the numerical method for the validation of proposed methods.

Haberman and Renshaw [13] define graduation as *the principles and methods by which a set of observed (or raw) probabilities are adjusted in order to provide a smooth base that will allow us to make inferences and also practical calculations of bonuses, reserves, etc.* The reasons for graduating an initial sequence of estimates is

based [18] on the fact that a sequence of initial values can often present abrupt changes between adjacent ages in the same period or between the same age in adjacent periods. These facts may be due to the concreteness of the random behavior of mortality.

The subject of graduation has been covered widely in the literature which describes a number of different types of graduation techniques, the main ones being Parametric and non-Parametric. The parametric graduation adjusts the raw data to a family of functions that depend on one or more parameters. In this context, the accepted mortality laws are well-known, examples of which include: De Moivre law; Gompertz laws [12]. These laws are applied only to adult ages and many of them fail to properly represent the hump of accidents in adulthood. Often, the study of mortality must be used over the whole age range (demographic previsions) and, in this case, the Heligman and Pollard Model [14] is a recommended model. Other types of techniques, such as splines graduation [10], are known as semiparametric. The last group of techniques, non-parametric techniques, is characterized by assuming non-functional forms for the behavior of the data. In this case, the mortality rates are obtained by applying smoothing methods by combining adjacent death rates (for example kernel graduation [1] and [5]). Some classical examples can be found in ([3, 9] or [11]). Recently, a new wavelet graduation method (non-parametric technique) was proposed in [2], and improved in [22].

In [22], authors propose a method based on the multiresolution wavelet decomposition, combining it with thresholding and Piecewise Polynomial Harmonic (PPH) techniques in order to obtain the graduation of the true risk of death values. In this paper, we combine wavelet graduation with the bootstrap technique in order to solve the limitations of information and to generate several alternative scenarios of death when the true values are not known. We use the Heligman and Pollard law to validate this technique.

The paper is structured as follows. Section 28.2 presents the biometric model. Section 28.3 describes (briefly) the wavelet graduation approach used and introduces some problems in its application. Also, this section describes the bootstrap technique and how to combine it with the wavelet graduation. Section 28.4 presents the measures used in the validation process and outlines an application to real data. The paper ends with the conclusions section (Sect. 28.5).

28.2 The Biometric Model

The Biometric Model is a stochastic conceptual framework constructed around the random variable *Age at death*, X and, in particular, around q_x, the risk of death at age x. The model used is sustained by the hypotheses that follow:

H1. *The moment of death is independent for each person.*
H2. *The distribution probability function of the moment of death is the same for different persons in the group studied.*

The H2 hypothesis suggests that *the number of deaths at age x*, d_x, can be seen as a random variable. It is modeled using a binomial law, $d_x \sim Bi(l_x, q_x)$, where l_x represents the number of persons exposed to risk (population of exact age x), and q_x is the above defined probability.

A priori, the probability q_x is unknown and we can only observe the number of deaths, d_x and the exposed to risk, l_x. From this, we can estimate the observed rate of death, $\hat{q}_x = d_x/l_x$. This fact allows us to split $\hat{q}_x$ into two parts: the real probability of death, q_x, plus a random fluctuation, f_x, due to the random nature of the phenomenon considered, giving us: $\hat{q}_x = q_x + f_x$. In Insurance and Finance practice, it is usual to use the estimation of q_x for several purposes; for example, to calculate life expectancy and to determine premiums, quotes and age to pension etc. So, the precision of the estimations is important. With the aim of reducing the effect of the random fluctuations, graduation techniques are used; parametric or non-parametric (see [2] and [19]).

In this paper, we consider *observed data* from different nature:

- Observed data of the real population, from INE (see [17])-for male and female, and for ages between [0, 100], and
- *Observed data* generated synthetically, obtained using a numerical procedure that evolves the H2 hypothesis (via $d_x \sim Bi(l_x, q_x)$) and a survival model, the Heligmann and Pollard law [14]:

$$\frac{q_x}{p_x} = A^{(x+B)^C} + De^{-E(ln(x)-ln(F))^2} + GH^x.$$

A usual demographic interpretation of the model is made term by term and using the representation in logarithmic scale of q_x (view Fig. 28.1):

(1) The first term represents the *adaptation to the environment*. The parameter A measures the infant mortality, B represents the probability of death in the first year of life and C is an environmental adaptation factor.
(2) The second term deals with the *Social hump* or *accidents hump*. The parameter D represents the intensity of the Social hump, E the amplitude of the hump and F is the exact point (age) where the hump has a maximal intensity.
(3) Finally, the last term represents the *natural longevity*. G is interpreted as the *Mortality at older ages* and H is a measure of the increment of mortality at older ages.

We use the Heligman and Pollard model for several reasons: (a) it is valid for the whole range of ages, (b) it has been applied successfully in other regions (Italy, France, United States, United Kingdom, Spain...) and at a different time (Australian population of 1962), (c) its numerical implementation is simple.

Figure 28.2 shows similarities between values observed in different regions (Italy 2009 and France 2012) and the qualitative description made previously of the Heligman and Pollard model of the three widely acknowledged parts: (1) adaptation to the environment, (2) social or accidents hump and (3) natural longevity.

