Springer Optimization and Its Applications 173

Themistocles M. Rassias *Editor*

Nonlinear Analysis, Differential Equations, and Applications

Springer

Springer Optimization and Its Applications

Volume 173

Aims and Scope

Optimization has continued to expand in all directions at an astonishing rate. New algorithmic and theoretical techniques are continually developing and the diffusion into other disciplines is proceeding at a rapid pace, with a spot light on machine learning, artificial intelligence, and quantum computing. Our knowledge of all aspects of the field has grown even more profound. At the same time, one of the most striking trends in optimization is the constantly increasing emphasis on the interdisciplinary nature of the field. Optimization has been a basic tool in areas not limited to applied mathematics, engineering, medicine, economics, computer science, operations research, and other sciences.

The series **Springer Optimization and Its Applications (SOIA)** aims to publish state-of-the-art expository works (monographs, contributed volumes, textbooks, handbooks) that focus on theory, methods, and applications of optimization. Topics covered include, but are not limited to, nonlinear optimization, combinatorial optimization, continuous optimization, stochastic optimization, Bayesian optimization, optimal control, discrete optimization, multi-objective optimization, and more. New to the series portfolio include Works at the intersection of optimization and machine learning, artificial intelligence, and quantum computing.

Volumes from this series are indexed by Web of Science, zbMATH, Mathematical Reviews, and SCOPUS.

More information about this series at http://www.springer.com/series/7393

Themistocles M. Rassias
Editor

Nonlinear Analysis, Differential Equations, and Applications

Springer

Editor
Themistocles M. Rassias
Department of Mathematics, Zografou
Campus
National Technical University of Athens
Athens, Greece

ISSN 1931-6828 ISSN 1931-6836 (electronic)
Springer Optimization and Its Applications
ISBN 978-3-030-72562-4 ISBN 978-3-030-72563-1 (eBook)
https://doi.org/10.1007/978-3-030-72563-1

Mathematics Subject Classification: 26-XX, 30-XX, 34-XX, 35-XX, 47-XX

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

Preface

Nonlinear Analysis, Differential Equations, and Applications publishes research and research-survey papers devoted to a broad variety of topics on functional equations, ordinary differential equations, partial differential equations, stochastic differential equations, optimization theory, network games, generalized Nash equilibria, critical point theory, calculus of variations, nonlinear functional analysis, convex analysis, variational inequalities, topology, global differential geometry, curvature flows, perturbation theory, numerical analysis, mathematical finance, and a variety of applications in interdisciplinary topics. More specifically, the book chapters of this volume investigate compound superquadratic functions, the Hyers–Ulam stability of functional equations, edge degenerate pseudo-hyperbolic equations, Kirchhoff wave equation, BMO norms of operators on differential forms, equilibrium points of the perturbed R3BP, complex zeros of solutions to second order differential equations, a higher-order Ginzburg–Landau-type equation, multi-symplectic numerical schemes for differential equations, the Erdős–Rényi network model, strongly m-convex functions, higher order strongly generalized convex functions, factorization and solution of second order differential equations, generalized topologically open sets in relator spaces, graphical mean curvature flow, critical point theory in infinite dimensional spaces using the Leray–Schauder index, non-radial solutions of a supercritical equation in expanding domains, the semi-discrete method for the approximation of the solution of stochastic differential equations, homotopic metric-interval L-contractions in gauge spaces, Rhoades contractions theory, network centrality measures, the Radon transform in three space dimensions via plane integration and applications in positron emission tomography boundary perturbations on medical monitoring and imaging techniques, the KdV-B equation and biomedical applications.

We would like to express our warmest thanks to all the authors who contributed their valuable works for publication in this volume. We would also like to express our sincere appreciation to the staff of Springer for their help throughout the preparation of this book.

Athens, Greece — Themistocles M. Rassias

Contents

On Compound Superquadratic Functions

Shoshana Abramovich

Abstract By using compound superquadratic functions, the inequalities presented in this paper refine and extend Jensen and Jensen-Steffensen inequalities.

2010 Mathematics Subject Classification: 26D15, 26A51, 47A63, 47A64.

1 Introduction

By using compound superquadratic functions, the inequalities presented in this paper refine and extend Jensen and Jensen-Steffensen inequalities.

We start with some definitions, notations and lemmas that are used in this paper.

Definition 1 [3, Definition 2.1] A function $f : [0, \infty) \to \mathbb{R}$ is superquadratic provided that for all $x \in [0, \infty)$ there exists a constant $C_f(x) \in \mathbb{R}$ such that the inequality

$$f(y) - f(x) - C_f(x)(y - x) - f(|y - x|) \geq 0 \tag{1}$$

holds for all $y \in [0, \infty)$.

f is called subquadratic if $-f$ is superquadratic.

Corollary 1 *The functions $f(x) = x^p$, $x \geq 0$ are superquadratic for $p \geq 2$, subquadratic for $0 < p \leq 2$, for which $C_f(x) = px^{p-1} = f'(x)$. When $p = 2$ (1) is an equality.*

Lemma 1 ([3, Lemma 2.1, Lemma 3.2]) *Let f be a superquadratic function with $C_f(x)$ as in Definition 1.*

S. Abramovich (✉)
Department of Mathematics, University of Haifa, Haifa, Israel
e-mail: abramos@math.haifa.ac.il

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_1

Then:

(i) $f(0) \leq 0$,
(ii) *if* $f(0) = f'(0) = 0$ *then* $C_f(x) = f'(x)$ *whenever* f *is differentiable at* $x > 0$, *and* $\frac{f(x)}{x^2}$ *is non-decreasing.*
(iii) *if* $f \geq 0$, *then* f *is convex and* $f(0) = f'(0) = 0$.

Lemma 2 ([3, Lemma 3.1]) *Suppose that* $\varphi : [0, b) [\to \mathbb{R}$ *is continuously differentiable and* $\varphi(0) \leq 0$. *If* $\varphi\prime$ *is superadditive or* $\frac{\varphi\prime(x)}{x}$ *is non-decreasing, then* φ *is superquadratic.*

Corollary 2 ([3]) *Suppose that* f *is superquadratic. Let* $\xi_i \geq 0$, $i = 1, \ldots, m$, *and let* $\overline{\xi} = \sum_{i=1}^{m} p_i \xi_i$ *where* $p_i \geq 0$, $i = 1, \ldots, m$, *and* $\sum_{i=1}^{m} p_i = 1$. *Then*

$$\sum_{i=1}^{m} p_i f(\xi_i) - f\left(\overline{\xi}\right) \geq \sum_{i=1}^{m} p_i f\left(\left|\xi_i - \overline{\xi}\right|\right),$$

and in the special case that $m = 2$, $0 \leq t \leq 1$ *and* $0 \leq a < b < \infty$

$$\begin{aligned} &(1-t) f(a) + t f(b) \\ &\geq f((1-t)a + tb) + (1-t) f(t(b-a)) + t f((1-t)(b-a)) \end{aligned}$$

hold.

Definition 2 ([4]) The real n-tuple that satisfies

$$0 \leq P_j \leq P_n, \quad j = 1, \ldots, n, \quad P_n = 1, \tag{2}$$

$$P_j = \sum_{i=1}^{j} \rho_i, \quad \overline{P}_j = \sum_{i=j}^{n} \rho_i, \quad j = 1, \ldots, n.$$

are called Steffensen's coefficients.

On Jensen-Steffensen's inequality which extends Jensen's inequality for not necessarily positive coefficients see for example [7, page 57]. It states that:

Theorem 1 *Let* $\varphi : I \to \mathbb{R}$ *be convex, then*

$$\varphi\left(\sum_{i=1}^{n} \rho_i \zeta_i\right) \leq \sum_{i=1}^{n} \rho_i \varphi(\zeta_i)$$

holds, where I *is an interval in* $\mathbb{R}$, $\boldsymbol{\zeta} = \left(\zeta_{1,\ldots,}\zeta_n\right)$ *is any monotonic* n*-tuple in* I^n *and* $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)$ *is a real* n*-tuple that satisfies (2).*

After the introduction we obtain in Section 2 refinements of Jensen and Jensen-Steffensen's inequalities via compound superquadratic functions.

On compound functions related to superquadracity see [1] and [2] where the properties of

$$\varphi^{-1}\left(\varphi\left(\sum_{i=1}^{n} t_i a_i\right)+\sum_{i=1}^{n} t_i \varphi\left(\left|a_i-\sum_{j=1}^{n} t_j a_j\right|\right)\right)$$

when $\sum_{i=1}^{n} t_i = 1, t_i \geq 0, a_i \geq 0, i = 1, \ldots, n$, are investigated.

In Section 3 similar to the discussion in [1] and [2] we deal briefly with the behavior of

$$\varphi^{-1}\left(\sum_{i=1}^{n} t_i \varphi\left(a_i\right)-\sum_{i=1}^{n} t_i \varphi\left(\left|a_i-\sum_{j=1}^{n} t_j a_j\right|\right)\right)$$

where $\sum_{i=1}^{n} t_i = 1, t_i \geq 0, a_i \geq 0, i = 1, \ldots, n$.

On compound functions related to convexity see for instance [6, Theorems 3 and 4], [5], and [7] and their references.

2 Refinements of Jensen's Inequality via Compound Functions

In this section we deal with compound superquadratic functions. These functions lead to new inequalities that refine Jensen's inequality and Jensen-Steffensen's inequality.

We start with a theorem which guaranties that the compound function $f = g \circ \varphi$ is non-negative superquadratic.

Theorem 2 *Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be twice differentiable, non-decreasing, convex function and $g(0) = 0$. Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be twice differentiable, superquadratic function, and let f be the compound function defined as $f = g \circ \varphi$. Then, the function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is convex, superquadratic and satisfies $\left(\frac{f'(x)}{x}\right)' \geq 0$ when $\left(\frac{\varphi'(x)}{x}\right)' \geq 0$, or when $g(x) = x^m$, $m \in [2, \infty)$.*

Proof The function f satisfies:

A : $f'(x) = \varphi'(x)\, g'(\varphi(x)) \geq 0$, because φ' and $g' \geq 0$.

B : $f''(x) = \varphi''(x)\, g'(\varphi(x)) + \left(\varphi'(x)\right)^2 g''(\varphi(x)) \geq 0$,

C : $\left(\frac{f'(x)}{x}\right)' = g'(\varphi(x))\left(\frac{\varphi'(x)}{x}\right)' + \frac{1}{x}\left(\varphi'(x)\right)^2 g''(\varphi(x)) \geq 0$, and when $g(x) = x^m$, $m \geq 2$

$$D: \quad \left(\frac{f^{'}(x)}{x}\right)^{'}$$

$$= \frac{mx\varphi^{m-1}(x)\varphi^{"}(x)}{x^2} + \frac{m\varphi^{m-2}(x)\varphi^{'}(x)\left((m-1)x\varphi^{'}(x)-\varphi(x)\right)}{x^2}$$

$$\geq \frac{m\varphi^{m-2}(x)\varphi^{'}(x)\left((m-1)x\varphi^{'}(x)-\varphi(x)\right)}{x^2} \geq 0$$

Indeed, when $\left(\frac{\varphi^{'}(x)}{x}\right)^{'} \geq 0$, and φ is non-negative superquadratic, then by Lemma 1(iii) it is also convex and $\varphi(0) = \varphi^{'}(0) = 0$. The function g is convex, increasing, and non-negative, therefore according to the computations of A and B, f is convex, increasing, $f(0) = f^{'}(0) = 0$, and by C, $\left(\frac{f^{'}(x)}{x}\right)^{'} \geq 0$ holds. Therefore according to Lemma 2 f is also superquadratic.

When $m \geq 2$ the last inequality in D follows from $\left(x\varphi^{'}(x)-\varphi(x)\right)^{'} = x\varphi^{"}(x) \geq 0$, and $\varphi(0) = \varphi^{'}(0) = 0$. Hence $\left(\frac{f^{'}(x)}{x}\right)^{'} \geq 0$, and together with $f(0) = f^{'}(0) = 0$, by Lemma 2, f is superquadratic too. The proof is complete.

We emphasize that there are non-negative superquadratic functions φ which do not satisfy the condition $\left(\frac{\varphi^{'}(x)}{x}\right)^{'} \geq 0$ as shown in [3, Example 3.3].

The proofs in the sequel do not need this condition, but only the conditions that $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing and superquadratic, that $g : \mathbb{R}_+ \to \mathbb{R}_+$ is convex and $g(0) = 0$ and $f = g \circ \varphi$ is non-negative superquadratic or in some cases only convex.

Obviously, all the results in this section hold when the functions φ, f and g satisfy the conditions of Theorem 2.

In the following theorem we present refinements of Jensen's inequality for compound functions f where $f \circ \varphi^{-1}$ are convex. In [5] these functions are called **composite φ^{-1} convex functions.**

Theorem 3 *Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be twice differentiable convex function and $g(0) = 0$. Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be twice differentiable, strictly increasing superquadratic function, and $f = g \circ \varphi$. Then, the following refinements of Jensen's inequality hold when $t_i \geq 0$, $a_i \geq 0$, $i = 1, \ldots, n$, $\sum_{i=1}^{n} t_i = 1$ and $\overline{a} = \sum_{i=1}^{n} t_i a_i$:*

$$f\left(\sum_{i=1}^{n} t_i a_i\right) \leq f \circ \varphi^{-1}\left(\sum_{i=1}^{n} t_i \varphi(a_i) - \sum_{i=1}^{n} t_i \varphi(|a_i - \overline{a}|)\right) \tag{3}$$

$$\leq f \circ \varphi^{-1}\left(\sum_{i=1}^{n} t_i \varphi\left(a_i\right)\right) - f \circ \varphi^{-1}\left(\sum_{i=1}^{n} t_i \varphi\left(|a_i - \overline{a}|\right)\right)$$

$$\leq \sum_{i=1}^{n} t_i f\left(a_i\right) - f \circ \varphi^{-1}\left(\sum_{i=1}^{n} t_i \varphi\left(|a_i - \overline{a}|\right)\right)$$

$$\leq \sum_{i=1}^{n} t_i f\left(a_i\right) - f\left(\sum_{i=1}^{n} t_i \left(|a_i - \overline{a}|\right)\right) \leq \sum_{i=1}^{n} t_i f\left(a_i\right).$$

If f is also superquadratic then

$$f\left(\sum_{i=1}^{n} t_i a_i\right) \leq \sum_{i=1}^{n} t_i f\left(a_i\right) - \sum_{i=1}^{n} t_i f\left(|a_i - \overline{a}|\right) \tag{4}$$

$$\leq \sum_{i=1}^{n} t_i f\left(a_i\right).$$

In particular, when $\frac{\varphi'(x)}{x}$ is increasing or when $f \circ \varphi^{-1}(x) = x^m$, $m \geq 2$ the inequalities (3) and (4) hold.

Proof The function φ is strictly increasing and therefore φ^{-1} exists.

To prove (3) we rewrite f as $f = g \circ \varphi$ and get

$$f\left(\sum_{i=1}^{n} t_i a_i\right) = g \circ \varphi\left(\sum_{i=1}^{n} t_i a_i\right) \tag{5}$$

$$\leq g\left(\sum_{i=1}^{n} t_i \varphi\left(a_i\right) - \sum_{i=1}^{n} t_i \varphi\left(|a_i - \overline{a}|\right)\right)$$

$$\leq g\left(\sum_{i=1}^{n} t_i \varphi\left(a_i\right)\right) - g\left(\sum_{i=1}^{n} t_i \varphi\left(|a_i - \overline{a}|\right)\right)$$

$$= f \circ \varphi^{-1}\left(\sum_{i=1}^{n} t_i \varphi\left(a_i\right)\right) - g\left(\sum_{i=1}^{n} t_i \varphi\left(|a_i - \overline{a}|\right)\right)$$

$$\leq \sum_{i=1}^{n} t_i f\left(a_i\right) - g\left(\sum_{i=1}^{n} t_i \varphi\left(|a_i - \overline{a}|\right)\right)$$

$$\leq \sum_{i=1}^{n} t_i f\left(a_i\right) - f\left(\sum_{i=1}^{n} t_i \left(|a_i - \overline{a}|\right)\right)$$

$$\leq \sum_{i=1}^{n} t_i f(a_i).$$

Indeed, the first inequality in (5) is due to the monotonicity of g and the superquadracity of φ. The second inequality follows from the convexity of g and from $g(0) = 0$. The third inequality is because the inequality $f \circ \varphi^{-1}\left(\sum_{i=1}^{n} t_i \varphi(a_i)\right) \leq \sum_{i=1}^{n} t_i f(a_i)$ which is the same as $g\left(\sum_{i=1}^{n} t_i \varphi(a_i)\right) \leq \sum_{i=1}^{n} t_i g \circ \varphi(a_i)$ is deduced from the convexity of g. The fourth inequality follows from the convexity of φ, the monotonicity of g and because $f = g \circ \varphi$. The last inequality in (5) is due to positivity of f. Inequality (3) is proved because it is just a variation of (5). Inequalities (4) follow directly from the superquadracity of f, Corollary 2 and because of the positivity and convexity of $f = g \circ \varphi$ which is derived directly from the conditions on φ and g (see A and B in Theorem 2). Under the conditions of Theorem 2 f is superquadratic, hence (3) and (4) hold. The proof is complete.

Corollary 3 *It is evident that we can extend Theorem 3 as follows: Let* $g_k : \mathbb{R}_+ \to \mathbb{R}_+$, *be twice differentiable non-decreasing convex function and* $g_k(0) = 0$, $k = 1, \ldots, m$. *Let* $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ *be twice differentiable, strictly increasing superquadratic function. Let* f *be the compound function* $f = g_m \circ g_{m-1} \circ \ldots \circ g_1 \circ \varphi$. *Then, the following refinements of Jensen's inequality hold when* $k = 1, \ldots, m$, $t_i \geq 0$, $a_i \geq 0$, $i = 1, \ldots, n$, $\sum_{i=1}^{n} t_i = 1$ *and* $\overline{a} = \sum_{i=1}^{n} t_i a_i$:

$$\begin{aligned}
&f\left(\sum_{i=1}^{n} t_i a_i\right) \\
&= g_m \circ g_{m-1} \circ \ldots \circ g_1 \circ \varphi\left(\sum_{i=1}^{n} t_i a_i\right) \\
&\leq g_m \circ g_{m-1} \circ \ldots \circ g_1\left(\sum_{i=1}^{n} t_i \varphi(a_i) - \sum_{i=1}^{n} t_i \varphi(|a_i - \overline{a}|)\right) \\
&\leq g_m \circ g_{m-1} \circ \ldots \circ g_k\left(\sum_{i=1}^{n} t_i g_{k-1} \circ g_{k-2} \circ \ldots \circ g_1 \varphi(a_i)\right. \\
&\qquad \left. - g_{k-1} \circ g_{k-2} \circ \ldots \circ g_1\left(\sum_{i=1}^{n} t_i \varphi(|a_i - \overline{a}|)\right)\right) \\
&\leq \sum_{i=1}^{n} t_i g_m \circ g_{m-1} \circ \ldots \circ g_1 \circ \varphi(a_i)
\end{aligned}$$

$$-g_m \circ g_{m-1} \circ \ldots \circ g_1 \circ \varphi \left(\sum_{i=1}^{n} t_i \left(|a_i - \overline{a}| \right) \right)$$

$$= \sum_{i=1}^{n} t_i f\left(a_i\right) - f\left(\sum_{i=1}^{n} t_i \left(|a_i - \overline{a}| \right) \right)$$

$$\leq \sum_{i=1}^{n} t_i f\left(a_i\right).$$

Corollary 4 *Under the same conditions as in Theorem 3, if* $n = 2$, $t_1 = t_2 = \frac{1}{2}$, *and* $a_1 = a < a_n = b$, *the inequalities in (3) and (4) translate into*

$$f\left(\frac{a+b}{2}\right) \leq g\left(\frac{\varphi(a)+\varphi(b)}{2} - \varphi\left(\left|\frac{b-a}{2}\right|\right)\right) \tag{6}$$

$$\leq g\left(\frac{\varphi(a)+\varphi(b)}{2}\right) - g\left(\varphi\left(\left|\frac{b-a}{2}\right|\right)\right)$$

$$= f\left(\varphi^{-1}\left(\frac{\varphi(a)+\varphi(b)}{2}\right)\right) - f\left(\left|\frac{b-a}{2}\right|\right)$$

$$\leq \frac{f(a)+f(b)}{2} - f\left(\left|\frac{b-a}{2}\right|\right) \leq \frac{f(a)+f(b)}{2}.$$

Therefore, if φ *is strictly convex or if* g *is strictly increasing, the inequalities in (6) are strict too. Hence, for* t *close enough to* $\frac{1}{2}$ *the inequalities in*

$$g \circ \varphi\left((1-t)a + tb\right)$$

$$\leq g\left((1-t)\varphi(a) + t\varphi(b) - t\varphi\left((1-t)|b-a|\right) - (1-t)\varphi\left(t|b-a|\right)\right)$$

$$\leq (1-t)\, g \circ \varphi(a) + t g \circ \varphi(b)$$

$$-t g \circ \varphi\left((1-t)|b-a|\right) - (1-t)\, g \circ \varphi\left(t|b-a|\right)$$

$$\leq (1-t)\, g \circ \varphi(a) + t g \circ \varphi(b)$$

hold too.

Corollary 4 is the motivation for proving Theorem 4, in which we deal with superquadratic functions φ and functions f such that the functions $f \circ \varphi^{-1}$ are convex on $x \geq 0$ and satisfy

$$\varphi^{-1}\left(\sum_{i=1}^{2} t_i \varphi\left(a_i\right) - \sum_{i=1}^{2} t_i \varphi\left(\left|a_i - \sum_{j=1}^{2} t_j a_j\right|\right)\right)$$

$$\leq f^{-1}\left(\sum_{i=1}^{2} t_i f(a_i) - \sum_{i=1}^{2} t_i f\left(\left|a_i - \sum_{j=1}^{2} t_j a_j\right|\right)\right)$$

when $\sum_{i=1}^{2} t_i = 1, t_i \geq 0, a_i \geq 0, i = 1, 2$.

Theorem 4 *Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be twice differentiable non-decreasing convex function and $g(0) = 0$. Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be twice differentiable, strictly increasing superquadratic function. Let f defined as $f = g \circ \varphi$ be a superquadratic function. If*

$$\begin{aligned} &g((1-t)\varphi(a) + t\varphi(b)) \\ &\leq (1-t)\, g \circ \varphi(a) + t g \circ \varphi(b) - t(1-t)\, g \circ \varphi(b-a) \end{aligned} \tag{7}$$

when $0 \leq t \leq 1$, $0 \leq a < b < \infty$ is satisfied, then the inequalities in

$$\begin{aligned} &f((1-t)a + tb) = g \circ \varphi((1-t)a + tb) \\ &\leq g((1-t)\varphi(a) + t\varphi(b) - t\varphi((1-t)(b-a)) - (1-t)\varphi(t(b-a))) \\ &\leq (1-t)\, g \circ \varphi(a) + t g \circ \varphi(b) \\ &\qquad - t g \circ \varphi((1-t)(b-a)) - (1-t)\, g \circ \varphi(t(b-a)) \\ &\leq (1-t)\, g \circ \varphi(a) + t g \circ \varphi(b) = (1-t) f(a) + t f(b) \end{aligned} \tag{8}$$

hold. In particular, if also $\frac{\varphi'(x)}{x}$ is increasing, or if $g(x) = x^m$, $m \geq 2$ and (7) holds then (8) holds.

Proof The function φ is non-negative and superquadratic, therefore it is convex and the inequalities in

$$\begin{aligned} 0 &\leq \varphi((1-t)a + tb) \\ &\leq (1-t)\varphi(a) + t\varphi(b) - t\varphi((1-t)|b-a|) - (1-t)\varphi(t|b-a|) \\ &\leq (1-t)\varphi(a) + t\varphi(b) \end{aligned} \tag{9}$$

hold. The function g is non-negative and increasing too and therefore from (9) we get

$$\begin{aligned} 0 &\leq g((1-t)\varphi(a) + t\varphi(b) - t\varphi((1-t)|b-a|) - (1-t)\varphi(t|b-a|)) \\ &\leq g((1-t)\varphi(a) + t\varphi(b)). \end{aligned} \tag{10}$$

The function $f = g \circ \varphi$ is also superquadratic and non-negative and $f(0) = f'(0) = 0$. According to Lemma 1(ii) $\frac{f(kx)}{x^2}$ is increasing when $k > 0$, $x \geq 0$ therefore the inequalities in

$$
\begin{aligned}
0 &\leq tf\left((1-t)(b-a)\right)+(1-t)f\left(t(b-a)\right) \qquad (11)\\
&= t(1-t)^2\frac{f\left((1-t)(b-a)\right)}{(1-t)^2}+(1-t)t^2\frac{f\left(t(b-a)\right)}{t^2}\\
&\leq t(1-t)^2 f(b-a)+(1-t)t^2 f(b-a)\\
&= t(1-t)f(b-a)
\end{aligned}
$$

hold. From (9), (10) and (11), and (7), the inequalities

$$
\begin{aligned}
&(1-t)g\circ\varphi(a)+tg\circ\varphi(b)\\
&-tg\circ\varphi\left((1-t)(b-a)\right)-(1-t)g\circ\varphi\left(t(b-a)\right)\\
&-g\left((1-t)\varphi(a)+t\varphi(b)-t\varphi\left((1-t)(b-a)\right)-(1-t)\varphi\left(t(b-a)\right)\right)\\
\geq\ &(1-t)g\circ\varphi(a)+tg\circ\varphi(b)\\
&-t(1-t)g\circ\varphi(b-a)-g\left((1-t)\varphi(a)+t\varphi(b)\right)\\
\geq\ &0
\end{aligned}
$$

hold from which (8) follows. In the special case that also $\frac{\varphi'(x)}{x}$ is increasing, or $g(x)=x^m$, $m\geq 2$, then according to Theorem 2 the function f is superquadratic and therefore (8) holds. The proof of the theorem is complete.

Lemma 3 *Let $\varphi:\mathbb{R}_+\to\mathbb{R}_+$ be an increasing function and let $0\leq a<b<\infty$. Then for $m=2,3,\ldots$, and $0\leq t\leq 1$ we get that*

$$
\begin{aligned}
&(1-t)\varphi^m(a)+t\varphi^m(b)-\left((1-t)\varphi(a)+t\varphi(b)\right)^m \qquad (12)\\
&=\left(\varphi(b)-\varphi(a)\right)^m t\left(1-t^{m-1}\right)+\sum_{k=2}^{m-1}\varphi^{m-k}(a)\left(\varphi(b)-\varphi(a)\right)^k\left(t-t^k\right)\\
&=t(1-t)\left(\varphi(b)-\varphi(a)\right)^m\sum_{k=0}^{m-2}t^k+\sum_{k=2}^{m-1}\varphi^{m-k}(a)\left(\varphi(b)-\varphi(a)\right)^k\sum_{j=0}^{k-2}t^j\\
&\geq t(1-t)\left(\varphi(b)-\varphi(a)\right)^m.
\end{aligned}
$$

In the special cases that $m=2$ and $m=3$

$$
\begin{aligned}
&(1-t)\varphi^2(a)+t\varphi^2(b)-\left((1-t)\varphi(a)+t\varphi(b)\right)^2\\
&=t(1-t)\left(\varphi(b)-\varphi(a)\right)^2
\end{aligned}
$$

and

$$
(1-t)\varphi^3(a)+t\varphi^3(b)-\left((1-t)\varphi(a)+t\varphi(b)\right)^3
$$

$$= t(1-t)\left((\varphi(b)-\varphi(a))^2(\varphi(b)+2\varphi(a))+t(\varphi(b)-\varphi(a))^3\right)$$
$$\geq t(1-t)(\varphi(b)-\varphi(a))^3$$

respectively hold.

Proof We denote $\varphi(a)=A$, $\varphi(b)-\varphi(a)=x$ and define

$$F(t,x)=(1-t)A^m+t(A+x)^m-(A+tx)^m$$

and using the Newton Binomial Expansion of $(A+x)^m$ and $(A+tx)^m$, we get that

$$F(t,x)=(1-t)A^m+\sum_{k=0}^{m}A^{m-k}x^k\left(t-t^k\right)$$
$$=x^m t\left(1-t^{m-1}\right)+\sum_{k=2}^{m-1}A^{m-k}x^k\left(t-t^k\right)$$
$$=t(1-t)\left(x^m\sum_{k=0}^{m-2}t^k+\sum_{k=2}^{m-1}Ax^k\sum_{j=0}^{k-2}t^j\right)$$
$$\geq t(1-t)x^m\sum_{j=0}^{m-2}t^j\geq t(1-t)x^m,$$

which means that (12) holds.

From D in Theorem 2, Theorem 4 and Lemma 3 we get that:

Theorem 5 *Let $\varphi:\mathbb{R}_+\to\mathbb{R}_+$ be twice differentiable, superquadratic function and let $f(x)=\varphi^m(x)$, $m\geq 2$. Then, f is superquadratic and when $m=2,3,\ldots,$ the following refinements of Jensen's inequality hold for $0\leq t\leq 1$*

$$\begin{aligned}&\varphi^m((1-t)a+tb) \qquad (13)\\ &\leq((1-t)\varphi(a)+t\varphi(b)-t\varphi((1-t)|b-a|)-(1-t)\varphi(t|b-a|))^m\\ &\leq(1-t)\varphi^m(a)+t\varphi^m(b)-t\varphi^m((1-t)|b-a|)-(1-t)\varphi^m(t|b-a|)\\ &.\leq(1-t)\varphi^m(a)+t\varphi^m(b).\end{aligned}$$

Proof When $f(x)=g\circ\varphi(x)=(\varphi(x))^m$, $m\geq 2$ then by D in Theorem 2, f is superquadratic. According to (7) in Theorem 4 and (12) in Lemma 3, in order to satisfy (8) it is enough to prove that the inequality

$$(1-t)\varphi^m(a)+t\varphi^m(b)-((1-t)\varphi(a)+t\varphi(b))^m \qquad (14)$$

$$\geq t\,(1-t)\,\varphi^{m}\,(b-a)$$

holds. This means that according to (12) and Theorem 4, it is enough to prove when $m = 2, 3, \ldots$ that

$$t\,(1-t)\,(\varphi\,(b)-\varphi\,(a))^{m} \geq t\,(1-t)\,\varphi^{m}\,(b-a)\,. \tag{15}$$

holds. Inequality (15) holds because $m > 0$ and φ is a convex function that satisfies $\varphi\,(0) = 0$, hence $\varphi\,(b)-\varphi\,(a) \geq \varphi\,(b-a)-\varphi\,(0) = \varphi\,(b-a)$ and because $0 \leq t \leq 1$. Hence from Theorem 4 we get that for $g\,(x) = x^{m}$, $m = 2, 3, \ldots$, inequality (13) holds. The proof is complete.

Next we deal with two refinements of Theorem 1—the Jensen-Steffensen's inequality for convex functions. First we quote part of the theorem on this subject from [4, Theorem 1] when the functions involved are non-negative superquadratic.

Theorem 6 *Let $\varphi : \mathbb{R}_{+} \to \mathbb{R}_{+}$ be differentiable and superquadratic, let $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ be a non-negative increasing n-tuple in $\mathbb{R}^{n}$, and $\boldsymbol{\rho}$ be a real n-tuple satisfying Steffensen's coefficients, that is ρ_i, $i = 1, \ldots, n$ satisfy the condition in (2). Let $\overline{a}$ be defined by $\overline{a} = \sum_{i=1}^{n} \rho_i a_i$. Then*

$$\begin{aligned} &\sum_{i=1}^{n} \rho_i \varphi\,(a_i) - \varphi\,(\overline{a}) \\ &\geq \left(\sum_{i=1}^{k} P_i + \sum_{i=k+1}^{n} \overline{P_i}\right) \varphi\left(\frac{\sum_{i=1}^{n} \rho_i\,(|a_i-\overline{a}|)}{\sum_{i=1}^{k} P_i + \sum_{i=k+1}^{n} \overline{P_i}}\right) \\ &\geq (n-1)\,\varphi\left(\frac{\sum_{i=1}^{n} \rho_i\,(|a_i-\overline{a}|)}{n-1}\right), \end{aligned} \tag{16}$$

where $k \in \{1, \ldots, n-1\}$ satisfies

$$a_1 \leq a_2 \leq \ldots \leq a_k < \overline{a} < a_{k+1} \leq \ldots \leq a_n.$$

The theorem below is a refinement of Theorem 1 and is a Jensen-Steffensen's inequality for non-negative compound superquadratic functions. Steffensen's coefficients in the case $n = 2$, leads always to $\rho_i > 0$. Therefore we deal in this theorem with $n \geq 3$.

Using the compound function $f = g \circ \varphi$ we get a superquadratic Jensen-Steffensen's type inequality, and therefore we refine Jensen-Steffensen's inequality for convex functions. The proof is similar to the proof of Theorem 3.

Theorem 7 *Let $g : \mathbb{R}_{+} \to \mathbb{R}_{+}$ be twice differentiable non-decreasing convex function and $g\,(0) = 0$. Let $\varphi : \mathbb{R}_{+} \to \mathbb{R}_{+}$ be twice differentiable, strictly increasing superquadratic function, and f defined as $f = g \circ \varphi$ be superquadratic. If a_i and ρ_i, $i = 1, \ldots, n$, satisfy the same conditions as in Theorem 6, then the inequalities*

$$f\left(\sum_{i=1}^{n}\rho_i a_i\right) \le \sum_{i=1}^{n}\rho_i f(a_i) - (n-1) f\left(\frac{\sum_{i=1}^{n}\rho_i (|a_i-\overline{a}|)}{n-1}\right) \tag{17}$$
$$\le \sum_{i=1}^{n}\rho_i f(a_i) - f\left(\frac{\sum_{i=1}^{n}\rho_i (|a_i-\overline{a}|)}{n-1}\right) \le \sum_{i=1}^{n}\rho_i f(a_i)$$

and

$$f\left(\sum_{i=1}^{n}\rho_i a_i\right) \tag{18}$$
$$\le f\circ\varphi^{-1}\left(\sum_{i=1}^{n}\rho_i \varphi(a_i) - (n-1)\varphi\left(\frac{\sum_{i=1}^{n}\rho_i (|a_i-\overline{a}|)}{n-1}\right)\right)$$
$$\le f\circ\varphi^{-1}\left(\sum_{i=1}^{n}\rho_i \varphi(a_i)\right) - f\circ\varphi^{-1}\left((n-1)\varphi\left(\frac{\sum_{i=1}^{n}\rho_i (|a_i-\overline{a}|)}{n-1}\right)\right)$$
$$\le f\circ\varphi^{-1}\left(\sum_{i=1}^{n}\rho_i \varphi(a_i)\right) - f\circ\varphi^{-1}\left(\varphi\left(\frac{\sum_{i=1}^{n}\rho_i (|a_i-\overline{a}|)}{n-1}\right)\right)$$
$$\le \sum_{i=1}^{n}\rho_i f(a_i) - f\left(\frac{\sum_{i=1}^{n}\rho_i (|a_i-\overline{a}|)}{n-1}\right) \le \sum_{i=1}^{n}\rho_i f(a_i)$$

hold. In particular, if also $\frac{\varphi'(x)}{x}$ *is increasing, or if* $f\circ\varphi^{-1}(x) = x^m$, $m \ge 2$, *then (17) and (18) hold.*

Proof Inequality (17) follows from (16) and the superquadracity and positivity of f.

We use the fact that g is convex increasing and $f = g\circ\varphi$ to prove (18). We follow step by step the proof of Theorem 3 and get that:

$$f\left(\sum_{i=1}^{n}\rho_i a_i\right) \le g\left(\sum_{i=1}^{n}\rho_i \varphi(a_i) - (n-1)\varphi\left(\frac{\sum_{i=1}^{n}\rho_i (|a_i-\overline{a}|)}{n-1}\right)\right)$$
$$\le g\left(\sum_{i=1}^{n}\rho_i \varphi(a_i)\right) - g\left((n-1)\varphi\left(\frac{\sum_{i=1}^{n}\rho_i (|a_i-\overline{a}|)}{n-1}\right)\right)$$
$$\le g\left(\sum_{i=1}^{n}\rho_i \varphi(a_i)\right) - g\circ\varphi\left(\frac{\sum_{i=1}^{n}\rho_i (|a_i-\overline{a}|)}{n-1}\right)$$
$$\le \sum_{i=1}^{n}\rho_i g\circ\varphi(a_i) - g\circ\varphi\left(\frac{\sum_{i=1}^{n}\rho_i (|a_i-\overline{a}|)}{n-1}\right)$$

$$\leq \sum_{i=1}^{n} \rho_i f(a_i) - f\left(\frac{\sum_{i=1}^{n} \rho_i (|a_i - \overline{a}|)}{n-1}\right) \leq \sum_{i=1}^{n} \rho_i f(a_i).$$

Therefore (18) is proved. If also $\frac{\varphi'(x)}{x}$ is increasing or $f \circ \varphi^{-1}(x) = x^m$, $m \geq 2$, then according to Theorem 2 $f = g \circ \varphi$ is superquadratic and therefore (17) and (18) hold. The proof of Theorem 7 is complete.

3 The Quasi-Mean $F_f(\mathbf{x}, \boldsymbol{\lambda})$

In [1] and [2] the following quasi-mean

$$W_f(\mathbf{x}, \boldsymbol{\lambda}) = f^{-1}\left(f\left(\sum_{r=1}^{n} \lambda_r x_r\right) + \sum_{r=1}^{n} \lambda_r f\left(\left|x_r - \sum_{i=1}^{n} \lambda_i x_i\right|\right)\right)$$

$$\sum_{r=1}^{n} \lambda_r = 1, \qquad \lambda_r \geq 0,\ x_{r,} \geq 0,\ r = 1, \ldots, n,$$

is discussed. Similarly, the quasi-mean

$$F_f(\mathbf{x}, \boldsymbol{\lambda}) = f^{-1}\left(\sum_{i=1}^{n} \lambda_i f(x_i) - \sum_{i=1}^{n} \lambda_i f\left(\left|x_i - \sum_{j=1}^{n} \lambda_j x_j\right|\right)\right)$$

$$\sum_{r=1}^{n} \lambda_r = 1, \qquad \lambda_r \geq 0,\ x_{r,} \geq 0,\ r = 1, \ldots, n,$$

is discussed here.

Both $F_f(\mathbf{x}, \boldsymbol{\lambda})$ and $W_f(\mathbf{x}, \boldsymbol{\lambda})$ consist the building blocks of superquadratic functions (see Corollary 2).

First we quote from [1] about $W_f(x_1, x_2)$:

Definition 3 ([1, Definition 1]) Let a strictly increasing convex function f be defined on $[0, b)$, $0 < b \leq \infty$, and let $f(0) = 0$. For such f we define the quasi-mean $W_f(x_1, x_2)$ as

$$W_f(x_1, x_2) = f^{-1}\left(f\left(\frac{x_1 + x_2}{2}\right) + f\left(\left|\frac{x_1 - x_2}{2}\right|\right)\right).$$

In the special case that $f(x) = x^p$, $W_p(x_1, x_2)$ is defined as

$$W_p(x_1, x_2) = \left(\left(\frac{x_1 + x_2}{2}\right)^p + \left(\left|\frac{x_1 - x_2}{2}\right|\right)^p\right)^{\frac{1}{p}}$$

for $x_1, x_2 \in [0, b)$, $p \geq 1$.

In Lemma 4 it is proved that $W_f(x_1, x_2)$ is a quasi-mean:

Lemma 4 ([1, Lemma 1]) *Under the conditions of Definition 3 on f, $W_f(x_1, x_2)$ is symmetric and satisfies:*

(a) $W_f(x, x) = x$, $x \in [0, b)$,
(b) $x_1 \leq W_f(x_1, x_2) \leq x_2$, $0 \leq x_1 \leq x_2 \leq b$.
When $f(x) = x^p$, $p \geq 1$, $W_p(x_1, x_2)$ satisfies
(c) $W_p(\lambda x_1, \lambda x_2) = \lambda W_p(x_1, x_2)$, $\lambda \geq 0$, $0 \leq x_1, x_2 \leq b$.

Similar to Definition 3 and Lemma 4, we discuss $F_f(x_1, x_2)$ which is related to the theorems in this paper.

Definition 4 Let a strictly increasing convex function f be defined on $[0, b)$, $0 < b \leq \infty$, and let $f(0) = 0$. For such f we define the quasi-mean $F_f(x_1, x_2)$ as

$$F_f(x_1, x_2) = f^{-1}\left(\frac{f(x_1) + f(x_2)}{2} - f\left(\left|\frac{x_1 - x_2}{2}\right|\right)\right).$$

In the special case that $f(x) = x^p$, $F_p(x_1, x_2)$ is defined as

$$F_p(x_1, x_2) = \left(\frac{x_1^p + x_2^p}{2} - \left(\left|\frac{x_1 - x_2}{2}\right|\right)^p\right)^{\frac{1}{p}}$$

for $x_1, x_2 \in [0, b)$, $p \geq 1$.

Lemma 5 *Under the conditions of Definition 4 on f, $F_f(x_1, x_2)$ is symmetric and satisfies:*

(a) $F_f(x, x) = x$, $x \in [0, b)$,
(b) $x_1 \leq F_f(x_1, x_2) \leq x_2$, $0 \leq x_1 \leq x_2 \leq b$.
When $f(x) = x^p$, $p \geq 1$, $F_p(x_1, x_2)$ satisfies
(c) $F_p(\lambda x_1, \lambda x_2) = \lambda F_p(x_1, x_2)$, $\lambda \geq 0$, $0 \leq x_1, x_2 \leq b$.

Proof Properties (a) and (c) are obvious. To prove that property (b) holds we have to show that

$$x_1 \leq f^{-1}\left(\frac{f(x_1) + f(x_2)}{2} - f\left(\left|\frac{x_1 - x_2}{2}\right|\right)\right) \leq x_2. \tag{19}$$

As f is strictly increasing (19) is equivalent to

$$f(x_1) \leq \frac{f(x_1) + f(x_2)}{2} - f\left(\left|\frac{x_1 - x_2}{2}\right|\right) \leq f(x_2) \tag{20}$$

and because f is non-negative and increasing the right hand-side of (20) follows from

$$f(x_2) \geq f(x_1) - 2f\left(\frac{x_2 - x_1}{2}\right),$$

and therefore the right hand-side of (19) holds.

To prove the left hand-side of (19) we deal with the left hand-side of (20) which is the same as to prove the inequality

$$f(x_2) - f(x_1) \geq 2f\left(\frac{x_2 - x_1}{2}\right). \tag{21}$$

We show that

$$f(x_2) - f(x_1) \geq f(x_2 - x_1) \geq 2f\left(\frac{x_2 - x_1}{2}\right). \tag{22}$$

Indeed the right hand-side inequality of (22) follows from the fact that when f is twice differentiable, convex and $f(0) = 0$, then $\frac{f(kx)}{x}$ is increasing for $x > 0$. From (22) inequality (21) follows, and therefore the left hand-side inequality of (19) holds. The proof of the lemma is complete.

References

1. S. Abramovich, Quasi-arithmetic means and superquadracity. J. Math. Inequal. **9**, 1157–1168 (2015)
2. S. Abramovich, The behavior of difference of two means, in *Developments in Functional Equations and Related Topics*, ed. by J. Brzdęk, K. Ciepliński, Th.M. Rassias (Springer, Cham, 2017), pp. 1–15
3. S. Abramovich, G. Jameson, G. Sinnamon, Refining Jensen's inequality. Bull. Math. Soc. Math. Roum (Novel Series) **47**(95), 3–14 (2004)
4. S. Abramovich, S. Banić, M. Matić, J. Pečarić, Jensen Steffensen's and related inequalities, for superquadratic functions. Math. Inequal. Appl. **11**, 23–41 (2008)
5. S.S. Dragomir, Inequalities of Hermite-Hadamard type for composite convex functions. RGMIA. Res. Rap. Coll. **23**, Art. 39, 21 pp. (2018)
6. J. Matkowski, A functional inequality characterizing convex functions, conjugacy and generalization of Hölder's and Minkowski's inequalities. Aequat. Math. **40**, 168–180 (1990)
7. J.E. Pečarić, F. Proschan, Y.L. Tong, *Convex Functions, Partial Ordering and Statistical Applications* (Academic Press, New York, 1992)

Best Hyers–Ulam Stability Constants on a Time Scale with Discrete Core and Continuous Periphery

Douglas R. Anderson and Masakazu Onitsuka

Abstract Consider a time scale consisting of a discrete core with uniform step size, augmented with a continuous-interval periphery. On this time scale, we determine the best constants for the Hyers–Ulam stability of a first-order dynamic equation with complex constant coefficient, based on the placement of the complex coefficient in the complex plane, with respect to the imaginary axis and the Hilger circle. These best constants are then related to known results for the special cases of completely continuous and uniformly discrete time scales.

1 Introduction

In this paper we explore the Hyers–Ulam stability of a certain dynamic equation on a new time scale with a discrete, uniform core and continuous periphery. Ulam inaugurated this type of stability [37], followed by Hyers [22] and Rassias [34]. Since then, there has been wide-spread interest in this type of stability, including for difference equations, recurrence relations, h-difference equations, quantum equations, and dynamic equations on time scales. For early papers on difference equations, see Popa [31, 32]; more current works include Anderson and Onitsuka [5, 6], Baias and Popa [13], Brzdęk and Wójcik [16], Onitsuka [29, 30], Rasouli, Abbaszadeh, and Eshaghi [33], Xu and Brzdęk [38]. A related monograph is Brzdęk, Popa, Raşa, and Xu [17]. Quantum equations and Hyers–Ulam stability are investigated in Anderson and Onitsuka [7, 8]. For some work on matrix and nonlinear difference equations, see Jung and Nam [24, 25], and Nam [26–28]. For early papers on time scales, see András and Mészáros [12], Hua, Li, and Feng [21];

D. R. Anderson (✉)
Concordia College, Department of Mathematics, Moorhead, MN, USA
e-mail: andersod@cord.edu

M. Onitsuka
Okayama University of Science, Department of Applied Mathematics, Okayama, Japan
e-mail: onitsuka@xmath.ous.ac.jp

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_2

contemporary results include Anderson [2], Anderson, Jennissen, and Montplaisir [10], Anderson and Onitsuka [3, 4, 9], Shen [35], Shen and Li [36]. For recent papers with non-constant or periodic coefficients, see Anderson [1], Anderson, Onitsuka, and Rassias [11], Baias, Blaga, and Popa [14], Buşe, Lupulescu, and O'Regan [19], Buşe, O'Regan, and Saierli [18].

This work will proceed as follows. In Section 2, we will define the time scale with discrete core and continuous periphery, introduce the basic derivative and exponential function for this time scale, and define Hyers–Ulam stability for the dynamic equation with a complex constant coefficient. In Section 3, we establish the best Hyers–Ulam stability constants in Theorem 5, based on the location of the complex coefficient with respect to the imaginary axis, and for negative real part, with respect to the left Hilger circle. If we expand the discrete core to all of $h\mathbb{Z}$, or shrink it to recover the continuum $\mathbb{R}$, we are able to relate our new results with the current literature in the field. As we do this, an interesting case arises when the real part of the complex coefficient is negative but it lies outside the Hilger circle; this case is explored in Section 4. After that, we provide a brief conclusion and future direction.

2 Time Scale with Discrete Core and Continuous Periphery

Let $\mathbb{N}_0$ denote the non-negative integers $\{0, 1, 2, \ldots\}$, let $m \in \mathbb{N}_0$, and let $h > 0$. Define the time scale with discrete core and continuous periphery via

$$\mathbb{T}_{hm} := (-\infty, -hm) \cup \{-hm, \ldots, -h, 0, h, \ldots, hm\} \cup (hm, \infty).$$

Here, $h > 0$ is the uniform step size in the discrete core, with discrete spread $m \in \mathbb{N}_0$ out to the continuous periphery. Define the graininess function $\mu : \mathbb{T}_{hm} \to \mathbb{R}$ via

$$\mu(t) = \begin{cases} 0 & : t \in (-\infty, -hm) \cup [hm, \infty), \\ h & : t \in \{-hm, \ldots, -h, 0, h, \ldots, h(m-1)\}. \end{cases}$$

As $h \to 0$, or if $m = 0$, we have $\mathbb{T}_{0,m} = \mathbb{T}_{h,0} = \mathbb{R}$, and we recover results for classical differential equations; as $m \to \infty$ for fixed $h > 0$, we have $\mathbb{T}_{h,\infty} = h\mathbb{Z}$ and we recover results for standard h-difference equations.

In this section we introduce the first-order linear homogeneous equation with constant complex-valued coefficient

$$x^{\Delta}(t) - \lambda x(t) = 0, \quad \lambda \in \mathbb{C} \backslash \left\{ \frac{-1}{h} \right\}, \quad t \in \mathbb{T}_{hm}, \tag{1}$$

where

$$x^{\Delta}(t) := \begin{cases} \frac{d}{dt}x(t) & : t \in (-\infty, -hm) \cup [hm, \infty) \\ \frac{x(t+h)-x(t)}{h} & : t \in \{-hm, \ldots, -h, 0, h, \ldots, h(m-1)\}. \end{cases}$$

Lemma 1 (Exponential Function) *Fix $h > 0$. For $t \in \mathbb{T}_{hm}$, define the function*

$$e_{\lambda}(t, 0) := \begin{cases} (1 + h\lambda)^{-m} e^{\lambda(t+hm)} & : t \in (-\infty, -hm) \\ (1 + h\lambda)^{\frac{t}{h}} & : t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\} \\ (1 + h\lambda)^{m} e^{\lambda(t-hm)} & : t \in (hm, \infty). \end{cases} \tag{2}$$

Then, $x(t) = x_0 e_{\lambda}(t, 0)$ for $e_{\lambda}(t, 0)$ given in (2) is the unique solution of (1) satisfying $x(0) = x_0 \in \mathbb{C}$.

3 Best Constants for First-Order Equations with Constant Complex Coefficient

In this section, we consider on $\mathbb{T}_{hm}$ the Hyers–Ulam stability of (1), defined as follows.

Definition 1 (HUS) Let $\varepsilon > 0$ be arbitrary. Equation (1) has Hyers–Ulam stability (HUS) if and only if given $\phi : \mathbb{T}_{hm} \to \mathbb{C}$ satisfying $|\phi^{\Delta}(t) - \lambda\phi(t)| \leq \varepsilon$ for all $t \in \mathbb{T}_{hm}$, there exists a solution $x : \mathbb{T}_{hm} \to \mathbb{C}$ of (1) and a constant $K > 0$ such that $|\phi(t) - x(t)| \leq K\varepsilon$ for all $t \in \mathbb{T}_{hm}$. Such a constant K is called an HUS constant for (1) on $\mathbb{T}_{hm}$.

Theorem 1 *Let $\lambda \in \mathbb{C} \backslash \left\{\frac{-1}{h}\right\}$ with $\mathrm{Re}(\lambda) > 0$. Let $\varepsilon > 0$ be a fixed arbitrary constant, and let ϕ be a function on $\mathbb{T}_{hm}$ satisfying the inequality*

$$\left|\phi^{\Delta}(t) - \lambda\phi(t)\right| \leq \varepsilon, \qquad t \in \mathbb{T}_{hm}.$$

Then, $\lim_{t\to\infty} \frac{\phi(t)}{e_{\lambda}(t,0)}$ exists, and the function x given by

$$x(t) := \left(\lim_{t\to\infty} \frac{\phi(t)}{e_{\lambda}(t,0)}\right) e_{\lambda}(t, 0)$$

is the unique solution of (1) with

$$|\phi(t) - x(t)| \leq \varepsilon \left(\frac{1}{\mathrm{Re}(\lambda)}\right)$$

for all $t \in \mathbb{T}_{hm}$.

Proof Let $\lambda \in \mathbb{C}\backslash\{\frac{-1}{h}\}$ with $\mathrm{Re}(\lambda) > 0$. Throughout this proof, as $|\phi^{\Delta}(t) - \lambda\phi(t)| \le \varepsilon$ for all $t \in \mathbb{T}_{hm}$, there exists a function $q : \mathbb{T}_{hm} \to \mathbb{C}$ such that

$$\phi^{\Delta}(t) - \lambda\phi(t) = q(t), \quad |q(t)| \le \varepsilon$$

for all $t \in \mathbb{T}_{hm}$. The variation of constants formula then yields

$$\phi(t) = \phi_0 e_\lambda(t,0) + e_\lambda(t,0)\int_0^t \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau.$$

Since $\mathrm{Re}(\lambda) > 0$ and $|q(t)| \le \varepsilon$, we can rewrite ϕ as

$$\phi(t) = \left(\phi_0 + \int_0^\infty \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau\right)e_\lambda(t,0) - e_\lambda(t,0)\int_t^\infty \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau, \tag{3}$$

where

$$x_0 := \phi_0 + \int_0^\infty \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau \in \mathbb{C}$$

exists and is finite. Clearly

$$x(t) := x_0 e_\lambda(t,0), \quad t \in \mathbb{T}_{hm}$$

is a solution of (1), and

$$\lim_{t\to\infty}\frac{\phi(t)}{e_\lambda(t,0)} = \lim_{t\to\infty}\left(\phi_0 + \int_0^t \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau\right) = x_0$$

exists, so

$$x(t) = \left(\lim_{t\to\infty}\frac{\phi(t)}{e_\lambda(t,0)}\right)e_\lambda(t,0).$$

We take into account three cases based on the three branches of the exponential function in (2).

(a). For $\mathrm{Re}(\lambda) > 0$ and $t \in (hm, \infty)$, using (3) we have that

$$\begin{aligned}
|\phi(t) - x(t)| &= \left|-e_\lambda(t,0)\int_t^\infty \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau\right| \\
&\le \varepsilon|e_\lambda(t,0)|\int_t^\infty \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|} \\
&= \varepsilon|1+h\lambda|^m e^{\mathrm{Re}(\lambda)(t-hm)}\int_t^\infty \frac{d\tau}{|1+h\lambda|^m e^{\mathrm{Re}(\lambda)(\tau-hm)}} \\
&= \frac{\varepsilon}{\mathrm{Re}(\lambda)}
\end{aligned}$$

holds for all $t \in (hm, \infty)$.

(b). For $\mathrm{Re}(\lambda) > 0$ and $t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\}$, using (3) we have

$$\begin{aligned}
|\phi(t) - x(t)| &\le \varepsilon |e_\lambda(t,0)| \int_t^\infty \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|} \\
&= \varepsilon |1+h\lambda|^{\frac{t}{h}} \left(\int_t^{hm} + \int_{hm}^\infty \right) \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|} \\
&= \varepsilon |1+h\lambda|^{\frac{t}{h}} \left(\sum_{j=\frac{t}{h}}^{m-1} \frac{h}{|1+h\lambda|^{j+1}} + \int_{hm}^\infty \frac{d\tau}{|1+h\lambda|^m e^{\mathrm{Re}(\lambda)(\tau-hm)}} \right) \\
&= \varepsilon |1+h\lambda|^{\frac{t}{h}} \left(\frac{h\left(|1+h\lambda|^{-\frac{t}{h}} - |1+h\lambda|^{-m}\right)}{|1+h\lambda|-1} + \frac{1}{|1+h\lambda|^m \,\mathrm{Re}(\lambda)} \right) \\
&= \varepsilon \left(\frac{h}{|1+h\lambda|-1} + |1+h\lambda|^{\frac{t}{h}-m} \left(\frac{1}{\mathrm{Re}(\lambda)} - \frac{h}{|1+h\lambda|-1} \right) \right) \\
&\le \frac{\varepsilon}{\mathrm{Re}(\lambda)}
\end{aligned}$$

for all $t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\}$, as $\frac{t}{h} \le m$ and $\frac{1}{\mathrm{Re}(\lambda)} \ge \frac{h}{|1+h\lambda|-1}$ for $\mathrm{Re}(\lambda) > 0$ and $h > 0$.