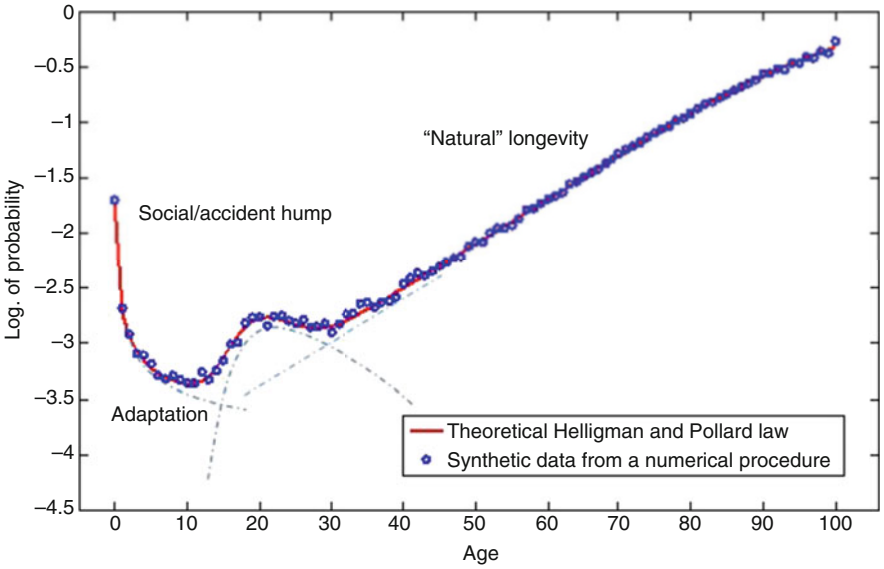


Fig. 28.1 Heligman and Pollard law and synthetic data obtained using a numerical procedure. Source: compiled by the authors

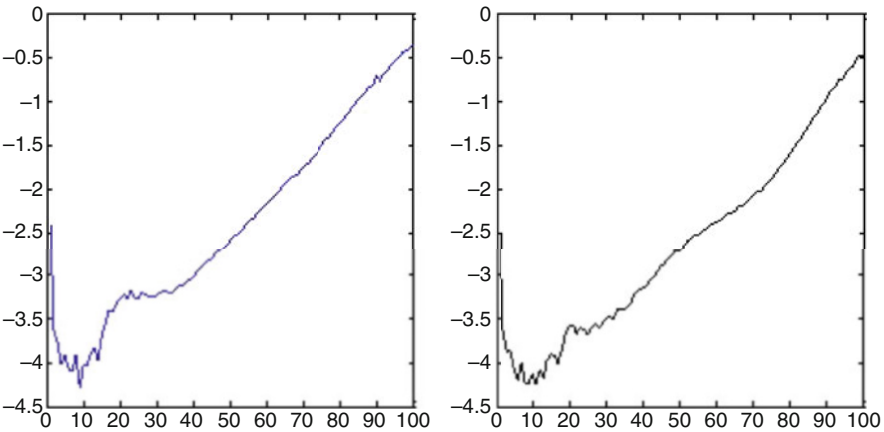


Fig. 28.2 Observed mortality rates: Italy 2009 (left) and France 2012 (right). Data Source: compiled by the authors using data from Human Mortality Data Base [15]

The Heligman and Pollard law is used in several steps of this research. Firstly, we use it to find out *true* values of the probability of death at each age, and we use these as *reference values*. Secondly, we use the reference values to generate (several) synthetic experiences of death: these values have random fluctuations. Thirdly, we graduate the synthetic experiences (applying the technique proposed)

and we compare the graduated values with the reference values. The differences obtained help us to validate the technique.

The numerical process, described below, is used to simulate the *observed values*. This procedure is repeated as many times as necessary to generate the different experiences required.

We start the process using theoretical probabilities of death given q_x derived from the Heligman and Pollard law (the reference values) and taking into account an initial number of persons, $l_0 = 100.000$. Follows:

- Given that d_x follows a binomial distribution, thus: $d_x \sim Bi(l_x, q_x)$, we generate a random number given by the distribution $Bi(l_0, q_0)$. This is the number of deaths at age 0, d_0, and we use it to estimate $l_1 = l_0 - d_0$.
- We then generate a random number that follows a distribution $Bi(l_1, q_1)$. We obtain the number of deaths at age $x = 1$, d_1, and we use it for the estimation of $l_2 = l_1 - d_1$.
- Iterating this process, we generate random numbers from a binomial law, $Bi(l_x, q_x)$, l_x being the estimated number of survivors in the previous stage $l_x = l_{x-1} - d_{x-1}$, and q_x, the risk of death at the age considered (and derived from the Heligman and Pollard law). In this way, we obtain d_x and l_{x+1}, the latter is used for the next step as input of a new random number of the distribution $Bi(l_{x+1}, q_{x+1})$.
- The process ends when we obtain the last value d_ω, ω being the age at which any person can survive.

The fact that q_x are *known* (synthetically via Heligman and Pollard law) enables us to obtain several scenarios of mortality. So, each of these scenarios can be used to apply the bootstrap method and then to construct/calculate several measures of the risk assumed by the insurers or the stakeholders for the population analyzed. In this research, we select, randomly, a unique scenario to apply the bootstrap method and to validate the combination of the wavelet graduation and the re-sampling technique.

28.3 Wavelet Graduation

As we can see in the literature [19], a wavelet (a family of them) is a set of functions that can be used to split a series of initial values into two parts with uniqueness: tendency and details. If we apply this procedure again in the tendency parts, we obtain a multiresolution analysis at different levels. It is known that if the values of the series contain random fluctuations, these fluctuations are manifested in the wavelet part. We can see these in Fig. 28.3.