(c). For $\mathrm{Re}(\lambda) > 0$ and $t \in (-\infty, -hm)$, using (3) we have

$$\begin{aligned}
|\phi(t) - x(t)| &\le \varepsilon |e_\lambda(t,0)| \int_t^\infty \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|} \\
&= \frac{\varepsilon\, e^{\mathrm{Re}(\lambda)(t+hm)}}{|1+h\lambda|^m} \left(\int_t^{-hm} + \int_{-hm}^{hm} + \int_{hm}^\infty \right) \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|} \\
&= \frac{\varepsilon\, e^{\mathrm{Re}(\lambda)(t+hm)}}{|1+h\lambda|^{2m}} \left(\frac{1}{\mathrm{Re}(\lambda)} + \frac{\left(e^{-\mathrm{Re}(\lambda)(t+hm)} - 1\right)}{|1+h\lambda|^{-2m}\,\mathrm{Re}(\lambda)} \right. \\
&\quad \left. + \frac{h\left(|1+h\lambda|^{2m} - 1\right)}{|1+h\lambda|-1} \right) \\
&= \varepsilon \left\{ \frac{1}{\mathrm{Re}(\lambda)} + e^{\mathrm{Re}(\lambda)(t+hm)} \left(\frac{1}{|1+h\lambda|^{2m}\,\mathrm{Re}(\lambda)} \right. \right. \\
&\quad \left. \left. + \left(\frac{h}{|1+h\lambda|-1} \right) \left(1 - \frac{1}{|1+h\lambda|^{2m}} \right) - \frac{1}{\mathrm{Re}(\lambda)} \right) \right\} \\
&\le \frac{\varepsilon}{\mathrm{Re}(\lambda)}
\end{aligned}$$

for all $t \in (-\infty, -hm)$, as $t < -hm$, and the expression inside the square brackets is negative.

We next show that x is the unique solution of (1) such that $|\phi(t) - x(t)| \le K\varepsilon := \frac{1}{\mathrm{Re}(\lambda)}\varepsilon$ for all $t \in \mathbb{T}_{hm}$. Suppose $\phi : \mathbb{T}_{hm} \to \mathbb{C}$ is an approximate solution of (1) such that

$$\left|\phi^{\Delta}(t) - \lambda\phi(t)\right| \le \varepsilon \text{ for all } t \in \mathbb{T}_{hm}$$

for some $\varepsilon > 0$. Suppose further that $x_1, x_2 : \mathbb{T}_{hm} \to \mathbb{C}$ are two different solutions of (1) such that $|\phi(t) - x_j(t)| \le K\varepsilon$ for all $t \in \mathbb{T}_{hm}$, for $j = 1, 2$. Then, we have for constants $c_j \in \mathbb{C}$ that

$$x_j(t) = c_j e_\lambda(t, 0), \quad c_1 \ne c_2,$$

and

$$|c_1 - c_2| \cdot |e_\lambda(t, 0)| = |x_1(t) - x_2(t)| \le |x_1(t) - \phi(t)| + |\phi(t) - x_2(t)| \le 2K\varepsilon;$$

letting $t \to \infty$ yields $\infty < 2K\varepsilon$, a contradiction. Consequently, x is the unique solution of (1) such that $|\phi(t) - x(t)| \le \frac{\varepsilon}{\mathrm{Re}(\lambda)}$ for all $t \in \mathbb{T}_{hm}$. This completes the proof. □

Theorem 2 *Let $\lambda \in \mathbb{C}\backslash\left\{\frac{-1}{h}\right\}$ with $\mathrm{Re}(\lambda) < 0$. Let $\varepsilon > 0$ be a fixed arbitrary constant, and let ϕ be a function on $\mathbb{T}_{hm}$ satisfying the inequality*

$$\left|\phi^{\Delta}(t) - \lambda\phi(t)\right| \le \varepsilon, \qquad t \in \mathbb{T}_{hm}.$$

Then, $\lim_{t\to-\infty} \frac{\phi(t)}{e_\lambda(t,0)}$ exists, and the function x given by

$$x(t) := \left(\lim_{t\to-\infty} \frac{\phi(t)}{e_\lambda(t,0)}\right) e_\lambda(t, 0)$$

is the unique solution of (1) with $|\phi(t) - x(t)| \le K\varepsilon$ for all $t \in \mathbb{T}_{hm}$, where

$$K := \begin{cases} \frac{-1}{\mathrm{Re}(\lambda)} + 2hm & : |1 + h\lambda| = 1 \\ \max\left\{\frac{-1}{\mathrm{Re}(\lambda)}, \frac{h}{1-|1+h\lambda|} + |1 + h\lambda|^{2m}\left(\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)}\right)\right\} & : |1 + h\lambda| \ne 1. \end{cases} \tag{4}$$

In particular, the following holds.

(i) If $t \in (-\infty, -hm)$, then

$$|\phi(t) - x(t)| \le \varepsilon\left(\frac{-1}{\mathrm{Re}(\lambda)}\right)$$

for all $t \in (-\infty, -hm)$.

(ii) If $|1+h\lambda| = 1$ *and* $t \in \{-hm, \dots, -h, 0, h, \dots, hm\} \cup (hm, \infty)$*, then*

$$|\phi(t) - x(t)| \leq \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} + 2hm \right)$$

for all $t \in \{-hm, \dots, -h, 0, h, \dots, hm\} \cup (hm, \infty)$.

(iii) If $0 < |1+h\lambda| < 1$ *and* $t \in \{-hm, \dots, -h, 0, h, \dots, hm\} \cup (hm, \infty)$*, then*

$$|\phi(t) - x(t)| \leq \varepsilon \left(\frac{h}{1-|1+h\lambda|} + |1+h\lambda|^{2m} \left(\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right)$$

for all $t \in \{-hm, \dots, -h, 0, h, \dots, hm\} \cup (hm, \infty)$.

(iv) If $|1+h\lambda| > 1$ *and* $t \in \{-hm, \dots, -h, 0, h, \dots, hm\} \cup (hm, \infty)$*, then*

$$|\phi(t) - x(t)| \leq \varepsilon \left(\frac{h}{1-|1+h\lambda|} + |1+h\lambda|^{2m} \left(\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right)$$

for all $t \in \{-hm, \dots, -h, 0, h, \dots, hm\} \cup (hm, \infty)$.

Proof Let $\lambda \in \mathbb{C} \backslash \{\frac{-1}{h}\}$ with $\mathrm{Re}(\lambda) < 0$. Supposing $|\phi^{\Delta}(t) - \lambda\phi(t)| \leq \varepsilon$ for all $t \in \mathbb{T}_{hm}$, there exists a function $q : \mathbb{T}_{hm} \to \mathbb{C}$ such that

$$\phi^{\Delta}(t) - \lambda\phi(t) = q(t), \quad |q(t)| \leq \varepsilon$$

for all $t \in \mathbb{T}_{hm}$. Then, we have

$$\phi(t) = \phi_0 e_\lambda(t, 0) + e_\lambda(t, 0) \int_0^t \frac{q(\tau)}{e_\lambda(\sigma(\tau), 0)} \Delta\tau.$$

Since $\mathrm{Re}(\lambda) < 0$ and $|q(t)| \leq \varepsilon$, we can rewrite ϕ as

$$\phi(t) = \left(\phi_0 - \int_{-\infty}^0 \frac{q(\tau)}{e_\lambda(\sigma(\tau), 0)} \Delta\tau \right) e_\lambda(t, 0) + e_\lambda(t, 0) \int_{-\infty}^t \frac{q(\tau)}{e_\lambda(\sigma(\tau), 0)} \Delta\tau, \tag{5}$$

where

$$x_0 := \phi_0 - \int_{-\infty}^0 \frac{q(\tau)}{e_\lambda(\sigma(\tau), 0)} \Delta\tau \in \mathbb{C}$$

exists and is finite. As in the previous case,

$$x(t) := x_0 e_\lambda(t, 0), \quad t \in \mathbb{T}_{hm}$$

is a solution of (1), and

$$\lim_{t \to -\infty} \frac{\phi(t)}{e_\lambda(t, 0)} = \lim_{t \to -\infty} \left(\phi_0 - \int_t^0 \frac{q(\tau)}{e_\lambda(\sigma(\tau), 0)} \Delta\tau \right) = x_0$$

exists, so

$$x(t) = \left(\lim_{t \to -\infty} \frac{\phi(t)}{e_\lambda(t,0)} \right) e_\lambda(t,0).$$

We again work our way through the three cases based on the three branches of the exponential function in (2).

(i). For $\mathrm{Re}(\lambda) < 0$ and $t \in (-\infty, -hm)$, using (5) we have

$$\begin{aligned} |\phi(t) - x(t)| &= \left| e_\lambda(t,0) \int_{-\infty}^{t} \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)} \Delta\tau \right| \\ &\le \varepsilon |e_\lambda(t,0)| \int_{-\infty}^{t} \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|} \\ &= \varepsilon |1+h\lambda|^{-m} e^{\mathrm{Re}(\lambda)(t+hm)} \int_{-\infty}^{t} \frac{|1+h\lambda|^m d\tau}{e^{\mathrm{Re}(\lambda)(\tau+hm)}} \\ &= -\frac{\varepsilon}{\mathrm{Re}(\lambda)} \end{aligned}$$

holds for $\mathrm{Re}(\lambda) < 0$ and for all $t \in (-\infty, -hm)$.

(ii) (a). For $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| = 1$ and $t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\}$, using (5) we have

$$\begin{aligned} |\phi(t) - x(t)| &\le \varepsilon |e_\lambda(t,0)| \int_{-\infty}^{t} \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|} \\ &= \varepsilon \left(\int_{-\infty}^{-hm} + \int_{-hm}^{t} \right) \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|} \\ &= \varepsilon \left(\int_{-\infty}^{-hm} \frac{d\tau}{e^{\mathrm{Re}(\lambda)(\tau+hm)}} + \sum_{j=-m}^{\frac{t-h}{h}} h \right) \\ &= \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} + hm + t \right) \\ &\le \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} + 2hm \right), \end{aligned}$$

for $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| = 1$ and $t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\}$.

(ii) (b). For $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| = 1$ and $t \in (hm, \infty)$, using (5) we have

$$|\phi(t) - x(t)| \le \varepsilon\, e^{\mathrm{Re}(\lambda)(t-hm)} \left(\int_{-\infty}^{-hm} + \int_{-hm}^{hm} + \int_{hm}^{t} \right) \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|}$$

$$= \varepsilon\, e^{\mathrm{Re}(\lambda)(t-hm)} \left(2hm - \frac{e^{-\mathrm{Re}(\lambda)(t-hm)}}{\mathrm{Re}(\lambda)}\right)$$

$$\leq \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} + 2hm\right),$$

as $t > hm$ and $\mathrm{Re}(\lambda) < 0$.

(iii) (a). For $\mathrm{Re}(\lambda) < 0$ with $0 < |1 + h\lambda| < 1$ and $t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\}$, using (5) we have

$$|\phi(t) - x(t)| \leq \varepsilon |e_\lambda(t, 0)| \int_{-\infty}^{t} \frac{\Delta\tau}{|e_\lambda(\sigma(\tau), 0)|}$$

$$= \varepsilon |1 + h\lambda|^{\frac{t}{h}} \left(\int_{-\infty}^{-hm} + \int_{-hm}^{t}\right) \frac{\Delta\tau}{|e_\lambda(\sigma(\tau), 0)|}$$

$$= \varepsilon |1 + h\lambda|^{\frac{t}{h}} \left(\int_{-\infty}^{-hm} \frac{|1 + h\lambda|^m d\tau}{e^{\mathrm{Re}(\lambda)(\tau+hm)}} + \sum_{j=-m}^{\frac{t-h}{h}} \frac{h}{|1 + h\lambda|^{j+1}}\right)$$

$$= \varepsilon |1 + h\lambda|^{\frac{t}{h}} \left(\frac{h\left(|1 + h\lambda|^m - |1 + h\lambda|^{-\frac{t}{h}}\right)}{|1 + h\lambda| - 1} - \frac{|1 + h\lambda|^m}{\mathrm{Re}(\lambda)}\right)$$

$$= \varepsilon \left(\frac{h}{1 - |1 + h\lambda|} + |1 + h\lambda|^{\frac{t}{h}+m} \left(\frac{h}{|1 + h\lambda| - 1} - \frac{1}{\mathrm{Re}(\lambda)}\right)\right)$$

$$\leq \varepsilon \left(\frac{h}{1 - |1 + h\lambda|} + |1 + h\lambda|^{2m} \left(\frac{h}{|1 + h\lambda| - 1} - \frac{1}{\mathrm{Re}(\lambda)}\right)\right),$$

as $\frac{t}{h} \leq m$ and $\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} \leq 0$ for $\mathrm{Re}(\lambda) < 0$ with $0 < |1 + h\lambda| < 1$ and $h > 0$.

(iii) (b). For $\mathrm{Re}(\lambda) < 0$ with $0 < |1 + h\lambda| < 1$ and $t \in (hm, \infty)$, using (5) we have

$$|\phi(t) - x(t)| \leq \frac{\varepsilon e^{\mathrm{Re}(\lambda)(t-hm)}}{|1 + h\lambda|^{-m}} \left(\int_{-\infty}^{-hm} + \int_{-hm}^{hm} + \int_{hm}^{t}\right) \frac{\Delta\tau}{|e_\lambda(\sigma(\tau), 0)|}$$

$$= \varepsilon e^{\mathrm{Re}(\lambda)(t-hm)} \left(\frac{1 - e^{-\mathrm{Re}(\lambda)(t-hm)} - |1 + h\lambda|^{2m}}{\mathrm{Re}(\lambda)}\right.$$

$$\left. + \frac{h\left(|1 + h\lambda|^{2m} - 1\right)}{|1 + h\lambda| - 1}\right)$$

$$= \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} + e^{\mathrm{Re}(\lambda)(t-hm)} \left(\frac{|1+h\lambda|^{2m}-1}{-\mathrm{Re}(\lambda)} + \frac{h\left(|1+h\lambda|^{2m}-1\right)}{|1+h\lambda|-1} \right) \right)$$

$$\leq \varepsilon \left(\frac{h}{1-|1+h\lambda|} + |1+h\lambda|^{2m} \left(\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right),$$

as $t > hm$, $\mathrm{Re}(\lambda) < 0$ with $0 < |1+h\lambda| < 1$, and the expression inside the square brackets is non-negative.

(iv) (a). For $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| > 1$ and $t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\}$, using the same calculation as in case (iii)(a), we get

$$|\phi(t) - x(t)| \leq \varepsilon \left(\frac{h}{1-|1+h\lambda|} + |1+h\lambda|^{\frac{t}{h}+m} \left(\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right)$$

$$\leq \varepsilon \left(\frac{h}{1-|1+h\lambda|} + |1+h\lambda|^{2m} \left(\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right),$$

as $\frac{t}{h} \leq m$ and $\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} > 0$ for $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| > 1$ and $h > 0$.

(iv) (b). For $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| > 1$ and $t \in (hm, \infty)$, using the same calculation as in case (iii)(b), we get

$$|\phi(t) - x(t)| \leq \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} + e^{\mathrm{Re}(\lambda)(t-hm)} \left(\frac{|1+h\lambda|^{2m}-1}{-\mathrm{Re}(\lambda)} + \frac{h\left(|1+h\lambda|^{2m}-1\right)}{|1+h\lambda|-1} \right) \right)$$

$$\leq \varepsilon \left(\frac{h}{1-|1+h\lambda|} + |1+h\lambda|^{2m} \left(\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right),$$

as $t > hm$, $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| > 1$, and the expression inside the square brackets is positive.

We next show that x is the unique solution of (1) such that $|\phi(t) - x(t)| \leq K\varepsilon$ for all $t \in \mathbb{T}_{hm}$, where K is given by (4). Suppose $\phi : \mathbb{T}_{hm} \to \mathbb{C}$ is an approximate solution of (1) such that

$$\left| \phi^{\Delta}(t) - \lambda\phi(t) \right| \leq \varepsilon \text{ forall } t \in \mathbb{T}_{hm}$$

for some $\varepsilon > 0$. Suppose further that $x_1, x_2 : \mathbb{T}_{hm} \to \mathbb{C}$ are two different solutions of (1) such that $|\phi(t) - x_j(t)| \leq K\varepsilon$ for all $t \in \mathbb{T}_{hm}$, for $j = 1, 2$. Then, we have for constants $c_j \in \mathbb{C}$ that

$$x_j(t) = c_j e_\lambda(t, 0), \quad c_1 \neq c_2,$$

and

$$|c_1 - c_2| \cdot |e_\lambda(t, 0)| = |x_1(t) - x_2(t)| \leq |x_1(t) - \phi(t)| + |\phi(t) - x_2(t)| \leq 2K\varepsilon;$$

letting $t \to -\infty$ yields $\infty < 2K\varepsilon$, a contradiction. Consequently, x is the unique solution of (1) such that $|\phi(t) - x(t)| \leq \varepsilon K$ for all $t \in \mathbb{T}_{hm}$. This completes the proof. □

Theorem 3 *Let* $\lambda \in \mathbb{C}\backslash\left\{\frac{-1}{h}\right\}$ *with* $\mathrm{Re}(\lambda) = 0$. *Then,* (1) *is not Hyers–Ulam stable on* $\mathbb{T}_{hm}$.

Proof Assume $\mathrm{Re}(\lambda) = 0$ for $\lambda \in \mathbb{C}$. Let arbitrary $\varepsilon > 0$ be given, and let $\lambda = i\beta$ for some $\beta \in \mathbb{R}$. Then,

$$\phi(t) := \frac{\varepsilon t e_{i\beta}(t,0)}{\left(1+h^2\beta^2\right)^{\frac{m+1}{2}}}, \qquad t \in \mathbb{T}_{hm}$$

satisfies the inequality

$$\left|\phi^{\Delta}(t) - i\beta\phi(t)\right| = \frac{\varepsilon\left|(1+i\beta\mu(t))e_{i\beta}(t,0)\right|}{\left(1+h^2\beta^2\right)^{\frac{m+1}{2}}} \leq \frac{\varepsilon\left|e_{i\beta}(t,0)\right|}{\left(1+h^2\beta^2\right)^{\frac{m}{2}}} \leq \varepsilon$$

for all $t \in \mathbb{T}_{hm}$. Since $x(t) = x_0 e_{i\beta}(t,0)$ is the general solution of (1) when $\lambda = i\beta$, then

$$|\phi(t) - x(t)| = \frac{\left|e_{i\beta}(t,0)\right|}{\left(1+h^2\beta^2\right)^{\frac{m+1}{2}}}\left|\varepsilon t - x_0\left(1+h^2\beta^2\right)^{\frac{m+1}{2}}\right| \to \infty$$

as $t \to \pm\infty$ for $t \in \mathbb{T}_{hm}$ and for any $x_0 \in \mathbb{C}$, $\beta \in \mathbb{R}$, $h > 0$. So, (1) lacks HUS on $\mathbb{T}_{hm}$ if $\lambda = i\beta$. □

Using the previous theorems, we can establish the following results.

Theorem 4 *Let* $\lambda \in \mathbb{C}\backslash\{\frac{-1}{h}\}$. *Equation* (1) *has HUS on* $\mathbb{T}_{hm}$ *if and only if* $\mathrm{Re}(\lambda) \neq 0$.

Proof By Theorems 1, 2 and 3, we obtain the result, immediately. □

Lemma 2 *Let* $\lambda \in \mathbb{C}\backslash\left\{\frac{-1}{h}\right\}$ *with* $\mathrm{Re}(\lambda) \neq 0$.

(i) If $\mathrm{Re}(\lambda) > 0$, *then the HUS constant K for* (1) *satisfies*

$$K \geq \frac{1}{\mathrm{Re}(\lambda)}.$$

(ii) If $\mathrm{Re}(\lambda) < 0$, *then the HUS constant K for* (1) *satisfies*

$$K \geq \begin{cases} \frac{-1}{\mathrm{Re}(\lambda)} + 2hm & : |1+h\lambda| = 1 \\ \max\left\{\frac{-1}{\mathrm{Re}(\lambda)}, \frac{h}{1-|1+h\lambda|} + |1+h\lambda|^{2m}\left(\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)}\right)\right\} & : |1+h\lambda| \neq 1. \end{cases}$$

Proof Since $\lambda \in \mathbb{C}\backslash\left\{\frac{-1}{h}\right\}$ with $\mathrm{Re}(\lambda) \neq 0$, Equation (1) has HUS by Theorem 4. We will proceed by cases.

(i). Let $\lambda = \alpha + i\beta \in \mathbb{C}\backslash\left\{\frac{-1}{h}\right\}$, and assume $\mathrm{Re}(\lambda) = \alpha > 0$; set

$$\phi(t) := \frac{-\varepsilon e_{i\beta}(t,0)}{\alpha\left(1+h^2\beta^2\right)^{\frac{m}{2}}} + \frac{\varepsilon}{\alpha} e_\lambda(t,0).$$

It follows that

$$\left|\phi^\Delta(t) - \lambda\phi(t)\right| = \frac{\left|\varepsilon\alpha e_{i\beta}(t,0)\right|}{\alpha\left(1+h^2\beta^2\right)^{\frac{m}{2}}} = \frac{\varepsilon\left|e_{i\beta}(t,0)\right|}{\left(1+h^2\beta^2\right)^{\frac{m}{2}}} \leq \varepsilon.$$

Since $x(t) = \frac{\varepsilon}{\alpha} e_\lambda(t,0)$ is a solution of (1),

$$|\phi(t) - x(t)| = \frac{\varepsilon\left|e_{i\beta}(t,0)\right|}{\alpha\left(1+h^2\beta^2\right)^{\frac{m}{2}}} \leq \frac{\varepsilon}{\alpha},$$

with equality at $t = hm$, so the minimal HUS constant K for (1) satisfies

$$K \geq \frac{1}{\alpha} = \frac{1}{\mathrm{Re}(\lambda)}.$$

This ends the proof of case (i).

(ii) (a). Assume $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| = 1$. Let

$$\phi(t) = e_\lambda(t,0)\int_0^t \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau, \quad q(\tau) = \frac{\varepsilon e_\lambda(\sigma(\tau),0)}{|e_\lambda(\sigma(\tau),0)|}, \tag{6}$$

for all $t \in \mathbb{T}_{hm}$. Then,

$$\phi^\Delta(t) - \lambda\phi(t) = q(t), \quad |q(t)| = \varepsilon,$$

and, employing (6), we see that ϕ takes the form

$$\phi(t) = \varepsilon\begin{cases}\left(\frac{1}{\mathrm{Re}(\lambda)} - hm\right)e_\lambda(t,0) - \frac{e^{i\,\mathrm{Im}(\lambda)(t+hm)}}{(1+h\lambda)^m\,\mathrm{Re}(\lambda)} & : t \in (-\infty, -hm)\\ te_\lambda(t,0) & : t \in \{-hm,\ldots,hm\}\\ \left(\frac{1}{\mathrm{Re}(\lambda)} + hm\right)e_\lambda(t,0) - \frac{e^{i\,\mathrm{Im}(\lambda)(t-hm)}}{(1+h\lambda)^{-m}\,\mathrm{Re}(\lambda)} & : t \in (hm,\infty).\end{cases}$$

If we take

$$x(t) := \varepsilon\left(\frac{1}{\mathrm{Re}(\lambda)} - hm\right)e_\lambda(t,0),$$

then x is a solution of (1), and

$$|\phi(t)-x(t)| = \begin{cases} \varepsilon \left| \frac{-e^{i\,\mathrm{Im}(\lambda)(t+hm)}}{(1+h\lambda)^m\,\mathrm{Re}(\lambda)} \right| = \frac{-\varepsilon}{\mathrm{Re}(\lambda)} & : t \in (-\infty, -hm) \\ \varepsilon \left| -\frac{1}{\mathrm{Re}(\lambda)} + t + hm \right| \le \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} + 2hm \right) & : t \in \{-hm, \dots, hm\} \\ \varepsilon \left| -\frac{1}{\mathrm{Re}(\lambda)} + 2hm e^{\mathrm{Re}(\lambda)(t-hm)} \right| \le \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} + 2hm \right) & : t \in (hm, \infty), \end{cases}$$

where we have equality at $t = hm$. This shows that the HUS constant K must satisfy

$$K \ge \left(\frac{-1}{\mathrm{Re}(\lambda)} + 2hm \right)$$

for $\mathrm{Re}(\lambda) < 0$ with $|1 + h\lambda| = 1$. Here ends the proof of case $(ii)(a)$.

(ii)(b). Assume $\mathrm{Re}(\lambda) < 0$ with $|1 + h\lambda| \neq 1$. Again, let ϕ be given by (6) for all $t \in \mathbb{T}_{hm}$. Then,

$$\phi^{\Delta}(t) - \lambda\phi(t) = q(t), \quad |q(t)| = \varepsilon,$$

and in this case ϕ takes the form

$$\phi(t) = \varepsilon \begin{cases} \left(\frac{|1+h\lambda|^m}{\mathrm{Re}(\lambda)} - \frac{h(|1+h\lambda|^m - 1)}{|1+h\lambda|-1} \right) e_\lambda(t,0) - \frac{|1+h\lambda|^m e^{i\,\mathrm{Im}(\lambda)(t+hm)}}{(1+h\lambda)^m\,\mathrm{Re}(\lambda)} & : t \in (-\infty, -hm) \\ \frac{h\left(|1+h\lambda|^{\frac{t}{h}} - 1 \right)}{|1+h\lambda|^{\frac{t}{h}}\,(|1+h\lambda|-1)} e_\lambda(t,0) & : t \in \{-hm, \dots, hm\} \\ \left(\frac{|1+h\lambda|^{-m}}{\mathrm{Re}(\lambda)} + \frac{h(|1+h\lambda|^m - 1)}{|1+h\lambda|^m\,(|1+h\lambda|-1)} \right) e_\lambda(t,0) - \frac{(1+h\lambda)^m e^{i\,\mathrm{Im}(\lambda)(t-hm)}}{|1+h\lambda|^m\,\mathrm{Re}(\lambda)} & : t \in (hm, \infty). \end{cases}$$

If we take

$$x(t) := \varepsilon \left(\frac{|1+h\lambda|^m}{\mathrm{Re}(\lambda)} - \frac{h\,(|1+h\lambda|^m - 1)}{|1+h\lambda| - 1} \right) e_\lambda(t,0), \tag{7}$$

then x is a solution of (1), and

$$|\phi(t) - x(t)| = \frac{-\varepsilon}{\mathrm{Re}(\lambda)}, \quad t \in (-\infty, -hm).$$

For $t \in \{-hm, \dots, -h, 0, h, \dots, hm\}$,

$$|\phi(t) - x(t)| = \varepsilon \left(\frac{h\left(|1+h\lambda|^{m+\frac{t}{h}} - 1 \right)}{|1+h\lambda| - 1} - \frac{|1+h\lambda|^{m+\frac{t}{h}}}{\mathrm{Re}(\lambda)} \right).$$

If $0 < |1 + h\lambda| < 1$, then as in the proof of Theorem 2 $(iii)(a)$, we have

$$|\phi(t) - x(t)| = \varepsilon \left(\frac{h}{1 - |1 + h\lambda|} + |1 + h\lambda|^{\frac{t}{h} + m} \left(\frac{h}{|1 + h\lambda| - 1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right)$$
$$\leq \varepsilon \left(\frac{h}{1 - |1 + h\lambda|} + |1 + h\lambda|^{2m} \left(\frac{h}{|1 + h\lambda| - 1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right),$$

as $\frac{t}{h} \leq m$ and $\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} \leq 0$ for $\mathrm{Re}(\lambda) < 0$ with $0 < |1 + h\lambda| < 1$ and $h > 0$, with equality at $t = hm$. If $|1 + h\lambda| > 1$, then as in the proof of Theorem 2 $(iv)(a)$,

$$|\phi(t) - x(t)| = \varepsilon \left(\frac{h}{1 - |1 + h\lambda|} + |1 + h\lambda|^{\frac{t}{h} + m} \left(\frac{h}{|1 + h\lambda| - 1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right)$$
$$\leq \varepsilon \left(\frac{h}{1 - |1 + h\lambda|} + |1 + h\lambda|^{2m} \left(\frac{h}{|1 + h\lambda| - 1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right),$$

as $\frac{t}{h} \leq m$ and $\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} > 0$ for $\mathrm{Re}(\lambda) < 0$ with $|1 + h\lambda| > 1$ and $h > 0$. For $t \in (hm, \infty)$ and $0 < |1 + h\lambda| < 1$, then as in the proof of Theorem 2 $(iii)(b)$,

$$|\phi(t) - x(t)| = \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} + e^{\mathrm{Re}(\lambda)(t-hm)} \left(\frac{|1 + h\lambda|^{2m} - 1}{-\mathrm{Re}(\lambda)} + \frac{h\left(|1 + h\lambda|^{2m} - 1\right)}{|1 + h\lambda| - 1} \right) \right)$$
$$\leq \varepsilon \left(\frac{h}{1 - |1 + h\lambda|} + |1 + h\lambda|^{2m} \left(\frac{h}{|1 + h\lambda| - 1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right),$$

as $t > hm$, $\mathrm{Re}(\lambda) < 0$ with $0 < |1 + h\lambda| < 1$, and the expression inside the square brackets is non-negative. For $t \in (hm, \infty)$ and $|1 + h\lambda| > 1$, then as in the proof of Theorem 2 $(iv)(b)$, we have

$$|\phi(t) - x(t)| = \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} + e^{\mathrm{Re}(\lambda)(t-hm)} \left(\frac{|1 + h\lambda|^{2m} - 1}{-\mathrm{Re}(\lambda)} + \frac{h\left(|1 + h\lambda|^{2m} - 1\right)}{|1 + h\lambda| - 1} \right) \right)$$
$$\leq \varepsilon \left(\frac{h}{1 - |1 + h\lambda|} + |1 + h\lambda|^{2m} \left(\frac{h}{|1 + h\lambda| - 1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right),$$

as $t > hm$, $\mathrm{Re}(\lambda) < 0$ with $|1 + h\lambda| > 1$, and the expression inside the square brackets is positive. This ends the proof of case $(ii)(b)$, and thus the overall result holds.

□

Theorem 5 *Let* $\lambda \in \mathbb{C} \backslash \{\frac{-1}{h}\}$. *If* $\mathrm{Re}(\lambda) \neq 0$, *then* (1) *has HUS on* $\mathbb{T}_{hm}$.

(i) If $\mathrm{Re}(\lambda) > 0$, *then*

$$K = \frac{1}{\mathrm{Re}(\lambda)}$$

is the best (minimal) HUS constant.

(ii) If $\mathrm{Re}(\lambda) < 0$*, then*

$$K = \begin{cases} \frac{-1}{\mathrm{Re}(\lambda)} + 2hm & : |1 + h\lambda| = 1 \\ \max\left\{ \frac{-1}{\mathrm{Re}(\lambda)}, \frac{h}{1-|1+h\lambda|} + |1 + h\lambda|^{2m} \left(\frac{h}{|1+h\lambda|-1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right\} & : |1 + h\lambda| \neq 1 \end{cases}$$

is the best (minimal) HUS constant.

Proof This result follows immediately from the definitions of HUS and HUS constant, Theorems 1–4, and Lemma 2. □

Remark 1 If $m = 0$, then $\mathbb{T}_{h,0} = \mathbb{R}$, and the results in Theorems 1 and 2 $(i) - (iv)$ match exactly the known results for $\mathbb{T} = \mathbb{R}$, namely that $x'(t) - \lambda x(t) = 0$ has HUS on $\mathbb{R}$, and

$$K = \frac{1}{|\,\mathrm{Re}(\lambda)|}$$

is the best possible HUS constant. If $h \to 0$, then $\mathbb{T}_{0,m} = \mathbb{R}$, and the results in Theorems 1 and 2 $(i) - (iv)$ also recover the known results for $\mathbb{T} = \mathbb{R}$, because

$$\lim_{h\to 0+} \left(\frac{-1}{\mathrm{Re}(\lambda)} + 2hm \right) = \frac{-1}{\mathrm{Re}(\lambda)}, \quad \lim_{h\to 0+} \frac{h}{|1 + h\lambda| - 1} = \lim_{h\to 0+} \frac{1}{\mathrm{Re}_h(\lambda)} = \frac{1}{\mathrm{Re}(\lambda)}$$

hold, where $\mathrm{Re}_h(\lambda)$ represents the Hilger real part [20] for h-difference equations.

For fixed $h > 0$, if $m \to \infty$, then $\mathbb{T}_{h,\infty} = h\mathbb{Z}$, and the results in Theorem 1 and Theorem 2 (i) and (iii) match exactly the known results for $\mathbb{T} = h\mathbb{Z}$, namely that $\Delta_h x(t) - \lambda x(t) = 0$ has HUS on $h\mathbb{Z}$, and

$$K = \frac{h}{\left|1 - |1 + h\lambda|\right|} = \frac{1}{|\,\mathrm{Re}_h(\lambda)|}$$

is the best possible HUS constant. Theorem 2 (ii) shows an interesting connection; as $m \to \infty$, the HUS constant in (ii) goes to infinity as well. This is accurate, as $\mathrm{Re}(\lambda) < 0$ with $|1 + h\lambda| = 1$ makes the h-difference equation version of (1) Hyers–Ulam unstable on $h\mathbb{Z}$, as $\lambda \in \mathbb{C}$ is then on the left Hilger circle [5]; see [15, Chapter 2.1], [20], and [23] for more on the Hilger complex plane, and [2, 5, 10] for more on the Hilger circle and HUS. On the other hand, in case (iv) $\mathrm{Re}(\lambda) < 0$ with $|1 + h\lambda| > 1$, a result that does not match is obtained, that is,

$$\lim_{m\to\infty} \left(\frac{h}{1 - |1 + h\lambda|} + |1 + h\lambda|^{2m} \left(\frac{h}{|1 + h\lambda| - 1} - \frac{1}{\mathrm{Re}(\lambda)} \right) \right) = \infty$$

when $\mathrm{Re}(\lambda) < 0$ with $|1 + h\lambda| > 1$, but, we know that $\Delta_h x(t) - \lambda x(t) = 0$ has HUS when $|1 + h\lambda| > 1$ (see [5]). Why does this logical gap occur? According to the information of Theorem 2.5 (ii) in [5], in this case, the unique solution x

is determined when $t \to \infty$. As you can see from the claim of Theorem 2, even in this case, the unique solution x is determined by the information of $t \to \infty$. Therefore, we can say that the case $\mathrm{Re}(\lambda) < 0$ with $|1 + h\lambda| > 1$ is a distinguishing characteristic of Hyers–Ulam stability on this time scale with discrete core and continuous periphery. We explore this anomaly in the next section.

4 Connection with h-Difference Equations in the Case $|1 + h\lambda| > 1$

The following result is effective for clarifying the connection with the h-difference equation $\Delta_h x(t) - \lambda x(t) = 0$ with $|1 + h\lambda| > 1$.

Theorem 6 *Let $\lambda \in \mathbb{C} \backslash \left\{ \frac{-1}{h} \right\}$ with $\mathrm{Re}(\lambda) < 0$ and $|1 + h\lambda| > 1$. Let $\varepsilon > 0$ be a fixed arbitrary constant, and let ϕ be a function on $\mathbb{T}_{hm}$ satisfying the inequality*

$$\left| \phi^{\Delta}(t) - \lambda \phi(t) \right| \leq \varepsilon, \qquad t \in \mathbb{T}_{hm}.$$

Then, the function x given by

$$x(t) := \left(\frac{\phi(hm)}{e_\lambda(hm, 0)} \right) e_\lambda(t, 0)$$

is a solution of (1) with

$$|\phi(t) - x(t)| \leq \varepsilon \max \left\{ \frac{h\left(1 - |1 + h\lambda|^{-2m}\right)}{|1 + h\lambda| - 1}, \frac{-1}{\mathrm{Re}(\lambda)} \right\}$$

for all $t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\} \cup (hm, \infty)$. In particular, the following holds.

(i) If $t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\}$, then

$$|\phi(t) - x(t)| \leq \varepsilon \left(\frac{h\left(1 - |1 + h\lambda|^{-2m}\right)}{|1 + h\lambda| - 1} \right)$$

for all $t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\}$.

(ii) If $t \in (hm, \infty)$, then

$$|\phi(t) - x(t)| \leq \varepsilon \left(\frac{-1}{\mathrm{Re}(\lambda)} \right)$$

for all $t \in (hm, \infty)$.

Proof Let $\lambda \in \mathbb{C}\backslash\left\{\frac{-1}{h}\right\}$ with $\mathrm{Re}(\lambda) < 0$ and $|1+h\lambda| > 1$. Suppose that $|\phi^{\Delta}(t) - \lambda\phi(t)| \leq \varepsilon$ for all $t \in \mathbb{T}_{hm}$, there exists a function $q : \mathbb{T}_{hm} \to \mathbb{C}$ such that

$$\phi^{\Delta}(t) - \lambda\phi(t) = q(t), \quad |q(t)| \leq \varepsilon$$

for all $t \in \mathbb{T}_{hm}$. Then, we have

$$\phi(t) = \phi_0 e_\lambda(t,0) + e_\lambda(t,0)\int_0^t \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau.$$

Let $x(t) = x_0 e_\lambda(t,0)$ be the solution of (1) with

$$x_0 := \phi_0 + \int_0^{hm} \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau \in \mathbb{C}.$$

It follows that

$$\phi(t) - x(t) = -e_\lambda(t,0)\left(\int_t^{hm} \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau\right). \tag{8}$$

(a) For $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| > 1$ and $t \in \{-hm, \ldots, -h, 0, h, \ldots, hm\}$, using (8) we have

$$|\phi(t) - x(t)| \leq \varepsilon|e_\lambda(t,0)|\int_t^{hm} \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|} = \varepsilon|1+h\lambda|^{\frac{t}{h}} \sum_{j=\frac{t}{h}}^{m-1} \frac{h}{|1+h\lambda|^{j+1}}$$

$$= \varepsilon\left(\frac{h\left(1-|1+h\lambda|^{-m+\frac{t}{h}}\right)}{|1+h\lambda|-1}\right) \leq \varepsilon\left(\frac{h\left(1-|1+h\lambda|^{-2m}\right)}{|1+h\lambda|-1}\right),$$

as $\frac{t}{h} \leq m$, $h > 0$, and $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| > 1$.

(b) For $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| > 1$ and $t \in [hm, \infty)$, using (8) we have

$$|\phi(t) - x(t)| = \left|e_\lambda(t,0)\left(\int_{hm}^t \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)}\Delta\tau\right)\right|$$

$$\leq \varepsilon|1+h\lambda|^m e^{\mathrm{Re}(\lambda)(t-hm)}\int_{hm}^t \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|}$$

$$= \varepsilon\left(\frac{-1+e^{\mathrm{Re}(\lambda)(t-hm)}}{\mathrm{Re}(\lambda)}\right) \leq \varepsilon\left(\frac{-1}{\mathrm{Re}(\lambda)}\right),$$

as $t \geq hm$ and $\mathrm{Re}(\lambda) < 0$ with $|1+h\lambda| > 1$. This completes the proof.

□

Remark 2 For fixed $h > 0$, if $m \to \infty$, then $\mathbb{T}_{h,\infty} = h\mathbb{Z}$, and the result in Theorem 6 (i) reproduces exactly the known result for $\mathbb{T} = h\mathbb{Z}$ (see, Theorem 2.5 (ii) in [5]). Actually, we will explain this fact. Since

$$\left| \int_0^{hm} \frac{q(\tau)}{e_\lambda(\sigma(\tau),0)} \Delta\tau \right| \le \varepsilon \int_0^{hm} \frac{\Delta\tau}{|e_\lambda(\sigma(\tau),0)|} \le \varepsilon \sum_{j=0}^{m-1} \frac{h}{|1+h\lambda|^{j+1}} < \frac{h}{|1+h\lambda|-1}$$

holds, we see that

$$\lim_{m\to\infty} \frac{\phi(hm)}{e_\lambda(hm,0)} = \lim_{t\to\infty} \frac{\phi(t)}{e_\lambda(t,0)}$$

exists. In addition, an HUS constant is

$$\lim_{m\to\infty} \frac{h\left(1-|1+h\lambda|^{-2m}\right)}{|1+h\lambda|-1} = \frac{h}{|1+h\lambda|-1}.$$

Theorem 6 (i) says that the function x given by

$$x(t) := \left(\lim_{t\to\infty} \frac{\phi(t)}{e_\lambda(t,0)} \right) e_\lambda(t,0)$$

is a solution of (1) with

$$|\phi(t) - x(t)| \le \frac{h\varepsilon}{|1+h\lambda|-1}$$

for all $t \in h\mathbb{Z}$. As $m \to \infty$, our exponential function $e_\lambda(t,0)$ corresponds to $(1+h\lambda)^{\frac{t}{h}}$ for all $t \in h\mathbb{Z}$. In this case, we can prove the uniqueness of the solution. Let $\varepsilon > 0$, and let $\phi : \{-hm, \dots, -h, 0, h, \dots, hm\} \to \mathbb{C}$ satisfy

$$\left|\phi^{\Delta}(t) - \lambda\phi(t)\right| \le \varepsilon \text{ for all } t \in \{-hm, \dots, -h, 0, h, \dots, hm\}.$$

Suppose that $x_1, x_2 : \{-hm, \dots, -h, 0, h, \dots, hm\} \to \mathbb{C}$ are two different solutions of (1) such that $|\phi(t) - x_j(t)| \le K\varepsilon := \frac{h}{|1+h\lambda|-1}\varepsilon$ for all $t \in \{-hm, \dots, -h, 0, h, \dots, hm\}$, for $j = 1, 2$. Then, we have for constants $c_j \in \mathbb{C}$ that

$$x_j(t) = c_j(1+h\lambda)^{\frac{t}{h}}, \quad c_1 \ne c_2,$$

and

$$|c_1 - c_2||1+h\lambda|^{\frac{t}{h}} = |x_1(t) - x_2(t)| \le |x_1(t) - \phi(t)| + |\phi(t) - x_2(t)| \le 2K\varepsilon;$$

letting $m \to \infty$ and $t \to \infty$ yields $\infty < 2K\varepsilon$, a contradiction. Consequently, x is the unique solution of $\Delta_h x(t) - \lambda x(t) = 0$ such that $|\phi(t) - x(t)| \leq \varepsilon K$ for all $t \in h\mathbb{Z}$.

5 Conclusion and Future Directions

In this paper we determined the best Hyers–Ulam stability constants for a first-order complex constant coefficient dynamic equation on a time scale with a discrete core and continuous periphery. In the future, we will study a time scale with a discrete periphery and continuous core, whose exponential function for $\lambda \in \mathbb{C}\backslash\{-\frac{1}{h}\}$ is

$$e_\lambda(t,0) := \begin{cases} (1+h\lambda)^{\frac{t}{h}+m} e^{-hm\lambda} & : t \in \{\ldots, -h(m+2), -h(m+1)\} \\ e^{\lambda t} & : t \in [-hm, hm] \\ (1+h\lambda)^{\frac{t}{h}-m} e^{hm\lambda} & : t \in \{h(m+1), h(m+2), \ldots\}. \end{cases}$$

on $\mathbb{T}_{hm} = \{\ldots, -h(m+2), -h(m+1)\} \cup [-hm, hm] \cup \{h(m+1), h(m+2), \ldots\}$, where $h > 0$ is the discrete step size and m is a non-negative integer.

Acknowledgments The second author was supported by JSPS KAKENHI Grant Number JP20K03668.

References

1. D.R. Anderson, Hyers–Ulam stability for a first-order linear proportional nabla difference operator, in *Frontiers in Functional Equations and Analytic Inequalities*, ed. by G. Anastassiou, J.M. Rassias (Springer Nature Switzerland AG, Cham, 2019)
2. D.R. Anderson, The discrete diamond-alpha imaginary ellipse and Hyers–Ulam stability. Int. J. Difference Equations **14**(1), 25–38 (2019)
3. D.R. Anderson, M. Onitsuka, Hyers–Ulam stability of first-order homogeneous linear dynamic equations on time scales. Demonstratio Math. **51**, 198–210 (2018)
4. D.R. Anderson, M. Onitsuka, Hyers–Ulam stability for a discrete time scale with two step sizes. Appl. Math. Comput. **344–345**, 128–140 (2019)
5. D.R. Anderson, M. Onitsuka, Best constant for Hyers–Ulam stability of second-order h-difference equations with constant coefficients. Results Math **74**, 151 (2019). https://doi.org/10.1007/s00025-019-1077-9
6. D.R. Anderson, M. Onitsuka, Hyers–Ulam stability and best constant for Cayley h-difference equations. Bull. Malaysian Math. Sci. Soc. (2020). https://doi.org/10.1007/s40840-020-00920-z
7. D.R. Anderson, M. Onitsuka, Hyers–Ulam stability for quantum equations of Euler type. Discrete Dyn. Nature Soc. **2020**, Article ID 5626481, 10 pp. (2020). https://doi.org/10.1155/2020/5626481
8. D.R. Anderson, M. Onitsuka, Hyers-Ulam stability for quantum equations. Aequationes Math. (2020). https://doi.org/10.1007/s00010-020-00734-1

9. D.R. Anderson, M. Onitsuka, Best constant for Hyers–Ulam stability of two step sizes linear difference equations. J. Math. Anal. Appl. (2021). https://doi.org/10.1016/j.jmaa.2020.124807
10. D.R. Anderson, A.J. Jennissen, C.J. Montplaisir, Hyers–Ulam stability for a continuous time scale with discrete uniform jumps. Int. J. Difference Equations **15**(2), 1–21 (2020)
11. D.R. Anderson, M. Onitsuka, J.M. Rassias, Best constant for Ulam stability of first-order h-difference equations with periodic coefficient. J. Math. Anal. Appl. **491**, 124363 (2020). https://doi.org/10.1016/j.jmaa.2020.124363
12. S. András, A.R. Mészáros, Ulam–Hyers stability of dynamic equations on time scales via Picard operators. Appl. Math. Computation. **219**, 4853–4864 (2013)
13. A.R. Baias, D. Popa, On Ulam stability of a linear difference equation in Banach spaces. Bull. Malays. Math. Sci. Soc. (2019). https://doi.org/10.1007/s40840-019-00744-6
14. A.R. Baias, F. Blaga, D. Popa, On the best Ulam constant of a first order linear difference equation in Banach spaces. Acta Math. Hungar. (2020). https://doi.org/10.1007/s10474-020-01098-3
15. M. Bohner, A. Peterson, *Dynamic Equations on Time Scales, An Introduction with Applications* (Birkhäuser, Boston, 2001)
16. J. Brzdęk, P. Wójcik, On approximate solutions of some difference equations. Bull. Australian Math. Soc. **95**(3), 76–481 (2017)
17. J. Brzdęk, D. Popa, I. Raşa, B. Xu, *Ulam Stability of Operators*, a volume in Mathematical Analysis and Its Applications (Academic Press, New York, 2018)
18. C. Buşe, D. O'Regan, O. Saierli, Hyers–Ulam stability for linear differences with time dependent and periodic coefficients. Symmetry **11**, 512 (2019). https://doi.org/10.3390/sym11040512
19. C. Buşe, V. Lupulescu, D. O'Regan, Hyers–Ulam stability for equations with differences and differential equations with time-dependent and periodic coefficients. Proc. R. Soc. Edinburgh A Math. **150**(5), 2175–2188 (2020). https://doi.org/10.1017/prm.2019.12
20. S. Hilger, Special functions, Laplace and Fourier transform on measure chains. Dynamic Syst. Appl. **8**(3–4), 471–488 (1999)
21. L. Hua, Y. Li, J. Feng, On Hyers–Ulam stability of dynamic integral equation on time scales. Mathematica Aeterna **4**(6), 559–571 (2014)
22. D.H. Hyers, On the stability of the linear functional equation. Proc. Nat. Acad. Sci. U. S. A. **27**, 222–224 (1941)
23. B.J. Jackson, J.M. Davis, D. Poulsen, A polar representation of the Hilger complex plane. Int. J. Difference Equations **15**(2), 419–427 (2020)
24. S.-M. Jung, Y.W. Nam, Hyers–Ulam stability of Pielou logistic difference equation. J. Nonlinear Sci. Appl. **10**, 3115–3122 (2017)
25. S.-M. Jung, Y.W. Nam, Hyers–Ulam stability of the first order inhomogeneous matrix difference equation. J. Comput. Anal. Appl. **23**(8), 1368–1383 (2017)
26. Y.W. Nam, Hyers–Ulam stability of hyperbolic Möbius difference equation. Filomat **32**(13), 4555–4575 (2018). https://doi.org/10.2298/FIL1813555N
27. Y.W. Nam, Hyers–Ulam stability of elliptic Möbius difference equation. Cogent Math. Stat. **5**(1), 1–9 (2018)
28. Y.W. Nam, Hyers–Ulam stability of loxodromic Möbius difference equation. Appl. Math. Comput. **356**(1), 119–136 (2019). https://doi.org/10.1016/j.amc.2019.03.033
29. M. Onitsuka, Influence of the step size on Hyers–Ulam stability of first-order homogeneous linear difference equations. Int. J. Difference Equations **12**(2), 281–302 (2017)
30. M. Onitsuka, Hyers–Ulam stability of second-order nonhomogeneous linear difference equations with a constant step size. J. Comput. Anal. Appl. **28**(1), 152–165 (2020)
31. D. Popa, Hyers–Ulam stability of the linear recurrence with constant coefficients. Adv. Differential Equations **2005**, 407076 (2005)
32. D. Popa, Hyers–Ulam–Rassias stability of a linear recurrence. J. Math. Anal. Appl. **309**, 591–597 (2005)
33. H. Rasouli, S. Abbaszadeh, M. Eshaghi, Approximately linear recurrences. J. Appl. Anal. **24**(1), 81–85 (2018)

34. Th.M. Rassias, On the stability of linear mapping in Banach spaces. Proc. Amer. Math. Soc. **72**, 297–300 (1978)
35. Y.H. Shen, The Ulam stability of first order linear dynamic equations on time scales. Results Math. **72**(4), 1881–1895 (2017)
36. Y.H. Shen, Y.J. Li, Hyers–Ulam stability of first order nonhomogeneous linear dynamic equations on time scales. Commun. Math. Res. **35**(2), 139–148 (2019). https://doi.org/10.13447/j.1674-5647.2019.02.05
37. S.M. Ulam, *A Collection of the Mathematical Problems* (Interscience, New York, 1960)
38. B. Xu, J. Brzdęk, Hyers–Ulam stability of a system of first order linear recurrences with constant coefficients. Discrete Dyn. Nat. Soc. **2015**, Article ID 269356, 5 pp. (2015)

Invariance Solutions and Blow-Up Property for Edge Degenerate Pseudo-Hyperbolic Equations in Edge Sobolev Spaces

Carlo Cattani and Morteza Koozehgar Kalleji

Abstract This article is dedicated to study of the initial-boundary value problem of edge pseudo-hyperbolic system with damping term on the manifold with edge singularity. First, we will discuss about the invariance of solution set of a class of edge degenerate pseudo-hyperbolic equations on the edge Sobolev spaces. Then, by using a family of modified potential wells and concavity methods, it is obtained existence and nonexistence results of global solutions with exponential decay and is shown the blow-up in finite time of solutions on the manifold with edge singularities.

1 Introduction

Initial-boundary value problems written for hyperbolic semilinear partial differential equations emerged in several applications to physics, mechanics and engineering sciences [9, 24, 25]. Interesting phenomena are often connected with geometric singularities, for instance, in mechanics or cracks in a medium are described by hypersurfaces with a boundary. In this cases, configurations of that kind belong to the category of spaces (manifolds) with geometric singularities, here with edges. Also, when one asks physics to calculate the self-energy of an electron, or the structure of space time at the center of a black hole, one encounter with mathematical bad behaviour, that is the singularities from the point view of mathematics. In recent years, from a mathematical point of view, the analysis on such (in general, stratified) spaces has become a mathematical structure theory with many deep relations with geometry, topology, and mathematical physics [10, 15, 23, 25]. In [21], Melrose,Vasy and Wunsch investigated the geometric propagation and

C. Cattani (✉)
Engineering School, University of Tuscia, Viterbo, Italy
e-mail: cattani@unitus.it

M. K. Kalleji
Department of Mathematics, Faculty of Sciences, Arak University, Arak, Iran
e-mail: m-koozehgarkalleji@araku.ac.ir

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_3

diffraction of singularities of solutions to the wave equation on manifolds with edge singularities. Let X be an n-dimensional manifold with boundary, where the boundary ∂X is endowed with a fibration $Z \to \partial X$ and $\partial X \to Y$ where Y, Z are without boundary. By an edge metric g on X, we mean a metric g on the interior of X which is a smooth 2-cotensor up to the boundary but which degenerates there in a way compatible with the fibration. A manifold with boundary equipped with such an edge metric also is called an edge manifold or a manifold with edge structure. If Z is point, then an edge metric on X is simply a metric in the usual sense, smooth up to the boundary, while if Y is a point, X is conic manifold [4]. A simple example of a more general edge metric is obtained by performing a real blowup on a submanifold B of a smooth, boundaryless manifold A. The blowup operation simply introduces polar coordinates near B, i.e., it replaces B by its spherical normal bundle, thus yielding a manifold X with boundary. The pullback of a smooth metric on A to X is then an edge metric [21].