So, to reduce the random fluctuations and recover the real values, we have to treat the wavelet part. To do this, we introduce another step in the procedure before the inverse transform. Specifically, we use a standard technique to treat the wavelet part known as thresholding, *μ-threshold method*. Values of less than a particular threshold *Thr* are *converted* to 0 (hard-thresholding). Then, we obtain *other* details

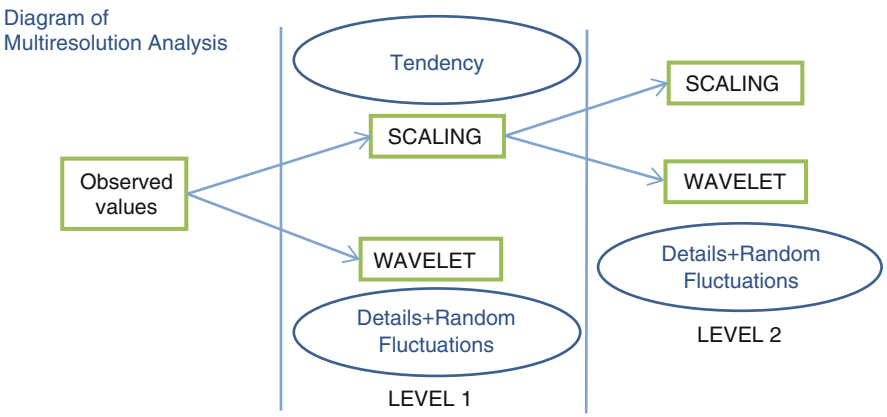


Fig. 28.3 Diagram of Multiresolution Analysis and fluctuations allocation

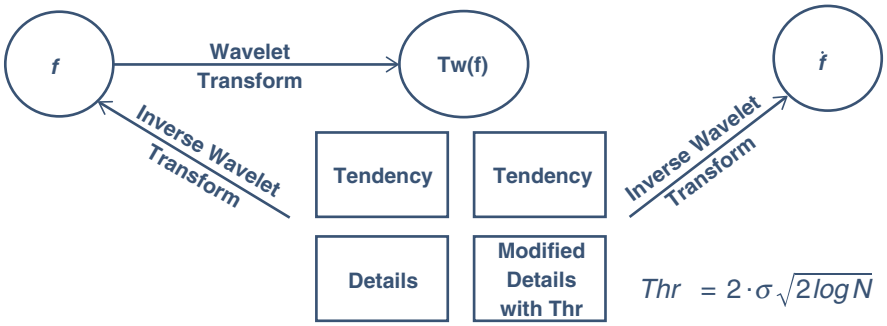


Fig. 28.4 Wavelet graduation scheme. Source: compiled by the authors

(a modification of the initially obtained ones). From here, it is a simple step to obtain the *graduated* values applying the inverse wavelet transform using the initial scaling part jointly with the modified wavelet part. In the thresholding method, the election of the threshold is a key point; [6] allows us to obtain the value $2\sigma\sqrt{2\log N}$ as threshold. In this expression, σ is the standard deviation of the $level$ 1 wavelet part. The scheme (mental framework) of wavelet graduation can be seen in Fig. 28.4. Here, via the thresholding technique and the inverse wavelet transform, it is possible to obtain a *pointwise estimation* for each age x of the biometric function, q_x.

28.3.1 Wavelet Graduation Problems

The wavelet graduation may have some drawbacks according to the information available or the functional relationship of the data. In the biometric model for younger ages (0–20 years), the mortality curve (logarithmic values) has a nonlinear

relationship that complicates the analysis because there are not enough data to recover its form. Other drawbacks are:

- The incorporation of symmetric information at the ends of the series introduces noise by discontinuity.
- The problem of discontinuity reappears if we use a wavelet family with a big support or if we make several scales of the process.
- An effect, similar to the Gibbs phenomenon, appears in the social hump projection (Fig. 28.1). We can see values that are smaller (greater) than the relative minimum (maximum) after (and before) the accident hump.

The Piecewise polynomial harmonic PPH interpolation used in [22] allows the incorporation of additional data preserving the concavity (or convexity) of the function to overcome the disadvantage of the limited information available. In addition, this introduces no spurious oscillations and it avoids some undesirable effects such as Gibbs phenomenon or noise by discontinuity. However, and by simplicity, we do not use this PPH-wavelet technique; in this paper we consider the usual wavelet transform. The results of the resampling method presented are good enough.

28.3.2 Combining Bootstrap and Wavelet Graduation

The proposed method starts with the wavelet graduation of a biometric function, q_x. This initial biometric function could be this or an original series like the one that can be seen in Fig. 28.5. First, the sequence of observed (or synthetic) values is decomposed into scaling part and wavelet part using the wavelet transform. Then, the modified wavelet part is obtained via the thresholding technique and the graduated values (point estimate), $\dot{q}_x$, are obtained via the inverse wavelet transform. These steps can be seen graphically in Fig. 28.6. Figure 28.7 The top left panel shows the residuals (difference between the original details and those obtained through thresholding) by age while the bottom left panel shows its density distribution. It is clear that the magnitude of the residuals depends on its corresponding age. This suggests that the residuals are not comparable; they have a strong dependency on age. It is therefore not appropriate to apply the resampling directly. To avoid this problem, the second step of this procedure is to use a transformation of the residuals, in particular we propose the Pearson residuals, [20]. The Pearson residuals are the initial residuals scaled by the estimated standard deviation, via binomial law and the previous wavelet (and pointwise) graduation. The Pearson residuals expression is:

$$r_x^{Pearson} = \frac{E_x q_x - E_x \dot{q}_x}{\sqrt{E_x \dot{q}_x (1 - \dot{q}_x)}}.$$

Figure 28.7 shows the Pearson residuals (top right panel). Now that they are comparable, it is not possible to associate the value of the residual to a particular

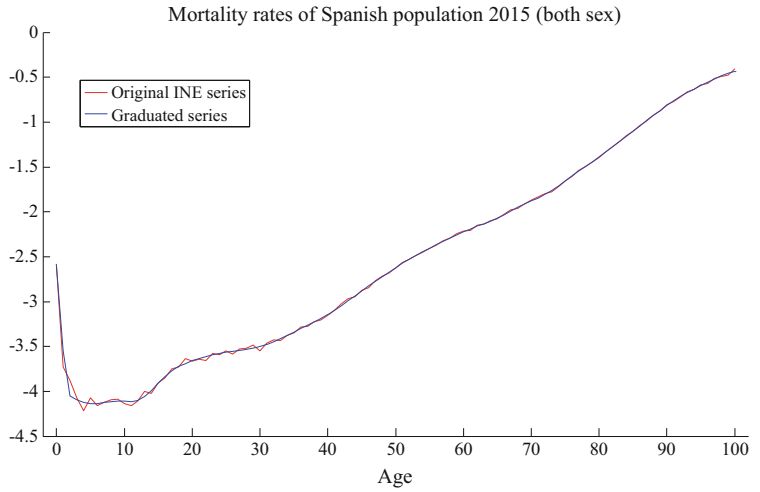


Fig. 28.5 Observed and graduated mortality rates series (logarithmic scale) of Spanish population 2015, both genders. Source: compiled by authors with data from INE

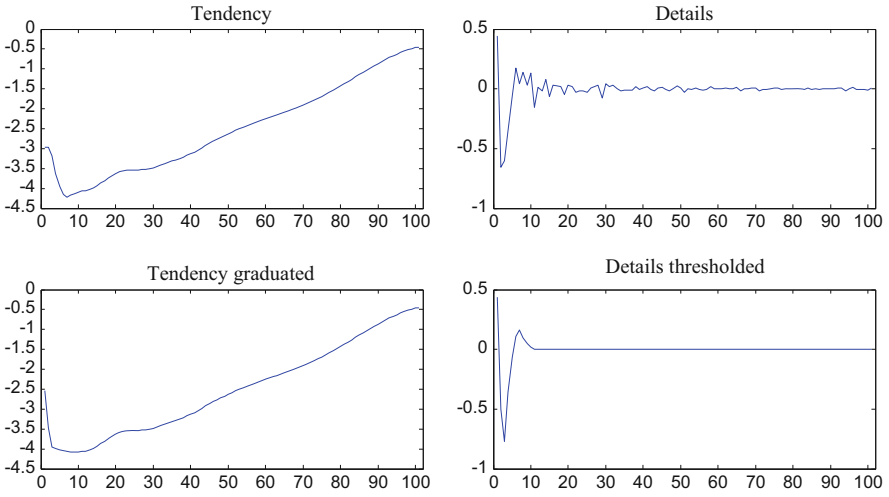


Fig. 28.6 Details of the graduation treatment of the wavelet part. Source: compiled by the authors

age. Also, Fig. 28.7 (bottom right panel) shows the density function of the Pearson residuals. We perform a goodness-of-fit to verify if the residuals have a Gaussian distribution. The answer is affirmative with a p-value equal to 0.9658 in the Kolmogorov-Smirnov test. From these considerations, it is appropriate to apply a bootstrap technique (view [7] and [8]).

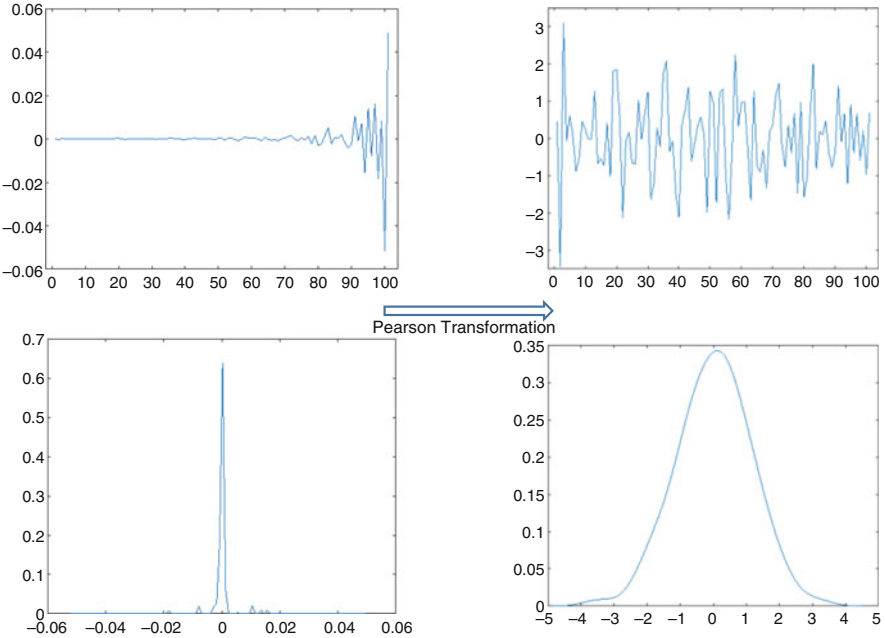


Fig. 28.7 Treatment of residuals. Original Residuals (left) and Pearson residuals (right), absolute values and density functions

In order to obtain several (a large quantity) bootstrap-sampled plausible (synthetic) values of residuals, the third step is to apply the re-sampling method on the Pearson residuals. The fourth step is to apply the Pearson inverse transformation in order to obtain the random fluctuations in the same magnitude as the initial residuals. Finally, the bootstrapping residuals are added to the point estimate obtained in the first step by the wavelet graduation. This procedure generates an arbitrary number of death scenario alternatives.