Up to now, elliptic boundary value problems in domains with point singularities have been thoroughly investigated [1–4, 7, 8, 14]. The natures of the solutions to these equations have been investigated by several means. For instance, problems with the Dirichlet boundary conditions were investigated in [1, 2, 7, 10, 14] in which the unique existence, the multiplicity, the regularity and the asymptotic behaviour near the conical points of the solutions are established. Finite time blow-up of solutions of generalized hyperbolic equations have been studied by many authors [1, 2, 5, 7, 18, 28]. In these references, the authors consider problems either for negative energy or for weaker conditions than a condition of negative initial energy. Other authors have assumed a condition of positive energy under other two conditions on the initial functions. However, the mentioned authors have not studied the compatibility of these conditions, which is come times hard to understand. These authors have used the classic concavity Levine's method [17]. In this article, we use the edge Sobolev inequality and Poincaré inequality and modified method in [7, 8] to prove on the global well-posedness of solutions to initial-boundary value problems for semilinear degenerate pseudo-hyperbolic equations with dissipative term on manifolds with edge singularities. More precisely, we study the following initial-boundary value problem for semilinear hyperbolic equation

$$\begin{cases} \partial_t^2 u - \Delta_{\mathbb{E}} u + V(z)u + \gamma \Delta_{\mathbb{E}} \partial_t u = g_t(z)|u|^{p-1}u, & z \in int\mathbb{E}, t > 0, \\ u(z,0) = u_0(z), \quad \partial_t u(z,0) = u_1(z), & z \in int\mathbb{E} \\ u(z,t) = 0, & z \in \partial\mathbb{E}, t \geq 0, \end{cases} \tag{1}$$

where, $2 < p+1 < \frac{2n}{n-2} = 2^*$ is the critical cone Sobolev exponents, $z = (r, x, y)$, $u = u(z,t)$ is unknown function and γ is a non-negative parameter. Also, $u_0 \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$, $u_1 \in L_2^{\frac{n+1}{2}}(\mathbb{E})$, $N = 1+n+q \geq 3$ is a dimension of $\mathbb{E}$ and coordinates $z := (r, x, y) = (r, x_1, \ldots, x_n, y_1, \ldots, y_q) \in \mathbb{E}$. Here the domain $\mathbb{E}$ is $[0, 1) \times X \times Y$, X is an $(n-1)$-dimensional closed compact manifold, $Y \subset \mathbb{R}^q$ is a bounded domain, which is regarded as the local model near the edge points on manifolds with edge singularities, and $\partial\mathbb{E} = \{0\} \times X \times Y$. Moreover, the operator $\Delta_{\mathbb{E}}$ in 1 is

defined by $(r\partial_r)^2+\partial_{x_1}^2+\ldots+\partial_{x_n}^2+(r\partial_{y_1})^2+\ldots+(r\partial_{y_q})^2$, which is an elliptic operator with totally characteristic degeneracy on the boundary $r=0$, we also call it Fuchsian type edge-Laplace operator, and the corresponding gradient operator by $\nabla_{\mathbb{E}}:=(r\partial_r,\partial_{x_1},\ldots,\partial_{x_n},r\partial_{y_1},\ldots,r\partial_{y_q})$. In the Equation 1, we assume that $V(z)\in L^{\frac{n+1}{4}}(int\mathbb{E})\cap C(int\mathbb{E})$ is a positive potential function such that $\inf_{z\in\mathbb{E}}V(z)>0$. For every $t\geq 0$, we suppose that $g_t:\mathbb{E}\to\mathbb{R}$ is a non-negative function which $g_t(z):=g(z,t)$ for every $z\in int\mathbb{E}$ and $g(z,t)\in L^{\infty}(int\mathbb{E})\cap C^1(int\mathbb{E})$. The through of this paper we consider the following constants:

$$C_*=\inf\left\{\frac{\|\sqrt{V(z)}u(z)\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})}}{\|\nabla_{\mathbb{E}}u\|_{L_2^{\frac{n+1}{2}}(\mathbb{B})}}\quad;\quad u\in\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\right\},$$

$$C_{**}=\sup\left\{\frac{\|g_t(z)^{\frac{1}{p+1}}u\|_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})}}{\|\nabla_{\mathbb{E}}u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})}}\quad;\quad u\in\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\right\}.$$

Our investigation is in fact provoked by the study of [20] and we shall apply a potential method which was established by Sattinger [26]. So based on edge Sobolev spaces [10, 27], we study the existence and non-existence global weak solutions for semilinear pseudo-hyperbolic differential equations with respect to variable time with a positive potential function and a non-negative weighted function. The well-known operator $(\Delta_{\mathbb{E}}+V(x)+\Delta_{\mathbb{E}}\partial_t)u$ and other special types of it (see [11]) appears naturally in the nonlinear heat and wave equations [25], nonlinear Schrödinger equation with potential function [12] and the references therein for a complete description of the model. In the sitting of parabolic type system, the authors [6, 18] studied global existence, exponential decay and finite time blow-up of solutions for a class of semilinear pseudo-parabolic equations with conical degeneration. Also, our problem can be seen as a class of degenerate hyperbolic type equations in case that $V(z)=0$ and $g_t(z)\equiv 1$ then the problem 1 is reduced to problem 1.1 in [13] and in the classical sense our problem include the classical problem

$$\begin{cases}\partial_t^2u-\Delta u+\gamma\partial_tu=f(u), & x\in\Omega,t>0,\\ u(x,0)=u_0(x), \quad \partial_tu(x,0)=u_1(x), & x\in\Omega\\ u(t,x)=0, & x\in\partial\Omega,t\geq 0,\end{cases}\tag{2}$$

where Ω is bounded domain of $\mathbb{R}^n$ with smooth boundary $\partial\Omega$ and Δ is the standard Laplace operator and f is a suitable function [13, 17, 19]. It is well-known that problem 2 has been studied by many authors, for example [19, 20] and the references therein.

In Section 2, we recall the definition of the edge Sobolev space and the corresponding properties. In Section 3, we will give some properties of potential

wells for problem 1 on the manifold with edge singularity, which is very useful in the process of our main results. In Section 4, we give the proofs of the results of global existence and non-existence, exponential decay and finite time blowing-up of problem 1.

2 Edge Sobolev Spaces

Consider X as a closed compact C^∞-manifold of dimension n of the unit sphere in $\mathbb{R}^{n+1}$. We define an infinite cone in $\mathbb{R}^{n+1}$ as a quotient space $X^\Delta = \frac{\bar{\mathbb{R}}_+ \times X}{\{0\} \times X}$, with base X. The cylindrical coordinates $(r, \theta) \in X^\Delta - \{0\}$ in $\mathbb{R}^{n+1} - \{0\}$ are the standard coordinates. This gives us the description of $X^\Delta - \{0\}$ in the form $\mathbb{R}_+ \times X$. Then the stretched cone can be defined as $\bar{\mathbb{R}}_+ \times X = X^\wedge$. Now, consider $B = X^\Delta$ with a conical point, then by the similar way in [8, 10, 27], one can define the stretched manifold $\mathbb{B}$ with respect to B as a C^∞-manifold with smooth boundary $\partial\mathbb{B} \cong X(0)$, where $X(0)$ is the cross section of singular point zero such that there is a diffeomorphism $B - \{0\} \cong \mathbb{B} - \partial\mathbb{B}$, the restriction of which to $U - \{0\} \cong V - \partial\mathbb{B}$ for an open neighborhood $U \subset B$ near the conic point zero and a collar neighborhood $V \subset \mathbb{B}$ with $V \cong [0, 1) \times X(0)$. Therefore, we can take $\mathbb{B} = [0, 1) \times X \subset \bar{\mathbb{R}}_+ \times X = X^\wedge$. In order to consider another type of a manifold with singularity of order one so-called wedge manifold, we consider a bounded domain Y in $\mathbb{R}^q$. Set $W = X^\Delta \times Y = B \times Y$. Then W is a corresponding wedge in $\mathbb{R}^{1+n+q}$. Therefore, the stretched wedge manifold $\mathbb{W}$ to W is $X^\wedge \times Y$ which is a manifold with smooth boundary $\{0\} \times X \times Y$. Set $(r, x) \in X^\wedge$. In order to define a finite wedge, it sufficient to consider the case $r \in [0, 1)$. Thus, we define a finite wedge as

$$E = \frac{[0, 1) \times X}{\{0\} \times X} \times Y \subset X^\Delta \times Y = W.$$

The stretched wedge manifold with respect to E is

$$\mathbb{E} = [0, 1) \times X \times Y = \mathbb{B} \times Y \subset X^\wedge \times Y = W^\wedge,$$

with smooth boundary $\partial\mathbb{E} = \{0\} \times X \times Y$.

Definition 1 For $(r, x, y) \in \mathbb{R}^N_+$ with $N = 1 + n + q$, assume that $u(r, x, y) \in \mathscr{D}'(\mathbb{R}^N_+)$. We say that $u(r, x, y) \in L_p(\mathbb{R}^N_+; d\mu)$ if

$$\|u\|_{L_p} = \left(\int_{\mathbb{R}^N_+} r^N |u(r, x, y)|^p d\mu \right)^{\frac{1}{p}} < +\infty,$$

where $d\mu = \frac{dr}{r} dx_1 \ldots .dx_n \frac{dy_1}{r} \ldots \frac{dy_q}{r}$ and for $1 \leq p < \infty$.

Moreover, the weighted L_p spaces with wight $\gamma \in \mathbb{R}$ is denoted by $L_p^{\gamma}(\mathbb{R}_+^N; d\mu)$, which consists of function $u(r, , y)$ such that

$$\|u\|_{L_p^{\gamma}} = \left(\int_{\mathbb{R}_+^N} r^N |r^{-\gamma} u(r, x, y)|^p d\mu \right)^{\frac{1}{p}} < +\infty.$$

Now, we can define the weighted p-Sobolev spaces with natural scale for all $1 \leq p < \infty$ on $\mathbb{R}_+^{N=1+n+q}$.

Definition 2 For $m \in \mathbb{N}$, $\gamma \in \mathbb{R}$ and $N = 1 + n + q$, the spaces

$$\mathscr{H}_p^{m,\gamma}(\mathbb{R}_+^N) = \left\{ u \in \mathscr{D}'(\mathbb{R}_+^N) \quad | \quad r^{\frac{N}{p}-\gamma} (r\partial_r)^k \partial_x^{\alpha} (r\partial_y)^{\beta} u \in L_p(\mathbb{R}_+^N; d\mu) \right\}$$

for $k \in \mathbb{N}$, multi-indices $\alpha \in \mathbb{N}^n$ and $\beta \in \mathbb{N}^q$ with $k + |\alpha| + |\beta| \leq m$. In other words, if $u(r, x, y) \in \mathscr{H}_p^{m,\gamma}(\mathbb{R}_+^N)$ then $(r\partial_r)^k \partial_x^{\alpha} (r\partial_y)^{\beta} u \in L_p^{\gamma}(\mathbb{R}_+^N; d\mu)$. Therefore, $\mathscr{H}_p^{m,\gamma}(\mathbb{R}_+^N)$ is a Banach space with the following norm

$$\|u\|_{\mathscr{H}_p^{m,\gamma}(\mathbb{R}_+^N)} = \sum_{k+|\alpha|+|\beta| \leq m} \left(\int_{\mathbb{R}_+^N} r^N |r^{-\gamma} (r\partial_r)^k \partial_x^{\alpha} (r\partial_y)^{\beta} u|^p d\mu \right)^{\frac{1}{p}}.$$

Moreover, the subspace $\mathscr{H}_{p,0}^{m,\gamma}(\mathbb{R}_+^N)$ of $\mathscr{H}_p^{m,\gamma}(\mathbb{R}_+^N)$ denotes the closure of $C_0^{\infty}(\mathbb{R}_+^N)$ in $\mathscr{H}_p^{m,\gamma}(\mathbb{R}_+^N)$. Now, similarly to the definitions above, we can introduce the following weighted p-Sobolev spaces on $X^{\wedge} \times Y$, where $X^{\wedge} = \mathbb{R}_+ \times X$ and $X^{\wedge} \times Y$ is an open stretched wedge.

$$\mathscr{H}_p^{m,\gamma}(X^{\wedge} \times Y) := \left\{ u \in \mathscr{D}'(X^{\wedge} \times Y) \quad | \quad r^{\frac{N}{p}-\gamma} (r\partial_r)^k \partial_x^{\alpha} (r\partial_y)^{\beta} u \in L_p(X^{\wedge} \times Y; d\mu) \right\}$$

for $k \in \mathbb{N}$, multi-indices $\alpha \in \mathbb{N}^n$ and $\beta \in \mathbb{N}^q$ with $k + |\alpha| + |\beta| \leq m$.

Then $\mathscr{H}_p^{m,\gamma}(X^{\wedge} \times Y)$ is a Banach space with the following norm

$$\|u\|_{\mathscr{H}_p^{m,\gamma}(X^{\wedge} \times Y)} = \sum_{k+|\alpha|+|\beta| \leq m} \left(\int_{X^{\wedge} \times Y} r^N |r^{-\gamma} (r\partial_r)^k \partial_x^{\alpha} (r\partial_y)^{\beta} u|^p d\mu \right)^{\frac{1}{p}}.$$

The subspace $\mathscr{H}_{p,0}^{m,\gamma}(X^{\wedge} \times Y)$ of $\mathscr{H}_p^{m,\gamma}(X^{\wedge} \times Y)$ is defined as the closure of $C_0^{\infty}(X^{\wedge} \times Y)$.

Definition 3 Let $\mathbb{E}$ be the stretched wedge to the finite wedge E, then $\mathscr{H}_p^{m,\gamma}(\mathbb{E})$ for $m \in \mathbb{N}$, $\gamma \in \mathbb{R}$ denotes the subset of all $u \in W_{loc}^{m,p}(int\mathbb{E})$ such that $\omega u \in$

$\mathscr{H}_p^{m,\gamma}(X^\wedge \times Y)$ for any cut-off function ω, supported by a collar neighborhood of $(0, 1) \times \partial\mathbb{E}$. Moreover, the subspace $\mathscr{H}_{p,0}^{m,\gamma}(\mathbb{E})$ of $\mathscr{H}_p^{m,\gamma}(\mathbb{E})$ is defined as follows

$$\mathscr{H}_{p,0}^{m,\gamma}(\mathbb{E}) := [\omega]\mathscr{H}_{p,0}^{m,\gamma}(X^\wedge \times Y) + [1-\omega]W_0^{m,p}(int\mathbb{E})$$

where the classical Sobolev space $W_0^{m,p}(int\mathbb{E})$ denotes the closure of $C_0^\infty(int\mathbb{E})$ in $W^{m,p}(\tilde{\mathbb{E}})$ for $\tilde{\mathbb{E}}$ that is a closed compact C^∞ manifold with boundary.

If $u \in L_p^{\frac{n+1}{p}}(\mathbb{E})$ and $v \in L_{p'}^{\frac{n+1}{p'}}(\mathbb{E})$ with $p, p' \in (1, \infty)$ and $\frac{1}{p} + \frac{1}{p'} = 1$, then one can obtain the following edge type Hölder inequality

$$\int_{\mathbb{E}} r^q |uv| d\mu \leq \left(\int_{\mathbb{E}} r^q |u|^p d\mu \right)^{\frac{1}{p}} \left(\int_{\mathbb{E}} r^q |v|^{p'} d\mu \right)^{\frac{1}{p'}}.$$

In the case $p = 2$, we have the corresponding edge type Schwartz inequality

$$\int_{\mathbb{E}} r^q |uv| d\mu \leq \left(\int_{\mathbb{E}} r^q |u|^2 d\mu \right)^{\frac{1}{2}} \left(r^q |v|^2 d\mu \right)^{\frac{1}{2}}.$$

In the sequel, for convenience we denote

$$(u, v)_2 = \int_{\mathbb{E}} r^q uv d\mu, \qquad \|u\|_{L_p^{\frac{n+1}{p}}(\mathbb{E})} = \left(\int_{\mathbb{E}} r^q |u|^p d\mu \right)^{\frac{1}{p}}.$$

Proposition 1 (Poincaré Inequality [7]) *Let $\mathbb{E} = [0, 1) \times X \times Y$ be a stretched edge manifold, $\gamma \in \mathbb{R}$ and $p \in (1, \infty)$. If $u \in \mathscr{H}_p^{1,\gamma}(\mathbb{E})$ then*

$$\|u(z)\|_{L_p^\gamma(\mathbb{E})} \leq c \|\nabla_{\mathbb{E}} u(z)\|_{L_p^\gamma(\mathbb{E})} \tag{3}$$

where $\nabla_{\mathbb{E}} := (r\partial_r, \partial_{x_1}, \ldots, \partial_{x_n}, r\partial_{y_1}, \ldots, r\partial_{y_q})$ and the constant c depending only on $\mathbb{E}$.

Proposition 2 ([7]) *For $1 < p < 2^*$ the embedding $\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \hookrightarrow \mathscr{H}_{p,0}^{0,\frac{n+1}{p}}(\mathbb{E})$ is continuous.*

Proposition 3 ([7]) *There exist $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots \leq \lambda_j \leq \ldots$, and $\lambda_j \to \infty$ such that for all $j \geq 1$, the following Dirichlet problem*

$$\begin{cases} -\Delta_{\mathbb{E}} \phi_j = \lambda_j \phi_j, & x \in int\mathbb{E}, \\ \phi_j = 0, & x \in \partial\mathbb{E}, \end{cases} \tag{4}$$

admits non-trivial solution in $\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$. *Moreover, we can choose positive* $\{\phi_j\}_{j\geq 1}$ *which constitute an orthonormal basis of Hilbert space* $\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$, *and the inequality*

$$\lambda_1^{\frac{1}{2}}\|u(z)\|_{L_2^{\frac{n}{2}}}(\mathbb{E}) \leq \|\nabla_{\mathbb{E}}u\|_{L_2^{\frac{n+1}{2}}}(\mathbb{E}),$$

holds.

3 Some Auxiliary Results

In this section we give some results about the potential wells for problem 1 and we obtain some properties of energy functional that we will use to prove the main results in Section 4.

Similar to the classical case, we introduce the following functionals on the cone Sobolev space $\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$:

$$J(u) = \frac{1}{2}\int_{\mathbb{E}} r^q|\nabla_{\mathbb{E}}u|^2d\mu + \frac{1}{2}\int_{\mathbb{E}} r^q V(z)|u|^2d\mu - \frac{1}{p+1}\int_{\mathbb{E}} r^q g_t(z)|u|^{p+1}d\mu,$$

$$K(u) = \int_{\mathbb{E}} r^q|\nabla_{\mathbb{E}}u|^2d\mu + \int_{\mathbb{E}} r^q V(x)|u|^2d\mu - \int_{\mathbb{E}} r^q g_t(z)|u|^{p+1}d\mu.$$

Then $J(u)$ and $K(u)$ are well-defined and belong to space $C^1\Big(\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}),\mathbb{R}\Big)$. Now we define

$$\mathscr{N} = \left\{u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \quad ; \quad K(u) = 0, \quad \int_{\mathbb{E}} r^q|\nabla_{\mathbb{E}}u|^2d\mu \neq 0\right\},$$

$$d = \inf\left\{\sup_{\lambda\geq 0} J(\lambda u) \quad ; \quad u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}), \quad \int_{\mathbb{E}} r^q|\nabla_{\mathbb{E}}u|^2d\mu \neq 0\right\}.$$

Thus, similar to the results in [20] we obtain that $0 < d = \inf_{u\in\mathscr{N}} J(u)$. For $0 < \delta$ we define

$$K_\delta(u) = \delta\left[\int_{\mathbb{E}} r^q|\nabla_{\mathbb{E}}u|^2d\mu + \int_{\mathbb{E}} r^q V(z)|u|^2d\mu\right] - \int_{\mathbb{E}} r^q g_t(z)|u|^{p+1}d\mu,$$

$$\mathcal{N}_\delta = \left\{ u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{B}) \quad ; \quad K_\delta(u) = 0, \quad \int_{\mathbb{E}} r^q |\nabla_{\mathbb{E}}|^2 d\mu \neq 0 \right\},$$

$$d(\delta) = \inf_{u \in \mathcal{N}_\delta} J(u).$$

Proposition 4 *If* $0 < \|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} < r(\delta)$ *where* $r(\delta) = \left(\frac{(C_*^2+1)\delta}{C_{**}^{p+1}}\right)^{\frac{1}{p-1}}$, *then* $K_\delta(u) > 0$. *In particular, if*

$$0 < \|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} < r(1)$$

then $K(u) > 0$.

Proof We conclude the following

$$\|g_t(z)^{\frac{1}{p+1}} u\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})} = \int_{\mathbb{E}} r^q |g_t(z)^{\frac{1}{p+1}} u(z)|^{p+1} d\mu = \int_{\mathbb{E}} r^q |g_t(z)||u(z)|^{p+1} d\mu \leq$$

$$\|g_t\|_{L^\infty} \int_{\mathbb{E}} r^q |u|^{p+1} d\mu \quad \Rightarrow$$

$$\|g_t(z)^{\frac{1}{p+1}} u\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})} \leq C_g \|u\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})}. \tag{5}$$

Also from definition of C_*:

$$\|V(z)^{\frac{1}{2}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \geq C_*^2 \|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}. \tag{6}$$

Then by definition of K_δ and using the assumption we get that

$$\begin{aligned} K_\delta(u) &= \delta \left[\int_{\mathbb{E}} r^q |\nabla_{\mathbb{E}} u|^2 d\mu + \int_{\mathbb{E}} r^q V(z) |u|^2 d\mu \right] - \int_{\mathbb{E}} r^q g_t(z) |u|^{p+1} d\mu \\ &\geq \delta(1 + C_*^2) \|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - C_{**}^{p+1} \|\nabla_{\mathbb{E}} u\|^{p+1}_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\ &= \left(\delta(1 + C_*^2) - C_{**}^{p+1} \|\nabla_{\mathbb{E}} u\|^{p-1}_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \right) \|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} > 0. \end{aligned}$$

In case that $\delta = 1$ then by definition of functional K we obtain that $K(u) > 0$.

Proposition 5 *If $K_\delta(u) < 0$, then $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} > r(\delta)$. In particular, if $K(u) < 0$, then $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} > r(1)$.*

Proof Since $K_\delta(u) < 0$, then by definition of $K_\delta(u)$, we get that $\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \neq 0$. Now, we have

$$\begin{aligned}\delta\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} &< \int_{\mathbb{E}} r^q g_t(z)|u(z)|^{p+1} d\mu - \delta \int_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu \\ &\leq \|g_t(x)^{\frac{1}{p+1}} u\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})} - \delta\|V(z)^{\frac{1}{2}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\ &< C_{**}^{p+1}\|\nabla_{\mathbb{E}} u\|^{p-1}_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - \delta C_*^2\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}.\end{aligned}$$

Therefore,

$$\|\nabla_{\mathbb{E}} u\|^{p-1}_{L_2^{\frac{n+1}{2}}(\mathbb{E})} > \left(\frac{\delta(1 + C_*^2)}{C_{**}^{p+1}}\right) = r^{p-1}(\delta).$$

Corollary 1 *Let $u \in \mathscr{H}^{1,\frac{n+1}{2}}_{2,0}(\mathbb{E})$, $K_\delta(u) = 0$ and $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \neq 0$. Then $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \geq r(\delta)$. In particular, if $K(u) = 0$ and $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \neq 0$, then $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \geq r(1)$.*

Lemma 1 *(i) The functional $J(\lambda u)$ admits its maximum for $\lambda = \lambda_*$ where*

$$\lambda_* = \left(\frac{\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu}{\int_{\mathbb{E}} r^q g_t(z)|u(z)|^{p+1} d\mu}\right)^{\frac{1}{p-1}}.$$

Also for $0 \leq \lambda < \lambda_*$, $J(\lambda u)$ is strictly increasing and for $\lambda_* < \lambda$, it is strictly decreasing.

(ii) $K(\lambda_* u) = 0$ and $K(\lambda u) > 0$ if $0 < \lambda < \lambda_*$. Also if $\lambda_* < \lambda$ then $K(\lambda u) < 0$.

(iii) By results in i and ii we obtain that

$$\begin{aligned}d &= \inf\left\{\sup_{\lambda \geq 0} J(\lambda u) \quad ; \quad u \in \mathscr{H}^{1,\frac{n+1}{2}}_{2,0}(\mathbb{E}), \quad \int_{\mathbb{E}} |\nabla_{\mathbb{E}} u|^2 d\mu \neq 0\right\} \\ &= \frac{p-1}{2(p+1)}(1 + C_*^2)^{\frac{p+1}{p-1}} C_{**}^{-2\frac{p+1}{1-p}} \|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}.\end{aligned}$$

Proof For proof of i and ii we obtain the following conclusions. Let $u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$ and $\int_{\mathbb{E}} |\nabla_{\mathbb{E}} u|^2 d\mu \neq 0$. Then by definition of J we obtain that

$$\begin{aligned}
\lim_{\lambda\to+\infty} J(\lambda u) &= \lim_{\lambda\to+\infty}\Big[\frac{1}{2}\int_{\mathbb{E}} r^q |\nabla_{\mathbb{E}}\lambda u|^2 d\mu + \frac{1}{2}\int_{\mathbb{E}} r^q V(z)|\lambda u(z)|^2 d\mu \\
&\quad - \frac{1}{p+1}\int_{\mathbb{E}} r^q g_t(z)|\lambda u(x)|^{p+1} d\mu\Big] \\
&= \lim_{\lambda\to+\infty}\Big[\frac{1}{2}\|\nabla_{\mathbb{E}}\lambda u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\|V(z)^{\frac{1}{2}}\lambda u(z)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - \frac{1}{p+1}\|g_t(z)^{\frac{1}{p+1}}\lambda u(z)\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})}\Big] \\
&= \lim_{\lambda\to+\infty}\Big[\frac{\lambda^2}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{\lambda^2}{2}\|V(z)^{\frac{1}{2}}u(z)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - \frac{\lambda^{p+1}}{p+1}\|g_t(z)^{\frac{1}{p+1}}u(z)\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})}\Big] \\
&\geq \lim_{\lambda\to+\infty}\Big[\frac{\lambda^2}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{\lambda^2}{2}C_*^2\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - \frac{\lambda^{p+1}}{p+1}C_{**}^{p+1}\|\nabla_{\mathbb{E}} u\|^{p+1}_{L_2^{\frac{n+1}{2}}(\mathbb{E})}\Big] \\
&= \lim_{\lambda\to+\infty}\Big[\frac{\lambda^2}{2} + \frac{\lambda^2}{2}C_*^2 - \frac{\lambda^{p+1}}{p+1}C_{**}^{p+1}\|\nabla_{\mathbb{E}} u\|^{p-1}_{L_2^{\frac{n+1}{2}}(\mathbb{E})}\Big]\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} = -\infty.
\end{aligned}$$

Also we have

$$\begin{aligned}
J(\lambda u) &= \frac{1}{2}\int_{\mathbb{E}} |\nabla_{\mathbb{E}}\lambda u|^2 d\mu + \frac{1}{2}\int_{\mathbb{E}} r^q V(z)|\lambda u(z)|^2 d\mu - \frac{1}{p+1}\int_{\mathbb{E}} r^q g_t(z)|\lambda u(z)|^{p+1} \\
&= \frac{\lambda^2}{2}\int_{\mathbb{E}} |\nabla_{\mathbb{E}} u|^2 d\mu + \frac{\lambda^2}{2}\int_{\mathbb{E}} V(z)|u(z)|^2 d\mu - \frac{\lambda^{p+1}}{p+1}\int_{\mathbb{E}} g_t(z)|u(z)|^{p+1} d\mu.
\end{aligned}$$

Then

$$\begin{aligned}
\frac{\partial J(\lambda u)}{\partial\lambda} &= \lambda\int_{\mathbb{E}} |\nabla_{\mathbb{E}} u|^2 d\mu + \lambda\int_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu - \lambda^p\int_{\mathbb{E}} r^q g_t(z)|u(z)|^{p+1} d\mu \\
&= \lambda\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \lambda\|V(z)^{\frac{1}{2}}u(z)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - \lambda^p\|g_t(z)^{\frac{1}{p+1}}u(z)\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})}.
\end{aligned}$$

Now, $\frac{\partial J(\lambda u)}{\partial\lambda} = 0$, it follows that

$$\lambda_* := \Big(\frac{\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu}{\int_{\mathbb{E}} r^q g_t(z)|u(z)|^{p+1} d\mu}\Big)^{\frac{1}{p-1}}$$

is a maximum of $J(\lambda u)$ since $\frac{\partial^2(J(\lambda u))}{\partial\lambda^2}|_{\lambda=\lambda_*} < 0$.

(iii) Using of i and ii $\sup\limits_{\lambda\geq 0} J(\lambda u) = J(\lambda_* u)$. Thus,

$$J(\lambda_* u) = \frac{1}{2}\int\limits_{\mathbb{E}} r^q |\nabla_{\mathbb{E}}\lambda_* u|^2 d\mu + \frac{1}{2}\int\limits_{\mathbb{E}} r^q V(z)|\lambda_* u(z)|^2 d\mu$$

$$- \frac{1}{p+1}\int\limits_{\mathbb{E}} r^q g_t(z)|\lambda_* u|^{p+1} d\mu$$

$$= \lambda_*^2\Big[\frac{1}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\int\limits_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu - \frac{\lambda_*^{p-1}}{p+1}\int\limits_{\mathbb{E}} r^q g_t(z)|u|^{p+1} d\mu\Big]$$

$$= \lambda_*^2\Big[\frac{1}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\int\limits_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu$$

$$- \frac{1}{p+1}\Big(\frac{\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int\limits_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu}{\int\limits_{\mathbb{E}} r6q g_t(z)|u(z)|^{p+1} d\mu}\Big)^{\frac{p-1}{p-1}} \Big(\int\limits_{\mathbb{E}} r^q g_t(z)|u|^{p+1} d\mu\Big)\Big]$$

$$= \lambda_*^2\Big[\frac{1}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\int\limits_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu - \frac{1}{p+1}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}$$

$$- \frac{1}{p+1}\int\limits_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu\Big]$$

$$= \lambda_*^2\Big[(\frac{1}{2} - \frac{1}{p+1})\,\|\nabla u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + (\frac{1}{2} - \frac{1}{p+1})\int\limits_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu\Big]$$

$$= \Big(\frac{\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int\limits_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu}{\int\limits_{\mathbb{E}} r^q g_t(z)|u(z)|^{p+1} d\mu}\Big)^{\frac{2}{p-1}}$$

$$\times \Big(\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int\limits_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu\Big) \times \frac{p-1}{2(p+1)}$$

$$\geq \Big(\frac{\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + C_*^2\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}}{\int\limits_{\mathbb{E}} r^q g_t|u|^{p+1} d\mu}\Big)^{\frac{2}{p-1}}\Big[\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + C_*^2\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}\Big]\frac{p-1}{2(p+1)}$$

$$\geq \frac{p-1}{2(p+1)}(1+C_*^2)^{\frac{p+1}{p-1}} C_{**}^{\frac{-2(p+1)}{p-1}}\,\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}.$$

Therefore,

$$d = \inf_{u\in\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}),\|\nabla_{\mathbb{E}}u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}\neq 0} J(\lambda_* u)$$

$$= \frac{p-1}{2(p+1)}(1+C_*^2)^{\frac{p+1}{p-1}}\, C_{**}^{\frac{-2(p+1)}{p-1}}\, \|\nabla_{\mathbb{E}}u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}.$$

Proposition 6 *Let* $0 < \delta < \frac{p+1}{2}$, *then* $d(\delta) \geq a(\delta)r^2(\delta)$ *where* $a(\delta) = \left(\frac{1}{2} - \frac{\delta}{p+1}\right)\left(1+C_*^2\right)$. *Moreover, we have*

$$d(\delta) = \inf_{u\in\mathscr{N}_\delta} J(u) = d\ \lambda(\delta)^2\ a(\delta)[1+c_*^2]^{-1}\ \frac{2(p+1)}{p-1}.$$

Proof Let $u \in \mathscr{N}_\delta$, so by Proposition 5 we get that $\|\nabla_{\mathbb{E}}u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} > r(\delta)$. Then by definition of J and K_δ we obtain that

$$J(u) = \frac{1}{2}\|\nabla_{\mathbb{E}}u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\int_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu - \frac{1}{p+1}\int_{\mathbb{E}} r^q g_t(z)|u(z)|^{p+1}d\mu$$

$$= \frac{1}{2}\|\nabla_{\mathbb{E}}u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\int_{\mathbb{E}} r^q V(z)|u(z)|^2 r^q$$

$$- \frac{1}{p+1}\left(\delta\|\nabla_{\mathbb{E}}u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - K_\delta(u) + \delta\int_{\mathbb{E}} r^q V(z)|u(z)|^2 d\mu\right).$$

Since $K_\delta(u) = 0$,

$$J(u) \geq (\frac{1}{2} - \frac{\delta}{p+1})\|\nabla_{\mathbb{E}}u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{\delta(p-1)}{2(p+1)}\|V(x)^{\frac{1}{2}}u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}$$

$$\geq (\frac{1}{2} - \frac{\delta}{p+1})\|\nabla_{\mathbb{E}}u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{\delta(p-1)}{2(p+1)}C_*^2\|\nabla_{\mathbb{E}}u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}$$

$$= \left(\frac{1}{2} - \frac{\delta}{p+1}\right)\left(1+C_*^2\right)\|\nabla_{\mathbb{E}}u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}.$$

Since $\|\nabla u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \geq r^2(\delta)$ then,

$$d(\delta) \geq a(\delta)r^2(\delta).$$

Now, we prove the second part of the assertion. By definition of $\mathcal{N}_\delta$ and $\mathcal{N}$, for $\bar{u} \in \mathcal{N}_\delta$ and $\lambda\bar{u} \in \mathcal{N}$, we obtain

$$\lambda^2 \|\nabla_{\mathbb{E}}\bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \lambda^2 \int_{\mathbb{E}} r^q V(z)|\bar{u}|^2 d\mu = \lambda^{p+1} \int_{\mathbb{E}} r^q g_t(z)|\bar{u}|^{p+1} d\mu, \tag{7}$$

and

$$\delta \|\nabla_{\mathbb{E}}\bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \delta \int_{\mathbb{E}} r^q V(z)|\bar{u}|^2 d\mu = \int_{\mathbb{E}} r^q g_t(z)|\bar{u}|^{p+1} d\mu. \tag{8}$$

Then 7 gives

$$\lambda = \left(\frac{\|\nabla_{\mathbb{E}}\bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int_{\mathbb{E}} r^q V(z)|\bar{u}|^2 d\mu}{\int_{\mathbb{E}} r^q g_t(z)|\bar{u}|^{p+1} d\mu} \right)^{\frac{1}{p-1}}, \tag{9}$$

and 8 gives that

$$\delta = \frac{\int_{\mathbb{E}} r^q g_t(z)|\bar{u}|^{p+1} d\mu}{\|\nabla_{\mathbb{E}}\bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int_{\mathbb{E}} r^q V(z)|\bar{u}|^2 d\mu}. \tag{10}$$

By 10 and 9, we define

$$\lambda = \lambda(\delta) = \left(\frac{1}{\delta}\right)^{\frac{1}{p-1}}. \tag{11}$$

Moreover, for such λ, $\lambda\bar{u} \in \mathcal{N}$, so by definition of d we get that

$$\begin{aligned} d \le J(\lambda\bar{u}) &= \frac{1}{2}\|\nabla_{\mathbb{E}}\lambda\bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\int_{\mathbb{E}} r^q V(z)|\lambda\bar{u}|^2 d\mu \\ &\quad - \frac{1}{p+1}\int_{\mathbb{E}} r^q g_t(z)|\lambda\bar{u}|^{p+1} d\mu \\ &= \frac{\lambda^2}{2}\|\nabla_{\mathbb{E}}\bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{\lambda^2}{2}\int_{\mathbb{E}} r^q V(z)|\bar{u}|^2 d\mu \\ &\quad - \frac{1}{p+1}\left[\|\nabla_{\mathbb{E}}\lambda\bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int_{\mathbb{E}} r^q V(z)|\lambda\bar{u}|^2 d\mu - K(\lambda\bar{u})\right] \end{aligned}$$

$$= \lambda^2 \Big[\frac{1}{2} \|\nabla_{\mathbb{E}} \bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2} \int_{\mathbb{E}} r^q V(z) |\bar{u}|^2 d\mu - \frac{1}{p+1} \|\nabla_{\mathbb{E}} \bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}$$

$$- \frac{1}{p+1} \int_{\mathbb{E}} r^q V(z) |\bar{u}|^2 d\mu \Big]$$

$$\leq (\frac{1}{\delta})^{\frac{2}{p-1}} \Big[\frac{p-1}{2(p+1)} \|\nabla_{\mathbb{E}} \bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - \frac{(1-p)C_*^2}{2(p+1)} \|\nabla_{\mathbb{E}} \bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \Big].$$

On the other hand,

$$d(\delta) = J(\bar{u}) = \frac{1}{2} \|\nabla_{\mathbb{E}} \bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2} \int_{\mathbb{E}} r^q V(z) |\bar{u}|^2 d\mu - \frac{1}{p+1} \int_{\mathbb{E}} r^q g_t(z) |\bar{u}|^{p+1} d\mu$$

$$= \frac{1}{2} \|\nabla_{\mathbb{E}} \bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2} \int_{\mathbb{E}} r^q V(z) |\bar{u}|^2 d\mu - \frac{1}{p+1} \Big(\delta \|\nabla_{\mathbb{E}} \bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}$$

$$+ \delta \int_{\mathbb{E}} r^q V(z) |\bar{u}|^2 d\mu - K_\delta(\bar{u}) \Big)$$

$$= (\frac{1}{2} - \frac{\delta}{p+1}) \|\nabla_{\mathbb{E}} \bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + (\frac{1}{2} - \frac{\delta}{p+1}) \int_{\mathbb{E}} r^q V(z) |\bar{u}|^2 d\mu$$

$$\geq \left(\frac{1}{2} - \frac{\delta}{p+1} \right) (1 + C_*^2) \|\nabla_{\mathbb{E}} \bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} = a(\delta) \|\nabla_{\mathbb{E}} \bar{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}.$$

Indeed,

$$d \leq J(\lambda \bar{u}) \leq \big(\frac{1}{\delta}\big)^{\frac{2}{p-1}} \left[\frac{p-1}{2(p+1)} \, (1 + C_*^2) \right] \frac{d(\delta)}{a(\delta)}.$$

Hence,

$$d(\delta) \geq a(\delta) \, \big(\frac{1}{\delta}\big)^{-\frac{2}{p-1}} \, [1 + C_*^2]^{-1} \, [\frac{2(p+1)}{p-1}] d.$$

Now, we let $0 < \delta$ and $\tilde{u} \in \mathcal{N}$ is minimizer of d that is

$$d = J(\tilde{u}) = \frac{1}{2} \|\nabla_{\mathbb{E}} \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2} \int_{\mathbb{E}} r^q V(z) |\tilde{u}|^2 d\mu - \frac{1}{p+1} \int_{\mathbb{E}} r^q g_t(z) |\tilde{u}|^{p+1} d\mu.$$

we define $\lambda = \lambda(\delta)$ by

$$\delta \|\nabla_{\mathbb{E}} \lambda \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \delta \int_{\mathbb{E}} r^q V(z) |\lambda \tilde{u}|^2 d\mu = \int_{\mathbb{E}} r^q g_t(z) |\lambda \tilde{u}|^{p+1} d\mu.$$

Then for any $0 < \delta$, there exists a unique λ which satisfies

$$\lambda = \delta^{\frac{1}{p-1}}.$$

Hence, for such λ, $\lambda \tilde{u} \in \mathscr{N}_\delta$ by definition of $d(\delta)$ we get that

$$\begin{aligned}
d &= \frac{1}{2} \|\nabla_{\mathbb{E}} \lambda \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2} \int_{\mathbb{E}} r^q V(z) |\lambda \tilde{u}|^2 d\mu \\
&\quad - \frac{1}{p+1} \left(\delta \|\nabla_{\mathbb{E}} \lambda \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \delta \int_{\mathbb{E}} r^q V(z) |\lambda \tilde{u}|^2 d\mu - K_\delta(\lambda \tilde{u}) \right) \\
&= (\frac{1}{2} - \frac{\delta}{p+1}) \|\nabla_{\mathbb{E}} \lambda \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + (\frac{1}{2} - \frac{\delta}{p+1}) \int_{\mathbb{E}} r^q V(z) |\lambda \tilde{u}|^2 d\mu \\
&\geq \left[\frac{1}{2} - \frac{\delta}{p+1} + C_*^2 (\frac{1}{2} - \frac{\delta}{p+1}) \right] \|\nabla_{\mathbb{E}} \lambda \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
d(\delta) &\leq J(\lambda \tilde{u}) = \frac{1}{2} \|\nabla_{\mathbb{E}} \lambda \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2} \int_{\mathbb{E}} r^q V(z) |\lambda \tilde{u}|^2 d\mu - \frac{1}{p+1} \int_{\mathbb{E}} r^q g_t(z) |\lambda \tilde{u}|^{p+1} d\mu \\
&= \frac{\lambda^2}{2} \|\nabla_{\mathbb{E}} \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{\lambda^2}{2} \int_{\mathbb{E}} r^q V(z) |\tilde{u}|^2 d\mu \\
&\quad - \frac{1}{p+1} \left(\delta \|\nabla_{\mathbb{E}} \lambda \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \delta \int_{\mathbb{E}} r^q V(z) |\lambda \tilde{u}|^2 d\mu - K_\delta(\lambda \tilde{u}) \right) \\
&= \lambda^2 \left[(\frac{1}{2} - \frac{\delta}{p+1}) \|\nabla_{\mathbb{E}} \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - (\frac{\delta}{p+1} - \frac{1}{2}) \int_{\mathbb{E}} r^q V(z) |\tilde{u}|^2 d\mu \right] \\
&\leq \delta^{\frac{2}{p-1}} \left[(\frac{1}{2} - \frac{\delta}{p+1}) \|\nabla_{\mathbb{E}} \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - (\frac{\delta}{p+1} - \frac{1}{2}) C_*^2 \|\nabla_{\mathbb{E}} \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \right] \\
&\leq \delta^{\frac{2}{p+1}} \left[\frac{1}{2} - \frac{\delta}{p+1} + (\frac{1}{2} - \frac{\delta}{p+1}) C_*^2 \right] \|\nabla_{\mathbb{E}} \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} = \delta^{\frac{2}{p-1}} a(\delta) \|\nabla_{\mathbb{E}} \tilde{u}\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}.
\end{aligned}$$

Then,

$$d(\delta) \leq \delta^{\frac{2}{p-1}}\, a(\delta)\, d\, [1 + C_*^2]^{-1}\, \frac{2(p+1)}{p-1}.$$

Therefore,

$$d(\delta) = \inf J(u)_{u \in \mathcal{N}_\delta} = \delta^{\frac{2}{p-1}}\, a(\delta)\, d\, [1 + C_*^2]^{-1}\, \frac{2(p+1)}{p-1}.$$

Remark 1 According to $d(\delta)$ in Proposition 6, we obtain that

(i) $\lim\limits_{\delta \to 0} d(\delta) = 0.$

(ii) $d(\delta) = d\frac{2(p+1)}{p-1}\left[\frac{1}{2}\delta^{\frac{2}{p-1}} - \frac{1}{p+1}\delta^{\frac{p+1}{p-1}}\right]$. Then

$$d'(\delta) = \frac{d2(p+1)}{(p-1)^2}\delta^{\frac{2}{p-1}}\left[\delta^{-1} - 1\right] = 0 \Rightarrow \delta = 1.$$

Hence, if $0 < \delta < 1$ then $d(\delta)$ is strictly increasing function and if $\delta > 1$ then $d(\delta)$ is strictly decreasing function.

4 Invariance of the Solutions

Now, we introduce the following potential wells

$$W = \left\{u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \quad ; \quad K(u) > 0, \quad J(u) < d\right\} \cup \{0\},$$

$$W_\delta = \left\{u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \quad ; \quad K_\delta(u) > 0, \quad J(u) < d(\delta)\right\} \cup \{0\},$$

for $0 < \delta$, and corresponding potentials outside of the set that defined as above

$$E = \left\{u \in u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \quad ; \quad K(u) < 0, \quad J(u) < d\right\},$$

$$E_\delta = \left\{u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \quad ; \quad K_\delta(u) < 0, \quad J(u) < d(\delta)\right\}$$

for any $0 < \delta$. According to the definition of potential wells W_δ and potential outside E_δ one can get the following inclusions:

(I) $W_{\delta_1} \subset W_{\delta_2}$ whenever $0 < \delta_1 < \delta_2 \leq 1$,

(II) $E_{\delta_1} \subset E_{\delta_2}$ whenever $1 \leq \delta_2 < \delta_1 < \frac{p+1}{2}$. Furthermore, from the above results on can define the following sets

$$V_\delta = \left\{u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \quad : \quad \|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} < r(\delta)\right\}$$

$$\bar{V}_\delta = V_\delta \cup \partial V_\delta = \left\{u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \quad : \quad \|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \leq r(\delta)\right\}$$

$$V_\delta^c = \left\{u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \quad : \quad \|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} > r(\delta)\right\}.$$

Then for every $0 < \delta < \frac{p+1}{2}$ one gets that

$$V_{t(\delta)} \subset W_\delta \subset V_{s(\delta)}, \quad E_\delta \subset V_\delta^c$$

where

$$V_{t(\delta)} = \left\{u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \quad : \quad \|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} < \min\{r^2(\delta), r_0^2(\delta)\}\right\}$$

$$V_{s(\delta)} = \left\{u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}) \quad : \quad \|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} < \frac{d(\delta)}{a(\delta)}\right\}$$

where $r_0(\delta)$ is the unique real root of equation $\frac{r^2}{2} = d(\delta)$.

Definition 4 Suppose that $u(t)$ is a weak solution of problem 1. $T_{\max}$ is called maximal existence time of solution $u(t)$ if one the following conditions hold:

(1) If $u(t)$ exists for every $0 \leq t < +\infty$ then $T_{\max} = +\infty$. In this case, we say that the solution is global.
(2) If there exists a $t_0 \in (0, \infty)$ such that $u(t)$ exists for every $0 \leq t < t_0$, but does not exist at $t = t_0$, then $T_{\max} = t_0$.

Definition 5 $u = u(z,t) \in L^\infty\Big(0, T_{\max}; \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\Big)$ with $\partial_t u \in L^\infty\Big(0, T_{\max}; L_2^{\frac{n+1}{2}}(\mathbb{E})\Big)$ is called a weak solution of the problem 1 on $int\mathbb{E} \times [0, T_{\max})$ if

$$\begin{aligned}(u_t, v)_2 + \gamma(\nabla_{\mathbb{E}} u, \nabla_{\mathbb{E}} v)_2 + \int_0^t (\nabla_{\mathbb{E}} u, \nabla_{\mathbb{E}} v)_2 d\tau + \int_0^t (V(x)u, v)_2 d\tau \\ = \int_0^t (g_t(z)|u|^{p-1}u, v)_2 d\tau \\ + (\gamma u_0, v)_2 + (u_1, v)_2 \qquad \forall v \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E}),\end{aligned}$$

$u(z,0) = u_0$ in $\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$ and hold the following energy inequality

$$I(t) + \gamma \int_0^t \|\nabla_{\mathbb{E}}(\partial_\tau u)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} d\tau \leq I(0), \quad \forall t \in (0, T_{\max}),$$

where $0 \leq T_{\max} \leq \infty$ and

$$I(t) = \frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{p+1}\int_0^t \|(\frac{d}{d\tau} g_\tau(z))^{\frac{1}{p+1}} u\|^{p+1}_{L_p^{\frac{n+1}{p+1}}(\mathbb{E})} d\tau + J(u).$$

We note, since $u \in L^\infty\Big(0, T_{\max}; \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\Big)$ and $\partial_t u \in L^\infty\Big(0, T_{\max}; L_2^{\frac{n+1}{2}}(\mathbb{E})\Big)$ from the first equation of the problem 1 as similar in [13], one can obtain that $\partial_t^2 u \in L^\infty\Big(0, T_{\max}; \mathscr{H}_{2,0}^{-1,\frac{n+1}{2}}(\mathbb{E})\Big)$.

Now we discuss the invariance of some sets corresponding to the problem 1.

Proposition 7 *Let $0 < J(u) < d$ for $u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$. Suppose that $\delta_1 < 1 < \delta_2$ be roots of equation $d(\delta) = J(u)$. Then $K_\delta(u)$ has no change in its sign for $\delta \in (\delta_1, \delta_2)$.*

Proof Since $0 < J(u) < d$ then by Propositions 1 and 2 we can assume that $\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \neq 0$. We assume that there exists a $\delta_0 \in (\delta_1, \delta_2)$ for which $K_{\delta_0}(u) = 0$. Hence, by definition of $d(\delta)$ we have $J(u) \geq d(\delta)$. But, we have two cases the following for δ_0

$$\begin{cases} \delta_1 < \delta_0 < 1 < \delta_2 \\ \delta_1 < 1 < \delta_0 < \delta_2 \end{cases}$$

Now, by Remark 1 We get that $d(\delta_1) < d(\delta_0)$ or $d(\delta_2) < d(\delta_0)$ then we obtain that $d(\delta_1) = d(\delta_2) = J(u) < d(\delta_0)$ that this is contradiction .

Theorem 1 *Let $u_0 \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$, $0 < e < d$. Suppose that $\delta_1 < \delta_2$ are roots of equations $d(\delta) = e$ then*

(i) all solutions of problem 1 with $0 < J(u_0) \leq e$ belong to set W_δ for $\delta_1 < \delta < \delta_2$ provided $K(u_0) > 0$ or $\|\nabla_{\mathbb{E}} u_0\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} = 0$.

(ii) all solutions of problem 1 with $0 < J(u_0) \leq e$ belong to E_δ for $\delta \in (\delta_1, \delta_2)$ provided $K(u_0) < 0$.

Proof

(i) Let $u(t)$ be a solution of the problem 1 with initial value u_0 for which satisfies in conditions $0 < J(u_0) \leq e < d$, $K(u_0) > 0$ or $\|\nabla_{\mathbb{E}} u_0\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} = 0$. Let T

be existence time for solution $u(t)$. If $\|\nabla_{\mathbb{E}} u_0\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} = 0$, then since u_0 has compact support $u_0 = 0$, so by definition of W_δ we obtain that $u_0 \in W_\delta$. If $K(u_0) > 0$ then by assumption we have

$$0 < J(u_0) \le e = d(\delta_1) = d(\delta_2) < d(\delta) \le d$$

for $\delta_1 < \delta < \delta_2$. Hence, $K_\delta(u_0(t)) > 0$ for $\delta_1 < \delta < \delta_2$, by Proposition 7. Therefore, by definition of W_δ, $u_0 \in W_\delta$ for $\delta_1 < \delta < \delta_2$. Now, we have to show that for $\delta_1 < \delta < \delta_2$ and $0 < t < T$, $u(t) \in W_\delta$. Suppose that, there exist $t_0 \in (0, T)$ such that for $\delta_1 < \delta < \delta_2$, $u(t_0) \in \partial W_\delta$. Then we can imply that, $K_\delta(u(t_0)) = 0$ and $\|\nabla_{\mathbb{E}} u_0\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \neq 0$, or by definition of W_δ, $J(u(t_0)) = d(\delta)$. Since $u(t_0)$ is a solution of problem 1, so it satisfies in energy inequality i.e.

$$\begin{aligned}\frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{p+1}\int_0^t \|(\frac{d}{d\tau} g_\tau(z))^{\frac{1}{p+1}} u\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{Z})} d\tau + J(u(t)) \\ + \gamma \int_0^t \|\nabla_{\mathbb{E}}(\partial_\tau u)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} d\tau \\ \le I(0) = J(u_0) \le e < d(\delta),\end{aligned}$$

for any $\delta \in (\delta_1, \delta_2)$ and $t \in (0, T)$. Therefore, the equality $J(u(t_0)) = d(\delta)$ for any $\delta \in (\delta_1, \delta_2)$ and $t \in (0, T)$ is not possible. If $K_\delta(u(t_0)) = 0$ and $\|\nabla_{\mathbb{E}} u_0\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \neq 0$, then by definition of $d(\delta)$ we get that $d(\delta) \le J(u_0(t))$, that is in contradiction with energy inequality. Therefore, $u(t) \in W_\delta$ for any $\delta \in (\delta_1, \delta_2)$ and $t \in (0, T)$.

(ii) similar to first case it can be prove that $u_0 \in E_\delta$ for $\delta \in (\delta_1, \delta_2)$ provided $K_\delta(u_0) < 0$. Now, we should prove $u(t) \in E_\delta$ for any $\delta \in (\delta_1, \delta_2)$ and $t \in (0, T)$. Suppose that there exist $t_0 \in (0, T)$, such that for $t \in [0\ ,\ t_0)$, $u(t) \in E_\delta$ and $u(t_0) \in \partial E_\delta$, that is, $K_\delta(u_0) = 0$ or $J(u(t_0)) = d(\delta)$ for $\delta \in (\delta_1, \delta_2)$. According to energy inequality the equality $J(u(t_0)) = d(\delta)$ is not possible similar to first case. Hence, we assume that $K_\delta(u(t_0)) = 0$, then $K_\delta(u(t)) < 0$ for $t \in (0, t_0)$, since for $t \in [0, t_0)$, $u(t) \in E_\delta$, then by definition of E_δ, $K_\delta(u(t)) < 0$. Now, using the Proposition 5 we obtain that $\|\nabla_{\mathbb{E}} u(t)\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} > r(\delta)$ and $\|\nabla_{\mathbb{E}} u(t_0)\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} > r(\delta) \neq 0$. Hence by definition of $d(\delta)$, $J(u(t_0)) \ge d(\delta)$ which is in contradiction with energy inequality.

Remark 2 suppose that all assumptions in Theorem 1 hold. Then for any $\delta \in (\delta_1, \delta_2)$ both seta W_δ and E_δ are invariant. Moreover, both sets

$$W_{\delta_1\delta_2} = \bigsqcup_{\delta_1<\delta<\delta_2} W_\delta, \qquad E_{\delta_1\delta_2} = \bigsqcup_{\delta_1<\delta<\delta_2} E_\delta$$

are invariant respectively under flow of the problem 1. Hence, we can get for all weak solutions of the problem 1

$$u(t) \notin \mathscr{N}_{\delta_1\delta_2} = \bigsqcup_{\delta_1<\delta<\delta_2} \mathscr{N}_\delta.$$

To discuss about the invariant of the solutions with negative level energy, we introduce the following results.

Proposition 8 *Let* $u_0 \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$ *and* $u_1 \in L_2^{\frac{n+1}{2}}(\mathbb{E})$. *Suppose that* $I(0) = 0$ *and* $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \neq 0$. *Then all weak solutions of the problem 1 satisfy*

$$\|\nabla_{\mathbb{E}} u\|^{p-1}_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \geq M = \frac{(p+1)(1+C_*^2)}{2C_{**}^{p+1}}.$$

Proof Let us consider $u \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$ as a weak solution of the problem 1. According to the Definition 5

$$I(t) + \gamma \int_0^t \|\nabla_{\mathbb{E}}(\partial_\tau u)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} d\tau \leq I(0) = 0.$$

Therefore, by definition of constants C_* and C_{**}

$$\begin{aligned}\frac{1}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{C_*^2}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} &\leq \frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\ &\quad + \frac{1}{2}\int_{\mathbb{E}} r^q V(z)|u|^2 d\mu \\ &\leq \frac{C_{**}^{p+1}}{p+1}\|\nabla_{\mathbb{E}} u\|^{p+1}_{L_2^{\frac{n+1}{2}}(\mathbb{E})}.\end{aligned}$$

Hence,

$$\|\nabla_{\mathbb{E}} u\|^{p-1}_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \geq \frac{(p+1)(1+C_*^2)}{2C_{**}^{p+1}} = M.$$

Theorem 2 *Let* $u_0 \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$ *and* $u_1 \in L_2^{\frac{n+1}{2}}(\mathbb{E})$. *Suppose that either* $I(0) < 0$ *or* $I(0) = 0$ *and* $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \neq 0$. *Then all weak solutions of the problem 1 belong to* E_δ *for any* $\delta > 0$.