28.4 Validation and Applications

In this section, we describe the process and perform a first test verifying the integrity of the process to check if it is robust. We also indicate the numerical features, parameters and data used.

MatLab R2017b is the software used to do the estimations and to implement the procedure. The synthetic (and *true*) values of mortality have been obtained from the Heligman and Pollard law. The values of the parameters used can be found in [14], the original work of Heligman and Pollard. This sequence of *true* values is the input used to obtain an arbitrary number of sequences of the *observed-*

synthetic experiences of death via the procedure described in Sect. 28.2. The number of synthetic experiences obtained is 1000. Randomly selected, we focus on the realization #498.

We graduate $log(q_x^{\#498})$ using wavelet graduation with wavelet family *Biorth3.3* and multiresolution scheme of $level$ 3. This graduated series of values enables us to estimate the residuals (with respect to $q_x^{\#498}$) and to apply the Pearson Transformation to them, obtaining the *Pearson Residuals*. Then, the Pearson transformations are re-sampled without replacement; for each bootstrap series of Pearson residuals, we apply the inverse Pearson transformation to each particular age considered to obtain the bootstrap residuals. Finally, we use these series of **B**-residuals to obtain the bootstrap series of death experiences.

Next, the **B**-series of death experiences are used to increase the information that the observed data (in this case, realization #498) give us about the phenomenon of the mortality. The bootstrap series allow us to obtain summaries of the death experiences; for example, for each age we can obtain the **B**-mean value, the B-standard deviation and a collection of **B**-quantiles. Here, the **B**-notation indicates estimations from the Bootstrap series of $q_x^{i=1,\ldots,B}$. The number of bootstrap series is the same as the initial realization from the Heligman and Pollard law, which is $B = 1000$ (equal to the number of synthetic experiences).

Figure 28.8a shows the theoretical Heligman and Pollard law, an arbitrary scenario derived from this (the number #498, randomly selected) and several (and alternative) B-scenarios of death obtained graduating the logarithmic of #498 via wavelet, and resampling the Pearson residuals. Figure 28.8b–d presents the histogram of all B-scenarios obtained at ages 12, 25 and 44 (ages from adaptation part, social hump part and natural longevity part, respectively). The graphical information must be completed using numerical measures to conclude if the procedure described is appropriate.

Table 28.1 summarizes the comparison criteria considered for the validation of the proposed procedure. Among these are (for ages 12, 25 and 44) the true values derived from the Heligman and Pollard model, and the homologous bootstrap values estimated from the bootstrap experiences of death using a unique realization, #498. In particular, the following are considered: (1) the expectation and the variance and (2) the $\alpha - quantiles$, with $\alpha = 0.025$ and $\alpha = 0.975$. The comparison is made verifying that the true values (from Heligman and Pollard law) and the estimated values (using the bootstrap technique proposed) are very similar. We denote them as HP-values and B-values.

Complementarily, theoretical bounds (upper and lower) are calculated using the reference HP-values $\{q_x^{HP}\}$ and $\{l_x\}$ and considering that $d_x \sim Bi(l_x, q_x^{HP})$ which can be approximated by $N(l_x q_x^{HP}, \sqrt{l_x q_x^{HP}(1 - q_x^{HP})}$- via a heuristic rule to apply that the binomial law can be approximated by a normal law. Then, the $\alpha - quantile$ of q_x has the expression $q_x^{HP} + z_\alpha \sqrt{\frac{q_x^{HP}(1-q_x^{HP})}{l_x}}$, where z_α indicates the $quantile$ of order α from the normal distribution, $N(0, 1)$. Figure 28.9 graphically shows the graduated values, the quantiles of the considered order (upper and lower bounds), and for the both cases, bootstrap quantiles (B-upper and B-lower bounds) and the

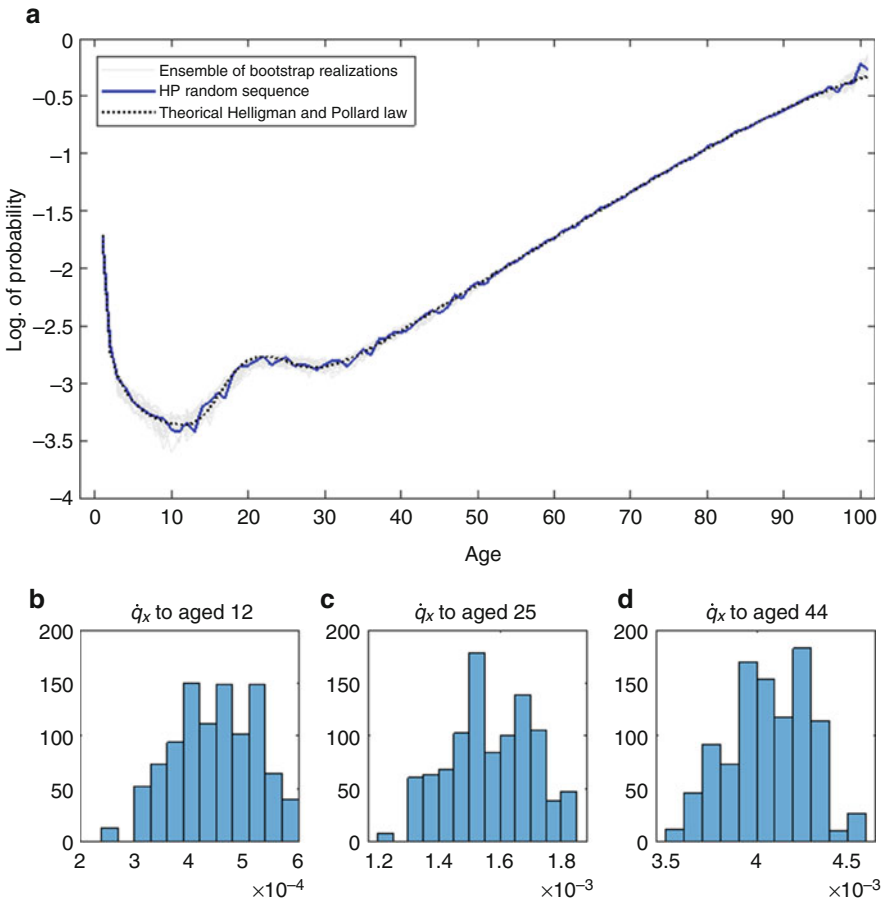


Fig. 28.8 Heligman and Pollard law, shown jointly with to #498-realization and B-scenarios of q_x (top). B-histogram (bottom) of selected ages (12, 25 and 44). Source: compiled by the authors

HP-quantiles (HP-upper and HP-lower bounds) from Heligman & Pollard law, as reference values; it illustrates what was highlighted in previous paragraph. We can see how close they are to both intervals. The bootstrap interval is usually wider except in a few cases, although hardly noticeable in the figure.

We use the next indicators or measures to compare the pointwise approximation:

1. Mean relative indicator by age x: $MRI_x(q) = \frac{|q_x - \hat{q}_x|}{q_x}$.
2. Mean squared relative indicator by age x: $MSRI_x(q) = \frac{|q_x - \hat{q}_x|^2}{q_x}$.

Table 28.1 Comparison criteria: summary

	Age 12	Age 25	Age 44
HP-value	0, 000437123	0, 001558924	0, 004186888
Mean B-realizations	0, 000594550	0, 001846275	0, 004525413
Mean HP-realizations	0, 000311250	0, 001315326	0, 003648085
Variance B-realizations	0, 000591281	0, 001852038	0, 004676685
Variance HP-realizations	0, 000304618	0, 001306981	0, 003765890
Upper B-realizations bound (97.5% percentile)	0, 000447843	0, 001569230	0, 004064932
Lower B-realizations bound (2.5% percentile)	0, 000451087	0, 001591419	0, 004235235
Upper HP-realizations bound (97.5% percentile)	$5,469243 \cdot 10^{-9}$	$1,861573 \cdot 10^{-8}$	$5,053358 \cdot 10^{-8}$
Lower HP-realizations bound (2.5% percentile)	$5,359965 \cdot 10^{-9}$	$2,008833 \cdot 10^{-8}$	$5,462567 \cdot 10^{-8}$
Upper theoretical bound (97.5% percentile)	0, 000568433	0, 001808512	0, 004603548
Lower theoretical bound (2.5% percentile)	0, 000305813	0, 001309336	0, 003770228