Proof Let $u(t)$ be an arbitrary weak solution of the problem 1 with expressed assumptions in face of the Theorem and T be the existence time of $u(t)$. From Definition 5, for every $\delta > 0$ and $t \in [0, T)$, we can obtain

$$
\begin{aligned}
&\frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + a(\delta)\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{p+1}K_\delta(u) \\
&\quad = \frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \left((\frac{1}{2} - \frac{\delta}{p+1})(1 + C_*^2)\right)\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
&\quad + \frac{1}{p+1}\left(\delta\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \delta\int_{\mathbb{E}} r^q V(z)|u|^2 d\mu \right. \\
&\qquad \left. - \int_{\mathbb{E}} r^q g_t(z)|u|^{p+1} d\mu\right) \\
&\quad = \frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - \frac{\delta C_*^2}{p+1}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
&\quad + \frac{C_*^2}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{\delta}{p+1}\int_{\mathbb{E}} r^q V(z)|u|^2 d\mu \\
&\qquad - \frac{1}{p+1}\int_{\mathbb{E}} r^q g_t(z)|u|^{p+1} d\mu \\
&\quad = \frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - \frac{\delta C_*^2}{p+1}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
&\quad + \frac{C_*^2}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - \left(\frac{1}{2} - \frac{\delta}{p+1}\right)\int_{\mathbb{E}} r^q V(z)|u|^2 d\mu \\
&\qquad + \frac{1}{2}\int_{\mathbb{E}} r^q V(z)|u|^2 d\mu \\
&\quad - \frac{1}{p+1}\int_{\mathbb{E}} r^q g_t(z)|u|^{p+1} d\mu \le \frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + J(u) \\
&\quad + \left(\frac{C_*^2}{2} - \frac{\delta C_*^2}{p+1} - C_*^2\left(\frac{1}{2} - \frac{\delta}{p+1}\right)\right)\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
&\quad + \frac{1}{p+1}\int_0^t \|(\frac{d}{d\tau} g_\tau(x))^{\frac{1}{p+1}} u\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})} \\
&\qquad + \gamma\int_0^t \|\nabla_{\mathbb{E}}(\partial_\tau u)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \le I(0). \qquad (12)
\end{aligned}
$$

If $I(0) < 0$, then 12 implies that $K_\delta(u) < 0$ and $J(u) < 0 < d(\delta)$ for every $\delta > 0$ and $t \in [0, T)$. If $I(0) = 0$ and $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \neq 0$, then Proposition 8 gives $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \geq M$ for $t \in [0, T)$. Again by relation 12 we get $K_\delta(u) < 0$ and $J(u) < 0 < d(\delta)$ for $\delta > 0$ and $t \in [0, T)$. Therefore, for two cases discussed above, for every $\delta > 0$ and $t \in [0, T)$, we have $u \in E_\delta$.

5 Global Existence and Finite-Time of the Solutions

In this section, we prove the global existence and nonexistence of solutions and give a sharp condition for global existence of solutions for problem 1 with $I(0) < d$.

Theorem 3 *Let $\gamma \geq 0$, $u_0 \in \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$ and $u_1 \in L_2^{\frac{n+1}{2}}(\mathbb{E})$. Suppose that $I(0) < d$, $K(u_0) > 0$ or $\|\nabla_{\mathbb{E}} u_0\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} = 0$. Then problem 1 admits a global weak solution $u(t) \in L^\infty\Big(0, \infty; \mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\Big)$ with $\partial_t u \in L_2^{\frac{n+1}{2}}(\mathbb{E})$ and $u(t) \in W$ for $t \in [0, \infty)$.*

Proof By Proposition 3 we can choose $\{w_j(z)\}$ as orthonormal basis of space $\mathscr{H}_{2,0}^{1,\frac{n}{2}}(\mathbb{B})$. Then we construct approximation solution $u_m(z, t)$ similar to [20] as following:

$$u_m(z, t) = \sum_{j=1}^{m} h_{jm}(t) w_j(z),$$

for $m = 1, 2, \ldots$ that satisfies in problem 1 then,

$$(\partial_t^2 u_m, w_k)_2 + (\nabla_{\mathbb{E}} u_m, \nabla_{\mathbb{E}} w_k)_2 + (V(z) u_m, w_k)_2 + \gamma\, (\nabla_{\mathbb{E}}(\partial_t u_m), \nabla_{\mathbb{E}} w_k)_2 = (g_t(z) u_m |u_m|^{p-1}, w_k)_2, \tag{13}$$

$$u_m(z, 0) = \sum_{j=1}^{m} h_{jm}(0) w_j(z) \to u_0(z), \tag{14}$$

in $\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$ and

$$\partial_t u_m(z, 0) = \sum_{j=1}^{m} h'_{jm}(0) w_j(z) \to u_1(z), \tag{15}$$

in $L_2^{\frac{n+1}{2}}(\mathbb{E})$. Multiplying 13, 14 and 15 by $h'_{km}(t)$ and forming the crossmarklogo sum on $k = 1, 2, \ldots,$

$$\sum_{k=1}^{m}(\partial_t^2 u_m, w_k)_2 h'_{km}(t) + (\nabla_{\mathbb{E}} u_m, \nabla_{\mathbb{E}} w_k)_2 h'_{km}(t) + (V(z)u_m, w_k)_2 h'_{km}(t)$$
$$+ \sum_{k=1}^{m}\gamma\,(\nabla_{\mathbb{E}}(\partial_t u_m), \nabla_{\mathbb{E}} w_k)_2 h'_{km}(t)$$
$$= \sum_{k=1}^{m}(g_t(z)u_m|u_m|^{p-1}, w_k)_2 h'_{km}(t),$$

for $m = 1, 2, 3, \ldots$. Therefore,

$$\int_{\mathbb{E}} r^q \partial_t^2 u_m \partial_t u_m d\mu + \int_{\mathbb{E}} r^q \nabla_{\mathbb{E}} u_m \partial_t \nabla_{\mathbb{E}} u_m d\mu + \int_{\mathbb{E}} r^q V(z) u_m \partial_t u_m d\mu$$
$$+ \gamma \int_{\mathbb{E}} r^q \nabla_{\mathbb{E}}(\partial_t u_m)\nabla_{\mathbb{E}}(\partial_t u_m) d\mu$$
$$= \int_{\mathbb{E}} r^q g_t(z) u_m |u_m|^{p-1} \partial_t u_m d\mu. \tag{16}$$

Using The Leibniz rule one can get

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbb{E}} r^q |\partial_t^2 u_m|^2 d\mu + \frac{1}{2}\frac{d}{dt}\int_{\mathbb{E}} r^q |\nabla_{\mathbb{E}} u_m|^2 d\mu + \frac{1}{2}\frac{d}{dt}\int_{\mathbb{E}} r^q V(z)|u_m|^2 d\mu$$
$$+ \gamma \int_{\mathbb{E}} r^q |\nabla_{\mathbb{E}}(\partial_t u_m)|^2 d\mu = \frac{1}{p+1}\frac{d}{dt}\int_{\mathbb{E}} r^q g_t(z)|u_m|^{p+1} d\mu$$
$$- \frac{1}{p+1}\int_{\mathbb{E}} r^q \left(\frac{d}{dt} g_t(z)\right)|u_m|^{p+1} d\mu. \tag{17}$$

By integration of the relation 17 with respect to t

$$\frac{1}{2}\|\partial_t u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\|\nabla_{\mathbb{E}} u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\int r^q V(z)|u_m|^2 d\mu + \gamma \int_0^t \|\nabla_{\mathbb{E}}(\partial_\tau u_m)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} d\tau$$
$$- \frac{1}{p+1}\int_{\mathbb{E}} r^q g_t(z)|u_m|^{p+1} d\mu + \frac{1}{p+1}\int_0^t \|\left(\frac{d}{d\tau} g_\tau(z)\right)^{p+1} u_m\|^{\frac{1}{p+1}} d\tau$$
$$= I(t) + \gamma \int_0^t \|\nabla_{\mathbb{E}}(\partial_\tau u_m)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} d\tau \le I(0) < d, \tag{18}$$

where the last equal is upon Definition 5. Hence, for sufficiently large m and $0 \le t < \infty$ we obtain that $u_m \in W$ by Proposition 1. Using 18 and definition of functional K,

$$
\begin{aligned}
J(u_m) &= \frac{1}{2}\|\nabla_{\mathbb{E}} u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\int_{\mathbb{E}} r^q V(z)|u_m|^2 d\mu - \frac{1}{p+1}\int_{\mathbb{E}} r^q g_t(z)|u_m|^{p+1} d\mu \\
&= \frac{1}{2}\|\nabla_{\mathbb{E}} u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{2}\int_{\mathbb{E}} r^q V(z)|u_m|^2 d\mu \\
&- \frac{1}{p+1}\left(\|\nabla_{\mathbb{E}} u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int_{\mathbb{E}} r^q V(z)|u_m|^2 d\mu - K(u_m)\right) \\
&\geq (\frac{p-1}{2(p+1)}\left[\|\nabla_{\mathbb{E}} u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int_{\mathbb{E}} r^q V(z)|u_m|^2 d\mu\right] \\
&\geq \frac{p-1}{2(p+1)}(1+C_*^2)\|\nabla_{\mathbb{E}} u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}.
\end{aligned}
$$

Then

$$
\begin{aligned}
\int_0^t \frac{1}{2}\|\partial_\tau u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} d\tau + \frac{1}{p+1}\int_0^t \|(\frac{d}{d\tau} g_\tau)^{\frac{1}{p+1}} u_m\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})} d\tau \\
+ \frac{p-1}{2(p+1)}(1+C_*^2)\|\nabla_{\mathbb{E}} u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
\leq I(t) + \gamma\int_0^t \|\nabla_{\mathbb{E}}(\partial_\tau u_m)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} d\tau \leq I(0) < d. \quad (19)
\end{aligned}
$$

for $t \in [0, \infty)$ and sufficiently large m. Now, by relation 19 we can get that

$$
\|\nabla_{\mathbb{E}} u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} < \frac{2(p+1)}{p-1}(1+C_*^2)^{-1} d, \quad (20)
$$

for $t \in [0, \infty)$ and

$$
\frac{1}{2}\int_0^t \|\partial_\tau u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} d\tau + \frac{1}{p+1}\int_0^t \|(\frac{d}{d\tau} g_\tau)^{\frac{1}{p+1}} u_m\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})} d\tau < d, \quad (21)
$$

for $t \in [0, \infty)$. Also we obtain that

$$
\begin{aligned}
\int_{\mathbb{E}} r^q |g_t(z)^{\frac{p}{p+1}} u_m |u_m|^{p-1}|^{\frac{p+1}{p}} d\mu = \int_{\mathbb{E}} r^q g_t(z)|u_m|^{p+1} d\mu \leq C_{**}^{p+1}\|\nabla_{\mathbb{E}} u_m\|^{p+1}_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
< C_{**}^{p+1}(\frac{2(p+1)}{p-1}(1+C_*^2)^{-1} d)^{\frac{p+1}{2}} \quad (22)
\end{aligned}
$$

and

$$\int_{\mathbb{E}} r^q |V(z)^{\frac{1}{2}} u_m|^2 d\mu = \int_{\mathbb{E}} r^q V(z)|u_m|^2 d\mu \le C_*^2 \|\nabla_{\mathbb{E}} u_m\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}$$

$$< C_*^2 \left(\frac{2(p+1)}{p-1}\,(1+C_*^2)^{-1}\, d\,\right)^2. \tag{23}$$

From 20, 21, 22 and 23, it follows that there exists u and a subsequence still denotes $\{u_m\}$ for which as $m \to \infty$, $u_m \to u$ in $L^\infty\Big(0,\infty\,;\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\Big)$ weakly star and a.e. in $int\mathbb{E}\times[0,\infty)$, $\partial_t u_m \to \partial_t u$ in $L^2\Big(0,\infty\,;L_2^{\frac{n+1}{2}}(\mathbb{E})\Big)$, weakly star. Also we have $V(z)|u_m|^2 \to V(z)|u|^2$ in $L^\infty\Big(0,\infty\,;\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\Big)$ weakly star and a.e. in $int\mathbb{E}\times[0,\infty)$, and $g_t(z)u_m|u_m|^{p-1} \to g_t(z)u|u|^{p-1}$ in $L^\infty\Big(0,\infty\,;\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\Big)$ weakly star and a.e. in $int\mathbb{E}\times[0,\infty)$. Therefore, in 13 for k fixed and $m\to\infty$ we get that

$$(\gamma u, w_k)_2 + (u_t, w_k)_2 + \int_0^t (\nabla_{\mathbb{E}} u, \nabla_{\mathbb{E}} w_k)_2 d\tau + \int_0^t (V(z)u, w_k)_2 d\tau$$

$$= \int_0^t (g_t(z)u|u|^{p-1}, w_k)_2 d\tau$$

$$+ (\gamma u_0, w_k)_2 + (u_1, w_k)_2.$$

On the other hand, from the relation 14, $u(z,0) = u_0(z)$ in $\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})$ and from 15 $\partial_t u(z,0) = u_1$ in $L_2^{\frac{n+1}{2}}(\mathbb{E})$. By density we obtain $u \in L^\infty\Big(0,\infty\,;\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\Big)$ with $\partial_t u \in L^2\Big(0,\infty\,;L_2^{\frac{n+1}{2}}(\mathbb{E})\Big)$ is global weak solution of problem 1. Since u satisfies problem 1, so by definition of K we have $K(u) = 0$. Hence, $u(t) \in W$ for $0 \le t < \infty$.

Corollary 2 *If we replace the assumption $I(0) < d$, $K(u_0) > 0$ by $0 < I(0) < d$, $K_{\delta_2}(u_0) > 0$ where (δ_1, δ_2) is the maximal interval including $\delta = 1$, (see Remark 1) such that $I(0) < d(\delta)$ for $\delta \in (\delta_1, \delta_2)$. Then problem 1 admits a global weak solution $u(t) \in L^\infty\Big(0,\infty;\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\Big)$ with $\partial_t u \in L^\infty\Big(0,\infty,L_2^{\frac{n+1}{2}}(\mathbb{E})\Big)$ and $u(t) \in W_\delta$ for $\delta \in (\delta_1,\delta_2)$, $t \in [0,\infty)$.*

Proof It is immediately implied form Theorems 1 and 3.

Corollary 3 *If we replace the assumption $K_{\delta_2}(u_0) > 0$ or $\|\nabla_{\mathbb{E}} u_0\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} = 0$, by $\|\nabla_{\mathbb{E}} u_0\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} < r(\delta_2)$, then problem 1 admits a global weak solution $u(t) \in L^\infty\Big(0,\infty;\mathscr{H}_{2,0}^{1,\frac{n+1}{2}}(\mathbb{E})\Big)$ with $\partial_t u(t) \in L^\infty\Big(0,\infty;L_2^{\frac{n+1}{2}}(\mathbb{E})\Big)$ satisfying*

$$\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \leq \frac{I(0)}{a(\delta_1)}, \quad \|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \leq 2I(0), \quad 0 \leq t \leq \infty \tag{24}$$

Proof From assumption $\|\nabla_{\mathbb{E}} u_0\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} < r(\delta_2)$, we can get that $K_{\delta_2}(u_0) > 0$ or $\|\nabla_{\mathbb{E}} u_0\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} = 0$. Then it follows from Corollary 2 that problem 1 admits a global weak solution such that for any $\delta_1 < \delta < \delta_2$, $0 \leq t < \infty$, $u(t) \in L^\infty\left(0, \infty; \mathscr{H}^{1,\frac{n+1}{2}}_{2,0}(\mathbb{E})\right)$ with $\partial_t u \in L^\infty\left(0, \infty; L_2^{\frac{n+1}{2}}(\mathbb{E})\right)$ and $u(t) \in W_\delta$. Moreover, similar of the proof Theorem 2 for every $\delta_1 < \delta < \delta_2$, $0 \leq t < \infty$,

$$\frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + a(\delta)\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{p+1}K_\delta(u) \leq I(0).$$

If we tend δ to δ_1 then we achieve 24.

Now we discuss the global non-existence of solutions of the problem 1.

Theorem 4 *Let* $0 \leq \gamma \leq (p-1)\sqrt{1 + C_*^2}\lambda_1^{\frac{1}{2}}$, $u_0 \in \mathscr{H}^{1,\frac{n+1}{2}}_{2,0}(\mathbb{E})$, $u_1 \in L_2^{\frac{n+1}{2}}(\mathbb{E})$. *Suppose that* $I(0) < d$ *and* $K(u_0) < 0$. *Then the existence time of solution for problem 1 is finite, where* λ_1 *is the first eigenvalue in Proposition 3 i.e.*

$$\lambda_1^{\frac{1}{2}} = \inf_{u \in \mathscr{H}^{1,\frac{n+1}{2}}_{2,0}(\mathbb{E}), \|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \neq 0} \frac{\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})}}{\|u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})}}.$$

Proof Let $u(t)$ be any weak solution of problem 1 with $I(0) < d$ and $K(u_0) < 0$, T be the maximal existence time of $u(t)$. We will prove $T < \infty$ by contradiction. Let $M(t) := \|u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}$, then

$$\dot{M}(t) = \frac{d}{dt}\int_E r^q |u(z,t)|^2 d\mu = 2(\partial_t u, u)_2,$$

from definition of functional K,

$$\ddot{M}(t) = 2\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + 2(\partial_t^2 u, u)_2 = 2\|\partial_t u\|^2_{L_2^{\frac{+1}{2}}(\mathbb{E})} - 2\gamma(\nabla_{\mathbb{E}}(\partial_t u), \nabla_{\mathbb{E}} u)_2 - 2K(u). \tag{25}$$

Using proof of Theorem 2 we can get,

$$\frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + a(1)\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{1}{p+1}K(u)$$

$$
\begin{aligned}
&= \frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \left((\frac{1}{2} - \frac{1}{p+1})(1+C_*^2)\right)\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
&+ \frac{1}{p+1}\left(\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \int_{\mathbb{E}} r^q V(z)|u|^2 d\mu - \int_{\mathbb{E}} r^q g_t(z)|u|^{p+1} d\mu\right) \\
&\le \frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \frac{1}{2}\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
&\quad + + \left[(\frac{1}{2} - \frac{1}{p+1}) + \frac{1}{p+1}\right]\int_{\mathbb{E}} r^q V(x)|u|^2 d\mu \\
&- \frac{1}{p+1}\int_{\mathbb{E}} r^q g_t(z)|u|^{p+1} d\mu \ \le \frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
&\quad + \frac{1}{p+1}\int_0^t \|(\frac{d}{d\tau} g_\tau(z))^{\frac{1}{p+1}} u\|^{p+1}_{L_{p+1}^{\frac{n+1}{p+1}}(\mathbb{E})} \\
&+ J(u) \ + \ \gamma \int_0^t \|\nabla_{\mathbb{E}}(\partial_\tau u)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
&\quad = I(t) + \gamma \int_0^t \|\nabla_{\mathbb{E}}(\partial_t u)\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \le I(0). \qquad (26)
\end{aligned}
$$

Thus inequality 26 implies that

$$
\begin{aligned}
\ddot{M}(t) &\ge 2\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
&- 2\gamma(\partial_t u, u)_2 - 2(p+1)\left[I(0) - \frac{1}{2}\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}\right. \\
&\quad \left. - \frac{p-1}{2(p+1)}(1+C_*^2)\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}\right] \\
&= (p+3)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + (p-1)(1+C_*^2)\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\
&- 2\gamma(\nabla_{\mathbb{E}}(\partial_t u), \nabla_{\mathbb{E}} u)_2 - 2(p+1)I(0). \qquad (27)
\end{aligned}
$$

In first, let us consider $I(0) \le 0$. Then,

$$
\ddot{M}(t) \ge (p+3)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + (p-1)(1+C_*^2)\lambda_1\|u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - 2\gamma(\nabla_{\mathbb{E}}(\partial_t u), \nabla_{\mathbb{E}} u)_2.
$$

condition $\gamma < (p-1)(1+C_*^2)\lambda_1$ implies that, there exists a constant $\epsilon \in \left(0 \quad , \quad (p-1)(1+C_*^2)\right)$ such that

$$\gamma^2 < (p-1-\epsilon)(1+C_*^2)\lambda_1.$$

Therefore,

$$\ddot{M}(t) \geq (4+\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + (p-1-\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - 2\gamma(\nabla_{\mathbb{E}}(\partial_t u), \nabla_{\mathbb{E}} u)_2 + (p-1)(1+C_*^2)\lambda_1^2\|u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}. \tag{28}$$

On the other hand,

$$\begin{aligned} 2\gamma(\nabla_{\mathbb{E}}(\partial_t u), \nabla_{\mathbb{E}} u)_2 &\leq (p-1-\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{\gamma^2}{p-1-\epsilon}\|u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} \\ &\leq (p-1-\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + (p-1)(1+C_*^2)\lambda_1^2\|u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}. \end{aligned} \tag{29}$$

From 28 and 29, we can get that

$$\ddot{M}(t) \geq (4+\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}. \tag{30}$$

By Edge Hölder inequality we get

$$M(t)\ddot{M}(t) - \frac{4+\epsilon}{4}\dot{M}(t) \geq (4+\epsilon)\left(\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}\|u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - (\nabla_{\mathbb{E}}(\partial_t u), \nabla_{\mathbb{E}} u)_2\right) \geq 0,$$

$$(M^{-\alpha})'' = \frac{-\alpha}{M^{\alpha+2}(t)}\left(M(t)\ddot{M}(t) - (\alpha+1)\dot{M}(t)^2\right) \leq 0,$$

for $\alpha = \frac{\epsilon}{4}$ and $0 \leq t < \infty$. Hence, there exists a $T_1 > 0$ such that

$$\lim_{t \to T_1} M^{-\alpha}(t) = 0$$

and $\lim_{t \to T_1} M(t) = +\infty$, which is contradicts $T_{\max} = +\infty$.

In second case, we consider $0 < I(0) < d$. In this case from Theorem 1 we have $u \in E_\delta$ for $0 \leq t < \infty$ and $\delta \in (1 \quad , \quad \delta_2)$ (see Remark 1) where $(\delta_1 \quad , \quad \delta_2)$ is the maximal interval including $\delta = 1$ such that $d(\delta) > I(0)$ for $\delta \in (\delta_1 \quad , \quad \delta_2)$.

Therefore, $K_\delta(u) < 0$ and $\|\nabla_{\mathbb{E}} u\|_{L_2^{\frac{n+1}{2}}(\mathbb{E})} > r(\delta)$ for $1 < \delta < \delta_2$, $0 \le t < \infty$. Consequent, $K_\delta(u) \le 0$ and $\|\nabla_{\mathbb{E}} u\| \ge r(\delta)$ for $0 \le t < \infty$. From 25,

$$\begin{aligned}\frac{d}{dt}(e^{\gamma t}\dot{M}(t)) &= e^{\gamma t}\Big(\gamma \dot{M}(t) + \ddot{M}(t)\Big) = 2e^{\gamma t}\Big(\|\partial_t u\|^2_{L_2^{\frac{n}{2}}(\mathbb{E})}\| - K(u)\Big)\\ &= 2e^{\gamma t}\Big(\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}\| + (\delta_2 - 1)\|\nabla_{\mathbb{E}} u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - K_{\delta_2}(u)\Big)\\ &\ge 2e^{\gamma t}(\delta_2 - 1)r^2\delta_2 = C\,\delta_2 e^{\gamma t}.\end{aligned}$$

Hence,

$$e^{\gamma t}\dot{M}(t) \ge C\,\delta_2\int_0^t e^{\gamma\tau}d\tau + \dot{M}(0) = \frac{C\,\delta_2}{\gamma}(e^{\gamma t} - 1) + \dot{M}(0),$$

$$\dot{M}(t) \ge \frac{C\,\delta_2}{\gamma}(1 - e^{-\gamma t}) + e^{-\gamma t}\dot{M}(0).$$

Hence there exists $t_0 > 0$ for which

$$\dot{M}(t) \ge \frac{C\,\delta_2}{2\gamma} \quad \forall t \ge t_0$$

and

$$M(t) \ge \frac{C\,\delta_2}{2\gamma}(t - t_0) + M(t_0) \ge \frac{C\,\delta_2}{2\gamma}(t - t_0), \qquad t \ge t_0. \tag{31}$$

From assumption $\gamma < (p-1)(1 + C_*^2)\lambda_1$, it follows there exists a constant

$$\epsilon \in \Big(0 \quad , \quad (p-1)(1 + C_*^2)\Big)$$

such that

$$\gamma^2 < (p - 1 - \epsilon)\Big[(p-1)(1 + C_*^2)\lambda_1 - \epsilon\Big].$$

From 27,

$$\begin{aligned}\ddot{M}(t) \ge\ & (p+3)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - 2\gamma(\nabla_{\mathbb{E}}(\partial_t u), \nabla_{\mathbb{E}} u)_2 + (p-1)(1 + C_*^2)\lambda_1^2\|u\|^2_{L_2^{\frac{n}{2}}(\mathbb{E})}\\ & -2(p+1)I(0)\end{aligned}$$

$$= \|(4+\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n}{2}}(\mathbb{E})} + (p-1-\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} - 2\gamma(\nabla_{\mathbb{E}}(\partial_t u), \nabla_{\mathbb{E}} u)_2$$
$$+ [(p-1)(1+C_*^2)\lambda_1^2 - \epsilon]\|\partial_t u\|^2_{L_2^{\frac{n}{2}}(\mathbb{E})} + \epsilon M(t) - 2(p+1)I(0). \tag{32}$$

Also we can obtain

$$2\gamma(\nabla_{\mathbb{E}}(\partial_t u), \nabla_{\mathbb{E}} u)_2 \leq (p-1-\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \frac{\gamma^2}{p-1-\epsilon}\|u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}$$
$$\leq (p-1-\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}$$
$$+[(p-1)(1+C_*^2)\lambda_1^2 - \epsilon]\|u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}. \tag{33}$$

From 32 and 33 we get

$$\ddot{M}(t) \geq (4+\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})} + \epsilon M(t) - 2(p+1)I(0). \tag{34}$$

From 31, it follows that there exists a $t_1 > 0$ such that

$$\epsilon M(t) > 2(p+1)I(0) \qquad \forall t > t_1,$$

and then

$$\ddot{M}(t) > (4+\epsilon)\|\partial_t u\|^2_{L_2^{\frac{n+1}{2}}(\mathbb{E})}, \qquad \forall t > t_1.$$

Now, similar to first case we can obtain a contradiction. Hence we always have $T_{\max} < \infty$.

From Theorems 13 and 4 we can obtain the following theorem for global existence and non-existence of solutions for problem 1.

Theorem 5 *Let* $0 \leq \gamma \leq (p-1)\sqrt{1+C_*^2}\lambda_1^{\frac{1}{2}}$, $u_0 \in \mathscr{H}^{1,\frac{n+1}{2}}_{2,0}(\mathbb{E})$ *and* $u_1 \in L_2^{\frac{n+1}{2}}(\mathbb{E})$. *Suppose that* $I(0) < 0$. *Then, when* $K(u_0) > 0$, *problem 1 admits a global weak solution and when* $K(u_0) < 0$, *problem 1 does not admits any global weak solution.*

References

1. M. Alimohammady, M.K. Kalleji, Existence results for a class of semilinear totally characteristic hypoelliptic equations with conicl degeneration. J. Func. Anal. **265**, 2331–2356 (2013)
2. M. Alimohammady, C. Carlo, M.K. Kalleji, Invariance and existence analysis for semilinear hyperbolic equations with damping and conical singularity. J. Math. Anal. Appl. **455**, 569–591 (2017)

3. M. Alimohammady, M.K. Kalleji, Gh. Karamali, Golobal results for semilinear hyperbolic equations with damping term on manifolds with conical singularity. Math. Methods Appl. Sci. **40**(11), 4160–4178 (2017)
4. G. AustinFord, J. Wunsch, The diffractive wave trace on manifolds with conic singularities. Adv. Math. **304**, 1330–1385 (2017)
5. G. Chen, F. Da, Blow-up of solution of Cauchy problem for three-dimensional damped nonlinear hyperbolic equation. Nonlinear. Anal. **71**(1–2), 358–372 (2009)
6. H. Chen, G. Liu, Global existence and nonexistence for semilinear parabolic equations with conical degeneration. J. Pseudo-Differ. Oper. Appl. **3**, 329–349 (2012)
7. H. Chen, X. Liu, Asymptotic stability and blow-upof solutions for semi-linear degenerate parabolic equations with singular potential. Discrete Contin. Dyn. Syst. **36**(2), 661–682 (2016)
8. H. Chen, X. Liu, Y. Wei, Dirichlet problem for semilinear edge-degenerate elliptic equations with singular potential term. J. Differ. Equ. **252**, 4289–4314 (2012)
9. N.J. Daras, Th.M. Rassias, *Computational Mathematics and Variational Analysis* (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-44625-3
10. Y.V. Egorov, B.W. Schulze, *Pseudo-differential Operators, Singularities, Applications* (Springer, Basel, 1997)
11. S. Ervedoza, Control and stabilization properties for a singular heat equation with an inverse-square potential. Commun. Partial Differential Equations **33**(10–12), 1996–2019 (2008)
12. V. Felli, E.M. Marchini, S. Terracini, On Schrödinger operators with multipolar inverse-square potentials. J. Funct. Anal. **250**(2), 265–316 (2007)
13. X. Jiang, R. Xu, Global well-posedness for semilinear hyperbolic equationswith dissipative term. J. Appl. Math. Comput. **38**, 467–687 (2012)
14. M.K. Kalleji, M. Alimohammady, A.A. Jafari, Multiple solutions for class of nonhomogeneous semilinear equations with critical cone Sobolev exponent. Proc. Amer. Math. Soc. **147**, 597–608 (2019)
15. A. Kashuri, T.M. Rassias, R. Liko, Some new integral inequalities via general fractional operators, in *Computational Mathematics and Variational Analysis*, ed. by N. Daras, T. Rassias. Springer Optimization and Its Applications, vol. 159 (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-44625-3_9
16. D.H. Lehmer, Euler constants for arithmetical progressions. Acta Arith. **27**, 125–142 (1975)
17. H.A. Levine, Instability and non-existence of global solutions to nonlinear wave equations of the form $Pu_{tt} = -Au + F(u)$. Trans. Am. Math. Soc. **192**, 1–21 (1974)
18. G. Li, J. Yu, W. Liu, Global existence, exponential decay and finite time blow-up of solutions for a class of semilinear pseudo-parabolic equations with conical degeneration. J. Pseudo-Differ. Oper. Appl. **8**(2–17), 629–660 (2017)
19. Y. Liu, On potential wells and vacumm isolating of solutions for semilinear wave equations. J. Differential Equations **192**, 155–169 (2003)
20. Y. Liu, J. Zhao, On potential wells and applications to semilinear hyperbolic and parabolic equations. Nonlinear Anal. **64**, 2665–2687 (2006)
21. R. Melrose, A. Vasy, J. Wunsch, Propagation of singularities for the wave equation on edge manifolds. Duke Math. J. **144**(1), 109–193 (2008)
22. M. Mikolás, On certain sums generating the Dedekind sums and their reciprocity laws. Pacific J. Math. **7**, 1167–1178 (1957)
23. M. Ramazannejad, M. Alimohammady, C. Cattani, On algorithms for difference of monotone operators, in *Computational Mathematics and Variational Analysis*, ed. by N. Daras, T. Rassias. Springer Optimization and Its Applications, vol. 159 (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-44625-3_21
24. T.M. Rassias, M. Pardalos, *Mathematical Analysis and Applications* (Springer International Publishing, Cham, 2019)
25. M. Reed, B. Simon, *Methods of Modern Mathematical Physics* (Academic Press, New York, 1980)
26. H.D. Sattinger, On global solutions of nonlinear hyperbolic equations. Arch. Ration. Mech. Anal. **30**, 148–172 (1975)

27. B.W. Schulze, *Boundary Value Problems and Singular Pseudo-Differential Operators* (Wiley, Chichester 1998)
28. Z. Yang, Global existence, asymptotic behavior and blow-up solutions for a class of nonlinear wave equations with dissipative term. J. Differential Equations **187**(2), 520–540 (2003)

ϕ^4 Solitons in Kirchhoff Wave Equation

Y. Contoyiannis, P. Papadopoulos, M. Kampitakis, S. M. Potirakis, and N. L. Matiadou

Abstract We express the Kirchhoff wave equation in terms of classic field theory. This permits us to introduce the spontaneous symmetry breaking phenomenon in the study of linear structures, such as strings in order to investigate the existence of solitons solutions. We find ϕ^4 solitons in the space of spatial gradient of lateral displacement of a string. This helps us detect stable states in deformations of strings.

1 Introduction

In the last 30 years important progress has been made in understanding of properties of certain non-linear differential equations which arise in many different areas of Physics, e.g., physics of plasma, solid state physics, biophysics, field theory etc. [1–5]. A common interesting feature is the occurrence of solitons, i.e., stable, non-dissipative and localized configurations behaving in many ways like particles. In the analysis of these equations many interesting mathematical structures have been discovered which surprisingly also appear in quantum mechanics and quantum field theory [6]. From a pragmatic point of view these completely soluble non-linear equations are a substantial extension of the 'tool kit' of a physicist which otherwise is mainly restricted to solving linear systems. They also serve as valuable source for intuition about the behavior of non-linear systems. In Mathematics and Physics, a soliton, or solitary wave, is a self-reinforcing wave-packet that maintains its shape while it propagates at a constant velocity. Solitons are caused by a cancellation of nonlinear and dispersive effects in the medium. Solitons are the solutions of

Y. Contoyiannis · P. Papadopoulos (✉) · S. M. Potirakis · N. L. Matiadou
Department of Electrical and Electronics Engineering, University of West Attica, Athens, Greece
e-mail: yiaconto@uniwa.gr; ppapadop@uniwa.gr; spoti@uniwa.gr; lmatiadou@uniwa.gr

M. Kampitakis
Hellenic Electricity Distribution Network Operator SA, Network Major Installations Department, N.Faliro, Greece
e-mail: m.kampitakis@deddie.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_4

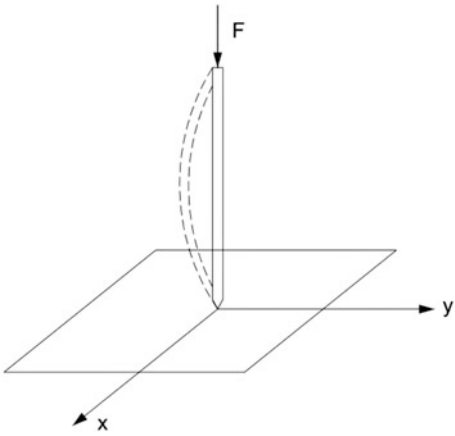


Fig. 1 The cylindrical symmetry of the system around z axis is broken by the buckling of the string

a widespread class of weakly nonlinear dispersive partial differential equations describing physical systems.

ϕ^4 solitons are stable solutions which appear in spontaneous symmetry breaking (SSB) in scalar field theories [7]. A category of systems for which SSB might happen are linear structures such as rob, string, needle etc. Thus, if a string is compressed by the application of a force F along its axis the obvious solution is that it stays in the configuration $x = y = 0$ (see Figure 1). However, if the force gets too large ($F > F_{cr}$), the string will jump into a bent position. It does this because the energy in this state is lower than in meta-stable state, where it stays aligned along the z axis. The cylindrical symmetry of the system around z axis is broken by the buckling of the string [7].

In a ϕ^4—scalar field the SSB is sourced from a concrete type of potential density. This SSB produces solitary waves which ensure the stable behavior of ϕ scalar field. The wave function of the lateral displacement u for a string such as the one in Figure 1 is the solution of the Kirchhoff wave equation [8]. In this work we attempt to reveal similarities between the potential density produced from Kirchhoff wave equation and ϕ^4—scalar field potential density. This would permit us to reveal the existence of soliton in the Kirchhoff description. This would help us in order to find stable states when the string suffers lateral deformation under the action of axial tensions. This is the main motivation of the present work.

2 SSB in ϕ^4 Scalar Field Theory: The Kink Solitons

The Lagrangian density of scalar field $\phi(x)$ with a ϕ^4 interaction is given as [9]:

$$\mathscr{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \left\{\frac{1}{2}\alpha\phi^2 + \frac{1}{4}\lambda\phi^4\right\}, \tag{1}$$

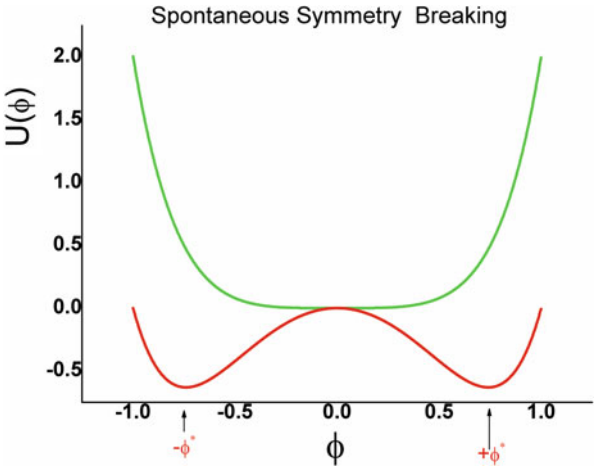


Fig. 2 The SSB in scalar field theory. The critical point (0.0) behaves as a saddle-point

where $\mu = 0, 1, 2, 3$ with $\partial_0\phi = \frac{\partial\phi}{\partial t}$, $\partial_1\phi = -\frac{\partial\phi}{\partial x}$, $\partial_2\phi = -\frac{\partial\phi}{\partial y}$, $\partial_3\phi = -\frac{\partial\phi}{\partial z}$

The term of kinetic density is $\frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi)$ and the potential density is:

$$U(\phi) = \frac{1}{2}\alpha\phi^2 + \frac{1}{4}\lambda\phi^4 \tag{2}$$

When α, λ have positive values we have the symmetric phase (green line in Figure 2). When $\alpha < 0$, $\lambda > 0$ we have the phase of symmetry breaking (SB) (red line in Figure 2). Thus, the ground state of energy has shifted from $\phi = 0$ to $\phi^* = \pm\sqrt{\frac{|\alpha|}{\lambda}}$. This is the SSB phenomenon where the system should select a new vacuum. In a thermal system the parameter α is a function of $\frac{T-T_c}{T_c}$ were T is the temperature and T_c is the critical temperature. In a similar way, in a string SSB the parameter α could be a function of $\frac{F-F_c}{F_c}$. A model which demonstrates SSB is the ϕ^4 theory [9] where the potential has the form:

$$U(\phi) = \frac{\lambda}{4}\left(\phi^2 - \frac{\alpha}{\lambda}\right)^2 = -\frac{1}{2}|\alpha|\,\phi^2 + \frac{1}{4}\lambda\phi^4 + \frac{\alpha^2}{4\lambda} \tag{3}$$

The above potential refers to the SB phase ($\alpha < 0, \lambda > 0$) and a constant term has been added. So, the critical state (0, 0) is excited to $(0, \frac{\alpha^2}{4\lambda})$. This state describes the meta-stable state of the string, before the cylindrical symmetry of the system around z axis is broken by the buckling of the string. The potential of Equation (3) has the same minima as the potential of Equation (2) namely $\phi^* = \pm\sqrt{\frac{|\alpha|}{\lambda}}$. This means that solitons solutions, if they exist, must asymptotically tend toward these values as $x \to \pm\infty$, that is:

$$\phi(|x| = \infty) = \pm\sqrt{\frac{|\alpha|}{\lambda}} \tag{4}$$

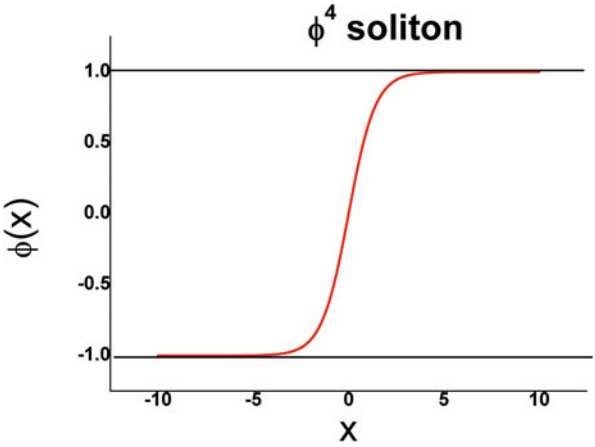


Fig. 3 The ϕ^4 soliton with $\alpha = \lambda = 1$ and $x_0 = 0$

We can integrate the ϕ^4 theory given by the Equation (3) to yield [9]: $x - x_0 = \pm \int_{\phi(x_0)}^{\phi(x)} \frac{d\bar{\phi}}{\sqrt{\frac{\lambda}{2}}} \frac{1}{(\bar{\phi}^2 - \frac{\alpha}{\lambda})}$ and inverting, we find that [9]:

$$\phi(x) = \pm\left(\sqrt{\frac{|\alpha|}{\lambda}}\right) \tanh\left(\frac{\sqrt{\alpha}}{\sqrt{2}}(x - x_0)\right) \tag{5}$$

This is the kink soliton of ϕ^4 scalar field (Figure 3).

The most important property of solitons, as it has already been mentioned, is that they are stable structures which behave as particles. The energy density of these solitons is [9]:

$$\varepsilon(x) = \left(\frac{\alpha^2}{2\lambda}\right) sech^4\left[\frac{m(x - x_0)}{\sqrt{2}}\right]. \tag{6}$$

The mass of particle-soliton is given by the integral over the energy density:

$$M = \int_{-\infty}^{\infty} \varepsilon(x)dx = \frac{2\sqrt{2}}{3} \frac{|\alpha|^{\frac{3}{2}}}{\lambda} \tag{7}$$

3 Kirchhoff Wave Function, Energy and Potential

The Kirchhoff wave equation without damping term is defined [8, 10, 11], as:

$$\frac{\partial^2 u}{\partial^2 t} - (1 + \int_{\Omega} |\nabla u|^2 \, dx)\nabla^2 u = 0 \tag{8}$$

Where $u(t, x)$ is the lateral displacement of a string at the space coordinate x and the time t, while Ω is a bounded domain in $\mathbb{R}^N$ with a smooth boundary $\partial\Omega$.

Here the energy is given as $E(t) = \int_\Omega \varepsilon dx$ where the energy density ε is [12–14]:

$$\varepsilon(t) = \frac{1}{2}\left|\frac{\partial u}{\partial t}\right|^2 + \left\{\frac{1}{2}\left|\frac{\partial u}{\partial x}\right|^2 + \frac{1}{4}\left|\frac{\partial u}{\partial x}\right|^2\left|\frac{\partial u}{\partial x}\right|^2\right\} \tag{9}$$

The first term in Equation (9) is the kinetic term and the term in the curly brackets is the Kirchhoff potential density (Kpd), $U_{Kirchhoff}$. So, we have that:

$$U_{Kirchhoff} = \frac{1}{2}\left|\frac{\partial u}{\partial x}\right|^2 + \frac{1}{4}\left|\frac{\partial u}{\partial x}\right|^2\left|\frac{\partial u}{\partial x}\right|^2 = \frac{1}{2}\left|\frac{\partial u}{\partial x}\right|^2 + \frac{1}{4}\left|\frac{\partial u}{\partial x}\right|^4 \tag{10}$$

4 The Kpd Produced from Classical Field Theory

In this section we will try to produce the Kpd through the classical field theory. Thus, the investigation of solitons in the wave equation is not just the result of comparing potentials but it has a fundamental origin. Let's start from the classical wave equation:

$$\partial_\mu^2\phi = \frac{\partial^2\phi}{\partial^2 t} - \nabla^2\phi = 0 \tag{11}$$

Note that in Equation (11) the wave speed constant factor c^2 has been omitted. This is done here since both Equation (8) and Equation (10) that represent Kirchhoff wave equation without damping term and Kpd, respectively, appear in the cited references without constant factors. However, we will restore the specific factor later, during the derivation of the complete form of Kpd (Equation (22)).

If we substitute the $\nabla^2\phi$ as $\nabla^2\phi(1 + \left(\frac{\partial\phi}{\partial x}\right)^2)$, then Equation (8) is written as:

$$\frac{\partial^2\phi}{\partial^2 t} - \nabla^2\phi\left(1 + \left(\frac{\partial\phi}{\partial x}\right)^2\right) = 0 \tag{12}$$

$$\partial_\mu^2\phi = \frac{\partial^2\phi}{\partial^2 t} - \nabla^2\phi = -\frac{\partial U}{\partial\phi} \tag{13}$$

For static solution $(\frac{\partial^2\phi}{\partial^2 t}) = 0$ the E-L equation of Equation (13) becomes:

$$-\nabla^2\phi = -\frac{\partial U}{\partial\phi} \tag{14}$$

where $U(\phi)$ the potential density of ϕ-field. From Equation (14) we can estimate the potential density U as follows:

We multiply Equation (14) by $\frac{\partial \phi}{\partial x}$ and we take the following:

$$-\nabla^2\phi \cdot \frac{\partial \phi}{\partial x} = -\frac{\partial U}{\partial \phi} \cdot \frac{\partial \phi}{\partial x}, \tag{15}$$

which can be integrated over x, yielding [9]:

$$U(\phi) = \frac{1}{2}\Big(\frac{\partial \phi}{\partial x}\Big)^2 \tag{16}$$

Following the above procedure, we estimate (see Appendix) the U' for the case of the wave described by Equation (12) as:

$$U'(\phi) = \frac{1}{2}\Big(\frac{\partial \phi}{\partial x}\Big)^2 + \frac{1}{4}\Big(\frac{\partial \phi}{\partial x}\Big)^4 \tag{17}$$

The potential density $U'(\phi)$ has the same form with Kpd, as expressed in Equation (10).

5 SSB in the Kpd

Now we will investigate the SSB in the Kpd. If we compare the Kpd from Equation (17) and the potential density of SSB in the ϕ^4 theory from Equation (3), we find out that Equation (3) refers to a scalar field ϕ while Equation (17) refers to gradient of ϕ, that is $\frac{\partial \phi}{\partial x}$ (or $\nabla\phi$). We face this by defining a new field ξ as $\xi \equiv \nabla\phi$. Thus, we can research solitons in ξ-field. This means for the string case, that the solitons solutions exist not at the lateral displacement space but in its spatial gradient space. The next thing we have to do, is to introduce coefficients in the terms of Equation (17).

The original equation of Kirchhoff without damping is written as [8]:

$$\frac{\partial^2 u}{\partial^2 t} = \nabla^2 u \cdot \Big(\frac{p_0}{p_h} + \frac{Y}{p2L}\int_0^L \Big(\frac{\partial u}{\partial x}\Big)^2 dx\Big), \tag{18}$$

where $0 < x < L$, with L the length of string, Y the Young modulus, p the mass density, h the cross-section area, p_0 the initial external force. In classical wave equation Equation (11), we normally have a coefficient $c^2 = \frac{p_0}{p_h}$ in front of the term $\nabla^2\phi$. This is in agreement with Equation (18). The substitution $\nabla^2\phi \rightarrow \frac{p_0}{p_h}\nabla^2\phi$ in Equation (11) transports the coefficient $\frac{p_0}{p_h}$ in Equation (17) in front of the first term. From Equation (18), we have that the second coefficient is the quantity $\frac{Y}{p2L} > 0$. Thus, the Kpd is written as:

$$U'(\phi) = \frac{1}{2}\alpha\left(\frac{\partial\phi}{\partial x}\right)^2 + \frac{1}{4}\lambda\left(\frac{\partial\phi}{\partial x}\right)^4, \tag{19}$$

where:

$$\alpha = \frac{p_0}{ph} \tag{20}$$

and

$$\lambda = \frac{Y}{p2L} \tag{21}$$

The next thing we have to do is to add the constant term in Equation (19). This quantity is $\frac{\alpha^2}{4\lambda} = \frac{p_0^2}{2ph^2Y}L$. We consider the p_0 as the resultant of axial forces. The quantity $\frac{p_0^2}{2ph^2Y}L > 0$ expresses the excited potential of string before the cylindrical symmetry of the system around z axis is broken by the buckling of the string. This meta-stable state due to the axial compression ΔL from the external force. We have to give an explanation for the negative sign of the coefficient α, which indicates the symmetry breaking whenever the string leaves the axis and goes to the lateral positions. The symmetry breaking is accomplished when the external axial force overcomes a critical value. Then the internal elasticity forces obtain measure greater than external forces and p_0 obtains negative sign. From Equation (20) the coefficient α becomes negative too. Therefore, the Kpd for the field ξ is written in the final form as:

$$U_{Kirchhoff}(\xi) = -\frac{1}{2}|\alpha|\,\xi^2 + \frac{1}{4}\lambda\xi^4 + \frac{\alpha^2}{4\lambda} \tag{22}$$

Now the Kpd has taken the form of the potential density of SSB (see Equation (3)), which provides the theoretical basis for the formation of kink solitons. The solitons solutions from Equation (5) and (Equations (20) and (21)) is written as:

$$\xi(x) = \pm\sqrt{\frac{p_0 2L}{Yh}}\tanh\left[\sqrt{\frac{p_0}{2ph}}(x - \Delta L_{cr})\right], \tag{23}$$

where ΔL_{cr} is the axial compression when the force overcomes its critical value. The existence of solitons dependents from the asymptotic behavior $\xi(\pm\infty) = \pm\sqrt{\frac{p_0 2L}{Yh}}$.

The mass of Kirchhoff soliton which is given in Equation (7) has the form:

$$M_{Kirchhoff} = \frac{4\sqrt{2}}{3}\frac{p_0^{3/2}}{Yh^{3/2}\sqrt{p}}L \tag{24}$$

The transmission length where the solitons survive is determined by their mass, according to the proportion which connects the mass of a particle and its transmission length in the field theory is:

$$R \sim \frac{1}{M} \tag{25}$$

Thus, for materials with high Young modulus Y, such as steel, from Equation (24) we obtain that the $M_{Kirchhoff}$ is small. This means from Equation (25) that transmission length R where the solitons survive is long and the range of stable state is long too. So, the steel string could be found in lateral positions (Figure 1) with greater stability and without breaking. Nevertheless, from Equation (24) one can determine values of parameters which are possible to give stable states.

6 Conclusions

In this work we have produced the potential density of Kirchhoff wave equation, for static solution without damping term, through the classical field theory. Thus, we attempt to study a linear structure such as a string, which suffers lateral deformation under the action of axial tensions through the spontaneous symmetry breaking phenomenon. The result is that ϕ^4 solitons in the space of gradient of lateral displacement of string, emerge. The existence of these stable solutions permits us to determine the stability of string deformation, through the extension of spatial range of solitons propagation. This approximation we applied on the issue of elasticity, is a new way to face the limits of elasticity for linear structures.

Appendix: Estimation of the Potential Density U′(ϕ)

For the Kirchhoff wave equation $\frac{\partial^2\phi}{\partial^2 t} - \nabla^2\phi\left(1 + \left(\frac{\partial\phi}{\partial x}\right)^2\right) = 0$ (Equation (12)) we initially consider that exists a potential density $U'(\phi)$ which satisfies a generalized "Euler-Lagrange" equation that could be written as:

$$-\nabla^2\phi\left(1 + \left(\frac{\partial\phi}{\partial x}\right)^2\right) = -\frac{\partial U'}{\partial\phi}, \tag{26}$$

by proceeding to the replacement $-\nabla^2\phi \rightarrow -\nabla^2\phi\left(1+\left(\frac{\partial\phi}{\partial x}\right)^2\right)$ in the E-L equation for static solution as presented in Equation (14).

We set as potential density $U'(\phi)$:

$$U'(\phi) = U(\phi) + U_1(\phi) \tag{27}$$

So, Equation (26) is written as:

$$-\nabla^2\phi - \nabla^2\phi \cdot \left(\frac{\partial\phi}{\partial x}\right)^2 = -\frac{\partial U}{\partial\phi} - \frac{\partial U_1}{\partial\phi}, \tag{28}$$

and by using the E-L of Equation (14) we obtain:

$$\nabla^2\phi \cdot \left(\frac{\partial\phi}{\partial x}\right)^2 = \frac{\partial U_1}{\partial\phi} \tag{29}$$

This equation tells us that the correction term in E-L equation corresponds to a potential density U_1 as the E-L equation corresponds to the potential density U.

We multiply Equation (29) by $\frac{\partial\phi}{\partial x}$ to take:

$$\nabla^2\phi \cdot \frac{\partial\phi}{\partial x}\left(\frac{\partial\phi}{\partial x}\right)^2 = \frac{\partial U_1}{\partial\phi}\frac{\partial\phi}{\partial x} \tag{30}$$

Moreover, Equation (30) can be integrated over x, yielding $\int \frac{\partial^2\phi}{\partial x^2} \cdot \frac{\partial\phi}{\partial x}\left(\frac{\partial\phi}{\partial x}\right)^2 dx = \int \frac{\partial U_1}{\partial x}\frac{\partial\phi}{\partial x}dx$.

Thus, we have that:

$$\int \frac{\partial}{\partial x}\left(\frac{\partial\phi}{\partial x}\right)\left(\frac{\partial\phi}{\partial x}\right)^3 dx = \int \frac{\partial U_1}{\partial x}dx = U_1(\phi) \tag{31}$$

The first part of Equation (31) is estimated as:

$$\begin{aligned}\int \frac{\partial}{\partial x}\left(\frac{\partial\phi}{\partial x}\right)\left(\frac{\partial\phi}{\partial x}\right)^3 dx &= \left(\frac{\partial\phi}{\partial x}\right)^4 - \int \left(\frac{\partial\phi}{\partial x}\right)\frac{\partial}{\partial x}\left(\frac{\partial\phi}{\partial x}\right)^3 dx = \left(\frac{\partial\phi}{\partial x}\right)^4 \\ &- \int \frac{\partial\phi}{\partial x} 3\left(\frac{\partial\phi}{\partial x}\right)^2 \frac{\partial}{\partial x}\left(\frac{\partial\phi}{\partial x}\right)dx \Rightarrow \\ 4\int \frac{\partial}{\partial x}\left(\frac{\partial\phi}{\partial x}\right)\left(\frac{\partial\phi}{\partial x}\right)^3 dx &= \left(\frac{\partial\phi}{\partial x}\right)^4 \Rightarrow \\ \int \frac{\partial}{\partial x}\left(\frac{\partial\phi}{\partial x}\right)\left(\frac{\partial\phi}{\partial x}\right)^3 &= \frac{1}{4}\left(\frac{\partial\phi}{\partial x}\right)^4\end{aligned} \tag{32}$$

Using Equations (16), (27), (31) and (32) we finally obtain:

$$U'(\phi) = \frac{1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2 + \frac{1}{4}\left(\frac{\partial\phi}{\partial x}\right)^4 \tag{33}$$

References

1. M. Remoissenet, *Waves Called Solitons: Concepts and Experiments* (Springer, Berlin-Heidelberg, 1999), p. 11
2. S.T. Cundiff, B.C. Collings, N.N. Akhmediev, J.M. Soto-Crespo, K. Bergman, W.H. Knox, Observation of polarization-locked vector solitons in an optical fiber. Phys. Rev. Lett. **82**, 3988 (1999)
3. A.S. Davydov, *Solitons in Molecular Systems. Mathematics and its Applications (Soviet Series)*, vol. 61, 2nd edn. (Kluwer Academic Publishers, Dordrecht, 1991)
4. T. Heimburg, A.D. Jackson, *On Soliton Propagation in Biomembranes and Nerves*. Proc. Natl. Acad. Sci. U. S. A. **102**(2), 9790–9795 (2005)
5. W. Craig, P. Guyenne, J. Hammack, D. Henderson, C. Sulem, Solitary water wave interactions. Phys Fluids **18**, 57106 (2006)
6. L.H. Ryder, *Quantum Field Theory* (Cambridge University Press, Cambridge, 1985)
7. F. Halzen, A.D. Martin *Quarks and Leptons: An Introductory Course in Modern Particle Physics* (Wiley, New York, 1983)
8. G. Kirchhoff, *Vorlesungen über mechanik* (B.G. Teubner, Leipzig, 1883)
9. M. Kaku, *Quantum Field Theory* (Oxford University Press, New York, 1993)
10. P. Papadopoulos, N. Stavrakakis, Central manifold theory for the generalized equation of Kirchhoff strings on $\mathbb{R}^N$. Nonlinear Anal. TMA **61**, 1343–1362 (2005)
11. P. Papadopoulos, N. Stavrakakis, Compact invariant sets for some quasilinear nonlocal Kirchhoff strings on $\mathbb{R}^N$. Appl. Anal. **87**, 133–148 (2008)
12. K. Ono, On global existence, asymptotic stability and blowing-up of solutions for some degenerate non-linear wave equations of Kirchhoff type with a strong dissipation. Math Methods Appl. Sci. **20**, 151–177 (1997)
13. P. Papadopoulos, N. Stavrakakis, Global existence and blow-up results for an equation of Kirchhoff type on $\mathbb{R}^N$. Topol. Methods Nonlinear Anal. **17**, 91–109 (2001)
14. P. Papadopoulos, M. Karamolengos, A. Pappas, Global existence and energy decay for mildly degenerate Kirchhoff's equations on $\mathbb{R}^N$. J. Interdisciplinary Math. **12**(6), 767–783 (2009)

Estimates for Lipschitz and BMO Norms of Operators on Differential Forms

Shusen Ding, Guannan Shi, and Yuming Xing

Abstract In this paper, we introduce the generalized Lipschitz and BMO norms of differential forms and establish the upper bound estimates for the generalized Lipschitz and BMO norms of operators applied to differential forms. We also demonstrate applications of our main results using examples.

1 Introduction

The main purpose of this paper is to establish the upper bound estimates for the generalized Lipschitz and BMO norms of the iterated operators $D^k G^k$ and $D^{k+1} G^k$ applied to differential forms u defined in $\mathbb{R}^n$ in terms of the L^p norms of u, where k is a positive integer; G is Green's operator and $D = d + d^\star$ is the Hodge-Dirac operator on differential forms. The Dirac operator D and Green's operator G are very well studied and widely used in many fields of mathematics and physics. They play a critical role in the study of the nonlinear problems in PDEs and nonlinear potential theory. For example, in the case $k = 1$, the composition $D^2 G$ is used to define the well-known Poisson's equation $D^2 G(u) = u - H(u)$ (or $\Delta G(u) = u - H(u)$), where H is the harmonic projection operator. In the same sense as the L^p theory, the estimates for the BMO norms of differential forms and the related operators are also decisive on the investigation of the solution properties of PDEs, especially on the study of Harnack's inequality for solutions

S. Ding
Department of Mathematics, Seattle University, Seattle, WA, USA
e-mail: sding@seattleu.edu

G. Shi (✉)
School of Mathematics and Statistics, Northeast Petroleum University, Daqing, China
e-mail: sgncx@163.com

Y. Xing
Department of Mathematics, Harbin Institute of Technology, Harbin, China
e-mail: xyuming@hit.edu.cn

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_5

to certain partial differential equations, see [1] for example. Some estimates for the BMO norm and local Lipschitz norm of differential forms or related operators can be found in [2–5]. We should notice that D^kG^k and $D^{k+1}G^k$ are more general operators which include the composite operator DG as a special case, see [5] where DG has been investigated. However, there is no systematic study on the BMO norm and local Lipschitz norm of the iterated operators D^kG^k and $D^{k+1}G^k$ for the case $k > 1$ in the literature. Hence, we are motivated to establish the upper bound estimates for the generalized Lipschitz and BMO norms of the composite operator in this paper. We first extend the definitions of the classical locLip_α and BMO norms into the generalized locLip_α^s and BMO^s norms, respectively. Then, we study the relationship between these two norms and L^p norms. The estimates for norms and comparisons of norms are very important in the investigation of the corresponding spaces in analysis. For example, it is well known that the BMO space, the dual of Hardy space, is a substitute of L^∞ space and has been playing a very indispensable role in harmonic analysis and exterior differential analysis, as well as in the study of the characterization of singular integral operators since it was set forth by John and Nirenberg in 1961. We refer the readers to Chapter IV in [6] and [1, 7] for the function case of the BMO space, and Chapter 9 in [2] and [8–11] for the case to differential forms. Our main results are presented and proved in Section 3. These results will enrich the theory of operators on differential forms.

Unless stated otherwise, we keep using the traditional notation and symbols throughout this paper. Let Ω be a smoothly bounded domain without the boundary in $\mathbb{R}^n$, $n \geq 2$, and $B = B(x, \rho)$ be the ball in $\mathbb{R}^n$ with radius ρ centered at x, which satisfies $\mathrm{diam}(\sigma B) = \sigma\,\mathrm{diam}(B)$. Let the direct sum $\Lambda = \Lambda(\mathbb{R}^n) = \oplus_{l=0}^n \Lambda^l(\mathbb{R}^n)$ be a graded algebra with respect to the exterior product, and $\Lambda^l = \Lambda^l(\mathbb{R}^n)$ be the space of l-covectors in $\mathbb{R}^n$, which is spanned by the dual orthogonal basis $dx_{i_1}, \cdots, dx_{i_l}$, where $x_{i_1}, \cdots, x_{i_l}$ are the coordinate functions on $\mathbb{R}^n$. For the set Λ, we denote the pointwise inner product by $< \cdot, \cdot >$ and the module by $|\cdot|$. Then, every differential form $u(x) \in \Lambda^l(\mathbb{R}^n)$ can be uniquely written as

$$u(x) = \sum_I u_I(x)dx_I = \sum u_{i_1 i_2 \cdots i_l}(x)dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_l},$$

where the coefficients $u_{i_1 i_2 \cdots i_l}(x)$ are differentiable functions and $I = (i_1, i_2, \cdots, i_l)$, $1 \leq i_1 < i_2 < \cdots < i_l \leq n$. Actually, differential forms are the generalizations of the functions, which include functions as their special cases (functions are called 0-forms). The Hodge-star operator $\star : \Lambda^l(\mathbb{R}^n) \to \Lambda^{n-l}(\mathbb{R}^n)$ is defined by the rule that $\star 1 = dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_l}$ and $\alpha \wedge \star\beta =< \alpha, \beta > (\star 1)$ for every $\alpha, \beta \in \Lambda^l$, $l = 0, 1, \cdots, n$. By this definition, it induces that $\star$ is an isometric isomorphism on Λ^l. The linear operator $d : \mathbb{D}'(\Omega, \Lambda^l) \to \mathbb{D}'(\Omega, \Lambda^{l+1})$, $0 \leq l \leq n-1$, is called the exterior differential and $d^\star = (-1)^{nl+1} \star d \star : \mathbb{D}'(\Omega, \Lambda^{l+1}) \to \mathbb{D}'(\Omega, \Lambda^l)$, the formal adjoint of d, is known as Hodge codifferential. The interested readers could see [10–13] for further introduction and appropriate properties. Also, we use $L^p(\Omega, \Lambda)$ to denote the classical L^p space for

differential forms, $1 < p < \infty$, equipped with the norm $\|u\|_{p,\Omega} = \left(\int_\Omega |u|^p dx\right)^{\frac{1}{p}} = \left(\int_\Omega (\sum_I |u_I|^2)^{\frac{p}{2}} dx\right)^{\frac{1}{p}}$. $W^{1,p}(\Omega, \Lambda)$ is the classical Sobolev space for differential forms with the norm $\|u\|_{W^{1,p}(\Omega)} = (diam(\Omega))^{-1}\|u\|_{p,\Omega} + \|\nabla u\|_{p,\Omega}$. $W_d^p(\Omega, \Lambda^l)$ is the space of differential l-forms such that $du \in L^p(\Omega, \Lambda^l)$. Analogously, $W_{d^\star}^p(\Omega, \Lambda^l)$ is the space of differential l-forms such that $d^\star u \in L^p(\Omega, \Lambda^l)$. Inspired by these classical spaces for differential forms, we generalize the BMO space and local Lipschitz space as follows.

Definition 1.1 For every $\omega \in L^s_{loc}(\Omega, \Lambda^l)$, $s \geq 1$, we say $\omega \in \text{BMO}^s(\Omega, \Lambda^l)$ with the norm defined by

$$\|\omega\|_{*,s,\Omega} = \sup_{\sigma Q \subset \Omega} |Q|^{-1/s} \|\omega - \omega_Q\|_{s,Q}, \tag{1.1}$$

if ω satisfies $\sup_{\sigma Q \subset \Omega} |Q|^{-1/s} \|\omega - \omega_Q\|_{s,Q} < \infty$, where $l = 0, 1, \cdots, n$ and $\sigma > 1$ is some expansion factor.

Definition 1.2 For every $\omega \in L^s_{loc}(\Omega, \Lambda^l)$, $s \geq 1$, $l = 0, 1, \cdots, n$ and $0 < \alpha \leq 1$, we call $\omega \in \text{loc}Lip^s_\alpha(\Omega, \Lambda^l)$ with the norm denoted by

$$\|\omega\|_{locLip^s_\alpha(\Omega)} = \sup_{\sigma Q \subset \Omega} |Q|^{-(n+\alpha s)/sn} \|\omega - \omega_Q\|_{s,Q}, \tag{1.2}$$

if ω satisfies $\sup_{\sigma Q \subset \Omega} |Q|^{-(n+\alpha s)/sn} \|\omega - \omega_Q\|_{s,Q} < \infty$, where $\sigma > 1$ is some expansion factor.

Especially, for the case $s = 1$, the BMO^s norm and locLip^s_α norm just reduce to the following classical BMO norm and locLip_α norm given in [10] by C. Nolder, respectively.

$$\|\omega\|_{*,1,\Omega} = \|\omega\|_{*,\Omega} = \sup_{\sigma Q \subset \Omega} |Q|^{-1} \|\omega - \omega_Q\|_{1,Q} \tag{1.3}$$

and

$$\|\omega\|_{locLip^1_\alpha(\Omega)} = \|\omega\|_{locLip_\alpha(\Omega)} = \sup_{\sigma Q \subset \Omega} |Q|^{-(n+\alpha)/n} \|\omega - \omega_Q\|_{1,Q} \tag{1.4}$$

Furthermore, notice that $|Q|^{\alpha/n} \leq |\Omega|^{1/n}$ since $0 < \alpha \leq 1$ and $n \geq 1$, which results in that

$$|Q|^{-1/s} = |Q|^{\alpha/n} |Q|^{-1/s-\alpha/n} \leq |\Omega|^{1/n} |Q|^{-1/s-\alpha/n}$$

So, similarly as the result in [14], we have that there is a constant $C > 0$, independent of ω, such that

$$\|\omega\|_{*,s,\Omega} \leq C\|\omega\|_{locLip^s_\alpha(\Omega)} \tag{1.5}$$

for every $\omega \in W^{1,s}(\Omega, \Lambda^l)$, which enables us to compare the locLip$^s_\alpha$ norm and the BMO^s norm for D^kG^k and $D^{k+1}G^k$ simply. In addition, from now on, we point out that the constants C and C_i employed in this paper, $i = 1, 2, \cdots$, may differ from one line to the next.

2 Local Poincaré-Type Inequalities

In this section, as preparation for the principle assertion, we show the explicit formulas of D^kG^k and $D^{k+1}G^k$ and the Poincaré-type inequalities of D^kG^k and $D^{k+1}G^k$ by applying the explicit representation in Lemma 2.4 and Lemma 2.5. First, let us start with the brief review of Green's operator G. For any fixed integer $l = 0, 1, \cdots, n$, let $\mathbb{H}$ be the harmonic l-field denoted by

$$\mathbb{H} = \{u \in W(\Omega, \Lambda) : du = d^\star u = 0,\ u \in L^p,\ \text{for some } 1 < p < \infty\}.$$

In the meantime, we take the operator $\delta : L^p(\Omega, \Lambda) \bigcap \mathbb{H}^\perp \to W^{1,p}(\Omega, \Lambda) \bigcap \mathbb{H}^\perp$ defined by Morrey in [15], which satisfies that for every $u \in L^p(\Omega, \Lambda) \bigcap \mathbb{H}^\perp$, $\delta(u)$ is the unique form in $W^{1,p}(\Omega, \Lambda) \bigcap \mathbb{H}^\perp$ such that $\Delta\delta(u) = u$, where $\Delta = D^2 = dd^\star + d^\star d$ is the Laplace operator, and $\mathbb{H}^\perp$ is the complement space of harmonic field $\mathbb{H}$. Therefore, we are given the definition as follows.

Definition 2.1 ([16]) Green's operator $G : L^p(\Omega, \Lambda) \to W^{1,p}(\Omega, \Lambda) \cap \mathbb{H}^\perp$, $1 < p < \infty$, is defined by

$$G(u) = \delta(u - H(u))$$

for every $u \in L^p(\Omega, \Lambda)$, where $H : L^p(\Omega, \Lambda) \to \mathbb{H}$ is the projection operator. Moreover, observe that $\Delta\delta(u) = u$, so we have that

$$\Delta G(u) = u - H(u). \tag{2.1}$$

By employing the classical dominated convergence theorem, C. Scott in [16] further gave the upper bound estimate of Green's operator G.

Lemma 2.2 *Let $u \in L^s(\Omega, \Lambda)$, $1 < s < \infty$, be a differential form defined in the domain Ω. Then, there exists a positive constant C, independent of u, such that*

$$\|dd^\star G(u)\|_{s,B} + \|d^\star dG(u)\|_{s,B} + \|dG(u)\|_{s,B} + \|d^\star G(u)\|_{s,B} + \|G(u)\|_{s,B} \leq C(s)\|u\|_{s,\sigma B} \tag{2.2}$$

for any ball $B \subset \sigma B \subset \Omega$ with some constant $\sigma > 1$, where Ω is a smoothly bounded domain without boundary.

Remark 1 For any $v \in L^p(\Omega, \Lambda) \bigcap \mathbb{H}^{\perp}$, by the definition of the projection operator H, it is easy to obtain that $H(v) = 0$. Since $G(u) \in W^{1,p}(\Omega, \Lambda) \cap \mathbb{H}^{\perp}$ for every $u \in L^p(\Omega, \Lambda)$, replacing v with $G(u)$ yields that $H(v) = HG(u) = 0$. In other words, the harmonic projection of Green's operator G on $L^p(\Omega, \Lambda)$ is always equal to zero.

Remark 2 Also, applying Lemma 2.2 repeatedly, it is obvious to achieve that there is a constant $C > 0$, independent of u, such that

$$\|G^m(u)\|_{p,B} \leq C\|u\|_{p,\sigma B}. \tag{2.3}$$

In particular, if $u \in W_d^p(\Omega, \Lambda)$ (or $u \in W_{d^\star}^p(\Omega, \Lambda)$), we know that Green's operator G can commute with d (or $d^\star$), which implies that

$$dG(u) = G(du) \ \text{ or } \ d^\star G(u) = G(d^\star u).$$

Similarly to the method employed in (2.3), we have that

$$\|dG^m(u)\|_{p,B} \leq C\|du\|_{p,\sigma B} \quad \text{or} \quad \|d^\star G^m(u)\|_{p,B} \leq C\|d^\star u\|_{p,\sigma B} \tag{2.4}$$

for any integer $m \geq 1$, where $\sigma > 1$ is some constant.

Meanwhile, to facilitate the upcoming argument about the Poincaré-type estimates in Theorem 2.6 and Theorem 2.7, we need the following results as well.

Lemma 2.3 ([17]) *Let $v \in L^p_{loc}(\Omega, \Lambda^l)$, $1 < p < \infty$, be a differential form defined in Ω and $T : L^p(\Omega, \Lambda^l) \to W^{1,p}(\Omega, \Lambda^{l-1})$ be the homotopy operator, $l = 1, 2, \cdots, n$. Then, we have*

$$v = d(Tv) + T(dv), \tag{2.5}$$

$$\|\nabla(Tv)\|_{p,\Omega} \leq C|\Omega|\|v\|_{p,\Omega} \ \text{ and } \ \|Tv\|_{p,\Omega} \leq C|\Omega|diam(\Omega)\|v\|_{p,\Omega} \tag{2.6}$$

hold for any bounded and convex domain Ω.

Before starting the primary argument in this section, it is worth to note that the explicit representations in Lemma 2.4 and Lemma 2.5 are the essential steps for the argument of the Poincaé-type inequalities. In precise, if our attention is only to estimate $\|D^k G^k(u)\|_{locLip^s_\alpha}$ (or $\|D^{k+1}G^k(u)\|_{locLip^s_\alpha(\Omega)}$) in terms of the L^p norm $\|u\|_{p,\Omega}$, we can prove it directly with the aid of the higher imbedding inequality given in [18]. Otherwise, while we are concerned on the upper boundedness of $\|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)}$ (or $\|D^{k+1}G^k(u)\|_{locLip^s_\alpha(\Omega)}$) in terms of the BMO^s norm $\|u\|_{*,s,\Omega}$, the higher imbedding result is not valid for this case any more. Thus, to overcome this difficulty, the key tools in our approach are Lemma 2.4 and Lemma 2.5, which are established by adapting the technique developed in [19] with the inductive method.

Lemma 2.4 *Let $u \in L^p_{loc}(\Omega, \Lambda)$, $1 < p < \infty$, be a differential form in the domain Ω, D be the Hodge-Dirac operator and G be Green's operator. Then, we have that*

$$D^k G^k(u) = G^m(u), \tag{2.7}$$

$$D^{k+1} G^k(u) = dG^m(u) + d^\star G^m(u), \tag{2.8}$$

for every even integer $k = 2m$ and $m = 1, 2, \cdots$.

Proof First, since $\Delta = D^2 = (d + d^\star)^2 = dd^\star + d^\star d$, we know that it holds

$$u = \Delta G(u) + H(u) = dd^\star G(u) + d^\star dG(u) + H(u) \tag{2.9}$$

for every $u \in L^p_{loc}(\Omega, \Lambda)$, which also implies that

$$dd^\star G(u) + d^\star dG(u) = u - H(u). \tag{2.10}$$

Due to the fact that $HG(u) = 0$ always holds by Remark 1, replacing u with $G^m(u)$ in (2.10) gives that

$$dd^\star G(G^m(u)) + d^\star dG(G^m(u)) = G^m(u) \tag{2.11}$$

whenever the positive integer $m \geq 1$.

Now, we will assert the representation (2.7) by using the inductive method. In the case of $k = 2$ and $m = 1$, we have

$$D^2 G^2(u) = (d + d^\star)^2 G^2(u) = dd^\star G(G(u)) + d^\star dG(G(u)). \tag{2.12}$$

Substituting 2.11 with $m = 1$ into (2.12) yields that $D^2 G^2(u) = G(u)$. Assume that the desired result holds for any $k = 2(m-1)$, $m = 2, 3, \cdots$, that is,

$$D^k G^k(u) = D^{2(m-1)} G^{2(m-1)}(u) = G^{m-1}(u). \tag{2.13}$$

Then, when k is taken as $2m$, it continues with (2.13) and (2.11) that

$$\begin{aligned} D^k G^k(u) &= D^2 D^{2(m-1)} G^{2(m-1)}(G^2(u)) = D^2 G^{m+1}(u) \\ &= dd^\star G(G^m(u)) + d^\star dG(G^m(u)) = G^m(u). \end{aligned} \tag{2.14}$$

So, the desired result (2.7) holds. Moreover, for the operator $D^{k+1} G^k(u)$, making use of (2.7) and the fact $D = d + d^\star$, we obtain that

$$D^{k+1} G^k(u) = D(D^k G^k(u)) = D(G^m(u)) = dG^m(u) + d^\star G^m(u).$$

Therefore, we finish the proof of Lemma 2.4. □

In analogue to the method developed in Lemma 2.4, we also derive the following results for the case $k = 2m + 1$.

Lemma 2.5 *Let $u \in L^p_{loc}(\Omega, \Lambda)$, $1 < p < \infty$, be a differential form defined in the domain Ω, D be the Hodge-Dirac operator and G be Green's operator. Then, we derive that*

$$D^k G^k(u) = dG^{m+1}(u) + d^\star G^{m+1}(u), \tag{2.15}$$

$$D^{k+1} G^k(u) = G^m(u) \tag{2.16}$$

for every odd integer $k = 2m + 1$ and $m = 1, 2, \cdots$.

Now, we are ready to give the local Poincaré-type estimates of the iterated operator $D^k G^k$ and $D^{k+1} G^k$ in terms of the L^p norms of du and $d^\star u$, respectively.

Theorem 2.6 *Assume that the differential form u is of the Sobolev class $W^{1,p}_{loc}(\Omega, \Lambda)$, $1 < p < \infty$, D is the Hodge-Dirac operator and G is Green's operator. Then, for any even integer $k = 2m$, $m = 1, 2, \cdots$, there exists a constant $C > 0$, independent of u, such that*

$$\|D^k G^k(u) - (D^k G^k(u))_B\|_{p,B} \le C|B|^{1+1/n} \|du\|_{p,\sigma B}, \tag{2.17}$$

$$\|D^{k+1} G^k(u) - (D^{k+1} G^k(u))_B\|_{p,B} \le C|B|^{1+1/n} \|d^\star u\|_{p,\sigma B} \tag{2.18}$$

for all balls $B \subset \sigma B \subset \Omega$ with some constant $\sigma > 1$.

Proof Initially, to prove (2.17), applying the decomposition (2.5) to $D^k G^k(u)$, we have

$$D^k G^k(u) = dT(D^k G^k(u)) + Td(D^k G^k(u)). \tag{2.19}$$

Since $dT(D^k G^k(u)) = (D^k G^k(u))_B$, for every $p > 1$, using (2.19), (2.7) and (2.6), it follows that

$$\begin{aligned} \|D^k G^k(u) - (D^k G^k(u))_B\|_{p,B} &= \|Td(D^k G^k(u))\|_{p,B} \\ &\le C_1|B|diam(B)\|d(D^k G^k(u))\|_{p,B} \\ &= C_1|B|diam(B)\|d(G^m(u))\|_{p,B} \\ &\le C_2|B|^{1+1/n}\|d(G^m(u))\|_{p,B}. \end{aligned} \tag{2.20}$$

Due to the definition of the Sobolev space and the facts that $\|du\|_{p,\Omega'} \le \|\nabla u\|_{p,\Omega'} < \infty$ and $\|d^\star u\|_{p,\Omega'} \le \|\nabla u\|_{p,\Omega'} < \infty$ for any $\Omega' \subset\subset \Omega$, one may readily see that Green's operator G can commute with d and $d^\star$. Then, combining (2.20) with (2.4) follows that

$$\|D^k G^k(u) - (D^k G^k(u))_B\|_{p,B} \le C_3 |B|^{1+1/n} \|du\|_{p,\sigma_1 B}$$

for any even integer $k > 0$. Thus, we have (2.17) always holds for all balls $B \subset \sigma B \subset \Omega$ with some constant $\sigma_1 > 1$.

Now, we turn to the proof of the inequality (2.18). First, applying the commute property between G and $d^\star$ and (2.2), we have

$$\|dd^\star G^k(u)\|_{p,B} = \|dG^k(d^\star(u))\|_{p,B} \le C_4 \|d^\star u\|_{p,\sigma_2 B}. \tag{2.21}$$

Making use of the similar treatment as in the proof of $D^k G^k$ with (2.8) and (2.21), we attain that

$$\begin{aligned}
\|D^{k+1} G^k(u) - (D^{k+1} G^k(u))_B\|_{p,B} &= \|Td(D^{k+1} G^k(u))\|_{p,B} \\
&\le C_5 |B| diam(B) \|d(D^{k+1} G^k(u))\|_{p,B} \\
&\le C_6 |B|^{1+1/n} \|d(dG^m(u) + d^\star G^m(u))\|_{p,B} \\
&= C_6 |B|^{1+1/n} \|dd^\star G^m(u)\|_{p,B} \\
&\le C_7 |B|^{1+1/n} \|d^\star u\|_{p,\sigma_2 B}
\end{aligned} \tag{2.22}$$

for every even integer $k > 0$ and some constant $\sigma_2 > 1$ with all balls $B \subset \sigma_2 B \subset \Omega$. Therefore, the proof of Theorem 2.6 is ended. □

Next, it is natural to take the case of the odd integer $k > 1$ into account. Using the same process as the case $k = 2m$ by Lemma 2.5 instead of Lemma 2.4, we derive the results for the odd integer $k = 2m + 1$. Considering the length of the paper, we only state the results of Theorem 2.7.

Theorem 2.7 *Assume that the differential form u is of the Sobolev class $W^{1,p}_{loc}(\Omega, \Lambda)$, $1 < p < \infty$, D is the Hodge-Dirac operator and G is Green's operator. Then, for any odd integer $k = 2m + 1$, $m = 1, 2, \cdots$, there exists a constant $C > 0$, independent of u, such that*

$$\|D^k G^k(u) - (D^k G^k(u))_B\|_{p,B} \le C |B|^{1+1/n} \|d^\star u\|_{p,\sigma B}, \tag{2.23}$$

$$\|D^{k+1} G^k(u) - (D^{k+1} G^k(u))_B\|_{p,B} \le C |B|^{1+1/n} \|du\|_{p,\sigma B} \tag{2.24}$$

for all balls $B \subset \sigma B \subset \Omega$ with some constant $\sigma > 1$.

Remark 3 It should be noticed that the results in Theorem 2.6 and Theorem 2.7 will play a significant role in latter discussion. Specifically, just because of the right terms du and $d^\star u$ in Theorem 2.6 and Theorem 2.7, it provides us an effective way to derive the upper boundedness of the iterated operators $D^k G^k$ and $D^{k+1} G^k$ in terms of the BMO^s norm for the conjugate A-harmonic tensors u and v.

3 Estimates for BMO^s and locLip^s_α Norms

In this section, we present our principal results about the estimates for BMO^s norm and locLip^s_α norm for $D^k G^k$ and $D^{k+1} G^k$ applied to differential forms u and v associated with some conjugate A-harmonic equation.

During the recent years, the study in the conjugate A-harmonic tensors is of growing interest and has made much progress, see [2, 10, 20, 21] for examples. Here, we consider the conjugate A-harmonic tensors of the form as follows.

Definition 3.1 ([10]) Differential forms $u \in W^{1,p}(\Omega, \Lambda)$ and $v \in W^{1,q}(\Omega, \Lambda)$ are called the conjugate A-harmonic tensors if u and v satisfy the conjugate A-harmonic equation of the form

$$A(du) = d^\star v, \tag{3.1}$$

where the operator $A : \Lambda(\Omega) \to \Lambda(\Omega)$ is restricted by the following structural assumptions:

(i) the mapping $\xi \to A(\xi)$ is continuous;
(ii) $|A(\xi)| \le a_1 |\xi|^{p-1}$, $< A(\xi), \xi > \ge b_1 |\xi|^p$;
(iii) $A(\lambda\xi) = \lambda|\lambda|^{p-2} A(\xi)$ whenever $\lambda \in \mathbb{R}$, $\lambda \neq 0$;
(iv) the monotonicity inequality: $| < A(\xi) - A(\eta), \xi - \eta > | \ge L_1 (|\xi|^2 + |\eta|^2)^{\frac{p-2}{2}} |\xi - \eta|^2$.

for all $\xi \in \Lambda(\mathbb{R}^n)$. Here, a_1, b_1 and $L_1 > 0$ are the positive constants and $1 < p, q < \infty$ are the conjugate exponents with $1/p + 1/q = 1$.

According to Definition 3.1, together with the facts $dd = 0$ and $d^\star d^\star = 0$, it is obvious to see that such a differential form u in (3.1) is also a solution to the A-harmonic equation

$$d^\star A(du) = 0. \tag{3.2}$$

Moreover, if the operator A is invertible, in view of the isometric property of the Hodge-star operator $\star$, there exists an operator B such that the differential form v in (3.1) meanwhile satisfies

$$d^\star B(d(\star v)) = 0, \tag{3.3}$$

where the operator $B : \Lambda(\Omega) \to \Lambda(\Omega)$ is given the similar conditions i)–iv) that

(b-i) the mapping $\xi \to B(\xi)$ is continuous on $\Lambda(\mathbb{R}^n)$;
(b-ii) $|B(\xi)| \le a_2 |\xi|^{q-1}$, $< B(\xi), \xi > \ge b_2 |\xi|^q$;
(b-iii) $B(\kappa\xi) = \kappa|\kappa|^{q-2} B(\xi)$ whenever $\kappa \in \mathbb{R}$, $\kappa \neq 0$;
(b-iv) the monotonicity inequality: $| < B(\xi) - B(\eta), \xi - \eta > | \ge L_2 (|\xi|^2 + |\eta|^2)^{\frac{q-2}{2}} |\xi - \eta|^2$.

for almost every $x \in \Omega$ and all $\xi \in \wedge^l(\mathbb{R}^n)$. Here, a_2, b_2 and L_2 are the positive constants and $1 < q < \infty$ is associated with (3.3).

Observe that A-harmonic equation is a special case of Dirac-harmonic equation, so we derive the Caccippoli inequality and the weak reverse Hölder inequality, respectively, by Corollary 2.3 and Theorem 4.3 in [22].

Lemma 3.2 *Let $u \in W^{1,p}(\Omega, \Lambda)$ and $v \in W^{1,q}(\Omega, \Lambda)$ satisfy the conjugate A-harmonic equation (3.1), and the operator A be invertible, where $1 < p, q < \infty$ are the given conjugate exponents with $1/p + 1/q = 1$. Then, there exists a constant $C > 0$, independent of u and v, such that*

$$\|du\|_{p,B} \leq C|B|^{-1/n}\|u - c\|_{p,\sigma B}, \tag{3.4}$$

$$\|d^\star v\|_{q,B} \leq C|B|^{-1/n}\| \star v - c^\star\|_{q,\sigma B} \tag{3.5}$$

for some constant $\sigma > 1$ and any ball $B \subset \sigma B \subset \Omega$, where c and $c^\star$ are both closed forms.

Lemma 3.3 *Let $\omega \in W^{1,p}(\Omega, \Lambda)$ be a solution to the homogenous A-harmonic equation, $1 < p < \infty$. Then, for every $0 < s, t < \infty$, there exists a constant $C > 0$, independent of ω, such that*

$$\|\omega\|_{s,B} \leq C|B|^{1/s-1/t}\|\omega\|_{t,\sigma B}, \tag{3.6}$$

where all balls $B \subset \sigma B \subset \Omega$ and $\sigma > 1$ is some constant.

In addition, the local higher order inequality is also necessary for our latter argument.

Lemma 3.4 *Let $u \in L^p_{loc}(\Omega, \Lambda)$, $1 < p < \infty$, be a differential form, D be the Hodge-Dirac operator and G be Green's operator. Then, for any positive integer $k \geq 1$, we have that*

(i) if $1 < p < n$, for any real number $0 < s < np/(n-p)$, there exists a constant $C > 0$, independent of u, such that

$$\|D^k G^k(u) - (D^k G^k(u))_B\|_{s,B} \leq C|B|^{1+1/n+1/s-1/p}\|u\|_{p,\sigma B}, \tag{3.7}$$

$$\|D^{k+1} G^k(u) - (D^{k+1} G^k(u))_B\|_{p,B} \leq C|B|^{1+1/n+1/s-1/p}\|u\|_{p,\sigma B} \tag{3.8}$$

(ii) if $p \geq n$, for any real number $s > 0$, there is a constant $C > 0$, independent of u, such that

$$\|D^k G^k(u) - (D^k G^k(u))_B\|_{s,B} \leq C|B|^{1+1/n+1/s-1/p}\|u\|_{p,\sigma B}, \tag{3.9}$$

$$\|D^{k+1} G^k(u) - (D^{k+1} G^k(u))_B\|_{p,B} \leq C|B|^{1+1/n+1/s-1/p}\|u\|_{p,\sigma B} \tag{3.10}$$

for all balls $B \subset \sigma B \subset \Omega$ with some constant $\sigma > 1$.

Now, with these facts in mind, let us first prove Theorem 3.5.

Theorem 3.5 *Let $u \in L^p(\Omega, \Lambda)$, $1 < p < n$, be a differential form defined on the smoothly bounded domain Ω, D be the Hodge-Dirac operator and G be Green's operator. Then, for any positive integer $k > 1$ and any real number $0 < s < np/(n-p)$, there exist two constants $C_1, C_2 > 0$, independent of u, such that*

$$\|D^k G^k(u)\|_{*,s,\Omega} \le C_1 \|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)} \le C_2 \|u\|_{p,\Omega}, \tag{3.11}$$

$$\|D^{k+1} G^k(u)\|_{*,s,\Omega} \le C_1 \|D^{k+1} G^k(u)\|_{locLip^s_\alpha(\Omega)} \le C_2 \|u\|_{p,\Omega}, \tag{3.12}$$

where $0 < \alpha \le 1$ is some constant.

Proof First, we notice that $1 + \frac{1}{n} - \frac{1}{p} - \frac{\alpha}{n} = \left(1 - \frac{1}{p}\right) + \left(\frac{1}{n} - \frac{\alpha}{n}\right) > 0$ because $0 < \alpha \le 1$ and $1 < p < \infty$. Then, for any ball $B \subset \Omega$, we have

$$|B|^{1+1/n-1/p-\alpha/n} \le |\Omega|^{1+1/n-1/p-\alpha/n}. \tag{3.13}$$

In the meantime, by replacing ω with $D^k G^k(u)$ and $D^{k+1} G^k(u)$ in (1.5), respectively, it is immediate to achieve that

$$\|D^k G^k(u)\|_{*,s,\Omega} \le C_1 \|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)}, \tag{3.14}$$

$$\|D^{k+1} G^k(u)\|_{*,s,\Omega} \le C_1 \|D^{k+1} G^k(u)\|_{locLip^s_\alpha(\Omega)}. \tag{3.15}$$

Thus, to estimate (3.11), applying (3.7) and (3.13) gives

$$\begin{aligned}
\|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)} &= \sup_{\sigma_2 B \subset \Omega} |B|^{-\frac{n+\alpha s}{sn}} \|D^k G^k(u) - (D^k G^k(u))_B\|_{s,B} \\
&\le \sup_{\sigma_2 B \subset \Omega} |B|^{-1/s-\alpha/n} C_2 |B|^{1+1/s+1/n-1/p} \|u\|_{p,\sigma_1 B} \\
&= \sup_{\sigma_2 B \subset \Omega} C_2 |B|^{1+1/n-1/p-\alpha/n} \|u\|_{p,\sigma_1 B} \\
&\le \sup_{\sigma_2 B \subset \Omega} C_2 |\Omega|^{1+1/n-1/p-\alpha/n} \|u\|_{p,\sigma_1 B} \\
&\le C_3 \sup_{\sigma_2 B \subset \Omega} \|u\|_{p,\sigma_1 B} \\
&\le C_4 \|u\|_{p,\Omega},
\end{aligned} \tag{3.16}$$

where the constants $\sigma_2 > \sigma_1 > 1$ and all balls $B \subset \sigma_1 B \subset \sigma_2 B \subset \Omega$. So, according to (3.14) and (3.16), we have that (3.11) holds as desired. Moreover, using the same treatment to the operator $D^{k+1} G^k(u)$ with (3.8) and (3.15), the inequality (3.12) holds as well. Therefore, the proof of Theorem 3.5 is finished. □

For the case $p \geq n$, repeating the process as in Theorem 3.5 with (3.9) and (3.10), we obtain the analogue results.

Theorem 3.6 *Let $u \in L^p(\Omega, \Lambda)$, $p \geq n$, be a differential form defined on the smoothly bounded domain Ω, D be the Hodge-Dirac operator and G be Green's operator. Then, for any positive integer $k > 1$ and any real number $s > 0$, there exist two constants $C_1, C_2 > 0$, independent of u, such that*

$$\|D^k G^k(u)\|_{*,s,\Omega} \leq C_1 \|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)} \leq C_2 \|u\|_{p,\Omega}, \tag{3.17}$$

$$\|D^{k+1} G^k(u)\|_{*,s,\Omega} \leq C_1 \|D^{k+1} G^k(u)\|_{locLip^s_\alpha(\Omega)} \leq C_2 \|u\|_{p,\Omega}, \tag{3.18}$$

where $0 < \alpha \leq 1$ is some constant.

Next, we begin to establish our principle relationship between BMO^s norm and locLip^s_α norm of the iterated operators in terms of the norms of the conjugate harmonic tensors u and v. From Theorem 3.7 and Theorem 3.8 to Corollary 3.9 and Corollary 3.10 below, we always assume that $\Omega \subset \mathbb{R}^n$ is smoothly bounded domain without boundary, the operator A in (3.1) is invertible.

Theorem 3.7 *Let $u \in W^{1,p}(\Omega, \Lambda)$ and $v \in W^{1,q}(\Omega, \Lambda)$, $1 < p, q < \infty$ with $1/p + 1/q = 1$, be the conjugate A-harmonic tensors satisfying the Equation (3.1), D be the Hodge-Dirac operator and G be Green's operator. Then, for every integer $k = 2m$ and any real number $s > 0$, $m = 1, 2, \cdots$, there are two constants $C_1, C_2 > 0$, independent of u and v, such that*

$$\|D^k G^k(u)\|_{*,s,\Omega} \leq C_1 \|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)} \leq C_2 \|u\|_{*,p,\Omega}, \tag{3.19}$$

$$\|D^{k+1} G^k(v)\|_{*,s,\Omega} \leq C_1 \|D^{k+1} G^k(v)\|_{locLip^s_\beta(\Omega)} \leq C_2 \| \star v\|_{*,p,\Omega}, \tag{3.20}$$

where $0 < \alpha, \beta \leq 1$ are the expansion factors.

Proof First, without loss of generality, we assume that the conjugate A-harmonic tensor u is a solution to the A-harmonic equation (3.2). Then, it is natural to view the corresponding v as a solution to Equation (3.3). Next, we will divide our proof into two parts.

(i) For every $1 < p < \infty$, applying (2.17) into Definition 1.2, we have that

$$\begin{aligned} \|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)} &= \sup_{\sigma_1 B \subset \Omega} |B|^{-1/s-\alpha/n} \|D^k G^k(u) - (D^k G^k(u))_B\|_{s,B} \\ &\leq \sup_{\sigma_1 B \subset \Omega} C_1 |B|^{-1/s-\alpha/n} |B|^{1+1/n} \|du\|_{s,\sigma_2 B} \\ &\leq C_1 \sup_{\sigma_1 B \subset \Omega} |B|^{1+1/n-1/s-\alpha/n} \|du\|_{s,\sigma_2 B} \end{aligned} \tag{3.21}$$

Observe that du is a solution for the A-harmonic equation since du is a closed form. Then, for any real number $s > 0$, using Lemma 3.3 yields that

$$\|du\|_{s,\sigma_2 B} \leq C_2 |B|^{1/s-1/p} \|du\|_{p,\sigma_3 B}. \tag{3.22}$$

where $\sigma_3 > \sigma_2 > 1$. Under the assumption, we know that u satisfies the Caccippoli inequality (3.4). Especially, choosing $c = u_B$ in (3.4) follows

$$\|du\|_{p,\sigma_3 B} \leq C_3 |B|^{-1/n} \|u - u_B\|_{p,\sigma_4 B} \tag{3.23}$$

for some constant $\sigma_4 > \sigma_3 > 1$ with any ball $\sigma_3 B \subset \sigma_4 B \subset \Omega$. Moreover, combining (3.22) and (3.23) gives

$$\|du\|_{s,\sigma_2 B} \leq C_4 |B|^{1/s-1/n-1/p} \|u - u_B\|_{p,\sigma_4 B} \tag{3.24}$$

So, substituting (3.24) into (3.21), together with Definition 1.1, yields that

$$\begin{aligned}
\|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)} &\leq C_5 \sup_{\sigma_1 B \subset \Omega} |B|^{1+1/n-1/s-\alpha/n} |B|^{1/s-1/n-1/p} \|u-u_B\|_{p,\sigma_4 B} \\
&\leq C_5 \sup_{\sigma_1 B \subset \Omega} |B|^{1-1/p-\alpha/n} \|u - u_B\|_{p,\sigma_4 B} \\
&\leq C_5 \sup_{\sigma_1 B \subset \Omega} |\Omega|^{1-\alpha/n} |B|^{-1/p} \|u - u_B\|_{p,\sigma_4 B} \\
&\leq C_6 \sup_{\sigma_1 B \subset \Omega} |B|^{-1/p} \|u - u_B\|_{p,\sigma_4 B} \\
&\leq C_6 \|u\|_{*,p,\Omega},
\end{aligned} \tag{3.25}$$

where the constants $\sigma_1 > \sigma_4 > 1$. Therefore, we have that (3.19) holds for any even integer $k > 1$ and any real number $s > 0$.

The proof of (3.20) is similar to that of (3.19). Next, we only present the different steps.

(ii) For every conjugate A-harmonic tensor $v \in W^{1,q}(\Omega, \Lambda)$, employing the same treatment used in the proof of (3.19), along with (2.18), we have that

$$\|D^k G^k(v)\|_{locLip^s_\beta(\Omega)} \leq C_7 \sup_{\eta_1 B \subset \Omega} |B|^{1+1/n-1/s-\beta/n} \|d^\star v\|_{s,\eta_2 B}. \tag{3.26}$$

According to the isometric property of the Hodge-star operator $\star$, we know that $|d^\star v| = |d \star v|$. Notice that $d \star v$ is a closed form satisfying A-harmonic equation. So, for any real number $s > 0$, using Lemma 3.3 again, we derive that

$$\|d^\star v\|_{s,\eta_2 B} = \|d \star v\|_{s,\eta_2 B} \leq C_8 |B|^{1/s-1/q} \|d \star v\|_{q,\eta_3 B} \tag{3.27}$$

Also, by the comments after Definition 3.1, it implies that $\star v$ is a solution to the A-harmonic equation (3.3). Then, by Lemma 3.2, letting $c^\star = (\star v)_B$ shows that

$$\|d \star v\|_{q,\eta_3 B} \leq C_9 |\eta_3 B|^{-1/n} \| \star v - (\star v)_B \|_{q,\eta_4 B}. \tag{3.28}$$

So, combining (3.27) with (3.28) and plugging it into (3.26), we have that

$$\begin{aligned}
\|D^{k+1} G^k(v)\|_{locLip^s_\beta(\Omega)} &\leq \sup_{\eta_1 B \subset \Omega} |B|^{1+1/n-1/s-\beta/n} \|d^\star v\|_{s,\eta_2 B} \\
&\leq \sup_{\eta_1 B \subset \Omega} C_{10} |B|^{1+1/n-1/s-\beta/n} |B|^{1/s-1/q-1/n} \\
&\qquad \| \star v - (\star v)_B \|_{q,\eta_4 B} \\
&\leq \sup_{\eta_1 B \subset \Omega} C_{10} |\Omega|^{1-\beta/n} |B|^{-1/q} \| \star v - (\star v)_B \|_{q,\eta_4 B} \\
&\leq C_{11} \sup_{\eta_1 B \subset \Omega} |B|^{-1/q} \| \star v - (\star v)_B \|_{q,\eta_4 B} \\
&= C_{11} \| \star v \|_{*,q,\Omega}
\end{aligned} \tag{3.29}$$

as desired, where the constants $\eta_1 > \eta_4 > \eta_3 > \eta_2 > 1$.

□

Now, in the odd case $k = 2m + 1$, we have the similar estimates as follows. It should be pointed out that the proof of Theorem 3.8 is the analogue of Theorem 3.7, so we only state the results and leave the proof of the odd case $k > 1$ to the readers.

Theorem 3.8 *Let $u \in W^{1,p}(\Omega, \Lambda)$ and $v \in W^{1,q}(\Omega, \Lambda)$, $1 < p, q < \infty$ with $1/p + 1/q = 1$, be the conjugate A-harmonic tensors satisfying the Equation (3.1), D be the Hodge-Dirac operator and G be Green's operator. Then, for every odd integer $k = 2m + 1$ and any real number $s > 0$, $m = 1, 2, \cdots$, there are two constants $C_1, C_2 > 0$, independent of u and v, such that*

$$\|D^k G^k(v)\|_{*,s,\Omega} \leq C_1 \|D^k G^k(v)\|_{locLip^s_\beta(\Omega)} \leq C_2 \| \star v\|_{*,q,\Omega}, \tag{3.30}$$

$$\|D^{k+1} G^k(u)\|_{*,s,\Omega} \leq C_1 \|D^{k+1} G^k(u)\|_{locLip^s_\alpha(\Omega)} \leq C_2 \|u\|_{*,p,\Omega}, \tag{3.31}$$

where $0 < \alpha, \beta \leq 1$ are the expansion factors.

In particular, if $p(\alpha - 1) = q(\eta - 1)$, as a consequence of Theorem 3.7 and 3.8, the following estimates are established simply by means of Theorem 6.6 in [10]. It is worth to notice that the treatment applied in Corollary 3.9 and Corollary 3.10 are very similar, so we only give the complete proof of Corollary 3.9 in details.

Corollary 3.9 *Let $u \in W^{1,p}(\Omega, \Lambda)$ and $v \in W^{1,q}(\Omega, \Lambda)$, $1 < p, q < \infty$ with $1/p + 1/q = 1$, be the conjugate A-harmonic tensors satisfying the Equation (3.1),*

D be the Dirac operator and G be the Green's operator. If $0 < \alpha, \beta \leq 1$ *satisfy* $p(\alpha-1) = q(\beta-1)$*, for any real* $s > 0$*, then there exist two constants* $C_1, C_2 > 0$*, independent of u and v, such that*

$$\|D^k G^k(u)\|_{*,s,\Omega} \leq C_1 \|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)} \leq C_2 \|\star v\|^{q/p}_{locLip^q_\beta(\Omega)}, \tag{3.32}$$

$$\|D^{k+1} G^k(v)\|_{*,s,\Omega} \leq C_1 \|D^{k+1} G^k(v)\|_{locLip^s_\beta(\Omega)} \leq C_2 \|u\|^{p/q}_{locLip^p_\alpha(\Omega)}, \tag{3.33}$$

whenever $k = 2m$*,* $m = 1, 2, \cdots$*.*

Proof First, combining (1.5) and Theorem 6.6 in [10], we have

$$\|u\|_{*,s,\Omega} \leq C_1 \|u\|_{locLip^s_\alpha(\Omega)} \leq C_2 \|\star v\|^{q/p}_{locLip^q_\beta(\Omega)}, \tag{3.34}$$

$$\|\star v\|_{BMO,\Omega} \leq C_3 \|\star v\|_{locLip_\beta(\Omega)} \leq C_4 \|u\|^{p/q}_{locLip_\alpha(\Omega)}. \tag{3.35}$$

Then, substituting (3.34) into (3.19) and (3.35) into (3.20), respectively, it yields that

$$\|D^k G^k(u)\|_{*,s,\Omega} \leq C_5 \|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)} \leq C_6 \|\star v\|^{q/p}_{locLip^q_\beta(\Omega)},$$

$$\|D^{k+1} G^k(v)\|_{*,s,\Omega} \leq C_7 \|D^{k+1} G^k(v)\|_{locLip^s_\beta(\Omega)} \leq C_8 \|u\|^{p/q}_{locLip^p_\alpha(\Omega)}$$

as desired. □

Corollary 3.10 *Suppose that* $0 < \alpha, \beta \leq 1$ *satisfy* $p(\alpha-1) = q(\beta-1)$*, for any real number* $s > 0$*, then there exist two constants* $C_1, C_2 > 0$*, independent of u and v, such that*

$$\|D^k G^k(v)\|_{*,s,\Omega} \leq C_1 \|D^k G^k(v)\|_{locLip^s_\beta(\Omega)} \leq C_2 \|u\|^{p/q}_{locLip^p_\alpha(\Omega)}, \tag{3.36}$$

$$\|D^{k+1} G^k(u)\|_{*,s,\Omega} \leq C_1 \|D^{k+1} G^k(u)\|_{locLip^s_\alpha(\Omega)} \leq C_2 \|\star v\|^{q/p}_{locLip^q_\beta(\Omega)}, \tag{3.37}$$

whenever $k = 2m+1$*,* $m = 1, 2, \cdots$*.*

What is more, for each pair of conjugate A-harmonic tensors u and v, in accord to the facts that $|du|^p \leq |d^\star v|^q \leq a_1^q |du|^p$ and $|d^\star v| = |d \star v|$, one may easily establish such a useful L^p-equivalence with respect to u and v as follows:

$$\|du\|_{p,\Omega'} \leq \|d \star v\|^{q/p}_{q,\Omega'} \leq a_1^{q/p} \|du\|_{p,\Omega'}, \tag{3.38}$$

whenever $\Omega' \subset \Omega$, where $1 < p, q < \infty$ are the conjugate Hölder exponents. In view of the equivalence (3.38), if u and v are the conjugate A-harmonic tensors,

it further reveals the relations (3.39)–(3.42) below. Namely, when k is any positive even integer, there exist two constants $C_1, C_2 > 0$, independent of u and v, such that

$$\|D^k G^k(u)\|_{*,s,\Omega} \leq C_1 \|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)} \leq C_2 \|d \star v\|^{q/p}_{q,\Omega}, \tag{3.39}$$

$$\|D^{k+1} G^k(v)\|_{*,s,\Omega} \leq C_1 \|D^{k+1} G^k(v)\|_{locLip^s_\beta(\Omega)} \leq C_2 \|du\|^{p/q}_{p,\Omega}. \tag{3.40}$$

for any real number $s > 0$. As such, when $k > 1$ is any odd integer, there also exist two constants $C_1, C_2 > 0$, independent of u and v, such that

$$\|D^k G^k(v)\|_{*,s,\Omega} \leq C_1 \|D^k G^k(v)\|_{locLip^s_\beta(\Omega)} \leq C_2 \|du\|^{p/q}_{p,\Omega}, \tag{3.41}$$

$$\|D^{k+1} G^k(u)\|_{*,s,\Omega} \leq C_1 \|D^{k+1} G^k(u)\|_{locLip^s_\alpha(\Omega)} \leq C_2 \|d \star v\|^{q/p}_{q,\Omega}. \tag{3.42}$$

It should be pointed out that the proof of the above assertions are parallel to the those of Theorem 3.7. Therefore, we omit the details.

4 Applications

In this section, we use some concrete examples to illustrate the applications of the main results obtained in Section 3.

Let the mapping $f : \Omega \to \mathbb{R}^n$, $f = (f^1, \cdots, f^n)$, be of Sobolev class $W^{1,p}_{loc}(\Omega, \Lambda)$ and $J(x, f) = det(Df(x))$ be the Jacobian determinant of f. Then, we have that

$$u = J(x_{i_1}, x_{i_2}, \cdots, x_{i_l}; f^{j_1}, f^{j_2}, \cdots, f^{j_l}) dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_l}, \tag{4.1}$$

is a differential l-form, where $J(x_{i_1}, x_{i_2}, \cdots, x_{i_l}; f^{j_1}, f^{j_2}, \cdots, f^{j_l})$ is the subdeterminant of $J(x, f)$ of the form:

$$J(x_{i_1}, x_{i_2}, \cdots, x_{i_l}; f^{j_1}, f^{j_2}, \cdots, f^{j_l}) = \begin{vmatrix} f^{j_1}_{x_{i_1}} & f^{j_1}_{x_{i_2}} & \cdots & f^{j_1}_{x_{i_l}} \\ f^{j_2}_{x_{i_1}} & f^{j_2}_{x_{i_2}} & \cdots & f^{j_2}_{x_{i_l}} \\ \cdots & \cdots & \cdots & \cdots \\ f^{j_l}_{x_{i_1}} & f^{j_l}_{x_{i_2}} & \cdots & f^{j_l}_{x_{i_l}} \end{vmatrix}$$

Referring to Chapter 1 in [2], we find that Theorem 3.5 and Theorem 3.6 are applicable to such sort of the differential form u. Here, take a special case of 2-dimensional Euclidean space for instance.

Example 4.1 Assume that $u = J(x, y; f^1, f^2)dx \wedge dy$ is the differential 2-form defined on the domain $\Omega = \{(x, y) \in \mathbb{R}^2 : 0 < x^2 + y^2 < r^2\}$, where the mapping $f : \Omega \to \mathbb{R}^2$ is of the Sobolev class $W^{1,p}_{loc}(\Omega, \Lambda)$ denoted by

$$f(x, y) = (f^1(x, y), f^2(x, y)) = \left(\frac{x}{(x^2 + y^2)^{1/8}}, \frac{y}{(x^2 + y^2)^{1/8}}\right) \tag{4.2}$$

for any $r > 0$ and $p > 1$. After a simple calculation, one may derive that

$$u = J(x, y; f^1, f^2)dx \wedge dy = \frac{3}{4}(x^2 + y^2)^{-1/4}dx \wedge dy.$$

Thus, by the spherical coordinate transformation, it is easy to see that $u \in L^p(\Omega, \Lambda^2)$ for any $p < 4$. For example, choosing $p = 3/2$, we know that $u \in L^{3/2}(\Omega, \Lambda^2)$. However, by the direct integral calculation with Definition 1.1 and Definition 1.2, it is quite hard to infer the higher order boundedness of BMO^s norm and locLip^s_α norm with respect to $D^kG^k(u)$ and $D^{k+1}G^k(u)$. Then, applying Theorem 3.5 to D^kG^k and $D^{k+1}G^k$, for any $0 < s < np/(n - p) = \frac{2\cdot 3/2}{2-3/2} = 6$, we have that $D^kG^k(u) \in \mathrm{BMO}^s(\Omega, \Lambda^2)$ and $D^{k+1}G^k(u) \in \mathrm{BMO}^s(\Omega, \Lambda^2)$. Moreover, there are two constants $C_1, C_2 > 0$, independent of u, such that

$$\|D^kG^k(u)\|_{*,s,\Omega} \le C_1\|D^kG^k(u)\|_{locLip^s_\alpha(\Omega)} \le C_2 r^{5/4}, \tag{4.3}$$

$$\|D^{k+1}G^k(u)\|_{*,s,\Omega} \le C_1\|D^{k+1}G^k(u)\|_{locLip^s_\alpha(\Omega)} \le C_2 r^{5/4} \tag{4.4}$$

for every $0 < \alpha \le 1$ and all positive integer $k \ge 1$.