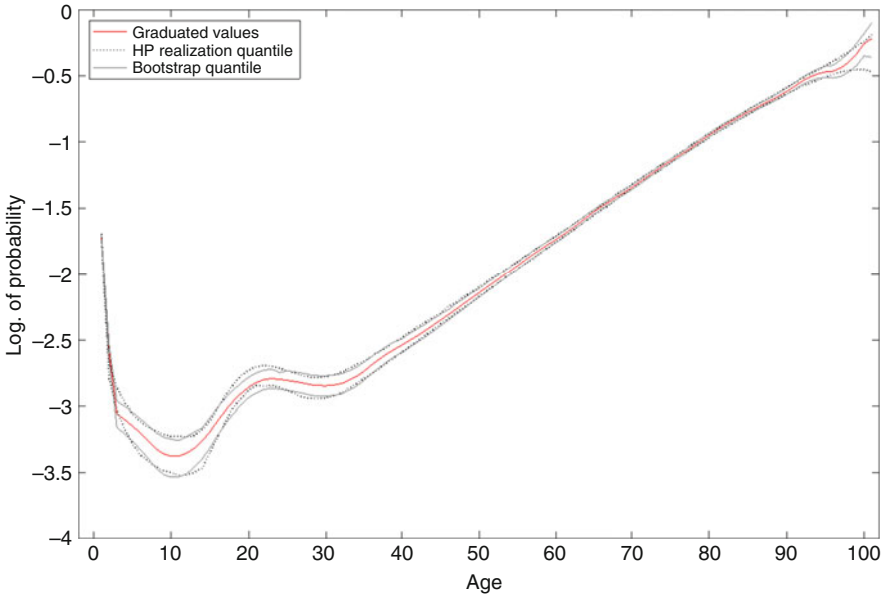


Fig. 28.9 HP-quantiles and B-quantiles. Source: compiled by the authors

Table 28.2 Indicators of differences in interval bounds

$MRI_x(q)$ upperbound	0, 005529244	0, 003111714	0, 032346040
$MRI_x(q)$ lowerbound	0, 021771653	0, 006384491	0, 031281947
$MSRI_x(q)$ upperbound	$1,807696358 \cdot 10^{-8}$	$1,793284626 \cdot 10^{-8}$	$4,893057642 \cdot 10^{-6}$
$MSRI_x(q)$ lowerbound	$1,443906219 \cdot 10^{-7}$	$5,327482761 \cdot 10^{-8}$	$3,685149757 \cdot 10^{-6}$

Table 28.3 Aggregate Indicators (considering all ages)

	$MRI(q)$	$MSRI(q)$
Upper bound	0, 034214835	$0,543317049 \cdot 10^{-3}$
Lower bound	0, 040460117	$0,596278098 \cdot 10^{-3}$

In the above expressions, q_x denotes the reference value, from Heligman and Pollard law, and $\hat{q}_x$ denotes an estimation obtained, in this case, via wavelet-bootstrap technique.

Table 28.2 gives us a comparison of the differences between the bootstrap and the theoretical bounds. This table shows differences less than 3, 1% in the sense of the MRI_x indicator. We observe that the value of the MRI_x is depending of age, from this, we note that for age 25, the differences are less than 0, 64%.

To compare the full sequence of the values we can generalize the indicators previously used:

1. Mean relative indicator: $MRI(q) = \frac{1}{\omega+1} \sum_{x=0}^{\omega} \frac{|q_x - \hat{q}_x|}{q_x}$
2. Mean squared relative indicator: $MSRI(q) = \frac{1}{\omega+1} \sum_{x=0}^{\omega} \frac{|q_x - \hat{q}_x|^2}{q_x}$

Table 28.3 gives us the values of the *MRI* and *MSRI* indicators. From these, we observe that the value of *MRI* is less (for all age) than 4%. Then, we can conclude that the bootstrap technique proposed can be valid and it increases the initial information.

28.4.1 Application to Real Data

In this section, we show an application to real data. In particular, we consider the observed values (real values) of the Spanish population (both genders) in the year 2015. The data are provided by INE [17]. We then apply the proposed technique to estimate, firstly, the subjacent risk of the population by age and; secondly, the bootstrap 95% (for example) confidence intervals (by quantiles) for these data and for each age.

To obtain these intervals, five thousand bootstrap realizations are obtained after discarding those that are impossible. For example, in the observed data we find that the probability at age 4 is equal to 0.000060946 (similar values at range of ages 3-14). Considering that a B-residual can be negative, then it can provide a negative value of the re-sampled probability, which would make this scenario impossible. It should be noted that only 2.5% of B-realizations are discarded.

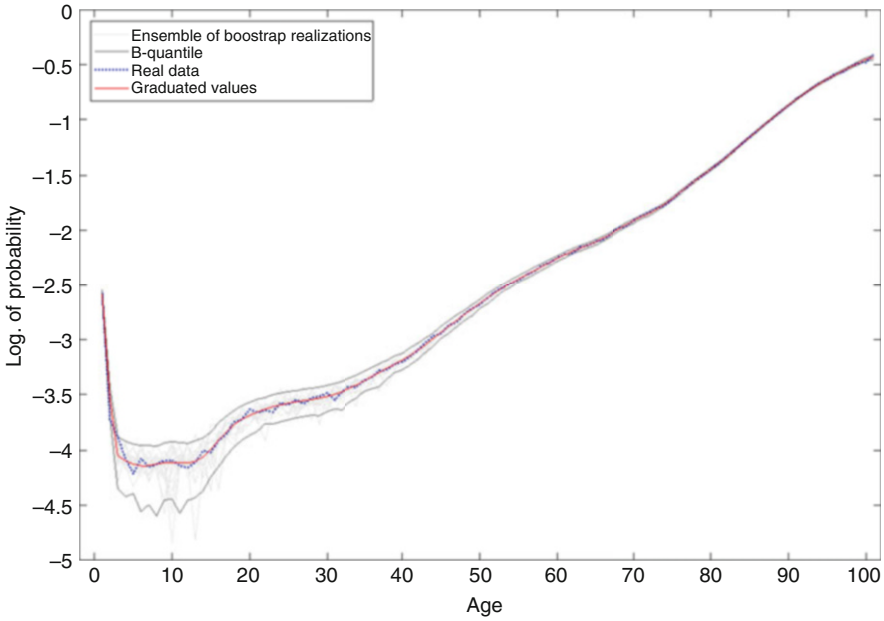


Fig. 28.10 Observed data and graduated values of the 2015 Spanish population (both genders); bootstrap confidence interval and ensemble of B-realizations. Source: compiled by the authors with data from INE

Figure 28.10 shows the bounds of bootstrap confidence intervals at 95% for the Spanish population of 2015 along with some realizations, the real data and its graduation.

28.5 Conclusions

Life tables (or mortality tables) are tools widely used in the study of mortality. These summarize the experience of mortality observed in a region (and time period). It is common for the death risks values $\{q_x\}$, $x = 0, \ldots, \omega$ (ω being the highest age considered in the study) not to be known. But these values, or their estimates, are necessary to calculate different policy premiums or to estimate the Sustainability Factor used in public pension systems and other scenarios.