Especially, if the homeomorphism $f : \Omega \to \mathbb{R}^n$ of Sobolev class $W^{1,n}_{loc}(\Omega, \mathbb{R}^n)$, as mentioned above, is the K-quasiregular mapping, $K \ge 1$. From [23], we know that

$$u = f^l df^1 \wedge \cdots \wedge df^{l-1} \quad \text{and} \quad v = \star f^{l+1} df^{l+2} \wedge \cdots \wedge df^n$$

are the conjugate A-harmonic tensors, whenever $l = 1, 2, \cdots, n-1$. Here, consider the 4-dimensional space as an example.

Example 4.2 Let $f = (f^1, f^2, f^3, f^4)$ be the K-quasiregular mapping defined on the domain $\Omega = \{(x_1, x_2, x_3, x_4) : |x_i| < a, i = 1, 2, 3, 4\} \subset \mathbb{R}^4$, and choose the conjugate A-harmonic tensors as follows:

$$u = f^2 df^1 \quad \text{and} \quad v = \star f^3 df^4.$$

where $0 < a < \infty$ is some real number. If $u \in \mathrm{BMO}^p(\Omega, \Lambda)$ and $\star v \in \mathrm{BMO}^q(\Omega, \Lambda)$, where p and q are conjugate exponents with $1/p + 1/q = 1$, by applying Theorem 3.7 and Theorem 3.8, respectively, we have that for any even

integer $k = 2m$ and any real number $s > 0$, $m = 1, 2, \cdots$, there are two constants $C_1, C_2 > 0$, such that

$$\|D^k G^k(u)\|_{*,s,\Omega} \le C_1 \|D^k G^k(u)\|_{locLip^s_\alpha(\Omega)} \le C_2 \|f^2 df^1\|_{*,p,\Omega},$$

$$\|D^{k+1} G^k(v)\|_{*,s,\Omega} \le C_1 \|D^{k+1} G^k(v)\|_{locLip^s_\beta(\Omega)} \le C_2 \|f^3 df^4\|_{*,q,\Omega},$$

where $0 < \alpha, \beta \le 1$ are two factors. While the integer $k = 2m + 1$, $m = 1, 2, \cdots$, it holds that

$$\|D^k G^k(v)\|_{*,s,\Omega} \le C_1 \|D^k G^k(v)\|_{locLip^s_\beta(\Omega)} \le C_2 \|f^3 df^4\|_{*,q,\Omega},$$

$$\|D^{k+1} G^k(u)\|_{*,s,\Omega} \le C_1 \|D^{k+1} G^k(u)\|_{locLip^s_\alpha(\Omega)} \le C_2 \|f^2 df^1\|_{*,p,\Omega},$$

for any real number $s > 0$, where $0 < \alpha, \beta \le 1$ are two factors.

Remark 4 In general, all results we establish here provide us an impressive description about the relation between BMO^s norm and locLip^s_α norm for the iterated operators. Also, from the results, one may realize that locLip^s_α-norm estimates for differential forms are fairly essential for the process to derive the BMO^s estimate with respect to $D^k G^k$ and $D^{k+1} G^k$ for differential forms.

Acknowledgments The author G. Shi are very grateful to the Chinese Scholarship Council (CSC) for its financial support on her visit in USA, and sincerely appreciates the Department of Mathematics at Seattle University for providing the precious opportunity for her academic visit.

References

1. G. Lu, Embedding theorems into Lipschitz and BMO spaces and applications to quasilinear subelliptic differential equations. Publ. Mat. **40**, 301–329 (1996)
2. R.P. Agarwal, S. Ding, C.A. Nolder, *Inequalities for Differential Forms* (Springer, New York, 2009)
3. X. Li, Y. Wang, Y. Xing, Lipschitz and BMO norm inequalities for the composite operator on differential forms. J. Inequalities Appl. (2015). https://doi.org/10.1186/s13660-015-0896-9
4. Y. Xing, Y. Wang, BMO and Lipschitz norm estimates for composite operators. Potential Anal. **31**, 335–344 (2009)
5. D. Deng, X. Duong, A. Sikora, L. Yan, Comparison of the classical BMO spaces associated with operators and applications. Rev. Mat. Iberoamericana **24**, 267–296 (2008)
6. S. Stein, *Harmonic Analysis: Real-variable Methods, Orthogonality and Oscillatory Integrals* (Princeton University Press, New Jersey, 1993)
7. A. Bonami, T. Iwaniec, P. Jones, M. Zinsmeister, On the product of functions in BMO and H^1. Annu. Inst. Fourier (Grenoble) **57**, 1405–1439 (2007)
8. A. Carbonaro, A. McIntosh, A.J. Morris, Local Hardy spaces of differential forms on Riemannian manifolds. J. Geom. Anal. **23**, 106–169 (2013)
9. Y. Xing, S. Ding, Norm comparison inequalities for the composite operator. J. Inequalities Appl. (2009). https://doi.org/10.1155/2009/212915

10. C.A. Nolder, Hardy-Littlewood theorems for A-harmonic tensors. Ill. J. Math. **43**, 613–632 (1999)
11. G.F.D. Duff, D.C. Spencer, Harmonic tensors on Riemannian manifolds with boundary. Ann. Math. **56**, 128–156 (1952)
12. B. Stroffolini, On weakly A-harmonic tensors. Stud. Math. **114**, 289–301 (1995)
13. V. Gol'dshtein, M. Troyanov, Sobolev inequalities for differential forms and $L_{q,p}$-cohomology. J. Geom. Anal. **16**, 597–631 (2006)
14. S. Ding, Lipschitz and BMO norm inequalities for operators. Nonlinear Anal. Theory Methods Appl. **71**, 2350–2357 (2009)
15. C.B. Morrey, *Multiple Integrals in the Calculus of Variations* (Springer, Berlin, 1966)
16. C. Scott, L^p theory of differential forms on manifolds. Trans. Am. Math. Soc. **347**, 2075–2096 (1995)
17. T. Iwaniec, A. Lutoborski, Integral estimates for null Lagrangians. Arch. Ration. Mech. Anal. **125**, 25–79
18. S. Ding, G. Shi, Y. Xing, Higher integrability of iterated operators on differential forms. Nonlinear Anal. Theory Methods Appl. **145**, 83–96 (2016)
19. S. Ding, G. Shi, Y. Xing, Representation and regularity of higher order Dirac and Green's operators. Preprint
20. R.P. Agarwal, S. Ding, Advances in differential forms and the A-harmonic equation. Math. Comput. Model. **37**, 1393–1426 (2003)
21. S. Ding, Weighted Hardy-Littlewood inequality for A-harmonic tensors. Proc. Am. Math. Soc. **125**, 1727–1735 (1997)
22. S. Ding, B. Liu, Dirac-harmonic equations for differential forms. Nonlinear Anal. Theory Methods Appl. **122**, 43–57 (2015)
23. T. Iwaniec, G. Martin, Quasiregular mappings in even dimensions. Acta Math. **170**, 29–81 (1933)

Application of Boundary Perturbations on Medical Monitoring and Imaging Techniques

M. Doschoris, A. Papargiri, V. S. Kalantonis, and P. Vafeas

Abstract We present an overview on the application of boundary perturbations for Electroencephalography and Magnetoencephalography predominantly for the spherical geometry. With the mathematical tools produced, both forward and inverse problems can be tackled providing explicit computationally efficient solutions. Utilizing perturbation analysis in the framework of medical monitoring and imaging techniques, possesses the advantage introducing geometric variations without limiting the installation of analytic, or at least semi-analytic solutions, in view of complicated surfaces. In our example, surfaces which do not allow an analytic mathematical treatment can be handled if considered as small deviations from the sphere. In that setting, irregularities in head shapes, e.g. craniofacial alterations can be investigated theoretically.

MSC 78A25, 35Q61, 92C05, 35Q92, 45K05, 35R30

M. Doschoris
Institute of Genetics and Biometry, Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany
e-mail: doschoris@fbn-dummerstorf.de

A. Papargiri · V. S. Kalantonis (✉)
Department of Electrical and Computer Engineering, University of Patras, Patras, Greece
e-mail: apapargiri@upatras.gr; kalantonis@upatras.gr

P. Vafeas
Department of Chemical Engineering, University of Patras, Patras, Greece
e-mail: vafeas@chemeng.upatras.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_6

1 Introduction

Medical monitoring techniques such as Electroencephalography (EEG) and functional neuroimaging techniques such as Magnetoencephalography (MEG) provide the possibility of recognizing brain structures and map cortical activity at its core functional level. Both consist noninvasive diagnostic techniques, playing a crucial role in many aspects of today's clinical research either as the principal investigative tool or providing supportive evidence. As Ilmoniemi and Näätänen assert (for MEG) [1] "[. . .] tool to probe into the real-time operation of the human brain in experimental conditions that are suitable for studying sensory and cognitive brain functions as well as their disturbances." As an illustration, EEG is practised in reporting schizophrenia [2], in detection and analysis of epileptic activity [3], in the study of language processing [4] and many others. On the other hand, MEG is used tracing the dynamics and connectivity of large-scale brain activity [5].

The core question of EEG and MEG is to gain insight of the working human brain by recognizing the precise position and strength of the underlying neuronal activity. In the case of EEG, measurements are based on the post-synaptic potentials, which originate in the pyramidal neurons recorded on the scalp [6], whereas MEG measures oscillatory magnetic fields [7]. The identification of which brain areas have been activated, based on either EEG or MEG measurements, is termed the inverse problem or source reconstruction, while the obverse is labeled the forward problem [8]. Historically, the forward and inverse EEG problems have been extensively scrutinized since the 1950s when Wilson and Bayley [9] attempted to quantify the interplay between neuronal activity and the potentials, generated at the scalp [10, 11]. Similar hold for MEG, since David Cohen first measured magnetic fields produced by Humans in the late 1960s [12].

From the mathematical point of view, the formulation for the Electroencephalography and Magnetoencephalography problems are derived from the quasi-static theory of electromagnetism [13, 14]. We recall that microscopic currents are largely approximated by equivalent dipole sources and that an exclusive source configuration for each measurement of either EEG or MEG does not exist, comprising the corresponding inverse problems ill-posed. Regarding non-uniqueness, Albanese and Monk [15] demonstrated that the reconstruction of a three-dimensional current based on EEG measurements is irrealizable and has been practically demonstrated by Dassios and Fokas [16]. Nevertheless, if the source representing neuronal activity has dimensionality less than three, the inverse problems at hand admit a unique solution. The classic example is the dipole approximation of zero dimensionality. The same is true stipulating that neuronal activity is characterized via continuously distributed currents in one [17] and two dimensions [18]. Evidently, discarding the dipole hypothesis, allows the investigation and analysis of complicated activation patterns in terms of distributed currents.

Another interesting aspect regards the complementarity of EEG and MEG, a consequence due to sensitivity of EEG and MEG to radial and tangential sources. The notion of complementarity has been mathematically defined spherical model

of the brain and a continuously distributed neuronal current, demonstrating that the information which is missing in EEG data is contained in MEG data and vice versa [19]. Importantly, it has been proved in [20] that, independent of the geometry, the inverse EEG problem is not capable of reconstructing more than 1/3 of any neuronal activity, whereas the inverse MEG problem cannot recover the 2/3 of any brain activity, including the 1/3 of EEG.

Scope of the paper is to provide an analytic overview on the application of boundary perturbations for EEG and MEG predominantly for the spherical geometry, allowing a systematic presentation of the mathematical complexity associated. Extensive use is made of special functions and especially spherical harmonics. Details and technicalities regarding manipulation of formulas involving spherical harmonics can be found in [21, 22]. Our work is organized as follows: In the next subsection, we provide a brief discussion on boundary perturbations. In Section 2, the mathematical formulation of medical monitoring techniques are presented and the influence of geometric variations on the forward as well as inverse problem to EEG and MEG are studied. Finally, in Section 3, two specific examples are provided for the obtained solutions of both EEG and MEG, while in Section 4 we discuss our results and conclude.

1.1 Boundary Perturbations. A Short Introduction

Consider the action of the operator $\mathscr{L}(\boldsymbol{\tau})$ on a function $w(\boldsymbol{\tau})$ within a suitable domain Ω, such that

$$(\mathscr{L}w)(\boldsymbol{\tau}) = \phi(\boldsymbol{\tau}), \quad \boldsymbol{\tau} \in \Omega, \tag{1}$$

where $\phi(\boldsymbol{\tau})$ is the outcome of the operation. In the setting of Boundary Value Problems (BVPs) the latter must be accompanied by appropriate boundary conditions, namely

$$(\mathscr{B}w)(\boldsymbol{\tau}) = \psi(\boldsymbol{\tau}), \quad \boldsymbol{\tau} \in \partial\Omega. \tag{2}$$

In above setting everything is known except $w(\boldsymbol{\tau})$. For a variety of reasons, closed form solutions to BVP (1) and (2) in its present form are not feasible. The operator $\mathscr{L}(\boldsymbol{\tau})$ may be linear but of high dimensions or even worst, non-linear. On the other hand, for complex boundaries even simple linear operators do not possess eigenfunctions.

We approach BVP (1) and (2) by substituting Ω and its boundary by an auxiliary domain D and boundary ∂D such that (1) and (2) can be solved in terms of the corresponding eigenfunctions of the operator in question, which we consider to be linear for simplicity. In order to proceed, a connection between boundaries has to established. To this end, any point $\boldsymbol{\tau}$ lying on $\partial\Omega$ can be written as the sum of the position vector $\mathbf{r}$ plus their difference, measuring the deviation between the so-

called perturbed boundary $\partial\Omega$ and the corresponding unperturbed ∂D. Denoting by $f(\hat{\mathbf{r}})$ the aforementioned difference and introducing a scaling factor ϵ, we have that

$$\hat{\boldsymbol{\tau}} = \hat{\mathbf{r}} + \epsilon f(\hat{\mathbf{r}}). \tag{3}$$

Note that for ϵ to be dimensionless, unit vectors have to be introduced. The initial BVP now reads as

$$(\mathscr{L}w)(\mathbf{r},\epsilon) = \phi(\mathbf{r},\epsilon), \quad \mathbf{r} \in D, \tag{4}$$

$$(\mathscr{B}w)(\mathbf{r},\epsilon) = \psi(\mathbf{r},\epsilon), \quad \mathbf{r} \in \partial D, \tag{5}$$

from which it is clear that any perturbation of the boundary is transferred onto terms involving powers of ϵ. In general, perturbations affect the functions $w(\boldsymbol{\tau})$, $\phi(\boldsymbol{\tau})$, $\psi(\boldsymbol{\tau})$ involved, as well as the operators $\mathscr{L}$ and $\mathscr{B}$. However, if the operator $\mathscr{L}$ describes the physical properties of the problem it remains unaltered. The sought perturbation terms are evaluated by expanding involved quantities in powers of ϵ, given that the perturbation parameter ϵ is considered small and the coefficients of the expansion are independent of ϵ. These coefficients, on the other hand, are evaluated solving specific expressions, produced after substituting the expansions into (1), (2) and collecting coefficients of ϵ.

For example, replacing the Poincaré type expansion $h(\mathbf{r},\epsilon) = \sum_{n=0}^{\infty} \epsilon^n h_n(\mathbf{r})$, where h represents any of the functions of interest, into (1) and (2), respectively, gives

$$\sum_{n=0}^{\infty} \epsilon^n \mathscr{L} w_n(\mathbf{r}) = \sum_{n=0}^{\infty} \epsilon^n \phi_n(\mathbf{r}), \quad \mathbf{r} \in D \tag{6}$$

and

$$\sum_{n=0}^{\infty} \epsilon^n \sum_{i=0}^{n} (\mathscr{B}_n w_{n-i})(\mathbf{r}) = \sum_{n=0}^{\infty} \epsilon^n \psi_n(\mathbf{r}), \quad \mathbf{r} \in \partial D, \tag{7}$$

respectively, the latter indicating the Cauchy product of $\left(\sum_{n=0}^{\infty} \epsilon^n \mathscr{B}_n(\mathbf{r})\right)\left(\sum_{n=0}^{\infty} \epsilon^n w_n(\mathbf{r})\right)$.

Collecting now coefficients of ϵ, furnishes the BVPs which have to be solved, e.g.

$$(\mathscr{L}w_0)(\mathbf{r}) = \phi_0(\mathbf{r}), \quad \mathbf{r} \in D, \qquad (\mathscr{B}_0 w_0)(\mathbf{r}) = \psi_0(\mathbf{r}), \quad \mathbf{r} \in \partial D, \tag{8}$$

$$(\mathscr{L}w_1)(\mathbf{r}) = \phi_1(\mathbf{r}), \quad \mathbf{r} \in D, \qquad (\mathscr{B}_0 w_1 + \mathscr{B}_1 w_0)(\mathbf{r}) = \psi_1(\mathbf{r}), \quad \mathbf{r} \in \partial D, \tag{9}$$

$$(\mathscr{L}w_2)(\mathbf{r}) = \phi_2(\mathbf{r}), \quad \mathbf{r} \in D, \quad (\mathscr{B}_0 w_2 + \mathscr{B}_1 w_1 + \mathscr{B}_2 w_0)(\mathbf{r}) = \psi_2(\mathbf{r}), \quad \mathbf{r} \in \partial D, \tag{10}$$

$\ldots$ $\ldots$

providing, the unperturbed instance ($n = 0$) (8), the first (9) and second (10) correction and so on. It is evident that for each correction, the solution to the previous one must be obtained in order to proceed. Further details can be found in [23, 24].

2 Mathematical Formulation of Medical Monitoring Techniques. Electroencephalography and Magnetoencephalography

Consider a homogeneous conductor modeling the brain denoted by $\Omega \subset \mathbb{R}^3$ and let $\partial\Omega$ be its boundary. Let the exterior, not conductive, to the brain region denoted by Ω^c. The activity of the brain, represented by a neuronal current $\mathbf{J}^p$, can be recorded either as electrical impulses along the surface (EEG), or measuring the magnetic fields generated by $\mathbf{J}^p$, the neuronal current a few centimeters above the surface (MEG). The basic assumption is that neuronal current $\mathbf{J}^p$ is represented either as a discrete or as a continuous distribution of dipoles with specified moments $\mathbf{Q}$.

In 1967, Plonsey and Heppner [13] illustrated that the electromagnetic activity of the brain is governed by the quasi-static theory of Maxwell's equations, namely

$$\nabla \times \mathbf{E} = 0, \tag{11}$$

$$\nabla \times \mathbf{B} = \mu_0(\mathbf{J}^p + \sigma\, \mathbf{E}), \tag{12}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{13}$$

where $\mathbf{E}$ and $\mathbf{B}$ denote the electric and magnetic field, respectively, corresponding to the neuronal current $\mathbf{J}^p$, σ is the conductivity of the medium occupying Ω and μ_0 is the magnetic permeability, assumed constant everywhere in $\mathbb{R}^3$ and equal to the magnetic permeability of the free space. After some straightforward calculations (for details see [25]) one can show that, if the medium is homogeneous, the electric potential U, introduced as $\mathbf{E} = -\nabla U$, is related to the primary current $\mathbf{J}^p$ by Poisson's equation

$$\Delta U = \frac{1}{\sigma} \nabla \cdot \mathbf{J}^p. \tag{14}$$

Similar, in the current free, non-conductive domain Ω^c the magnetic field, due to (13), can be represented as a gradient of a harmonic function, namely

$$\mathbf{B}(\boldsymbol{\tau}, \mathbf{r}_0) = \frac{\mu_0}{4\pi} \nabla\, W(\boldsymbol{\tau}, \mathbf{r}_0), \tag{15}$$

where $W(\boldsymbol{\tau}, \mathbf{r}_0)$ denotes the scalar magnetic potential, vanishing at infinity.

Another way is connected to Geselowitz's integral formula [26], which is geometry independent, but relies on knowledge of the electrical potential U

$$\mathbf{B}(\boldsymbol{\tau};\mathbf{r}_0)=\frac{\mu_0}{4\pi}\,\mathbf{Q}(\mathbf{r}_0)\times\frac{\boldsymbol{\tau}-\mathbf{r}_0}{|\boldsymbol{\tau}-\mathbf{r}_0|^3}-\frac{\mu_0\sigma}{4\pi}\int_{\Omega}\nabla_{\mathbf{r}'}U(\mathbf{r}';\mathbf{r}_0)\times\frac{\boldsymbol{\tau}-\mathbf{r}'}{|\boldsymbol{\tau}-\mathbf{r}'|^3}\mathrm{d}\upsilon(\mathbf{r}'). \tag{16}$$

Since

$$\nabla_{\mathbf{r}_0}\frac{1}{|\boldsymbol{\tau}-\mathbf{r}_0|}=-\nabla_{\boldsymbol{\tau}}\frac{1}{|\boldsymbol{\tau}-\mathbf{r}_0|}=\frac{\boldsymbol{\tau}-\mathbf{r}_0}{|\boldsymbol{\tau}-\mathbf{r}_0|^3}, \tag{17}$$

(16) becomes

$$\mathbf{B}(\boldsymbol{\tau};\mathbf{r}_0)=\frac{\mu_0}{4\pi}\,\mathbf{Q}(\mathbf{r}_0)\times\nabla_{\mathbf{r}_0}\frac{1}{|\boldsymbol{\tau}-\mathbf{r}_0|}-\frac{\mu_0\sigma}{4\pi}\int_{\Omega}\nabla_{\mathbf{r}'}U(\mathbf{r}';\mathbf{r}_0)\times\nabla_{\mathbf{r}'}\frac{1}{|\boldsymbol{\tau}-\mathbf{r}'|}\mathrm{d}\upsilon(\mathbf{r}'). \tag{18}$$

In view of the identity

$$\int_{\Omega}(\nabla f)\times(\nabla g)\,\mathrm{d}\upsilon=\int_{\Omega}\nabla\times(f\,\nabla g)\,\mathrm{d}\upsilon=\oint_{\partial\Omega}\hat{\boldsymbol{\nu}}\times f\,\nabla g\,\mathrm{d}S, \tag{19}$$

the integral on the RHS of equation (18) simplifies as

$$\int_{\Omega}\nabla_{\mathbf{r}'}U(\mathbf{r}';\mathbf{r}_0)\times\nabla_{\mathbf{r}'}\frac{1}{|\boldsymbol{\tau}-\mathbf{r}'|}\mathrm{d}\upsilon(\mathbf{r}')=\int_{\Omega}\nabla_{\mathbf{r}'}\times\left(U(\mathbf{r}';\mathbf{r}_0)\nabla_{\mathbf{r}'}\frac{1}{|\boldsymbol{\tau}-\mathbf{r}'|}\right)\mathrm{d}\upsilon(\mathbf{r}') \tag{20}$$

$$=\oint_{\partial\Omega}U(\mathbf{r}';\mathbf{r}_0)\hat{\boldsymbol{\nu}}\times\frac{\boldsymbol{\tau}-\mathbf{r}'}{|\boldsymbol{\tau}-\mathbf{r}'|^3}\mathrm{d}S(\hat{\mathbf{r}}'). \tag{21}$$

2.1 The Influence of Geometric Variations on the Forward Problem

2.1.1 EEG

Adopt that the neuronal current is represented by a single equivalent dipole at the point $\mathbf{r}_0$ with moment $\mathbf{Q}$, which is sufficient when modeling smaller cortical neuronal sources [27]. Then, the primary current can be represented as $\mathbf{J}^p = \mathbf{Q}\delta(\boldsymbol{\tau}-\mathbf{r}_0)$, $\delta(\mathbf{r})$ denoting the Dirac measure. Equation (14) then simplifies as

$$\Delta_{\boldsymbol{\tau}}\, U(\boldsymbol{\tau}, \mathbf{r}_0) = \frac{1}{\sigma}\mathbf{Q}(\mathbf{r}_0)\cdot\nabla_{\boldsymbol{\tau}}\,\delta(\boldsymbol{\tau}-\mathbf{r}_0), \quad \boldsymbol{\tau}\in\Omega, \tag{22}$$

accompanied by the condition

$$\hat{\boldsymbol{\nu}}\cdot\nabla_{\boldsymbol{\tau}}\, U(\boldsymbol{\tau}, \mathbf{r}_0) = 0, \quad \boldsymbol{\tau}\in\partial\Omega, \tag{23}$$

where $\hat{\boldsymbol{\nu}}$ denotes the unit vector normal to the boundary. The solution to BVP (22) and (23) is

$$U(\boldsymbol{\tau};\mathbf{r}_0) = \frac{1}{4\pi\sigma}\left(\mathbf{Q}\cdot\nabla_{\mathbf{r}_0}\right)\left(\frac{1}{|\boldsymbol{\tau}-\mathbf{r}_0|}+u(\boldsymbol{\tau};\mathbf{r}_0)\right), \tag{24}$$

where $-\frac{1}{4\pi}|\boldsymbol{\tau}-\mathbf{r}_0|^{-1}$ is the fundamental solution and $u(\boldsymbol{\tau};\mathbf{r}_0)$ is known as we will see shortly.

Due to the arbitrary shape of the boundary, arriving at analytic solutions for the problems (22) and (23) is in general not feasible. Notwithstanding, introducing an auxiliary boundary ∂D say, in the shape of a sphere of radius R, it is possible to transform (22) and (23) to approximate problems with reference to the specific symmetrical case. Following the steps outlined in Section 1.1, we connect the boundaries as

$$\boldsymbol{\tau} = R\hat{\mathbf{r}} + \epsilon\, f(\hat{\mathbf{r}}) \tag{25}$$

with $f(\hat{\mathbf{r}})$ explicitly carrying units of [L]. The corresponding electric potentials are then straightforwardly computed by replacing

$$U(\boldsymbol{\tau}) = \sum_{n=0}^{\infty}\epsilon^n\, U_n(\mathbf{r})\,, \tag{26}$$

into expression (22). After collecting coefficients of ϵ we obtain

$$\sigma\,\Delta_{\mathbf{r}}\, U_0(\mathbf{r},\mathbf{r}_0) = \mathbf{Q}\cdot\nabla_{\mathbf{r}}\,\delta(\mathbf{r}-\mathbf{r}_0), \quad \mathbf{r}\in D, \tag{27}$$

$$\Delta_{\mathbf{r}}\, U_n(\mathbf{r},\mathbf{r}_0) = 0, \quad n\geq 1 \quad \mathbf{r}\in D. \tag{28}$$

In order to identify the implication of the deformed boundary on the solution process, expansion (26) is assumed to be valid all the way to the boundary. The Neumann condition (23) now reads

$$\tau^2\,\frac{\partial U(\boldsymbol{\tau})}{\partial\tau} - \frac{\partial\tau}{\partial\theta}\,\frac{\partial U(\boldsymbol{\tau})}{\partial\theta} - \frac{1}{\sin^2\theta}\,\frac{\partial\tau}{\partial\phi}\,\frac{\partial U(\boldsymbol{\tau})}{\partial\phi} = 0, \tag{29}$$

combined with expressions (25), (26) and bearing in mind that $\partial_r\, f = 0$, gives rise to the system

$$\hat{\mathbf{r}} \cdot \nabla U_n(R\hat{\mathbf{r}}, \mathbf{r}_0) = \sum_{k=0}^{n-1} (-1)^k (k+1) \left(\frac{f}{R}\right)^k \Big(\nabla f \cdot \nabla U_{n-k-1}(R\hat{\mathbf{r}}, \mathbf{r}_0) \Big), \tag{30}$$

the latter holding for every n, where by convention $\sum_{n=0}^{-1} = 0$.

The electric potential of a deformed spherical conductor due to a single dipole excitation can be presented in terms of spherical harmonics as

$$U(\boldsymbol{\tau}, \mathbf{r}_0) = \sum_{k=1}^{\infty} \sum_{\ell=-k}^{k} \mathbb{B}_k^{\ell}(\tau, \mathbf{r}_0, \mathbf{Q})\, \mathrm{Y}_k^{\ell}(\hat{\boldsymbol{\tau}}), \quad \boldsymbol{\tau} \in \Omega. \tag{31}$$

In a similar fashion, each nth-order correction specified by relation (26) is given by

$$U_n(\mathbf{r}, \mathbf{r}_0) = \sum_{k=1}^{\infty} \sum_{\ell=-k}^{k} \mathbb{A}_{n,k}^{\ell}(r, \mathbf{r}_0, \mathbf{Q})\, \mathrm{Y}_k^{\ell}(\hat{\mathbf{r}}), \quad \mathbf{r} \in D. \tag{32}$$

Combining relations (26) and (32) yields

$$\mathbb{B}_k^{\ell}(r, \mathbf{r}_0, \mathbf{Q}; \epsilon) = \sum_{n=0}^{\infty} \epsilon^n \mathbb{A}_{n,k}^{\ell}(r, \mathbf{r}_0, \mathbf{Q})\,. \tag{33}$$

The electric potential in regard to the spherical conductor $D \subset \mathbb{R}^3$ is then provided via the zeroth-order correction (32), a solution to (27) together with $\partial_r U_0(R\hat{\mathbf{r}}, \mathbf{r}_0) = 0$. It is not hard to show that on the surface of the non-deformed conductor we have

$$\mathbb{A}_{0,k}^{\ell}(R, \mathbf{r}_0, \mathbf{Q}) = \frac{\mathbf{Q}}{k\sigma R^{k+1}} \cdot \nabla_{\mathbf{r}_0} r_0^k\, \overline{\mathrm{Y}}_k^{\ell}(\hat{\mathbf{r}}_0), \tag{34}$$

where Y represent the spherical harmonics and an over-line denotes complex conjugation and we made use of the fact that in spherical coordinates the function U in (24) enjoys the expansion

$$U(\mathbf{r}; \mathbf{r}_0) = 4\pi \sum_{n=1}^{\infty} \sum_{m=-n}^{n} \frac{n+1}{n(2n+1)} \frac{r^n r_0^n}{R^{2n+1}} \mathrm{Y}_n^m(\hat{\mathbf{r}}) \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}_0). \tag{35}$$

Conveniently, in conjecture with the generating function

$$\frac{1}{|\mathbf{r} - \mathbf{r}_0|} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{4\pi}{2n+1} \frac{r_0^n}{r^{n+1}} \mathrm{Y}_n^m(\hat{\mathbf{r}}) \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}_0), \tag{36}$$

it is possible to express the surface potential U_0 in closed form (see [28] for details) as

$$U_0(R\hat{\mathbf{r}}, \mathbf{r}_0) = \frac{\mathbf{Q}}{4\pi\sigma} \cdot \left(2\frac{R\hat{\mathbf{r}} - \mathbf{r}_0}{|R\hat{\mathbf{r}} - \mathbf{r}_0|^3} + \frac{1}{R|R\hat{\mathbf{r}} - \mathbf{r}_0|} \frac{|R\hat{\mathbf{r}} - \mathbf{r}_0|\hat{\mathbf{r}} + (R\hat{\mathbf{r}} - \mathbf{r}_0)}{|R\hat{\mathbf{r}} - \mathbf{r}_0| + \hat{\mathbf{r}} \cdot (R\hat{\mathbf{r}} - \mathbf{r}_0)} \right). \tag{37}$$

On the other hand, the first-order correction U_1 is derived as the solution to (28) for $n = 1$ joint by (30), namely

$$R^2 \frac{\partial U_1(R\hat{\mathbf{r}}, \mathbf{r}_0)}{\partial r} = \frac{\partial f(\hat{\mathbf{r}})}{\partial \theta} \frac{\partial U_0(R\hat{\mathbf{r}}, \mathbf{r}_0)}{\partial \theta} + \frac{1}{\sin^2\theta} \frac{\partial f(\hat{\mathbf{r}})}{\partial \phi} \frac{\partial U_0(R\hat{\mathbf{r}}, \mathbf{r}_0)}{\partial \phi}. \tag{38}$$

Replacing into the latter the expressions for U_0 and U_1 through (32) and integrating the resulting relation over the unit ball, determines the coefficients $\mathbb{A}^{\ell}_{1,k}(R, \mathbf{r}_0, \mathbf{Q})$ as

$$\mathbb{A}^{\ell}_{1,k} = \frac{1}{kR} \sum_{n=1}^{\infty} \sum_{m=-n}^{n} \mathbb{A}^{m}_{0,n} \left\{ \oint_{S^2} \left[\frac{1}{\sin\theta} \frac{\partial f}{\partial \theta} \left(n\, j^m_{n+1}\, \mathrm{Y}^m_{n+1}(\hat{\mathbf{r}}) - (n+1)\, j^m_n\, \mathrm{Y}^m_{n-1}(\hat{\mathbf{r}}) \right) + i \frac{m}{\sin^2\theta} \frac{\partial f}{\partial \phi} \mathrm{Y}^m_n(\hat{\mathbf{r}}) \right] \mathrm{Y}^{\ell}_k(\hat{\mathbf{r}}) \mathrm{d}S(\hat{\mathbf{r}}) \right\}, \tag{39}$$

where the coefficients $\mathbb{A}^m_{0,n}$ are given by (34) and we further employed the recurrence relations

$$\sin\theta \frac{\partial}{\partial \theta} \mathrm{Y}^m_n(\hat{\mathbf{r}}) = n\, j^m_{n+1}\, \mathrm{Y}^m_{n+1}(\hat{\mathbf{r}}) - (n+1)\, j^m_n\, \mathrm{Y}^m_{n-1}(\hat{\mathbf{r}}), \quad j^m_n = \sqrt{\frac{n^2 - m^2}{4n^2 - 1}}, \tag{40}$$

$$\frac{\partial}{\partial \phi} \mathrm{Y}^m_n(\hat{\mathbf{r}}) = im\, \mathrm{Y}^m_n(\hat{\mathbf{r}}). \tag{41}$$

By the fact that the geometry is fixed and parameters corresponding to the geometry of choice (here the radius R) and medium (here the conductivity σ) are known, the surface potential U_0 is calculated via (37), which leads to knowledge of the coefficients $\mathbb{A}^{\ell}_{0,k}$ through (32) for $n = 0$, since

$$\mathbb{A}^{\ell}_{0,k} = \oint_{S^2} U_0(R\hat{\mathbf{r}}, \mathbf{r}_0)\, \overline{\mathrm{Y}}^{\ell}_k(\hat{\mathbf{r}})\, \mathrm{d}S(\hat{\mathbf{r}}). \tag{42}$$

After an approximate function $f(\hat{\mathbf{r}})$ has been derived capable to fit the perturbed surface, the coefficients $\mathbb{A}^{\ell}_{n,k}$ corresponding to the nth-order correction ($n \geq 1$) are computed as demonstrated above. Subsequently, either the corresponding potentials

$U_n, n \geq 1$ via (32) or the coefficients (33) are estimated, leading to the perturbed surface potential $U(\boldsymbol{\tau})$.

The electric potential due to a dipolar source can be represented by Geselowitz's integral representation [29]

$$U(\mathbf{r}) = \frac{1}{4\pi\sigma}\mathbf{Q}\cdot\frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|^3} - \frac{1}{4\pi}\oint_{S^2} U(\boldsymbol{\tau})\hat{\boldsymbol{\nu}}\cdot\frac{\mathbf{r}-\boldsymbol{\tau}}{|\mathbf{r}-\boldsymbol{\tau}|^3}\mathrm{d}S(\hat{\boldsymbol{\tau}}), \quad \mathbf{r}\notin S, \tag{43}$$

where $\mathbf{r}$ is the point of observation (measurement site). In the case of a spherical conductor of radius R, the integral becomes

$$U(\mathbf{r}) = \frac{1}{4\pi\sigma}\mathbf{Q}\cdot\frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|^3} - \frac{R^2}{4\pi}\oint_{\partial D} U(\mathbf{r}')\frac{\hat{\mathbf{r}}'\cdot\mathbf{r}-R}{|\mathbf{r}-R\hat{\mathbf{r}}'|^3}\mathrm{d}\Omega(\hat{\mathbf{r}}'), \quad \mathbf{r}\notin S, \tag{44}$$

where Ω denotes the solid angle. The latter is a Fredholm integral equation of the second kind and can be solved accordingly. By expanding all relevant expressions in (43) an integral representation for the perturbed instance is obtained leading again to Fredholm equations of the second kind for both the zeroth U_0 and first order correction U_1. This scenario will be presented elsewhere.

2.1.2 MEG

From (13) and (15) it is evident that the scalar magnetic potential W is harmonic and can be expanded as

$$W(\mathbf{r};\mathbf{r}_0) = \sum_{n=1}^{\infty}\sum_{m=-n}^{n}\mathbb{C}_n^m\frac{\mathrm{Y}_n^m(\hat{\mathbf{r}})}{r^{n+1}}, \tag{45}$$

where

$$\mathbb{C}_n^m = \frac{4\pi}{(n+1)(2n+1)}(\mathbf{Q}\times\mathbf{r}_0)\cdot\nabla_{\mathbf{r}_0}\mathbf{r}_0^n\overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}_0), \tag{46}$$

or

$$\frac{\mathbb{C}_n^m}{r^{n+1}} = \oint_{S^2} W(\mathbf{r};\mathbf{r}_0)\overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}_0)\mathrm{d}S(\hat{\mathbf{r}}). \tag{47}$$

Clearly, knowledge of W provides the values of $\mathbb{C}_n^m$ via (47) and thus the solution to the forward MEG problem.

The scalar magnetic potential W introduced in (15) can be computed, using integration along a ray from $\mathbf{r}$ to infinity as

$$W(\mathbf{r};\mathbf{r}_0) = -\frac{4\pi}{\mu_0}\int_r^{+\infty}\hat{\boldsymbol{\tau}}\cdot\mathbf{B}(\boldsymbol{\tau};\mathbf{r}_0)\mathrm{d}\tau, \tag{48}$$

which, combined with expression

$$\hat{\mathbf{r}} \cdot \mathbf{B}(\mathbf{r}; \mathbf{r}_0) = -\frac{\mu_0}{4\pi} \mathbf{Q} \times \frac{\mathbf{r}_0 \cdot \hat{\mathbf{r}}}{|\mathbf{r} - \mathbf{r}_0|^3}, \tag{49}$$

becomes

$$W(\mathbf{r}; \mathbf{r}_0) = \mathbf{Q} \times \mathbf{r}_0 \cdot \hat{\mathbf{r}} \int_{\mathbf{r}}^{+\infty} \frac{\mathrm{d}\tau}{|\tau \hat{\boldsymbol{\tau}} - \mathbf{r}_0|^3}. \tag{50}$$

Performing straightforward calculations (details are provided in [28]) the magnetic potential assumes the form

$$W(\mathbf{r}; \mathbf{r}_0) = \mathbf{Q} \cdot \frac{\hat{\mathbf{r}} \times \hat{\mathbf{P}}}{P(1 + \hat{\mathbf{r}} \cdot \hat{\mathbf{P}})}, \quad \mathbf{P} = \mathbf{r} - \mathbf{r}_0, \tag{51}$$

implying that

$$\mathbf{B}(\mathbf{r}; \mathbf{r}_0) = \frac{\mu_0}{4\pi} \frac{\mathbf{Q} \times (\mathbf{r} - \mathbf{P})}{r P^2 (1 + \hat{\mathbf{r}} \cdot \hat{\mathbf{P}})} - \frac{\mu_0}{4\pi} \frac{U(\mathbf{r}; \mathbf{r}_0)}{P} \left[\hat{\mathbf{P}} + \frac{r - P}{r(1 + \hat{\mathbf{r}} \cdot \hat{\mathbf{P}})} (\hat{\mathbf{r}} - \hat{\mathbf{P}}) \right]. \tag{52}$$

Based on Geselowitz's formula (16), the magnetic field can also be computed as

$$\mathbf{B}(\mathbf{r}; \mathbf{r}_0) = \frac{\mu_0}{4\pi} \mathbf{Q}(\mathbf{r}_0) \times \nabla_{\mathbf{r}_0} \frac{1}{|\mathbf{r} - \mathbf{r}_0|} - \frac{\mu_0}{(4\pi)^2} \left(\mathbf{Q} \cdot \nabla_{\mathbf{r}_0} \right) \mathbf{K}(\mathbf{r}; \mathbf{r}_0), \tag{53}$$

where

$$\mathbf{K}(\mathbf{r}; \mathbf{r}_0) = \oint_{\partial D} \left(\frac{1}{|\mathbf{r}' - \mathbf{r}_0|} + \sum_{n=1}^{\infty} \sum_{m=-n}^{n} 4\pi \frac{n+1}{n(2n+1)} \frac{r^n r_0^n}{R^{2n+1}} \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}) \mathrm{Y}_n^m(\hat{\mathbf{r}}_0) \right) \mathbf{r}' \times \nabla_{\mathbf{r}'} \frac{1}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}s(\mathbf{r}'). \tag{54}$$

It can be shown that the magnetic field $\mathbf{B}$ for the perturbed sphere equals [30]

$$\begin{aligned} \mathbf{B}(\boldsymbol{\tau}; \mathbf{r}_0) &= \sum_{n=0}^{\infty} \epsilon^n \mathbf{B}_n(\mathbf{r}; \mathbf{r}_0) \\ &= \frac{\mu_0}{4\pi} \mathbf{Q} \times \frac{\mathbf{r} - \mathbf{r}_0}{|\mathbf{r} - \mathbf{r}_0|^3} - \frac{\mu_0 \sigma}{4\pi} \oint_{\partial D} \boldsymbol{\alpha}_0(\hat{\mathbf{r}}, \mathbf{r}) \, U_0(R\hat{\mathbf{r}}, \mathbf{r}) \frac{\mathrm{d}\Omega(\hat{\mathbf{r}})}{|\mathbf{r} - R\hat{\mathbf{r}}|^3} \end{aligned}$$

$$-\epsilon\frac{\mu_0\sigma}{4\pi}\oint_{\partial D}\Big(\boldsymbol{\alpha}_0(\hat{\mathbf{r}},\mathbf{r})\,U_1(R\hat{\mathbf{r}},\mathbf{r})+\boldsymbol{\alpha}_1(\hat{\mathbf{r}},\mathbf{r})\,U_0(R\hat{\mathbf{r}},\mathbf{r})\Big)\frac{\mathrm{d}\Omega(\hat{\mathbf{r}})}{|\mathbf{r}-R\hat{\mathbf{r}}|^3}$$
$$+\mathcal{O}(\epsilon^2), \tag{55}$$

given that

$$\boldsymbol{\alpha}_0(\hat{\mathbf{r}},\mathbf{r})=R^2(\hat{\mathbf{r}}\times\mathbf{r}), \tag{56}$$

$$\boldsymbol{\alpha}_1(\hat{\mathbf{r}},\mathbf{r})=\Big[2|\mathbf{r}-R\hat{\mathbf{r}}|^2-3R(R-\mathbf{r}\cdot\hat{\mathbf{r}})\Big]\frac{f(\hat{\mathbf{r}})\boldsymbol{\alpha}_0(\hat{\mathbf{r}},\mathbf{r})}{R|\mathbf{r}-R\hat{\mathbf{r}}|^2}-R\nabla f(\hat{\mathbf{r}})\times(\mathbf{r}-R\hat{\mathbf{r}}). \tag{57}$$

2.2 The Influence of Geometric Variations on the Inversion Algorithm

2.2.1 EEG

In order to pinpoint the position $\mathbf{r}_0=\begin{pmatrix}x_{01}\ x_{02}\ x_{03}\end{pmatrix}^\top$ and moment $\mathbf{Q}=\begin{pmatrix}Q_1\ Q_2\ Q_3\end{pmatrix}^\top$ of the dipole, six equations are required in general. These relations are acquired based on formulas translating the spherical harmonics into Cartesian form (for example [21, pp. 155–156]). In the unperturbed case the position and moment of the equivalent dipole are expressed through (34) for k up to two as

$$\mathbf{r}_0=\frac{R}{\sqrt{5}}\left(\frac{\mathbb{A}^{-2}_{0,2}}{\mathbb{A}^{-1}_{0,1}}-\frac{\mathbb{A}^{2}_{0,2}}{\mathbb{A}^{1}_{0,1}}\ -i\left(\frac{\mathbb{A}^{-2}_{0,2}}{\mathbb{A}^{-1}_{0,1}}+\frac{\mathbb{A}^{2}_{0,2}}{\mathbb{A}^{1}_{0,1}}\right)\frac{1}{\mathbb{A}^{\pm1}_{0,1}}\left(2\mathbb{A}^{\pm1}_{0,2}-\sqrt{2}\frac{\mathbb{A}^{0}_{0,1}\mathbb{A}^{\pm2}_{0,2}}{\mathbb{A}^{\pm1}_{0,1}}\right)\right)^\top \tag{58}$$

and

$$\mathbf{Q}=\sigma\,R^2\sqrt{\frac{\pi}{3}}\left(\sqrt{2}\left(\mathbb{A}^{-1}_{0,1}-\mathbb{A}^{1}_{0,1}\right)\ -i\sqrt{2}\left(\mathbb{A}^{-1}_{0,1}+\mathbb{A}^{1}_{0,1}\right)\ 2\mathbb{A}^{0}_{0,1}\right)^\top, \tag{59}$$

respectively.

In order to analyze the effect of perturbations on the inversion algorithm, our starting point is boundary condition (38). Substituting the matching potentials from (32) and expanding the known function $f(\hat{\mathbf{r}})$ in terms of conjugate spherical harmonics

$$f(\hat{\mathbf{r}})=\sum_{p=0}^{\infty}\sum_{q=-p}^{p}\overline{\mathbb{C}}^{q}_{p}\,\overline{\mathrm{Y}}^{q}_{p}(\hat{\mathbf{r}}) \tag{60}$$

we obtain

$$\sum_{k=1}^{\infty}\sum_{\ell=-k}^{k} k\,R^{k+1}\,\mathbb{A}_{1,k}^{\ell}\,\mathrm{Y}_k^{\ell}(\hat{\mathbf{r}}) = \sum_{k=1}^{\infty}\sum_{\ell=-k}^{k}\sum_{p=0}^{\infty}\sum_{q=-p}^{p}\mathbb{A}_{0,k}^{\ell}\,\overline{\mathbb{C}}_p^q\left(\frac{\partial\overline{\mathrm{Y}}_p^q(\hat{\mathbf{r}})}{\partial\theta}\frac{\partial\mathrm{Y}_k^{\ell}(\hat{\mathbf{r}})}{\partial\theta} + \frac{\ell q}{\sin^2\theta}\overline{\mathrm{Y}}_p^q(\hat{\mathbf{r}})\mathrm{Y}_k^{\ell}(\hat{\mathbf{r}})\right). \tag{61}$$

Integrating above first with respect to $\phi \in (0, 2\pi]$ and then with respect to $\theta \in [0, \pi]$ and having in mind that

$$\int_{-1}^{1}\left((1-x^2)\frac{\mathrm{dP}_n^m(x)}{\mathrm{d}x}\frac{\mathrm{dP}_k^{\ell}(x)}{\mathrm{d}x} + m^2\frac{\mathrm{P}_n^m(x)\,\mathrm{P}_k^{\ell}(x)}{1-x^2}\right)\mathrm{d}x = \frac{2n(n+1)}{2n+1}\,\frac{(n+m)!}{(n-m)!}\,\delta_{n,k}\,, \tag{62}$$

where we make use of the Kronecker symbol $\delta_{n,n} = 1$, else 0, yields

$$\sum_{k=1}^{\infty}\sum_{\ell=-k}^{k}\frac{k(k+1)}{R^{k+1}}\left(\mathbf{Q}\cdot\nabla_{\mathbf{r}_0} r_0^k\,\overline{\mathrm{Y}}_k^{\ell}(\hat{\mathbf{r}}_0)\right)\overline{\mathbb{C}}_k^{\ell} = 0\,, \tag{63}$$

providing a criterion for the solvability of the BVP at hand. A crucial aspect is the fact that due to relation (63), which is an immediate consequence of the compatibility condition regarding the Neumann problem for the first-order correction, a restriction between the surface deformation $f(\hat{\mathbf{r}})$, the dipole's position $\mathbf{r}_0$ and moment $\mathbf{Q}$ is imposed. The fine points will be demonstrated in the Example section.

2.2.2 MEG

As aforementioned, the key in calculating the sought position and moments of the dipole is to translate the spherical harmonics in Cartesian coordinates. For example, in the unperturbed case the coefficients are given by (46) and for $k = 2$ and $\ell = 0$ we have

$$\mathbb{C}_2^0 = \frac{4\pi}{15}\,(\mathbf{Q}\times\mathbf{r}_0)\cdot\nabla_{\mathbf{r}_0}\mathbf{r}_0^2\overline{\mathrm{Y}}_2^0(\hat{\mathbf{r}}_0) = \frac{2\pi}{15}\sqrt{\frac{5}{\pi}}\,(\mathbf{Q}\times\mathbf{r}_0)\cdot\left(-x_{01}\hat{\mathbf{x}}_1 - x_{02}\hat{\mathbf{x}}_2 + 2x_{03}\hat{\mathbf{x}}_3\right). \tag{64}$$

Repeating the procedure and after some algebra, we obtain the following 3×3 system

$$\begin{pmatrix} \mathbb{C}_1^1 - \mathbb{C}_1^{-1} & i(\mathbb{C}_1^1 + \mathbb{C}_1^{-1}) & 2\sqrt{2}\mathbb{C}_1^0 \\ 0 & i\sqrt{2}\mathbb{C}_1^0 & \mathbb{C}_1^1 + \mathbb{C}_1^{-1} \\ -\sqrt{2}\mathbb{C}_1^0 & 0 & \mathbb{C}_1^{-1} - \mathbb{C}_1^1 \end{pmatrix} \begin{pmatrix} x_{01} \\ x_{02} \\ x_{03} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{15}{2}}\mathbb{C}_2^0 \\ \frac{\sqrt{5}}{2}\mathbb{C}_2^{-1} + \mathbb{C}_2^1 \\ \frac{\sqrt{5}}{2}\mathbb{C}_2^{-1} - \mathbb{C}_2^1 \end{pmatrix}, \tag{65}$$

from which we can pinpoint the source $\mathbf{r}_0 = \begin{pmatrix} x_{01} & x_{02} & x_{03} \end{pmatrix}^\top$. On the other hand, using higher-order coefficients, we obtain $\mathbf{r}_0$ in an easier fashion as

$$\mathbf{r}_0 = \frac{\sqrt{5}}{4}\left(\frac{\mathbb{C}_2^{-2}}{\mathbb{C}_1^{-1}} - \frac{\mathbb{C}_2^2}{\mathbb{C}_1^1} - i\left(\frac{\mathbb{C}_2^{-2}}{\mathbb{C}_1^{-1}} + \frac{\mathbb{C}_2^2}{\mathbb{C}_1^1} \right) \sqrt{3}\frac{\mathbb{C}_2^0}{\mathbb{C}_1^0} - \frac{1}{\sqrt{2}}\left(\frac{\mathbb{C}_2^{-2}\mathbb{C}_1^1}{\mathbb{C}_1^{-1}\mathbb{C}_1^0} + \frac{\mathbb{C}_2^2\mathbb{C}_1^{-1}}{\mathbb{C}_1^1\mathbb{C}_1^0} \right) \right)^\top. \tag{66}$$

Clearly, equating the relations for x_{0i}, $i = 1, 2, 3$, uniqueness conditions emerge as we will see shortly. Similar, by properly rearranging coefficients and by the fact that

$$\mathbf{Q} \times \mathbf{r}_0 = \hat{\mathbf{i}}(x_{03}Q_2 - x_{02}Q_3) - \hat{\mathbf{j}}(x_{03}Q_1 - x_{01}Q_3) + \hat{\mathbf{k}}(x_{02}Q_1 - x_{01}Q_2), \tag{67}$$

we find for the moments

$$\begin{pmatrix} 0 & x_{03} & -x_{02} \\ -x_{03} & 0 & x_{01} \\ x_{02} & -x_{01} & 0 \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{3}{2\pi}}\left(\mathbb{C}_1^{-1} - \mathbb{C}_1^1 \right) \\ -i\sqrt{\frac{3}{2\pi}}\left(\mathbb{C}_1^{-1} + \mathbb{C}_1^1 \right) \\ \sqrt{\frac{3}{\pi}}\mathbb{C}_1^0 \end{pmatrix}. \tag{68}$$

However, the determinant of the latter vanishes and thus more information is needed.

Before we proceed calculating the moments of the dipole, we note that if $\mathbb{C}_1^0 = 0$ and one of $\mathbb{C}_1^{\pm 1} = 0$, results in $\mathbf{Q} \times \mathbf{r}_0 = \mathbf{0}$. In the presence of neuronal activity the latter implies $\mathbf{Q}$ is parallel to $\mathbf{r}_0$, i.e. the moment $\mathbf{Q}$ is always radial in the spherical coordinate system. This leads to that all radial dipoles inside a homogeneous sphere are not visible by MEG and localization of the dipole $\mathbf{r}_0$, $\mathbf{Q}$ is restrained finding the position $\mathbf{r}_0$ as well as the tangential components Q_{θ_0} and Q_{ϕ_0} as $\mathbf{Q} = Q_{\mathbf{r}_0}\hat{\mathbf{r}}_0 + Q_{\theta_0}\hat{\boldsymbol{\theta}}_0 + Q_{\boldsymbol{\phi}_0}\hat{\boldsymbol{\phi}}_0$. Following the same procedure, we have

$$Q_{\boldsymbol{\theta}_0} = \sqrt{\frac{3}{2\pi}}\frac{\mathbb{C}_1^{-1} - \mathbb{C}_1^1}{r_0 \sin\phi_0} + \sqrt{\frac{3}{\pi}}\frac{\mathbb{C}_1^0}{r_0 \tan\theta_0 \tan\phi_0}, \tag{69}$$

$$Q_{\boldsymbol{\phi}_0} = -\sqrt{\frac{3}{\pi}}\frac{\mathbb{C}_1^0}{r_0 \sin\theta_0}. \tag{70}$$

Moreover, employing a different combination of coefficients would lead, for example, to

$$Q_{\boldsymbol{\theta}_0} = -\sqrt{\frac{3}{\pi}}\frac{\tan\phi_0}{r_0 \tan\theta_0}\mathbb{C}_1^0 - i\sqrt{\frac{3}{2\pi}}\frac{1}{r_0 \cos\phi_0}(\mathbb{C}_1^{-1} + \mathbb{C}_1^1). \tag{71}$$

Evidently, above relations regarding $Q_{\boldsymbol{\theta}_0}$ must coincide and by equating them we obtain

$$\frac{i}{\sqrt{2}}\left(e^{-i\phi_0}\mathbb{C}_1^{-1}+e^{i\phi_0}\mathbb{C}_1^{1}\right)=-\frac{1}{\tan\theta_0}\mathbb{C}_1^{0}. \tag{72}$$

namely the condition for $Q_{\boldsymbol{\theta}_0}$ and $Q_{\boldsymbol{\phi}_0}$ to be unique and is consistent with the system (68). Additional details can be found in [31].

In the perturbed case the scalar magnetic potential can be express as

$$W(\boldsymbol{\tau})=\sum_{n=0}^{\infty}\epsilon^n W_n(\mathbf{r}) \tag{73}$$

and expanding both $W(\boldsymbol{\tau})$ and $W_n(\mathbf{r})$ as in (45), gives

$$\mathbb{D}_k^{\ell}(\boldsymbol{\tau},\mathbf{r}_0,\mathbf{Q};\epsilon)=\sum_{n=0}^{\infty}\epsilon^n\mathbb{E}_{n,k}^{\ell}(\mathbf{r},\mathbf{r}_0,\mathbf{Q}),\quad \mathbb{E}_{0,k}^{\ell}(\mathbf{r},\mathbf{r}_0,\mathbf{Q})=\mathbb{C}_k^{\ell}(\mathbf{r},\mathbf{r}_0,\mathbf{Q}). \tag{74}$$

The coefficients $\mathbb{E}_{n,k}^{\ell}$ are computed with the aid of (48) by properly expanding the inner product $\hat{\mathbf{r}}'\cdot\mathbf{B}(\hat{\mathbf{r}}',\mathbf{r}_0)$, yielding

$$\begin{aligned}W(\mathbf{r},\mathbf{r}_0;\epsilon)=&\int_r^{+\infty}\frac{\mathbf{Q}\times\mathbf{r}_0\cdot\hat{\mathbf{r}}'}{|\mathbf{r}'-\mathbf{r}_0|^3}\mathrm{d}r'\\&+\epsilon\sigma R^2\int_r^{+\infty}\left[\oint_{\partial D}U_1(\mathbf{v})\frac{\hat{\mathbf{r}}'\cdot\nabla f\times\hat{\mathbf{v}}}{|\mathbf{r}'-R\hat{\mathbf{v}}|^3}\mathrm{d}\Omega(\hat{\mathbf{v}})\right]\mathrm{d}r'.\end{aligned} \tag{75}$$

By the fact that

$$\nabla_{\mathbf{v}}\frac{1}{|r'\hat{\mathbf{r}}'-\mathbf{v}|}=\frac{r'\hat{\mathbf{r}}'-\mathbf{v}}{|r'\hat{\mathbf{r}}'-\mathbf{v}|^3},\quad \mathbf{v}=v\hat{\mathbf{v}} \tag{76}$$

and

$$\int_r^{\infty}\frac{1}{|r'\hat{\mathbf{r}}'-\mathbf{v}|}\frac{\mathrm{d}r'}{r'}=\sum_{k=0}^{\infty}\sum_{\ell=-k}^{k}\frac{4}{(k+1)(2k+1)}\frac{v^k}{r^{k+1}}\mathrm{Y}_k^{\ell}(\hat{\mathbf{r}}')\overline{\mathrm{Y}}_k^{\ell}(\hat{\mathbf{v}}), \tag{77}$$

the first order coefficients, in view of (40) and (41), are

$$\begin{aligned}\mathbb{E}_{1,k}^{\ell}=&\frac{4\pi\sigma R^{k+1}}{(k+1)(2k+1)}\sum_{p=1}^{\infty}\sum_{q=-p}^{p}\mathbb{B}_p^q\oint_{\partial D}\left[\frac{\partial f}{\partial\phi}\frac{1}{\sin\theta}\left(k\,j_{k+1}^{\ell}\,\overline{\mathrm{Y}}_{k+1}^{\ell}(\hat{\mathbf{r}})\mathrm{Y}_p^q(\hat{\mathbf{r}})\right.\right.\\&\left.\left.-(k+1)\,j_k^{\ell}\,\overline{\mathrm{Y}}_{k-1}^{\ell}(\hat{\mathbf{r}})\mathrm{Y}_p^q(\hat{\mathbf{r}})\right)+ik\frac{\partial f}{\partial\theta}\overline{\mathrm{Y}}_k^{\ell}(\hat{\mathbf{r}})\mathrm{Y}_p^q(\hat{\mathbf{r}})\right]\mathrm{d}\theta\mathrm{d}\phi\end{aligned} \tag{78}$$

where $\mathbb{B}_p^q$ are the coefficients corresponding to the surface electric potential and are given by (33). The coefficients $\mathbb{E}_{0,k}^{\ell}$ for the unperturbed case are given by (47).

The analysis so far holds for a single dipole located at $\mathbf{r}_0$. Following Fokas [32], let us assume that the neuronal activity is distributed inside the spherical brain as $\mathbf{J}^p$ at $\mathbf{r}_0$. This instance is analyzed replacing $\mathbf{Q}(\mathbf{r}_0)$ by $\int_D \mathbf{J}^p(\mathbf{r}_0)\mathrm{d}\upsilon(\mathbf{r}_0)$ in Geselowitz's formula (53), i.e.

$$\frac{4\pi}{\mu_0}\mathbf{B}(\mathbf{r};\mathbf{r}_0)=\int_D \mathbf{J}^p(\mathbf{r}_0)\times\nabla_{\mathbf{r}_0}\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0)-\frac{1}{4\pi}\int_D \mathbf{J}^p(\mathbf{r}_0)\cdot\nabla_{\mathbf{r}_0}\mathbf{K}(\mathbf{r};\mathbf{r}_0)\mathrm{d}\upsilon(\mathbf{r}_0). \tag{79}$$

Although the analysis is presented in [32], we will reproduce part of it providing all necessary details in order to facilitate the transition to the perturbed case. Utilizing the identities $\nabla\times(f\mathbf{g})=\nabla f\times\mathbf{g}+f\nabla\times\mathbf{g}$ and $\nabla\cdot(\mathbf{f}\otimes\mathbf{g})=(\nabla\cdot\mathbf{f})\otimes\mathbf{g}+\mathbf{f}\cdot\nabla\otimes\mathbf{g}$, $\otimes$ denoting the tensor product, gives

$$\mathbf{J}^p(\mathbf{r}_0)\times\nabla_{\mathbf{r}_0}\frac{1}{|\mathbf{r}-\mathbf{r}_0|}=\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\nabla_{\mathbf{r}_0}\times\mathbf{J}^p(\mathbf{r}_0)-\nabla_{\mathbf{r}_0}\times\frac{\mathbf{J}^p(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|} \tag{80}$$

and

$$\nabla_{\mathbf{r}_0}\cdot(\mathbf{J}^p(\mathbf{r}_0)\otimes\mathbf{K}(\mathbf{r};\mathbf{r}_0))=(\nabla_{\mathbf{r}_0}\cdot\mathbf{J}^p(\mathbf{r}_0))\otimes\mathbf{K}(\mathbf{r};\mathbf{r}_0)+\mathbf{J}^p(\mathbf{r}_0)\cdot\nabla_{\mathbf{r}_0}\otimes\mathbf{K}(\mathbf{r};\mathbf{r}_0). \tag{81}$$

Replacing (80) and (81) into (79) yields

$$\int_D \mathbf{J}^p(\mathbf{r}_0)\times\nabla_{\mathbf{r}_0}\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0)=\int_D\left(\nabla_{\mathbf{r}_0}\times\mathbf{J}^p(\mathbf{r}_0)\right)\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0), \tag{82}$$

since, according to (19)

$$\int_D\nabla_{\mathbf{r}_0}\times\left(\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathbf{J}^p(\mathbf{r}_0)\right)\mathrm{d}\upsilon(\mathbf{r}_0)=\int_{\partial D}\hat{\mathbf{n}}\times\left(\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathbf{J}^p(\mathbf{r}_0)\right)\mathrm{d}s(\mathbf{r}_0)=0, \tag{83}$$

which vanishes bearing in mind that $\mathbf{J}^p(\mathbf{r}_0)=\mathbf{0}$ on the boundary ∂D. Similar,

$$\begin{aligned}\int_D \mathbf{J}^p(\mathbf{r}_0)\cdot\nabla_{\mathbf{r}_0}\mathbf{K}(\mathbf{r};\mathbf{r}_0)\mathrm{d}\upsilon(\mathbf{r}_0)&=\int_D\nabla_{\mathbf{r}_0}\cdot\left(\mathbf{J}^p(\mathbf{r}_0)\mathbf{K}(\mathbf{r};\mathbf{r}_0)\right)\mathrm{d}\upsilon(\mathbf{r}_0)\\&\quad-\int_D\left(\nabla_{\mathbf{r}_0}\cdot\mathbf{J}^p(\mathbf{r}_0)\right)\mathbf{K}(\mathbf{r};\mathbf{r}_0)\mathrm{d}\upsilon(\mathbf{r}_0).\end{aligned} \tag{84}$$

Applying the divergence theorem $\int_D\nabla\otimes\mathbf{f}\mathrm{d}\upsilon=\int_{\partial D}\mathbf{f}\otimes\hat{\boldsymbol{\nu}}\mathrm{d}S$, we get

$$\int_D\nabla_{\mathbf{r}_0}\cdot\left(\mathbf{J}^p(\mathbf{r}_0)\mathbf{K}(\mathbf{r};\mathbf{r}_0)\right)\mathrm{d}\upsilon(\mathbf{r}_0)=\int_{\partial D}\hat{\boldsymbol{\nu}}\cdot\left(\mathbf{J}^p(\mathbf{r}_0)\mathbf{K}(\mathbf{r};\mathbf{r}_0)\right)\mathrm{d}s(\mathbf{r}_0). \tag{85}$$

Therefore, (79) reads as

$$\frac{4\pi}{\mu_0}\mathbf{B}(\mathbf{r}) = \int_D \left(\nabla_{\mathbf{r}_0} \times \mathbf{J}^p(\mathbf{r}_0)\right) \frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|} + \frac{1}{4\pi}\int_D \left(\nabla_{\mathbf{r}_0} \cdot \mathbf{J}^p(\mathbf{r}_0)\right) \mathbf{K}(\mathbf{r};\mathbf{r}_0)\mathrm{d}\upsilon(\mathbf{r}_0). \tag{86}$$

The fact that the conductor D is star shaped and letting the neuronal current $\mathbf{J}^p$ be $\mathscr{C}^1$ allows for $\mathbf{J}^p$ to be represented by Helmholtz's decomposition, namely

$$\mathbf{J}^p(\mathbf{r}_0) = \nabla\Psi(\mathbf{r}_0) + \nabla \times \mathbf{A}(\mathbf{r}_0), \quad \nabla \cdot \mathbf{A}(\mathbf{r}_0) = 0, \quad \mathbf{r}_0 \in D \tag{87}$$

Above in mind, we have $\nabla_{\mathbf{r}_0} \times \mathbf{J}^p(\mathbf{r}_0) = -\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)$ and $\nabla_{\mathbf{r}_0} \cdot \mathbf{J}^p(\mathbf{r}_0) = \Delta_{\mathbf{r}_0}\Psi(\mathbf{r}_0)$ and equation (86) becomes

$$\frac{4\pi}{\mu_0}\mathbf{B}(\mathbf{r}) = -\int_D \left(\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\right) \frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|} + \frac{1}{4\pi}\int_D \left(\Delta_{\mathbf{r}_0}\Psi(\mathbf{r}_0)\right) \mathbf{K}(\mathbf{r};\mathbf{r}_0)\mathrm{d}\upsilon(\mathbf{r}_0). \tag{88}$$

Now, the first integral of (88) writes as

$$\mathbf{r} \cdot \int_D \Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|} = \int_D \Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0) \cdot \frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0) + \int_D \Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0) \cdot \frac{\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0), \tag{89}$$

where

$$\int_D \Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0) \cdot \frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0) = \int_D |\mathbf{r}-\mathbf{r}_0|^2 \left(\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0) \cdot \nabla_{\mathbf{r}_0}\frac{1}{|\mathbf{r}-\mathbf{r}_0^3|}\right)\mathrm{d}\upsilon(\mathbf{r}_0) \tag{90}$$

and

$$\begin{aligned}\int_D \nabla_{\mathbf{r}_0} \cdot \left[\left(|\mathbf{r}-\mathbf{r}_0|^2\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\right)\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\right]\mathrm{d}\upsilon(\mathbf{r}_0) = \\ \int_D \left[\nabla_{\mathbf{r}_0} \cdot \left(|\mathbf{r}-\mathbf{r}_0|^2\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\right)\right]\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0) + \int_D \left(|\mathbf{r}-\mathbf{r}_0|^2\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\right) \\ \cdot \nabla_{\mathbf{r}_0}\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0)\end{aligned} \tag{91}$$

so that

$$\begin{aligned}\int_D \Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0) \cdot \frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0) = \oint_{\partial D} \hat{\boldsymbol{\nu}}(\mathbf{r}_0) \cdot |\mathbf{r}-\mathbf{r}_0|\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\mathrm{d}s(\mathbf{r}_0) \\ - \int_D \left[\nabla_{\mathbf{r}_0} \cdot \left(|\mathbf{r}-\mathbf{r}_0|^2\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\right)\right]\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0)\end{aligned} \tag{92}$$

By the fact that $\mathbf{J}^p(\mathbf{r}_0)$ vanishes on ∂D so does $\mathbf{A}(\mathbf{r}_0)$. Thus

$$\int_D \Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\cdot\frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0) = -\int_D \left(\nabla_{\mathbf{r}_0}|\mathbf{r}-\mathbf{r}_0|^2\right)\cdot\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0) - \int_D |\mathbf{r}-\mathbf{r}_0|^2\left(\nabla_{\mathbf{r}_0}\cdot\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\right)\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0). \tag{93}$$

Also $\nabla_{\mathbf{r}_0}\cdot\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0) = \Delta_{\mathbf{r}_0}\left(\nabla_{\mathbf{r}_0}\cdot\mathbf{A}(\mathbf{r}_0)\right) = 0$, $\nabla_{\mathbf{r}_0}|\mathbf{r}-\mathbf{r}_0|^2 = -2(\mathbf{r}-\mathbf{r}_0)$ and therefore

$$\int_D \Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\cdot\frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0) = 2\int_D (\mathbf{r}-\mathbf{r}_0)\cdot\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\mathrm{d}\upsilon(\mathbf{r}_0) \tag{94}$$

so that

$$\mathbf{r}\cdot\int_D \Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|} = \int_D \mathbf{r}_0\cdot\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|}. \tag{95}$$

In the sequel, employing the identity

$$\Delta(\mathbf{f}\cdot\mathbf{g}) = (\Delta\mathbf{f})\cdot\mathbf{g} + \mathbf{f}\cdot(\Delta\mathbf{g}) + 2(\nabla\otimes\mathbf{f})^{\top}:(\nabla\otimes\mathbf{g}), \tag{96}$$

$\top$ denoting transposition and : the double dot product, yields

$$\begin{aligned}\Delta_{\mathbf{r}_0}(\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0)) &= \Delta_{\mathbf{r}_0}\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0) + \mathbf{r}_0\cdot\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0) + 2(\nabla\otimes\mathbf{r}_0)^{\top}:\left(\nabla_{\mathbf{r}_0}\otimes\mathbf{A}(\mathbf{r}_0)\right)\\ &= \mathbf{0}\cdot\mathbf{A}(\mathbf{r}_0) + \mathbf{r}_0\cdot\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0) + 2\tilde{\mathbf{I}}:\nabla_{\mathbf{r}_0}\otimes\mathbf{A}(\mathbf{r}_0)\\ &= \mathbf{r}_0\cdot\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0) + 2\nabla_{\mathbf{r}_0}\cdot\mathbf{A}(\mathbf{r}_0)\\ &= \mathbf{r}_0\cdot\Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0).\end{aligned} \tag{97}$$

Combining (97) and (95) gives

$$\mathbf{r}\cdot\int_D \Delta_{\mathbf{r}_0}\mathbf{A}(\mathbf{r}_0)\frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|} = \int_D \Delta_{\mathbf{r}_0}(\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0))\frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|} \tag{98}$$

and finally, (88) reads

$$\frac{4\pi}{\mu_0}\mathbf{r}\cdot\mathbf{B}(\mathbf{r}) = -\int_D \Delta_{\mathbf{r}_0}(\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0))\frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|} + \frac{1}{4\pi}\int_D \left(\Delta_{\mathbf{r}_0}\Psi(\mathbf{r}_0)\right)\mathbf{r}\cdot\mathbf{K}(\mathbf{r};\mathbf{r}_0)\mathrm{d}\upsilon(\mathbf{r}_0). \tag{99}$$

Next, consider the quantity $\mathbf{r}\cdot\mathbf{K}(\mathbf{r};\mathbf{r}_0)$ present in the second integral of the RHS of (99). In view of (54) we have

$$\mathbf{r}\cdot\mathbf{r}'\times\nabla_{\mathbf{r}'}\frac{1}{|\mathbf{r}-\mathbf{r}'|}=\mathbf{r}\cdot\mathbf{r}'\times\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}=\mathbf{r}\cdot\frac{\mathbf{r}'\times\mathbf{r}}{|\mathbf{r}-\mathbf{r}'|^3}=0 \tag{100}$$

and thus $\mathbf{r}\cdot\mathbf{K}(\mathbf{r};\mathbf{r}_0)=0$ so that (99) simplifies as

$$\mathbf{r}\cdot\mathbf{B}(\mathbf{r})=-\frac{\mu_0}{4\pi}\int_{\partial D}\Delta_{\mathbf{r}_0}\left(\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0)\right)\frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|}. \tag{101}$$

Introducing the expansion

$$\frac{1}{|\mathbf{r}-\mathbf{r}_0|}=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{4\pi}{2n+1}\frac{r_0^n}{r^{n+1}}\mathrm{Y}_n^m(\hat{\mathbf{r}})\overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}_0) \tag{102}$$

and (101) reads

$$\mathbf{r}\cdot\mathbf{B}(\mathbf{r})=-\mu_0\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{1}{2n+1}\frac{\mathrm{Y}_n^m(\hat{\mathbf{r}})}{r^{n+1}}\int_{r_0<R}\Delta_{\mathbf{r}_0}\left(\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0)\right)r_0^n\overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}_0)\mathrm{d}\upsilon(\mathbf{r}_0). \tag{103}$$

Further, assume

$$\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}r_0^2\,a_n^m(r_0)\mathrm{Y}_n^m(\hat{\mathbf{r}}_0),\quad r_0\in[0,R) \tag{104}$$

so that

$$\Delta_{\mathbf{r}_0}\left(\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0)\right)=\left(\frac{\partial^2}{\partial r_0^2}+\frac{2}{r_0}\frac{\partial}{\partial r_0}+\frac{1}{r_0^2}\mathscr{B}(\hat{\mathbf{r}}_0)\right)\left(\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0)\right), \tag{105}$$

where the Beltrami or surface Laplacian operator $\mathscr{B}(\hat{\mathbf{r}}_0)$ is

$$\mathscr{B}(\hat{\mathbf{r}}_0)=\frac{\partial^2}{\partial\theta_0^2}+\frac{\cos\theta_0}{\sin\theta_0}\frac{\partial}{\partial\theta_0}+\frac{1}{\sin^2\theta_0}\frac{\partial^2}{\partial\phi_0^2}, \tag{106}$$

fulfilling

$$\mathscr{B}(\hat{\mathbf{r}}_0)\mathrm{Y}_n^m(\hat{\mathbf{r}}_0)=-n(n+1)\mathrm{Y}_n^m(\hat{\mathbf{r}}_0). \tag{107}$$

Hence,

$$\Delta_{\mathbf{r}_0}\left(\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0)\right)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left[r_0\frac{\mathrm{d}^2a_n^m}{\mathrm{d}r_0^2}+4\frac{\mathrm{d}a_n^m}{\mathrm{d}r_0}-\frac{1}{r_0}(n-1)(n+2)a_n^m\right]\mathrm{Y}_n^m(\hat{\mathbf{r}}_0). \tag{108}$$

Replacing above into (103) yields

$$\begin{aligned}\mathbf{r}\cdot\mathbf{B}(\mathbf{r})=-\mu_0\sum_{n=1}^{\infty}\sum_{m=-n}^{n}&\frac{1}{2n+1}\frac{\mathrm{Y}_n^m(\hat{\mathbf{r}})}{r^{n+1}}\\&\times\int_0^R\left(r_0^{n+3}\frac{\mathrm{d}^2a_n^m}{\mathrm{d}r_0^2}+4r_0^{n+2}\frac{\mathrm{d}a_n^m}{\mathrm{d}r_0}-r_0^{n+1}(n-1)(n+2)a_n^m\right)\mathrm{d}r_0,\end{aligned} \tag{109}$$

since $\mathrm{d}\upsilon(\mathbf{r}_0)=r_0^2\mathrm{d}r_0\mathrm{d}\Omega(\hat{\mathbf{r}}_0)$ and

$$\int_{S^2(\hat{\mathbf{r}}_0)}\mathrm{Y}_p^q(\hat{\mathbf{r}}_0)\overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}_0)\mathrm{d}\Omega(\hat{\mathbf{r}}_0)=\delta_{pn}\delta_{qm}. \tag{110}$$

Note, that since $\mathbf{B}(\mathbf{r})=\mathscr{O}\left(r^{-2}\right)$ as $r\to\infty$, n must begin with one. Simple calculations lead to

$$\mathbf{r}\cdot\mathbf{B}(\mathbf{r})=-\mu_0\sum_{n=1}^{\infty}\sum_{m=-n}^{n}\frac{1}{2n+1}\frac{R^{n+2}}{r^{n+1}}\left(R\frac{\mathrm{d}a_n^m}{\mathrm{d}r_0}\Big|_{r_0=R}-(n-1)a_n^m(R)\right)\mathrm{Y}_n^m(\hat{\mathbf{r}}). \tag{111}$$

In what follows, consider the quantity $\mathbf{r}\cdot\mathbf{B}(\mathbf{r})$ known from measurements, i.e. if we expand

$$\mathbf{r}\cdot\mathbf{B}(\mathbf{r})=\sum_{n=1}^{\infty}\sum_{m=-n}^{n}d_n^m(r)\mathrm{Y}_n^m(\hat{\mathbf{r}}), \tag{112}$$

then the coefficients $d_n^m(r)$ are known as well. Combining (111) and (112) we have

$$R\frac{\mathrm{d}a_n^m}{\mathrm{d}r_0}\Big|_{r_0=R}-(n-1)a_n^m(R)=-\frac{2n+1}{\mu_0}\frac{r^{n+1}}{R^{n+2}}d_n^m(r) \tag{113}$$

from which the coefficients a_n^m are calculated.

The remaining two components of $\mathbf{A}(\mathbf{r})$, namely $A_\theta(\mathbf{r})$ and $A_\phi(\mathbf{r})$, are computed with the aid of $\nabla\cdot\mathbf{A}(\mathbf{r})=0$, i.e.

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2A_r(\mathbf{r})\right)+\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta A_\theta(\mathbf{r})\right)+\frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}A_\phi(\mathbf{r})=0, \tag{114}$$

since

$$\mathbf{A}(\mathbf{r}) = A_r(\mathbf{r})\hat{\mathbf{r}} + A_\theta(\mathbf{r})\hat{\boldsymbol{\theta}} + A_\phi(\mathbf{r})\hat{\boldsymbol{\phi}}. \tag{115}$$

Further, considering that $\mathbf{r}\cdot\mathbf{A}(\mathbf{r}) = rA_r(\mathbf{r})$, we find, with the help of (104), that

$$r^2 A_r(\mathbf{r}) = \sum_{n=1}^{\infty}\sum_{m=-n}^{n} r^2\, a_n^m(r)\mathrm{Y}_n^m(\hat{\mathbf{r}}), \tag{116}$$

which is known. Moreover, let

$$\frac{\partial}{\partial\theta}\left(\sin\theta A_\theta(\mathbf{r})\right) = \sin\theta\sum_{n=1}^{\infty}\sum_{m=-n}^{n}\zeta_n^m(r)\mathrm{Y}_n^m(\hat{\mathbf{r}}), \tag{117}$$

$$\frac{\partial}{\partial\phi}A_\phi(\mathbf{r}) = \sin\theta\sum_{n=1}^{\infty}\sum_{m=-n}^{n}\eta_n^m(r)\mathrm{Y}_n^m(\hat{\mathbf{r}}), \tag{118}$$

so that (114) gives

$$\zeta_n^m + \eta_n^m = \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 a_n^m\right). \tag{119}$$

Switching to the perturbed case, we see from equation (54) that the quantity $\mathbf{K}(\boldsymbol{\tau};\mathbf{r}_0)$ is the only one affected by boundary deformations. Therefore, on the deformed boundary,

$$\hat{\boldsymbol{\nu}}(\boldsymbol{\tau})\mathrm{d}S(\boldsymbol{\tau}) = (R+\epsilon f)\left[(R+\epsilon f)\sin\theta\hat{\mathbf{r}} - \epsilon\sin\theta\nabla f\right]\mathrm{d}\theta\mathrm{d}\phi. \tag{120}$$

Also, employing Taylor's expansion gives

$$\hat{\mathbf{n}}(\boldsymbol{\tau})\times\frac{\mathbf{r}-\boldsymbol{\tau}}{|\mathbf{r}-\boldsymbol{\tau}|^3}\mathrm{d}S(\boldsymbol{\tau}) = \left(\sum_{n=0}^{\infty}\epsilon^n\boldsymbol{\alpha}_n(\hat{\mathbf{r}},\mathbf{r})\right)\frac{\mathrm{d}S(\hat{\mathbf{r}})}{|\mathbf{r}-R\hat{\mathbf{r}}|^3}, \tag{121}$$

where the first two coefficients $\boldsymbol{\alpha}$ are given by (56) and (57), respectively and further

$$\frac{1}{|\boldsymbol{\tau}-\mathbf{r}_0|} = \frac{1}{|R\hat{\mathbf{r}}-\mathbf{r}_0|} - \epsilon f\frac{R-\hat{\mathbf{r}}\cdot\mathbf{r}_0}{|R\hat{\mathbf{r}}-\mathbf{r}_0|^3} + \ldots \tag{122}$$

On the other hand, $\mathbf{r}\cdot\mathbf{K}(\mathbf{r};\mathbf{r}_0)$ implies $\mathbf{r}\cdot\boldsymbol{\alpha}_n(\hat{\mathbf{r}},\mathbf{r})$, $n\in\mathbb{N}$, namely

$$\mathbf{r}\cdot\boldsymbol{\alpha}_0(\hat{\mathbf{r}},\mathbf{r}) = 0, \tag{123}$$

$$\mathbf{r}\cdot\boldsymbol{\alpha}_1(\hat{\mathbf{r}},\mathbf{r}) = R^2\nabla_s f(\hat{\mathbf{r}})\times\hat{\mathbf{r}}\cdot\mathbf{r}. \tag{124}$$

Combining all results, finally gives

$$\mathbf{r}\cdot\mathbf{K}(\mathbf{r};\mathbf{r}_0)=\sum_{n=1}^{\infty}\epsilon^n \mathrm{I}_n(\mathbf{r};\mathbf{r}_0), \tag{125}$$

provided that

$$\mathrm{I}_1(\mathbf{r};\mathbf{r}_0)=R^2\oint_{S^2}\left(\frac{1}{|R\hat{\mathbf{r}}-\mathbf{r}_0|}+U_0(R\hat{\mathbf{r}};\mathbf{r}_0)\right)\frac{\nabla_s f(\hat{\mathbf{r}})\times\hat{\mathbf{r}}\cdot\mathbf{r}}{|\mathbf{r}-R\hat{\mathbf{r}}|^3}\mathrm{d}S(\hat{\mathbf{r}}). \tag{126}$$

Note that the integrals $\mathrm{I}_n(\mathbf{r};\mathbf{r}_0)$ do not depend on $\mathbf{J}^p$. Replacing everything back, gives

$$\begin{aligned}\frac{4\pi}{\mu_0}\mathbf{r}\cdot\mathbf{B}(\mathbf{r})=&-\int_D \Delta_{\mathbf{r}_0}\left(\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0)\right)\frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|}\\&+\frac{1}{4\pi}\sum_{n=1}^{\infty}\epsilon^n\int_D\left(\Delta_{\mathbf{r}_0}\Psi(\mathbf{r}_0)\right)\mathrm{I}_n(\mathbf{r};\mathbf{r}_0)\mathrm{d}\upsilon(\mathbf{r}_0).\end{aligned} \tag{127}$$

Following previous analysis we have

$$\begin{aligned}&\int_D \Delta_{\mathbf{r}_0}\left(\mathbf{r}_0\cdot\mathbf{A}(\mathbf{r}_0)\right)\frac{\mathrm{d}\upsilon(\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|}\\&\quad=4\pi\sum_{n=1}^{\infty}\sum_{m=-n}^{n}\frac{1}{2n+1}\frac{R^{n+2}}{r^{n+1}}\left(R\frac{\mathrm{d}a_n^m}{\mathrm{d}r_0}\Big|_{r_0=R}-(n-1)a_n^m(R)\right)\mathrm{Y}_n^m(\hat{\mathbf{r}}).\end{aligned} \tag{128}$$

On the other hand, letting

$$\Psi(\mathbf{r}_0)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\psi_n^m(r_0)\mathrm{Y}_n^m(\hat{\mathbf{r}}_0), \tag{129}$$

we find

$$\Delta_{\mathbf{r}_0}\Psi(\mathbf{r}_0)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(\ddot{\psi}_n^m(r_0)+\frac{2}{r_0}\dot{\psi}_n^m(r_0)-\frac{n(n+1)}{r_0^2}\psi_n^m(r_0)\right)\mathrm{Y}_n^m(\hat{\mathbf{r}}_0). \tag{130}$$

The surface integral I_1 provided by (126) is evaluated as follows. First we note that the quantity inside the parenthesis enjoys the expansion

$$\frac{1}{|R\hat{\mathbf{r}}-\mathbf{r}_0|}+u(R\hat{\mathbf{r}};\mathbf{r}_0)=4\pi\sum_{n=1}^{\infty}\sum_{m=-n}^{n}\frac{n+1}{n(2n+1)}\frac{r_0^n}{R^{n+1}}\mathrm{Y}_n^m(\hat{\mathbf{r}})\overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}_0). \tag{131}$$

Moreover, by the fact that $\nabla f(\hat{\mathbf{r}})\times\hat{\mathbf{r}}\cdot\mathbf{r}'=0$, we have

$$\frac{\nabla f(\hat{\mathbf{r}}) \times \hat{\mathbf{r}} \cdot \mathbf{r}'}{|\mathbf{r}' - R\hat{\mathbf{r}}|^3} = \left(\nabla f(\hat{\mathbf{r}}) \times \hat{\mathbf{r}}\right) \cdot \nabla_{\mathbf{r}'} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{r'^n}{r^{n+1}} \mathrm{Y}_n^m(\hat{\mathbf{r}}) \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}'). \tag{132}$$

Expand the functions in terms of spherical harmonics as

$$\frac{1}{\sin\theta} \frac{\partial f(\hat{\mathbf{r}})}{\partial \phi} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^m \mathrm{Y}_n^m(\hat{\mathbf{r}}), \quad \frac{\partial f(\hat{\mathbf{r}})}{\partial \theta} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} B_n^m \mathrm{Y}_n^m(\hat{\mathbf{r}}) \tag{133}$$

to find

$$\nabla f(\hat{\mathbf{r}}) \times \hat{\mathbf{r}} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathbf{C}_n^m \mathrm{Y}_n^m(\hat{\mathbf{r}}), \quad \mathbf{C}_n^m = A_n^m \hat{\boldsymbol{\theta}} - B_n^m \hat{\boldsymbol{\phi}}, \tag{134}$$

given that

$$A_n^m = \oint_{S^2} \frac{1}{\sin\theta} \frac{\partial f(\hat{\mathbf{r}})}{\partial \phi} \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}) \mathrm{d}S(\hat{\mathbf{r}}), \tag{135}$$

$$B_n^m = \oint_{S^2} \frac{\partial f(\hat{\mathbf{r}})}{\partial \theta} \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}) \mathrm{d}S(\hat{\mathbf{r}}). \tag{136}$$

Replacing (134) into (132), gives

$$\begin{aligned} &\sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathbf{C}_n^m \mathrm{Y}_n^m(\hat{\mathbf{r}}) \cdot \nabla_{\mathbf{r}} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{r'^n}{r^{n+1}} \mathrm{Y}_n^m(\hat{\mathbf{r}}) \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}) \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{R^{n-1}}{r^{n+1}} \mathrm{Y}_n^m(\hat{\mathbf{r}}) \mathrm{Y}_n^m(\hat{\mathbf{r}}) \left(A_n^m \frac{\partial \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}})}{\partial \theta} - \frac{B_n^m}{\sin\theta} \frac{\partial \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}})}{\partial \phi} \right) \end{aligned} \tag{137}$$

Substituting (131), (137) into the surface integral (126) yields

$$\begin{aligned} \mathrm{I}_1(\mathbf{r}; \mathbf{r}_0) = R^2 \oint_{S^2} &\left(4\pi \sum_{n=1}^{\infty} \sum_{m=-n}^{n} \upsilon_n^m(r_0) \mathrm{Y}_n^m(\hat{\mathbf{r}}) \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}_0) \right) \\ &\times \left[\sum_{p=1}^{\infty} \sum_{q=-p}^{p} \frac{\tau^{p-1}}{r^{p+1}} \mathrm{Y}_p^q(\hat{\mathbf{r}}) \mathrm{Y}_p^q(\hat{\mathbf{r}}) \left(A_p^q \frac{\partial \overline{\mathrm{Y}}_p^q(\hat{\mathbf{r}})}{\partial \theta} - \frac{B_p^q}{\sin\theta} \frac{\partial \overline{\mathrm{Y}}_p^q(\hat{\mathbf{r}})}{\partial \phi} \right) \right] \mathrm{d}S(\hat{\mathbf{r}}), \end{aligned} \tag{138}$$

where the second expansion now begins with $p = 1$, since for $p = 0$ the corresponding spherical harmonic is a constant and the derivatives present vanish. Employing (40) and (41), we find

$$A_p^q \frac{\partial \overline{\mathrm{Y}}_p^q(\hat{\mathbf{r}})}{\partial \theta} - \frac{B_p^q}{\sin\theta} \frac{\partial \overline{\mathrm{Y}}_p^q(\hat{\mathbf{r}})}{\partial \phi'} = \frac{1}{\sin\theta'} \Big[p\, j_{p+1}^q\, A_p^q \overline{\mathrm{Y}}_{p+1}^q(\hat{\mathbf{r}}) + iq\, B_p^q \overline{\mathrm{Y}}_p^q(\hat{\mathbf{r}}) - (p+1)\, j_p^q\, A_p^q \overline{\mathrm{Y}}_{p-1}^q(\hat{\mathbf{r}}) \Big] \tag{139}$$

so that the integral now reads

$$\mathrm{I}_1(\mathbf{r};\mathbf{r}_0) = 4\pi R^2 \sum_{n,m} \sum_{p,q} \frac{R^{p-1}}{r^{p+1}} \frac{n+1}{n(2n+1)} \frac{r_0^n}{R^{2n+1}} \mathrm{Y}_p^q(\hat{\mathbf{r}}) \overline{\mathrm{Y}}_n^m(\hat{\mathbf{r}}_0) \mathbb{J}_{n,p}^{m,q}, \tag{140}$$

where the indices n and p start with 1 and

$$\mathbb{J}_{n,p}^{m,q} = p\, j_{p+1}^q\, A_p^q\, \mathscr{I}_{1,n,p}^{m,q} + iq\, B_p^q\, \mathscr{I}_{2,n,p}^{m,q} - (p+1)\, j_p^q\, A_p^q\, \mathscr{I}_{3,n,p}^{m,q}, \tag{141}$$

while

$$\mathscr{I}_{1,n,p}^{m,q} = \oint_{S^2} \frac{1}{\sin\theta} \mathrm{Y}_n^m(\hat{\mathbf{r}}) \mathrm{Y}_p^q(\hat{\mathbf{r}}) \overline{\mathrm{Y}}_{p+1}^q(\hat{\mathbf{r}}) \mathrm{d}S(\hat{\mathbf{r}}), \tag{142}$$

$$\mathscr{I}_{2,n,p}^{m,q} = \oint_{S^2} \frac{1}{\sin\theta} \mathrm{Y}_n^m(\hat{\mathbf{r}}) \mathrm{Y}_p^q(\hat{\mathbf{r}}) \overline{\mathrm{Y}}_p^q(\hat{\mathbf{r}}) \mathrm{d}S(\hat{\mathbf{r}}), \tag{143}$$

$$\mathscr{I}_{3,n,p}^{m,q} = \oint_{S^2} \frac{1}{\sin\theta} \mathrm{Y}_n^m(\hat{\mathbf{r}}) \mathrm{Y}_p^q(\hat{\mathbf{r}}) \overline{\mathrm{Y}}_{p-1}^q(\hat{\mathbf{r}}) \mathrm{d}S(\hat{\mathbf{r}}). \tag{144}$$

With above relations, as well as (130), the second integral on the RHS of (127) in the linear regime becomes

$$\int_D \big(\Delta_{\mathbf{r}_0} \Psi(\mathbf{r}_0)\big)\, \mathrm{I}_1(\mathbf{r};\mathbf{r}_0) \mathrm{d}\upsilon(\mathbf{r}_0) = \int_D \left[\sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left(\ddot{\psi}_n^m(r_0) + \frac{2}{r_0} \dot{\psi}_n^m(r_0) - \frac{n(n+1)}{r_0^2} \psi_n^m(r_0) \right) \mathrm{Y}_n^m(\hat{\mathbf{r}}_0) \right] \times 4\pi R^2 \sum_{k,l} \sum_{p,q} \frac{R^{p-1}}{r^{p+1}} \frac{k+1}{k(2k+1)} \frac{r_0^k}{R^{2k+1}} \mathrm{Y}_p^q(\hat{\mathbf{r}}) \overline{\mathrm{Y}}_k^l(\hat{\mathbf{r}}_0) \mathbb{J}_{k,p}^{l,q} \mathrm{d}\upsilon(\mathbf{r}_0). \tag{145}$$

Moreover, due to orthogonality the integral corresponding to $n = 0$ vanishes, namely

$$\oint_{S^2} \mathrm{Y}_0^0(\hat{\mathbf{r}}_0) \mathrm{Y}_k^l(\hat{\mathbf{r}}_0) \mathrm{d}S(\hat{\mathbf{r}}_0) = 0, \quad \forall k \geq 1 \tag{146}$$

and we are left with

$$\int_D \left(\Delta_{\mathbf{r}_0}\Psi(\mathbf{r}_0)\right) \mathrm{I}_1(\mathbf{r};\mathbf{r}_0)\mathrm{d}\upsilon(\mathbf{r}_0) = 4\pi \sum_{n,m}\sum_{p,q} \frac{p+1}{p(2p+1)}\left(\frac{R}{r}\right)^{n+1} \mathbb{J}_{p,n}^{q,m} \left(R\dot{\psi}_p^q(R) - p\psi_p^q(R)\right) \mathrm{Y}_n^m(\hat{\mathbf{r}}). \tag{147}$$

Replacing everything back into (127) and following the procedure demonstrated, yields

$$\sum_{n,m} d_n^m \mathrm{Y}_n^m(\hat{\mathbf{r}}) = \frac{\mu_0}{4\pi}\sum_{n,m}\Bigg\{ -\frac{4\pi}{2n+1}\frac{R^{n+2}}{r^{n+1}}\left(R\dot{a}_n^m(R) - (n-1)a_n^m(R)\right) + \epsilon \sum_{p,q}\frac{p+1}{p(2p+1)}\left(\frac{R}{r}\right)^{n+1}\mathbb{J}_{p,n}^{q,m} \times \left(R\dot{\psi}_p^q(R) - p\psi_p^q(R)\right)\Bigg\} \mathrm{Y}_n^m(\hat{\mathbf{r}}). \tag{148}$$

Since a_n^m are known via (113), we deduce that the coefficients ψ_n^m are calculated as

$$d_n^m = \mu_0\left(\frac{R}{r}\right)^{n+1}\Bigg[-\frac{R}{2n+1}\left(R\dot{a}_n^m(R) - (n-1)a_n^m(R)\right) + \frac{\epsilon}{4\pi}\sum_{p=1}^{\infty}\sum_{q=-p}^{p}\frac{p+1}{p(2p+1)}\mathbb{J}_{p,n}^{q,m}\left(R\dot{\psi}_p^q(R) - p\psi_p^q(R)\right)\Bigg]. \tag{149}$$

The remaining two components are given by equations (119).

3 Example

Let us now showcase the particulars of the algorithms presented in an analytical fashion by referring to the simplest example possible, namely that of an univariate function $f(\theta)$. In order to facilitate calculations, but without loss of generality, let $f(\hat{\mathbf{r}}) = \cos\theta$. The procedure regarding complex functions $f(\hat{\mathbf{r}})$ can be found in [33].

3.1 EEG

When $f(\hat{\mathbf{r}}) = \cos\theta$, the corresponding coefficients $\overline{\mathbb{C}}_n^m$ are computed from (60) by integrating over the unit ball, i.e. $\overline{\mathbb{C}}_n^m = \oint_{S^2} \cos\theta \, \mathrm{Y}_n^m(\hat{\mathbf{r}}) \mathrm{d}S(\hat{\mathbf{r}})$. Forasmuch as, $\cos\theta = \sqrt{4\pi/3}\mathrm{Y}_1^0(\hat{\mathbf{r}})$ only coefficients with $m = 0$ will survive, namely $\overline{\mathbb{C}}_n^0 = \sqrt{4\pi/(2n+1)}\,\delta_{1,n}$ and thus (63) simplifies to $\mathbf{Q}\cdot\nabla_{\mathbf{r}_0} r_0 \, \overline{\mathrm{Y}}_1^0(\hat{\mathbf{r}}_0) = 0$, implying $Q_3 = 0$. Hence, for if $f(\hat{\mathbf{r}}) = \cos\theta$, we cannot retrieve the z-coordinate of the moment for the first correction.

Expanding the dipoles position and moment in powers of ϵ, we obtain the following approximations (keeping only linear corrections)

$$\mathbf{r}_0 = \mathbf{r}_{0,0} + \epsilon\, \mathbf{r}_{0,1}\,, \tag{150}$$

$$\mathbf{Q} = \mathbf{Q}_0 + \epsilon\, \mathbf{Q}_1\,, \tag{151}$$

respectively. The zeroth-order corrections $\mathbf{r}_{0,0}$ and $\mathbf{Q}_0$ are provided via relations (58) and (59), respectively. On the other hand, the first-order corrections $\mathbf{r}_{0,1}$ and $\mathbf{Q}_1$ depend upon the surface deformation, namely the function $f(\hat{\mathbf{r}})$.

Nevertheless, relation (39) simplifies accordingly as

$$\mathbb{A}_{1,k}^{\ell} = -\frac{4\pi}{2k+1}\frac{N_k^{\ell}}{kR}\frac{(k+\ell)!}{(k-\ell)!}\sum_{n=1}^{\infty}\mathbb{A}_{0,n}^{\ell}\left(n\, j_{n+1}^{\ell} N_{n+1}^{\ell}\delta_{n+1,k} - (n+1)\, j_n^{\ell} N_{n-1}^{\ell}\delta_{n-1,k}\right), \tag{152}$$

where N_k^{ℓ} denote the corresponding normalization constants and $\delta_{n,m}$ the Kronecker symbol. Computing the latter for $k = 1, 2$, furnishes the following eight relations

$$\mathbb{A}_{1,1}^{\ell} = \frac{3 j_2^{\ell}}{R^2}\mathbb{A}_{0,2}^{\ell}, \quad \ell = -1, 0, 1, \tag{153}$$

$$\mathbb{A}_{1,2}^{\ell} = \frac{1}{2R^3}\left(4 j_3^{\ell}\mathbb{A}_{0,3}^{\ell} - j_2^{\ell}\mathbb{A}_{0,1}^{\ell}\right), \quad \ell = -2, -1, 0, 1, 2\,. \tag{154}$$

However, only six of them are needed in order to identify $\mathbf{r}_{0,1}$ and $\mathbf{Q}_1$, those being

$$\mathbb{A}_{1,1}^{\pm 1} = \mp\frac{3}{4\sigma R^5}\sqrt{\frac{3}{2\pi}}x_{1,3}\left(Q_{1,1} \mp i\, Q_{1,2}\right), \tag{155}$$

$$\mathbb{A}_{1,1}^{0} = -\frac{3}{2\sigma R^5}\sqrt{\frac{1}{3\pi}}\left(Q_{1,1}x_{1,1} + Q_{1,2}x_{1,2}\right), \tag{156}$$

$$\mathbb{A}_{1,2}^{\pm 2} = \mp\frac{8\sqrt{5}}{9R^2}\mathbb{B}_1^{\pm 1}\left(x_{1,1} \mp i x_{1,2}\right), \tag{157}$$

$$\mathbb{A}_{1,2}^{0} = \frac{2}{R^2}\mathbb{B}_1^0 x_{1,3}. \tag{158}$$

Solving above system (155)–(158), yields

$$\mathbf{r}_{0,1} = \frac{R^2}{2}\left(\frac{9}{8\sqrt{5}}\left(\frac{\mathbb{A}^{-2}_{1,2}}{\mathbb{A}^{-1}_{1,1}} - \frac{\mathbb{A}^{2}_{1,2}}{\mathbb{A}^{1}_{1,1}}\right) - i\frac{9}{8\sqrt{5}}\left(\frac{\mathbb{A}^{-2}_{1,2}}{\mathbb{A}^{-1}_{1,1}} + \frac{\mathbb{A}^{2}_{1,2}}{\mathbb{A}^{1}_{1,1}}\right)\frac{\mathbb{A}^{0}_{1,2}}{\mathbb{A}^{0}_{1,1}}\right)^{\top}, \tag{159}$$

$$\mathbf{Q}_1 = \frac{4\sigma R^3}{3}\sqrt{\frac{2\pi}{3}}\frac{\mathbb{A}^{0}_{1,1}}{\mathbb{A}^{0}_{1,2}}\left(\mathbb{A}^{-1}_{1,1} - \mathbb{A}^{1}_{1,1} \; -i(\mathbb{A}^{-1}_{1,1} + \mathbb{A}^{1}_{1,1})\; 0\right)^{\top}. \tag{160}$$

3.2 MEG

Substituting $f(\hat{\boldsymbol{\tau}}) = \cos\theta$ into (78) we find

$$\mathbb{E}^{\ell}_{1,k} = -i4\pi\sigma R^{k+1}\frac{k}{(k+1)(2k+1)}\mathbb{B}^{\ell}_{k}. \tag{161}$$

As seen, forming a 3×3 system from six equations for $k = 1, 2$ and utilizing the corresponding Cartesian forms, furnishes

$$x_{1,1} = i\frac{\sqrt{5}\pi}{6}\sigma R^2\left[\frac{\mathbb{B}^{-1}_{1}\,\mathbb{E}^{-2}_{1,2}}{(\mathbb{E}^{-1}_{1,1})^2} + \frac{\mathbb{B}^{1}_{1}\,\mathbb{E}^{2}_{1,2}}{(\mathbb{E}^{1}_{1,1})^2} - \frac{4R}{5}\left(\frac{\mathbb{B}^{2}_{2}}{\mathbb{E}^{1}_{1,1}} + \frac{\mathbb{B}^{-2}_{2}}{\mathbb{E}^{-1}_{1,1}}\right)\right], \tag{162}$$

$$x_{1,2} = \frac{\sqrt{5}\pi}{6}\sigma R^2\left[\frac{\mathbb{B}^{-1}_{1}\,\mathbb{E}^{-2}_{1,2}}{(\mathbb{E}^{-1}_{1,1})^2} - \frac{\mathbb{B}^{1}_{1}\,\mathbb{E}^{2}_{1,2}}{(\mathbb{E}^{1}_{1,1})^2} + \frac{4R}{5}\left(\frac{\mathbb{B}^{2}_{2}}{\mathbb{E}^{1}_{1,1}} - \frac{\mathbb{B}^{-2}_{2}}{\mathbb{E}^{-1}_{1,1}}\right)\right], \tag{163}$$

$$x_{1,3} = i\frac{\sqrt{5}\pi}{3\sqrt{2}}\sigma R^2\frac{1}{\mathbb{E}^{0}_{1,1}}\left[\frac{\mathbb{B}^{-1}_{1}\,\mathbb{E}^{-2}_{1,2}}{\mathbb{E}^{-1}_{1,1}} - \frac{\mathbb{B}^{1}_{1}\,\mathbb{E}^{2}_{1,2}}{\mathbb{E}^{1}_{1,1}} + \frac{2R}{5}\left(\mathbb{B}^{2}_{2} - \mathbb{B}^{-2}_{2}\right)\right]. \tag{164}$$

The moment equals

$$Q_{r_0} = 0, \tag{165}$$

$$Q_{\theta_0} = -i\sqrt{\frac{2\pi}{3}}\sigma R^2\frac{\mathbb{B}^{1}_{1} - \mathbb{B}^{-1}_{1}}{r_0\,\sin\theta_0}, \tag{166}$$

$$Q_{\phi_0} = -\sqrt{\frac{3}{\pi}}\frac{\mathbb{E}^{0}_{1,1}}{r_0\,\sin\theta_0}. \tag{167}$$

4 Conclusions and Discussion

We presented a short review on the application of boundary perturbations for the forward and inverse problems in EEG and MEG. In all four cases, the surface of reference has been the homogeneous sphere, leading to solutions with corrections incorporating the deviation from the spherical geometry, included up to the first correction. Higher order corrections are feasible, but mathematically tedious to obtain and seldom provide more accuracy. A major impediment are the constrains, which are brought in by the Neumann condition (30), demonstrated in the example (Section 3), leading to the fact that components of either the moment of the dipole or its location or both, remain concealed. As a rule of thumb, complex functions introduce complex constrains. However, since the Neumann condition, rising from Green's theorem, cannot be avoided in the framework of the specific BVP, it is an integral part of the presented analysis. Nevertheless, one can still gain valuable insight into the problem at hand, namely if and how strong surface deformations affect measurements.

Advanced analytical or semi-analytical solutions and formulae attaining closed-type forms in EEG and MEG have quite important advantages compared with the pure numerical methods. Indeed, the validity of numerical solutions can be verified by such techniques, nevertheless the basic outcome gained is the facility in incorporating with the inverse problem, knowing the mathematical tools of the forward one itself. On the other hand, bearing in mind that very important physical laws can be derived from analytical methods, we can understand the necessity of tackling with a confident mathematical basis before getting involved with an algorithmic procedure. Therefore, even nowadays, there is always room for such kind of methods that coexist with pure numerical codes and aim to the solution of boundary value problems in physical applications of major importance.

Mathematical and computational work is currently in progress and involves research into several directions, such as the introduction of more complicated geometries for representing the head's shape or the accomplishment of actually difficult inversion algorithms, taking profit from the proposed framework.

References

1. R.J. Ilmoniemi, R.J. Näätänen, Magnetoencephalography, in *International Encyclopedia of the Social & Behavioral Sciences*, ed. by N.J. Smelser, P.B. Baltes (2001), pp. 9131–9137
2. I.M. Turan, Qualitative and quantitative EEG findings in Schizophrenia. Schizophrenia Bull. **3**(1), 61–79 (1977)
3. S. Noachtar, J. Rémi, The role of EEG in epilepsy: a critical review. Epilepsy Behav. **15**(1), 22–33 (2009)
4. A.M. Beres, Time is of the essence: a review of electroencephalography (EEG) and event-related brain potentials (ERPs) in language research. Appl. Psychophysiol. Biofeedback **42**(4), 247–255 (2017)

5. S. Baillet, Magnetoencephalography for brain electrophysiology and imaging. Nat. Neurosci. **20**, 327–339 (2017)
6. T. Kirschstein, R. Köhling, What is the Source of the EEG? Clin. EEG Neurosci. **40**(3), 146–149 (2009)
7. M. Hämämäinen, R. Hari, R. Ilmoniemi, J. Knuutila, O. Lounasmaa, Magnetoencephalography: theory, instrumentation and applications to the noninvasive study of human brain function. Rev. Mod. Phys. **65**, 413–497 (1993)
8. S. Baillet, Forward and Inverse Problems of MEG/EEG, in *Encyclopedia of Computational Neuroscience*, ed. by D. Jaeger, R. Jung (Springer, New York, 2014)
9. F.N. Wilson, R.H. Bayley, The electric field of an eccentric dipole in a homogeneous spherical conducting medium. Circulation **1**(1), 84–92 (1950)
10. H. Hallez, B. Vanrumste, R. Grech, J. Muscat, W. De Clercq, A. Vergult, Y. D'Asseler, K.P. Camilleri, S.G. Fabri, S. Van Huffel, I. Lemahieu, Review on solving the forward problem in EEG source analysis. J. NeuroEng. Rehabil. **4**, 46 (2007)
11. R. Grech, T. Cassar, J. Muscat, K.P. Camilleri, S.G. Fabri, M. Zervakis, P. Xanthopoulos, V. Sakkalis, B. Vanrumste, Review on solving the inverse problem in EEG source analysis. J. NeuroEng. Rehabil. **5**, 25 (2008)
12. D. Cohen, Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. Science **161**(3843), 784–786 (1968)
13. R. Plonsey, D.B. Heppner, Considerations of quasi-stationarity in electrophysiological systems. Bull. Math. Biophys. **29**, 657–664 (1967)
14. J. Sarvas, Basic mathematical and electromagnetic concepts of the biomagnetic inverse problems. Phys. Med. Biol. **32**, 11–22 (1987)
15. R. Albanese, P.B. Monk, The inverse source problem for Maxwell's equations. Inverse Problems **22**(3), 1023–1035 (2006)
16. G. Dassios, A.S. Fokas, Electro-magneto-encephalography for a three-shell model: dipoles and beyond for the spherical geometry. Inverse Problems **25**(3), 035001 (2009)
17. G. Dassios, G. Fragoyiannis, K. Satrazemi, On the inverse EEG problem for a 1D current distribution. J. Appl. Math. **2014**, 715785 (2014)
18. G. Dassios, K. Satrazemi, Inversion of electroencephalography data for a 2-D current distribution. Math. Methods Appl. Sci. **38**(6), 1098–1105 (2014)
19. G. Dassios, A.S. Fokas, D. Hadjiloizi, On the complementarity of electroencephalography and magnetoencephalography. Inverse Problems **23**, 2541 (2007)
20. G. Dassios, A.S. Fokas, The definite non-uniqueness results for deterministic EEG and MEG data. Inverse Problems **29**(6), 065012 (2013)
21. D.A. Varshalovich, A.N. Moskalev, V.K. Khersonskii, *Quantum Theory of Angular Momentum: Irreducible Tensors, Spherical Harmonics, Vector Coupling Coefficients, 3nj Symbols* (1. repr. ed.) (World Scientific, Singapore, 1988)
22. F.W.J. Olver, D.W. Lozier, R.F. Boisvert, C.W. Clark (eds.), *NIST Handbook of Mathematical Functions* (Cambridge University Press, Cambridge, 2010)
23. R. Bellman, *Perturbation Techniques in Mathematics, Engineering and Physics* (Holt, Rinehart and Winston, New York, 1966)
24. N.G. De Bruijn, *Asymptotic Methods in Analysis* (North-Holland Publishing, Amsterdam, 1958)
25. M. Doschoris, F. Kariotou, Mathematical foundation of electroencephalography, in *Electroencephalography*, ed. by P. Sittiprapaporn (IntechOpen, Rijeka, 2017)
26. D.B. Geselowitz, On the magnetic field generated outside an inhomogeneous volume conductor by internal current sources. IEEE Trans. Mag. **6**, 346–347 (1970)
27. Z.Z. Acar, S. Makeig, Effects of forward model errors on EEG source localization. Brain Topogr. **26**, 378–396 (2013)
28. G. Dassios, Electric and magnetic activity of the brain in spherical and ellipsoidal geometry, in *Mathematical Modeling in Biomedical Imaging I. Lecture Notes in Mathematics*, ed. by H. Ammari, vol. 1983 (Springer, Berlin, 2009)

29. D.B. Geselowitz, On bioelectric potentials in an inhomogeneous volume conductor. Biophys. J. **7**(1), 1–11 (1967)
30. M. Doschoris, P. Vafeas, G. Fragoyiannis, The influence of surface deformations on the forward magnetoencephalographic problem. SIAM J. Appl. Math. **78**(2), 963–976 (2018)
31. G. Dassios, M. Doschoris and K. Satrazemi, On the resolution of synchronous dipolar excitations via MEG measurements. Q. Appl. Math. **76**, 39–45 (2018)
32. A.S. Fokas, Electro-magneto-encephalography for a three-shell model: distributed current in arbitrary, spherical and ellipsoidal geometries. J. R. Soc. Interface **6**(34), 479–488 (2009)
33. G. Dassios, M. Doschoris, G. Fragoyannis, Sensitivity analysis of the forward EEG problem depending on head shape variations. Math. Probl. Eng. **2015**, Article ID 612528 (2015)

Poynting–Robertson and Oblateness Effects on the Equilibrium Points of the Perturbed R3BP: Application on Cen X-4 Binary System

Aguda Ekele Vincent and Angela E. Perdiou

Abstract We examine the dynamical effects of Poynting–Robertson (P–R) drag and oblateness together with small perturbations in the Coriolis and centrifugal forces on the existence, location and stability of equilibrium points in the photogravitational restricted three-body problem. It is found that under constant P–R drag effect, collinear equilibrium points cease to exist numerically and of course analytically. The problem admits five non-collinear equilibrium points and it is found that the positions of these points depend on all the system parameters except small perturbation in the Coriolis force. Finally, we justify the relevance of the model in astronomy by applying it to Cen X-4 binary system, for which all the equilibrium points have been seen to be unstable.