In this paper, we introduce a technique that uses wavelet transform to obtain (1) a pointwise estimation of the true (and subjacent) probability of death and (2) other alternative scenarios. In actuarial life methodologies, this estimation impacts directly on the calculus of the best-estimate and can be used in the estimation of the risk-margin. The pointwise estimation is obtained using a wavelet-graduation technique (introduced in [2] and improvement in [22]). Thus, using re-sampling

techniques over the deviations between the graduated values and the observed values, we obtain several alternative scenarios. It is important to note that the Pearson Transformation must be used in order to treat the direct residuals to reduce the effect of the age; in other words, to reduce the heteroscedasticity of the observed data.

To verify the integrity of the process and to check if it is robust, we apply it to a random sequence (#498) obtained from the Heligman and Pollard law and, using a biometric model, produce one thousand scenarios. From the comparative measures, it has been possible to conclude that the data is close to the theoretical function from which the random sequence comes. The results obtained comparing the bootstrap quantiles and the quantiles obtained from the Heligman and Pollard law (as reference values) verify that this methodology provides an appropriate method for estimating the real and plausible scenarios.

As an application, we estimate the subjacent probability of death and we generate five thousand bootstrap realizations (discarding those that are impossible) as alternative scenarios, for the actual Spanish population in 2015. From these scenarios, we calculate bootstrap quantiles and then we estimate confidence intervals. It should be noted that quantiles of any order can be calculated. So, this method increases the information available to the analyst and it provides an interval estimate that can be used as input in a different process (reloaded life tables); it also provides other measures such as VAR (value-at-risk) that the *Solvency directive* requires. Although we obtain the bootstrap confidence intervals to approximate the true-confidence intervals, we do not need to know the real value of the risk of death nor the number of survivors by age.

Acknowledgements This paper is partially supported by *Ministerio de Economia y Competitividad* (Spanish Government), Grant MTM2016-74921-P.

Thanks are also due to M. Hodkinson for reviewing the English of the paper. The usual disclaimer applies.

References

1. Ayuso, M., Corrales, H., Guillen, M., Perez-Martín, A.M., Rojo, J.L.: Estadística Actuarial Vida. UBe, Barcelona (2007)
2. Baeza Sampere, I., Morillas Jurado, F.G.: Using wavelet to non-parametric graduation of mortality rates. Rev. Anales del Instituto de Actuarios Españoles, pp. 135–164 (2011)
3. Cairns, A.J.G., Blake, D., Dowd, K.: Modelling and management of mortality risk: a review. Scand. Actuar. J. **2**(3), 79–113 (2008)
4. CEIOPS: QIS5 Technical specifications. Technical Report. European Commissions-Internal market and services DG (2010)
5. Copas, J., Haberman, S.: Non parametric graduation using kernel methods. J. Inst. Actuar. **110**, 135–156 (1983)
6. Donoho, D., Johnstone, I.: Ideal spatial adaptation via wavelet shrinkage. Biometrika **81**, 425–455 (1994)
7. Efron, B.: Bootstrap methods: another look at the jackknife. Ann. Stat. **7**, 1–26 (1979)

8. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Chapman & Hall, Boca Raton, FL (1994)
9. Felipe, A., Guillen, M., Nielsen, J.: Longevity studies based on kernel hazard estimation. Insur. Math. Econ. **28**, 191–204 (2001)
10. Forfar, D., McCutcheon J., Wilkie, A.: On graduation by mathematical formulae. J. Inst. Actuar. **115**, 693–694 (1988)
11. Gavin, J., Haberman, S., Verrall, R.: Moving weighted average graduation using kernel estimation. Math. Econ. **12**(2), 113–126 (1993)
12. Gompertz, B.: On the nature of the function of the law of human mortality and on a new mode of determining the value of life contingencies. Trans. R. Soc. **115**, 513–585 (1825)
13. Haberman, S., Renshaw, A.: Generalized linear models and actuarial science. Statistician **4**(45), 113–126 (1996)
14. Heligman, L., Pollard, J.: The age pattern of mortality. J. Inst. Actuar. **107**, 49–80 (1980)
15. Human Mortality Database.: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (2017, April 15).
16. IFRS 17 Insurance Contracts. International Accounting Standards Boards (2017). International Financial Reporting Standards. Available at www.ifrs.org (2018, May 5).
17. Instituto Nacional de Estadística, INE (2017). Tablas de mortalidad de la población de España 1991–2012. http://www.ine.es, April 15 (2017)
18. London, D.: Graduation: The Revision of Estimates. ACTEX Publications, Connecticut (1985)
19. Mallat, S.G.: A Wavelet Tour of Signal Processing. Elsevier, Oxford (2009)
20. McCullagh, P., Nelder, J.A.: Generalized Linear Models. Chapman and Hall, New York (1983)
21. Meneu, R., Devesa, J.E., Devesa, M., Nagore, A.: El Factor de Sostenibilidad: Diseños alternativos y valoración financiero-actuarial de sus efectos sobre los parámetros del sistema. Economía Española y Protección Social, V, 63–96 (2013)
22. Morillas. F.G., Baeza I., Pavia, J.M.: Risk of death: a two-step method using wavelets and piecewise harmonic interpolation. Estadística Española **58**, 191 (2016)