MSC 70F07, 70F15, 70M20, 70K42

1 Introduction

The restricted three-body problem (R3BP) consists of two finite bodies, known as primaries which rotate in circular orbits around their common center of mass and a massless body which moves in the plane of motion of the primaries under their gravitational attraction and does not affect their motion. The study of the R3BP is still an active field of research because of its applications in dynamics of the solar and stellar systems, artificial satellites and lunar theory. The circular restricted three-body problem (CR3BP) has been the well known studied problem in celestial

A. E. Vincent
Department of Mathematics, School of Basic Sciences, Nigeria Maritime University, Okerenkoko, Delta State, Nigeria
e-mail: vincentekele@yahoo.com

A. E. Perdiou (✉)
Department of Civil Engineering, School of Engineering, University of Patras, Patras, Greece
e-mail: aperdiou@upatras.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*, Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_7

mechanics. In this problem there are five equilibrium points; three of them lie on the x-axis and are called collinear while the other two are away from the x-axis and are called triangular equilibrium points. The three collinear points are generally unstable while the triangular points are generally stable for the mass ratio $\mu \leqslant 0.03850\ldots$ [30]. These equilibrium points are extensively used in space mission (see [1, 3, 12, 31] and references therein).

In celestial mechanics, many scientists and astronomers over the years have made modifications to the classical CR3BP (e.g. [6, 8, 13, 17–19, 22, 23, 29, 32–34]). Some of the modifications made, include the consideration of one or both primaries as oblate spheroids and/or radiation sources with small change in Coriolis and centrifugal forces and/or under the effect of different kinds of dissipation (Stokes and/or Poynting–Robertson drags). The studying of these issues enable us to get real and accurate data about the dynamical features of the system. For example, Oberti and Vienne [15] showed that the addition of oblateness effects leads to improved approximations of real orbits of certain satellites in the Solar System. Singh [28] examined out-of-plane equilibria by considering effect of a small change in Coriolis and centrifugal forces, when the primaries are both radiating and oblate spheroids. Chernikov [4] studied the existence and stability of equilibrium points under the influence of radiation and Poynting–Robertson drag. He found that six equilibrium points exist at most and pointed out that the collinear points are not positioned on the axis connecting the primaries any more while the triangular points are not symmetrical with respect to this axis. It was found that the triangular points are unstable for P–R effect. Schuerman [21] studied the triangular points of the problem and found that the points are unstable due to P–R effect. Furthermore, Ragos and Zafiropoulos [20] extended the problem to the case that both main bodies are radiation sources and studied the existence and stability of the equilibrium points. The P–R effect renders unstable those equilibrium points which are conditionally stable in the classical case. Murray [14] discussed the dynamical effect of different kinds of dissipation (nebular drag, gas drag, and P–R drag) in the circular restricted three body problem and found the collinear points are not positioned on the axis joining the two masses while the displaced triangular points L_4 and L_5 are asymptotically stable for certain classes of drag forces.

Kushvah [11] studied numerically the existence of equilibrium points of the perturbed R3BP, where the bigger and smaller primaries are considered radiation sources and oblate spheroids, respectively, and discussed the P–R effect which is caused due to the radiation pressure. They observed that the collinear points deviate from the axis joining the two primaries, while the triangular points are not symmetrical due to radiation pressure. The P–R effect ruins the stability of equilibrium points known to be conditionally stable in the gravitational case. When the primaries are radiation sources, Singh and Aminu [24] investigated the influences of small perturbations in the Coriolis and centrifugal forces together with P–R drag from both primaries on the triangular points. They found that the positions of these points are affected from the radiation pressure, P–R drag and small perturbation in the centrifugal force. They also discovered that these perturbing forces do not influence the nature of the stability of the points in the presence of P–R

drag as they remain unstable for the binary systems Luyten 726–8 and Kruger 60.1. In the same vein, Singh and Amuda [25] studied the triangular equilibrium points when the effect of radiation pressure from the smaller primary and its Poynting Robertson (P–R) drag are taken into account and the bigger primary as an oblate spheroid. They found numerically that the equilibrium points of the binary RXJ 0450.1–5856 are unstable. Later, Singh and Amuda [26] investigated the three dimensional case of the problem studied in [25] and they pointed out that the out-of-plane equilibria of the binary Cen X-4 system are unstable. By taking into consideration the P–R effect and Stellar wind drag, Chakraborty and Narayan [5] investigated the photogravitational elliptic restricted three-body problem and found that the equilibrium points are unstable due to the effect of the drag. Recently, Kalantonis et al. [10] studied the stability of the triangular equilibrium points in the elliptic R3BP with radiation and oblateness and showed that the positions of the triangular equilibrium points are given by an analytical formulae in which the parameters of the problem are only involved.

In this work, we aim to make an extension to the work of Singh and Amuda [25] by also taking small perturbations in the Coriolis and centrifugal forces and continue to study numerically the existence and location of the equilibrium points. As an application in this study, we consider the Cen X-4 binary system. The paper is organized as follows: In Section 2, the dynamical equations that involve the parameters of the infinitesimal particle in the binary system under consideration are obtained. In Section 3, we determine the existence and locations of the equilibrium points numerically and verify them graphically for values of the parameters of the problem, while their linear stability is analyzed in Section 4. A numerical application of these results is given in Section 5 while Section 6 summarizes the discussion and conclusion of our study.

2 Equations of Motion

The dynamical system consists of two bodies (known as the primaries) which move on circular orbits. We consider a barycentric coordinate system Oxyz rotating relative to an inertial reference system with angular velocity ω about a common z-axis. The two finite bodies P_1 (bigger primary) and P_2 (smaller primary) have masses $m_1 = 1-\mu$ and $m_2 = \mu$ $(0 < \mu \leqslant 1/2)$, respectively, with μ being the mass ratio parameter while the test particle P is considered to have a mass m, which is significantly smaller than the masses of the primaries and therefore it does not affect their motion. Also, the bigger primary body is considered to be an oblate spheroid while the smaller one is a source of radiation with its P–R drag. The equations of motion of the test particle in the three-dimensional restricted three-body problem with the origin resting at the center of mass, in a barycentric rotating coordinate system take the form [25]:

$$
\begin{aligned}
\ddot{x} - 2n\dot{y} &= n^2 x - \frac{(1-\mu)(x+\mu)}{r_1^3} - \frac{\mu(x+\mu-1)q_2}{r_2^3} - \frac{3(1-\mu)(x+\mu)A_1}{2r_1^5} - \\
&\quad \frac{W_2}{r_2^2}\left\{\frac{x+\mu-1}{r_2^2}\left[(x+\mu-1)\dot{x} + y\dot{y} + z\dot{z}\right] + \dot{x} - ny\right\}, \\
\ddot{y} + 2n\dot{x} &= n^2 y - \frac{(1-\mu)y}{r_1^3} - \frac{\mu q_2 y}{r_2^3} - \frac{3(1-\mu)A_1 y}{2r_1^5} - \\
&\quad \frac{W_2}{r_2^2}\left\{\frac{y}{r_2^2}\left[(x+\mu-1)\dot{x} + y\dot{y} + z\dot{z}\right] + \dot{y} + n(x+\mu-1)\right\}, \\
\ddot{z} &= -\frac{(1-\mu)z}{r_1^3} - \frac{\mu q_2 z}{r_2^3} - \frac{3A_1(1-\mu)z}{2r_1^5} - \\
&\quad \frac{W_2}{r_2^2}\left\{\frac{z}{r_2^2}\left[(x+\mu-1)\dot{x} + y\dot{y} + z\dot{z}\right] + \dot{z}\right\},
\end{aligned}
\tag{1}
$$

with

$$
r_1^2 = (x+\mu)^2 + y^2 + z^2, \quad r_2^2 = (x+\mu-1)^2 + y^2 + z^2, \quad W_2 = \frac{\mu(1-q_2)}{c_d}, \tag{2}
$$

where r_i, $i = 1, 2$ are the distances of the test particle from the bigger and smaller primaries, respectively, $q_2 \in (0, 1]$, $W_2 \ll 1$ stand for radiation pressure and P–R drag of the smaller body, respectively, c_d is the dimensional velocity of light which depends on the physical masses of the two bodies and the distance between them, chosen to the value $c_d = 299792458$ (see [25]) while the dots denote differentiation with respect to time t. Also, A_1 is the oblateness coefficient of the bigger primary body defined by the formula $A_1 = (A_E^2 - A_P^2)/5R^2 \ll 1$ where A_E and A_P are the equatorial and polar radii of the said primary body, respectively, and R is the distance between the primaries. On account of the oblateness of the primary body m_1, the mean perturbed motion n is defined by $n^2 = 1 + \frac{3}{2}A_1$. Additionally, perturbations on the Coriolis and centrifugal forces are included with the help of the parameters α and β, respectively, such that $\alpha = 1 + \varepsilon_1$, $\beta = 1 + \varepsilon_2$, $|\varepsilon_i| \ll 1$, $i = 1, 2$. The unperturbed value of each is taken as unity. Restricting ourselves to the plane Oxy and following the work of Singh and Aminu [24], the pertinent equations of motion (1) are finally written in the form:

Table 1 Numerical data for the binary Cen X-4 system

Binary system	Mass ($M_\otimes$)	Radiation pressure	Binary separation	Dimensionless speed of light	Mass ratio
	m_1 m_2	q_2	a	c_d	μ
Cen X-4	1.9996 0.0801	0.993	4.31	988.323	0.038515

$$\begin{aligned}
\ddot{x} - 2n\alpha\dot{y} &= n^2\beta x - \frac{(1-\mu)(x+\mu)}{r_1^3} - \frac{\mu(x+\mu-1)q_2}{r_2^3} - \frac{3(1-\mu)(x+\mu)A_1}{2r_1^5} - \\
&\quad \frac{W_2}{r_2^2}\left\{\frac{x+\mu-1}{r_2^2}\left[(x+\mu-1)\dot{x} + y\dot{y}\right] + \dot{x} - ny\right\}, \\
\ddot{y} + 2n\alpha\dot{x} &= n^2\beta y - \frac{(1-\mu)y}{r_1^3} - \frac{\mu q_2 y}{r_2^3} - \frac{3(1-\mu)A_1 y}{2r_1^5} - \\
&\quad \frac{W_2}{r_2^2}\left\{\frac{y}{r_2^2}\left[(x+\mu-1)\dot{x} + y\dot{y}\right] + \dot{y} + n(x+\mu-1)\right\},
\end{aligned} \tag{3}$$

while now

$$r_1^2 = (x+\mu)^2 + y^2, \quad r_2^2 = (x+\mu-1)^2 + y^2. \tag{4}$$

The physical parameters of the binary Cen X-4 system are shown in Table 1 (see [2, 25, 26]).

3 Existence and Positions of Equilibrium Points

The equilibrium (or Lagrangian) points are obtained when the acceleration $(\ddot{x}, \ddot{y})$ and velocity $(\dot{x}, \dot{y})$ components of the test particle are zero. So, we obtain the coordinates (x_0, y_0) of equilibrium points as solutions of the equations:

$$\begin{aligned}
n^2\beta\, x - \frac{(1-\mu)(x+\mu)}{r_1^3} - \frac{\mu(x+\mu-1)q_2}{r_2^3} - \frac{3(1-\mu)(x+\mu)A_1}{2r_1^5} + \frac{W_2 ny}{r_2^2} &= 0, \\
n^2\beta\, y - \frac{(1-\mu)y}{r_1^3} - \frac{\mu y q_2}{r_2^3} - \frac{3y(1-\mu)A_1}{2r_1^5} - \frac{W_2 n(x+\mu-1)}{r_2^2} &= 0.
\end{aligned} \tag{5}$$

It is interesting to note that for $A_1 = 0$, $q_2 = \beta = 1$, the classical case of the R3BP is recovered while the case $\beta = 1$ leads to the equations of motion presented in [25]. It is well known that in the classical R3BP there are two types of equilibria or solutions, depending on whether $y = 0$ or $y \neq 0$. Points for which $y = 0$ are called collinear equilibrium points and they lie on the line connecting the primaries, the x-axis of the synodic system, while points for which $y \neq 0$ are called triangular

(non-collinear) equilibrium points and they lie away from the x-axis of the synodic system.

In the perturbed R3BP where the radiation pressure coupled with P–R drag terms appear, the existence of collinear equilibria as well as the total number of the equilibrium points depend on the particular values of the radiation pressure (see, for example, [4, 14]). Ragos and Zafiropoulos [20] have shown numerically that in the photogravitational CR3BP including the P–R effect there are at most five equilibrium points (with no collinear points), depending on the values of radiation factors q_1 and q_2. Following the lead of above paper, we resort to a numerical study in this case of the problem since the system of Equations (5) which provides the (x_0, y_0) coordinates of the points of equilibrium cannot be solved analytically. In this premise, the equilibrium points are obtained by solving Equations (5) simultaneously using any well-known iterative method for finding roots of non-linear algebraic equations. The aforementioned method has been successfully applied in [16, 27] and [7] (see also references therein) for the determination of equilibrium points in a different model problem of Celestial Mechanics. We observe that our problem admits five non-collinear equilibrium points, $L_i,\ i = 1, 2, \ldots, 5$, which positions are independent of the Coriolis force but dependent upon the centrifugal force and the remaining involved parameters.

Generally, to obtain the positions for the collinear equilibrium points we solve Equations (5) for $y = 0$ but due to the existence of the dissipative term defined by the P–R drag, it is obvious that collinear equilibrium solution does not exist anymore. This is also easy to show geometrically by plotting the contours of the two implicit functions presented in system (5) (see Figure 1). We observe from this figure that the y components of the equilibrium points $L_{1,2,3}$ are close to zero but not zero. Moreover, this can be easily seen from bottom-left and right frames in Figure 1 where we enlarge the area close to $L_{1,2}$ and L_3, respectively. Therefore, we can conclude that under the effect of P–R drag, induced by the radiation pressure of the smaller primary, there are no equilibrium points that lie exactly on the x-axis, called collinear equilibrium points. This result agrees with [14, 20] and [11].

So, for the non-collinear equilibrium points, the second Equation (5) holds and the equilibria are obtained by solving both Equations (5) simultaneously. Figure 1 depicts the five non-collinear equilibrium points, $L_i, i = 1, 2, \ldots, 5$ of the problem in the xy-plane, along with the associated primaries, which have been found by solving numerically the aforementioned system for assumed values of $\mu = 0.03852,\ \beta = 1.01,\ A_1 = 0.0005,\ q_2 = 0.9999$ and $c_d = 299792458$. We denote here that the equilibria in the xy-plane are given by the mutual intersections of the two coloured curves where blue and brown lines in the figure correspond to the first and second equation of (5), respectively. Here we also note that the intersection points of these curves show the coordinates (x_0, y_0) of the equilibria on the xy-plane. It is seen that under the combined effects of the parameters, there exist five non-collinear equilibrium points for which the ordinates of $L_1,\ L_2$ and L_3 are close to zero but not zero. Therefore, from Figure 1, it is observed that under the combined effects of radiating smaller primary with it P–R drag, and oblateness of

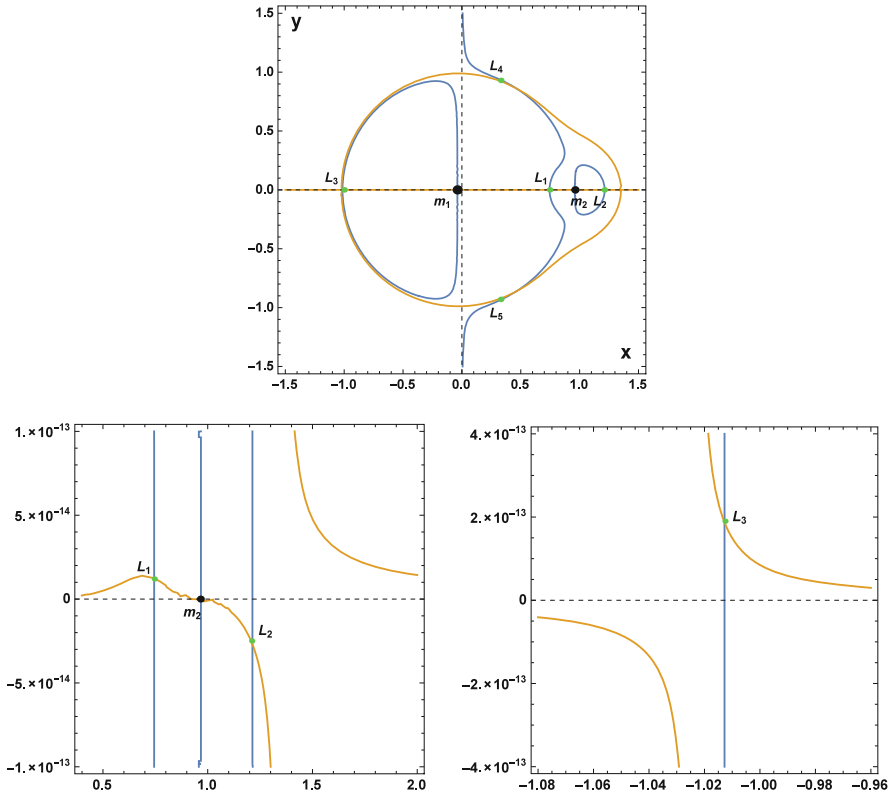


Fig. 1 The five non-collinear equilibrium points and the position of the primary bodies for $\mu = 0.03852$, $\beta = 1.01$, $A_1 = 0.0005$, $q_2 = 0.9999$ and $c_d = 299792458$. Bottom frames depict zoomed images of $L_{1,2}$ and L_3, respectively, with intersections of the curves. Black dots indicate the positions of the bodies m_i, $i = 1, 2$ while the positions of the equilibrium points L_i, $i = 1, 2, \ldots, 5$ are denoted by green dots

the bigger primary, the equilibria positions are different from those of the classical R3BP. All these results tally with [20].

4 Stability of the Non-collinear Equilibrium Points

To study analytically the solutions in the neighborhood of the non-collinear equilibrium points $L_i, i = 1, 2, \ldots, 5$, following Ragos and Zafiropoulos [20] as well as Singh and Amuda [25], we consider small displacements ξ and η given to the coordinates of an equilibrium point (x_0, y_0) such that $\xi = x - x_0$, $\eta = y - y_0$ and denote the right-hand side of equations of motion (3) by $\Omega_x = \partial\Omega/\partial x$ and $\Omega_y = \partial\Omega/\partial y$, respectively. Then the variational form of the equations of motion is

derived as:

$$\begin{aligned}\ddot{\xi} - 2n\alpha\dot{\eta} &= \Omega_{x\dot{x}}^{(0)}\dot{\xi} + \Omega_{x\dot{y}}^{(0)}\dot{\eta} + \Omega_{xx}^{(0)}\xi + \Omega_{xy}^{(0)}\eta,\\ \ddot{\eta} + 2n\alpha\dot{\xi} &= \Omega_{y\dot{x}}^{(0)}\dot{\xi} + \Omega_{y\dot{y}}^{(0)}\dot{\eta} + \Omega_{yx}^{(0)}\xi + \Omega_{yy}^{(0)}\eta,\end{aligned} \tag{6}$$

where the dots are the derivatives with respect to time t and only the linear terms in ξ and η have been taken. Now, we assume solutions of the variational equations of the form:

$$\xi = B_1 e^{\lambda t}, \qquad \eta = B_2 e^{\lambda t}, \tag{7}$$

where B_i, $i = 1, 2$, are arbitrary constants and λ is a parameter. Substituting Equations (7) in Equations (6) and simplifying, we obtain:

$$\begin{aligned}(\lambda^2 - \lambda\Omega_{x\dot{x}}^{(0)} - \Omega_{xx}^{(0)})B_1 + (-2n\alpha\lambda - \lambda\Omega_{x\dot{y}}^{(0)} - \Omega_{xy}^{(0)})B_2 = 0,\\ (2n\alpha\lambda - \lambda\Omega_{y\dot{x}}^{(0)} - \Omega_{yx}^{(0)})B_1 + (\lambda^2 - \lambda\Omega_{y\dot{y}}^{(0)} - \Omega_{yy}^{(0)})B_2 = 0.\end{aligned} \tag{8}$$

Now, for the nontrivial solution the determinant of the coefficients matrix of the above system must be zero, namely:

$$\begin{vmatrix} \lambda^2 - \lambda\Omega_{x\dot{x}}^{(0)} - \Omega_{xx}^{(0)} & -2n\alpha\lambda - \lambda\Omega_{x\dot{y}}^{(0)} - \Omega_{xy}^{(0)} \\ 2n\alpha\lambda - \lambda\Omega_{y\dot{x}}^{(0)} - \Omega_{yx}^{(0)} & \lambda^2 - \lambda\Omega_{y\dot{y}}^{(0)} - \Omega_{yy}^{(0)} \end{vmatrix} = 0. \tag{9}$$

Simplifying Equation (9) we obtain the characteristic polynomial corresponding to the system (6) as:

$$\lambda^4 + a\lambda^3 + b\lambda^2 + c\lambda + d = 0, \tag{10}$$

with

$$\begin{aligned}a &= -(\Omega_{y\dot{y}}^{(0)} + \Omega_{x\dot{x}}^{(0)}),\\ b &= 4n^2\alpha^2 + \Omega_{x\dot{x}}^{(0)}\Omega_{y\dot{y}}^{(0)} - \Omega_{xx}^{(0)} - \Omega_{yy}^{(0)} - [\Omega_{x\dot{y}}^{(0)}]^2,\\ c &= \Omega_{x\dot{x}}^{(0)}\Omega_{yy}^{(0)} + \Omega_{xx}^{(0)}\Omega_{y\dot{y}}^{(0)} + 2n\alpha\Omega_{xy}^{(0)} - 2n\alpha\Omega_{yx}^{(0)} - \Omega_{y\dot{x}}^{(0)}\Omega_{xy}^{(0)} - \Omega_{yx}^{(0)}\Omega_{x\dot{y}}^{(0)},\\ d &= \Omega_{xx}^{(0)}\Omega_{yy}^{(0)} - \Omega_{yx}^{(0)}\Omega_{xy}^{(0)},\end{aligned} \tag{11}$$

and the obtained eigenvalues determine the stability or instability of the respective equilibrium point. The second order partial derivative of Ω are denoted by subscripts while the superscript "0" means that the corresponding derivatives have been evaluated at the equilibrium points (x_0, y_0) and are given by the following analytical formulas:

$$\Omega_{xx}^{(0)} = n^2\beta - \frac{(1-\mu)}{r_{10}^3} - \frac{q_2\mu}{r_{20}^3} - \frac{3A_1(1-\mu)}{2r_{10}^5} + \frac{3(1-\mu)(x_0+\mu)^2}{r_{10}^5} + \frac{3q_2\mu(x_0+\mu-1)^2}{r_{20}^5} + \frac{15A_1(1-\mu)(x_0+\mu)^2}{2r_{10}^7} - \frac{2nW_2y_0(x_0+\mu-1)}{r_{20}^4}, \tag{12}$$

$$\Omega_{yy}^{(0)} = n^2\beta - \frac{(1-\mu)}{r_{10}^3} - \frac{q_2\mu}{r_{20}^3} + \frac{3(1-\mu)y_0^2}{r_{10}^5} - \frac{3(1-\mu)A_1}{2r_{10}^5} + \frac{3q_2\mu y_0^2}{r_{20}^5} + \frac{15A_1(1-\mu)y_0^2}{2r_{10}^7} + \frac{2nW_2y_0(x_0+\mu-1)}{r_{20}^4}, \tag{13}$$

$$\Omega_{xy}^{(0)} = \frac{nW_2}{r_{20}^2} - \frac{2nW_2y_0^2}{r_{20}^4} + \frac{3(1-\mu)(x_0+\mu)y_0}{r_{10}^5} + \frac{3q_2\mu(x_0+\mu-1)y_0}{r_{20}^5} + \frac{15A_1(1-\mu)(x_0+\mu)y_0}{2r_{10}^7}, \tag{14}$$

$$\Omega_{yx}^{(0)} = -\frac{nW_2}{r_{20}^2} + \frac{2nW_2(x_0+\mu-1)^2}{r_{20}^4} + \frac{3(1-\mu)(x_0+\mu)y_0}{r_{10}^5} + \frac{3q_2\mu(x_0+\mu-1)y_0}{r_{20}^5} + \frac{15A_1(1-\mu)(x_0+\mu)y_0}{2r_{10}^7}, \tag{15}$$

$$\Omega_{x\dot{x}}^{(0)} = -\frac{W_2}{r_{20}^2}(1+\frac{1}{r_{20}^2}) + \frac{W_2x_0}{r_{20}^4}(2-x_0) + \frac{W_2\mu}{r_{20}^4}(2(1-x_0)-\mu), \tag{16}$$

$$\Omega_{y\dot{y}}^{(0)} = -\frac{W_2}{r_{20}^2}(1+\frac{y_0^2}{r_{20}^2}), \qquad \Omega_{x\dot{y}}^{(0)} = \frac{W_2y_0}{r_{20}^4}(1-(x_0+\mu)) = \Omega_{y\dot{x}}^{(0)}, \tag{17}$$

with

$$r_{10}^2 = (x_0+\mu)^2 + y_0^2, \qquad r_{20}^2 = (x_0+\mu-1)^2 + y_0^2. \tag{18}$$

An equilibrium point (x_0, y_0) is said to be stable in the sense of Lyapunov if and only if all the four roots of the characteristic polynomial, given by Equation (10), are either negative real numbers or distinct imaginary; asymptotically stable if roots are complex with negative real parts and unstable, otherwise.

5 Numerical Application

In this section, we compute and examine graphically and numerically the positions of the non-collinear equilibrium points for the binary Cen X-4 system using the astrophysical parameters presented in Table 1 for some assumed oblateness and centrifugal force parameters. As pointed out in Section 2, the adjective non-collinear is due to the fact that L_1, L_2 and L_3 do not lie exactly on the x-axis. In order to visualize the evolution of the equilibria we consider region of the oblateness coefficient A_1 is [0, 0.2] (see [9]). The investigated region for the values of the Coriolis and centrifugal forces are α, $\beta \in [1, 1.2]$ (see, e.g. [28]) while the value of the dimensional velocity of light is kept fixed to $c_d = 988.323$ for all numerical calculations.

Solving Equations (5), using parameters in Table 1, we present in Figures 2 and 3 the positions of the equilibria for the binary system as the two parameters A_1 and β vary in the absence and presence of the P–R drag effect, respectively. For better understanding the evolution of the equilibria, in both figures, we use colour codes to indicate the set of pairs (A_1, β), while green dots signify the positions of the equilibria. So, the intersections of blue-magenta, black-magenta, and red-magenta curves correspond to three specific pairs of values of A_1 and β; particularly to (0, 1), (0.1, 1.04) and (0.2, 1.1), respectively. It is necessary to note that, although the curves are identical, their behaviours are different as we observe completely different results regarding the movement of the equilibrium points. From Figure 2 it can be observed that for varying oblateness factor and varying centrifugal force we have five equilibrium points (as in the classical restricted problem), three collinear $L_{1,2,3}$ and two triangular $L_{4,5}$, where equilibria L_1 and L_2 both approach the radiating primary m_2, while L_3 moves toward the oblate primary m_1 and point L_4 (the situation is same at the symmetric point L_5) moves closer to the point L_1. For clarity purposes, the top-right, bottom-left, and bottom-right frames are enlargements of the top-left frame of Figure 2 (first frame) close to $L_{1,2}$, L_3, and $L_{4(5)}$ points, respectively.

In Table 2, we have evaluated numerically the coordinates of the five equilibrium points for different values of the parameters A_1 and β for the binary system. One can observe in this table that the variational trend of the equilibria is similar to the scenario presented in Figure 2. However, the situation is different in the presence of P–R drag effect as we observe that for increasing values of the oblateness and centrifugal force parameters, there exist five non-collinear equilibrium points positioned off the Ox-axis. In addition $L_{1,3,4}$ have $y > 0$, while points $L_{2,5}$ have $y < 0$. It can be observed that the equilibria L_1 and L_2 approach the radiating

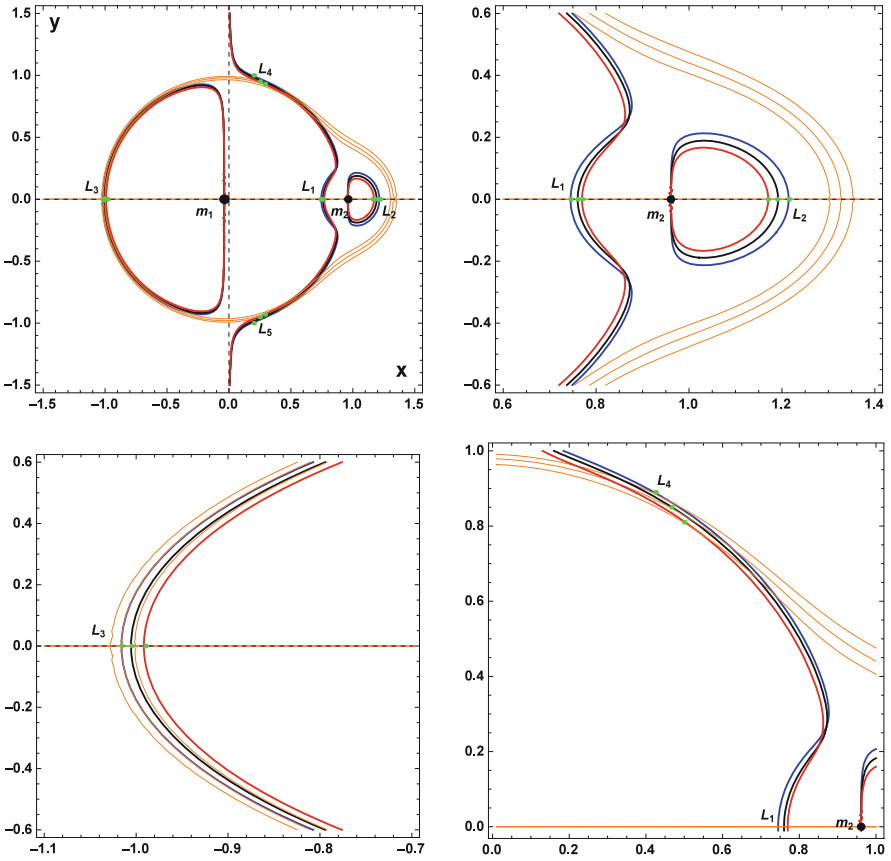


Fig. 2 Effect of oblateness and centrifugal force parameters on the collinear $L_{1,2,3}$ and the triangular $L_{4,5}$ points of Cen X-4 system without P–R effect (i.e., $q_2 = 1$, $W_2 = 0$) for $A_1 = 0$, $\beta = 1$ (blue, magenta); $A_1 = 0.1$, $\beta = 1.04$ (black, magenta) and $A_1 = 0.2$, $\beta = 1.1$ (red, magenta). Top-right, bottom-left and right frames: Zoomed areas close to $L_{1,2}$, L_3 and L_4 points, respectively. Black dots represent the primaries while green dots represent the positions of the equilibria

primary body m_2 in opposite directions while L_3 approaches the oblate primary m_1, and the two non symmetric equilibria L_4 and L_5 approach the displaced L_1 in opposite directions. Tables 3 and 4 provide the locations of the equilibrium points L_i, $i = 1, 2, \ldots, 5$ for varying oblateness and centrifugal force parameters in the presence of P–R drag for same fixed values of the parameters. One can see from these tables that the variational trend of the equilibria is similar to the behaviour presented in Figure 3.

Next, since we have already found the coordinates (x_0, y_0) of the equilibrium points (presented in Tables 2, 3, and 4), we can insert them into the characteristic Equation (10) and thus derive their linear stability numerically. In Tables 5 and 6,

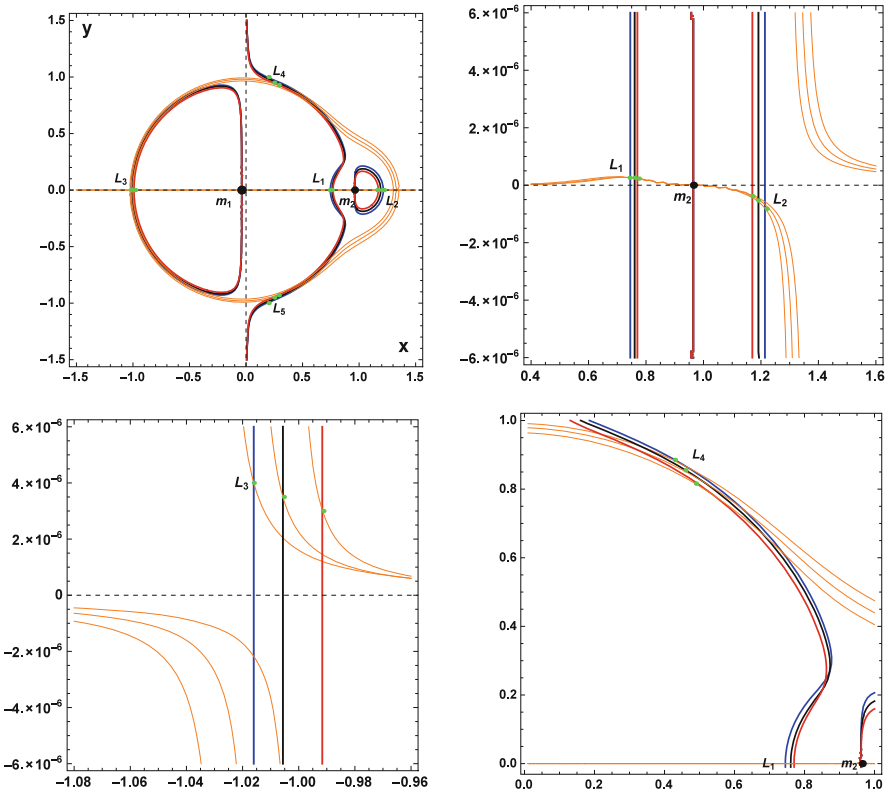


Fig. 3 Effect of oblateness and centrifugal force parameters on the non-collinear equilibrium points of Cen X-4 system with P–R effect (i.e., $q_2 = 0.993$, $W_2 = 2.72790 \times 10^{-7}$) for $A_1 = 0$, $\beta = 1$ (blue, magenta); $A_1 = 0.1$, $\beta = 1.04$ (black, magenta) and $A_1 = 0.2$, $\beta = 1.1$ (red, magenta). Top-right, bottom-left and right frames: Zoomed areas close to $L_{1,2}$, L_3 and L_4 points, respectively. Black dots represent the primaries while green dots represent the positions of the equilibria

Table 2 Positions of the five equilibrium points for varying oblateness and varying centrifugal force in the absence of P–R (i.e. $q_2 = 1, W_2 = 0$) for the binary Cen X-4 system

(A_1, β)	L_1	L_2	L_3	$L_{4,5}$
(0, 1)	(0.744951, 0)	(1.21443, 0)	(−1.01604, 0)	(0.461485, ±0.866025)
(0.025, 1.025)	(0.748593, 0)	(1.20566, 0)	(−1.00853, 0)	(0.473601, ±0.849585)
(0.05, 1.05)	(0.751774, 0)	(1.19750, 0)	(−1.00156, 0)	(0.484977, ±0.833890)
(0.075, 1.075)	(0.754572, 0)	(1.18988, 0)	(−0.99508, 0)	(0.495655, ±0.818874)
(0.1, 1.1)	(0.757045, 0)	(1.18275, 0)	(−0.98900, 0)	(0.505677, ±0.804479)

we show the nature of the stability of the equilibrium points for various values of oblateness, Coriolis and centrifugal forces in the absence and presence of the P–R effect, respectively, for the binary Cen X-4 system. Analysis of Tables 5 and 6

Table 3 Positions of $L_{1,2,3}$ non-collinear equilibrium points for varying oblateness and varying centrifugal force in the presence of P–R for the binary Cen X-4 system

(A_1, β)	L_1	L_2	L_3
(0, 1)	$(0.745409, 2.63737 \times 10^{-7})$	$(1.21380, -5.78026 \times 10^{-7})$	$(-1.01602, 4.01687 \times 10^{-6})$
(0.025, 1.025)	$(0.749036, 2.60924 \times 10^{-7})$	$(1.20505, -5.44399 \times 10^{-7})$	$(-1.00851, 3.80502 \times 10^{-6})$
(0.05, 1.05)	$(0.752204, 2.58813 \times 10^{-7})$	$(1.19690, -5.13966 \times 10^{-7})$	$(-1.00154, 3.61170 \times 10^{-6})$
(0.075, 1.075)	$(0.754990, 2.57277 \times 10^{-7})$	$(1.18930, -4.86306 \times 10^{-7})$	$(-0.99501, 3.43467 \times 10^{-6})$
(0.1, 1.1)	$(0.757452, 2.56225 \times 10^{-7})$	$(1.18218, -4.61069 \times 10^{-7})$	$(-0.98899, 3.27203 \times 10^{-6})$

Table 4 Positions of $L_{4,5}$ non-collinear equilibrium points for varying oblateness and varying centrifugal force in the presence of P–R for the binary Cen X-4 system (continuation of Table 3)

(A_1, β)	L_4	L_5
(0, 1)	(0.46382, 0.864673)	(0.463823, −0.864672)
(0.025, 1.025)	(0.47584, 0.848230)	(0.475845, −0.845228)
(0.05, 1.05)	(0.48713, 0.832534)	(0.487133, −0.832533)
(0.075, 1.075)	(0.49773, 0.817519)	(0.497729, −0.817517)
(0.1, 1.1)	(0.50767, 0.803125)	(0.507675, −0.803124)

Table 5 Stability of Cen X-4 system for small assumed values of oblateness and perturbations in Coriolis and centrifugal forces in the absence of P–R drag effect (see Table 2)

$L_i,\ i = 1, 2, \ldots, 5$	(x_0, y_0)	$\lambda_{1,2}$	$\lambda_{3,4}$
Case: $A_1 = 0,\ \beta = 1,\ \alpha = 1$			
L_1	(0.744951, 0)	± 3.145064	$\pm 2.469515i$
L_2	(1.21443, 0)	± 2.002264	$\pm 1.772113i$
L_3	(−1.01604, 0)	± 0.314525	$\pm 1.031797i$
$L_{4,5}$	(0.461485, ±0.866025)	$\pm 0.711480i$	$\pm 0.702705i$
Case: $A_1 = 0.05,\ \beta = 1.05,\ \alpha = 1.04$			
L_1	(0.751774, 0)	± 3.338678	$\pm 2.591204i$
L_2	(1.19750, 0)	± 2.224148	$\pm 1.956125i$
L_3	(−1.00156, 0)	± 0.351533	$\pm 1.083626i$
$L_{4,5}$	(0.484977, ±0.83389)	$-0.130094 \pm 0.755847i$	$0.130094 \pm 0.755847i$
Case: $A_1 = 0.1,\ \beta = 1.1,\ \alpha = 1.08$			
L_1	(0.757045, 0)	± 3.497120	$\pm 2.700120i$
L_2	(1.182750, 0)	± 2.452064	$\pm 2.146047i$
L_3	(−0.989004, 0)	± 0.386942	$\pm 1.141295i$
$L_{4,5}$	(0.505677, ±0.804479)	$-0.180429 \pm 0.806787i$	$0.180429 \pm 0.806787i$

reveals the non existence of pure imaginary roots except in the classical case (i.e. $q_2 = 1, W_2 = 0, \alpha = \beta = 1$). In all cases for all the assumed values of oblateness and perturbations in Coriolis and centrifugal forces with and without P–R effect, there exists at least a positive real root and/or a complex root with positive real part. Consequently the motion of the infinitesimal body is unbounded and thus unstable around all these equilibrium points.

6 Discussion and Conclusion

The location and stability of the equilibrium points in the photogravitational restricted three-body problem that accounts for Poynting–Robertson (P–R) drag force with oblateness of the first primary together with small perturbations in the Coriolis and centrifugal forces were studied. It was found both analytically and numerically that in the presence of P–R drag effect the well-known collinear

Table 6 Stability of Cen X-4 system for small assumed values of oblateness and perturbations in Coriolis and centrifugal forces in the presence of P–R drag effect (see Tables 3 and 4)

$L_i,\ i = 1, 2, \ldots, 5$	(x_0, y_0)	$\lambda_{1,2}$	$\lambda_{3,4}$
Case: $A_1 = 0,\ \beta = 1,\ \alpha = 1$			
L_1	$(0.745409, 2.63737 \times 10^{-7})$	$-3.143139, 3.143130$	$-3.892 \times 10^{-6} \pm 2.4683i$
L_2	$(1.2138, -5.78026 \times 10^{-7})$	$-2.003189, 2.003184$	$-3.721 \times 10^{-6} \pm 1.7726i$
L_3	$(-1.01602, 4.01687 \times 10^{-6})$	$-0.314652, 0.314652$	$-1.857 \times 10^{-7} \pm 1.0318i$
L_4	$(0.46382, 0.864673)$	$-0.01326 \pm 0.707256i$	$0.013264 \pm 0.707207i$
L_5	$(0.463823, -0.864672)$	$-0.01336 \pm 0.707257i$	$0.013358 \pm 0.707209i$
Case: $A_1 = 0.05,\ \beta = 1.05,\ \alpha = 1.04$			
L_1	$(0.752204, 2.58813 \times 10^{-7})$	$-3.336270, 3.336260$	$-4.186 \times 10^{-6} \pm 2.5898i$
L_2	$(1.1969, -5.13966 \times 10^{-7})$	$-2.225351, 2.225345$	$-4.183 \times 10^{-6} \pm 1.957i$
L_3	$(-1.00154, 3.6117 \times 10^{-6})$	$-0.351708, 0.351708$	$-1.932 \times 10^{-7} \pm 1.0837i$
L_4	$(0.48713, 0.832534)$	$-0.13090 \pm 0.755987i$	$0.130890 \pm 0.755983i$
L_5	$(0.487133, -0.832533)$	$-0.13090 \pm 0.75599i$	$0.130902 \pm 0.755983i$
Case: $A_1 = 0.1,\ \beta = 1.1,\ \alpha = 1.08$			
L_1	$(0.757452, 2.56225 \times 10^{-7})$	$-3.494287, 3.494280$	$-4.462 \times 10^{-6} \pm 2.6985i$
L_2	$(1.18218, -4.61069 \times 10^{-7})$	$-2.453508, 2.453500$	$-4.671 \times 10^{-6} \pm 2.1468i$
L_3	$(-0.98899, 3.27203 \times 10^{-6})$	$-0.387111, 0.387111$	$-1.993 \times 10^{-7} \pm 1.1413i$
L_4	$(0.507673, 0.803125)$	$-0.18105 \pm 0.806928i$	$0.181047 \pm 0.806923i$
L_5	$(0.507675, -0.803124)$	$-0.18105 \pm 0.806929i$	$0.181051 \pm 0.806924i$

equilibrium points of the circular restricted three-body problem cease to exist while the respective triangular equilibrium points do not form equilateral triangles.

Using the astrophysical parameters of the Cen X-4 binary system we performed a numerical study for its equilibrium points and showed that in the case where P–R drag was considered five non-collinear equilibrium points exist whereas in the absence of P–R drag force there are also five equilibrium points but three of them are located on the axis joining the primaries and the rest two form in the plane of motion equilateral triangles with the primaries, as in the circular restricted three-body problem. It was also found that the equilibrium points are independent of the effect of small perturbation in the Coriolis force but are affected by the small perturbation in centrifugal force. For the stability of the five equilibria, the four roots of the characteristic polynomial were determined numerically and found that are unstable due to the existence of at least one positive real root or a complex root with positive real part. The instability of the equilibrium points agrees with the results existing in the literature when the primaries are not oblate spheroids and small perturbations in the Coriolis and centrifugal forces are not considered (for details we refer to [4, 21])

References

1. E.I. Abouelmagd, F. Alzahrani, A. Hobiny, J.L.G. Guirao, M. Alhothuali, Periodic orbits around the collinear libration points. J. Nonlinear Sci. Appl. **9**, 1716–1727 (2016)
2. N. Bello, A. Umar, On the stability of triangular points in the relativistic R3BP when the bigger primary is oblate and the smaller one radiating with application on Cen X-4 binary system. Results Phys. **9**, 1067–1076 (2018)
3. L.R. Capdevila, K.C. Howell, A transfer network linking Earth, Moon, and the triangular libration point regions in the Earth-Moon system. Adv. Space. Res. **62**, 1826–1852 (2018)
4. J.A. Chernikov, The photogravitational restricted three-body problem. Sov. Astron. **14**, 176–181 (1970)
5. A. Chakraborty, A. Narayan, Effect of stellar wind and Poynting–Robertson drag on photogravitational elliptic restricted three-body problem. Sol. Syst. Res. **52**, 168–179 (2018)
6. S.M. Elshaboury, E.I. Abouelmagd, V.S. Kalantonis, E.A. Perdios, The planar restricted three-body problem when both primaries are triaxial rigid bodies: Equilibrium points and periodic orbits. Astrophys. Space Sci. **361**, 315 (2017)
7. D. Fakis, T. Kalvouridis, The Copenhagen problem with a quasi-homogeneous potential. Astrophys. Space Sci. **362**, 102 (2017)
8. F. Gao, R. Wang, Bifurcation analysis and periodic solutions of the HD 191408 system with triaxial and radiative perturbations. Universe **6**, 35 (2020)
9. V.S. Kalantonis, C.N. Douskos, E.A. Perdios, Numerical determination of homoclinic and heteroclinic orbits at collinear equilibria in the restricted three-body problem with oblateness. Celest. Mech. Dyn. Aston. **94**, 135–153 (2006)
10. V.S. Kalantonis, A.E. Perdiou, E.A. Perdios, On the stability of the triangular equilibrium points in the elliptic restricted three-body problem with radiation and oblateness, in *Mathematical Analysis and Applications*, ed. by T. Rassias, P. Pardalos. Springer Optim. Its Appl., vol. 154 (Springer, Cham, 2019), pp. 273–286
11. B.S. Kushvah, The effect of radiation pressure on the equilibrium points in the generalized photogravitational restricted three body problem. Astrophys. Space Sci. **315**, 231–241 (2008)

12. G. Mingotti, F. Topputo, F. Bernelli-Zazzera, Earth-Mars transfers with ballistic escape and low-thrust capture. Celest. Mech. Dyn. Astr. **110**, 169–188 (2011)
13. Z.E. Musielak, B. Quarles, The three-body problem. Rep. Progr. Phys. **77**, 065901 (2014)
14. C.D. Murray, Dynamical effects of drag in the circular restricted three-body problem. I: location and stability of the Lagrangian equilibrium points. Icarus **112**, 465–484 (1994)
15. P. Oberti, A. Vienne, An upgraded theory for Helene, Telesto, and Calypso. Astron. Astrophys. **397**, 353–359 (2003)
16. J.P. Papadouris, K.E. Papadakis, Equilibrium points in the photogravitational restricted four-body problem. Astrophys. Space Sci. **344**, 21–38 (2013)
17. E.A. Perdios, V.S. Kalantonis, Critical periodic orbits in the restricted three-body problem with oblateness. Astrophys. Space Sci. **305**, 331–336 (2006)
18. E.A. Perdios, V.S. Kalantonis, Self-resonant bifurcations of the Sitnikov family and the appearance of 3D isolas in the restricted three-body problem. Celest. Mech. Dyn. Astr. **113**, 377–386 (2011)
19. E.A. Perdios, V.S. Kalantonis, C.N. Douskos, Critical periodic orbits in the restricted three-body problem with oblateness. Astrophys. Space Sci. **314**, 199–208 (2008)
20. O. Ragos, F.A. Zafiropoulos, A numerical study of the influence of the Poynting–Robertson effect on the equilibrium points of the photogravitational restricted three-body problem. I. Coplanar case. Astron. Astrophys. **300**, 568–578 (1995)
21. D.W. Schuerman, Influence of the Poynting–Robertson effect on triangular points of the photogravitational restricted three-body problem. Astrophys. J., **238**, 337–342 (1980)
22. J. Singh, A.E. Perdiou, J.M. Gyegwe, V.S. Kalantonis, Periodic orbits around the collinear equilibrium points for binary Sirius, Procyon, Luhman 16, α-Centuari and Luyten 726–8 systems: the spatial case. J. Phys. Commun. **1**, 025008 (2017)
23. J. Singh, A.E. Perdiou, J.M. Gyegwe, E.A. Perdios, Periodic solutions around the collinear equilibrium points in the perturbed restricted three-body problem with triaxial and radiating primaries for binary HD 191408, Kruger 60 and HD 155876 systems. Appl. Math. Comput. **325**, 358–374 (2018)
24. J. Singh, A. Aminu, Instability of triangular libration points in the perturbed photogravitational R3BP with Poynting–Robertson (P–R) drag. Astrophys. Space Sci. **351**, 473–482 (2014)
25. J. Singh, T.O. Amuda, Poynting–Robertson (P–R) drag and oblateness effects on motion around the triangular equilibrium points in the photogravitational R3BP. Astrophys. Space Sci. **350**, 119–126 (2014)
26. J. Singh, T.O. Amuda, Out-of-plane equilibrium points in the photogravitational CR3BP with oblateness with P–R drag. J. Astrophys. Astr. **36**, 291–305 (2015)
27. J. Singh, A.E. Vincent, Equilibrium points in the restricted fourbody problem with radiation pressure. Few-Body Syst. **57**, 83–91 (2016)
28. J. Singh, Motion around the out of plane equilibrium points of the perturbed R3BP. Astrophys. Space Sci. **342**, 13–19 (2013)
29. Md.S. Suraj, R. Aggarwal, A. Mittal, Md C.A., The perturbed restricted three-body problem with angular velocity: Analysis of basins of convergence linked to the libration points. Int. J. Nonlinear Mech. **123**, 103494 (2020)
30. V. Szebehely, *Theory of Orbits. The Restricted Problem of Three-Bodies* (Academic Press, New York, 1967)
31. P. Verrier, T. Waters, J. Sieber, Evolution of the L_1 halo family in the radial solar sail circular restricted three-body problem. Celest. Mech. Dyn. Astron. **120**, 373–400 (2014)
32. X.Y. Zeng, H.X. Baoyin, J.F. Li, Updated rotating mass dipole with oblateness of one primary (I): equilibria in the equator and their stability. Astrophys. Space Sci. **361**, 14 (2015)
33. E.E. Zotos, F.L. Dubeibe, Orbital dynamics in the post Newtonian planar circular restricted Sun-Jupiter system. Int. J. Mod. Phys. D **27**, 1850036 (2018)
34. E.E. Zotos, D. Veras, The grain size survival threshold in one-planet post-main-sequence exoplanetary systems. Astron. Astrophys. **637**, A14 (2020)

Localization and Perturbation of Complex Zeros of Solutions to Second Order Differential Equations with Polynomial Coefficients. A Survey

Michael Gil'

Abstract This paper is a survey of the recent results of the author on the complex zeros of solutions to linear homogeneous second order ordinary differential equations with polynomial coefficients. In particular, estimates for the sums and products of the zeros are derived. These estimates give us bounds for the function counting the zeros of solutions and information about the zero-free domain. Some other applications of the obtained estimates for the sums and products of the zeros are also discussed. In addition, we investigate the variation of the zeros of solutions under perturbations of the coefficients. Illustrative examples are also presented. A part of the results presented in the paper is new.

AMS (MOS) Subject Classification 34C10, 34A30

1 Introduction

This paper is a survey of the recent results of the author on the zeros of solutions to a linear ordinary differential equation (ODE) with polynomial coefficients in the complex domain.

The literature devoted to the zeros of the solutions of such equations is very rich. Besides, the main tool is the Nevanlinna theory. The excellent exposition of the Nevanlinna theory and its applications to differential equations is given in the book [32]. In that book, in particular, the well-known results of Banks [4–6], Brűggemann [7, 8], Hellerstein and Rossi [26–28, 36], and other mathematicians are reflected. The classical comparison principle for zeros of ODEs in the complex plane is presented in [29].

The real zeros of solutions to equations with polynomial coefficients were investigated in the papers by Gundersen [25], Eremenko and Merenkov [11], and

M. Gil' (✉)
Department of Mathematics, Ben Gurion University of the Negev, Beer-Sheva, Israel
e-mail: gilmi@bezeqint.net

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_8

by C.Z. Huang [30]. The paper [37] studies the convergence of the zeros of a non-trivial (entire) solution to the linear differential equation

$$f'' + \left\{Q_1(z)e^{P_1(z)} + Q_2(z)e^{P_2(z)} + Q_3(z)e^{P_3(z)}\right\} f = 0,$$

where P_j are polynomials of degree $n \geq 1$ and $Q_j (\not\equiv 0)$ are entire functions of order less than n $(j = 1, 2, 3)$. In the paper [13], by certain separation and comparison results, estimates for the counting functions of the zeros of solutions to nth-order linear differential equations are deduced. These estimates generalize known results for the zeros of solutions to third- and fourth-order linear differential equations. The remarkable results on the zeros of a wide class of ordinary differential equations with polynomial coefficients, whose solutions are classical orthogonal polynomials, have been established by N. Anghel [2]. In addition, in the paper [3] N. Anghel investigated the following question: when is an entire function of finite order, the solution to a complex second order homogeneous linear differential equation with polynomial coefficients ? He gives two (equivalent) answers to this question, one of which involves certain Stieltjes-like relations for the zeros of solutions, the second one requires the vanishing of all but finitely many suitable expressions constructed via the relations of the sums of the zeros of the function derived in [17].

In connection with the recent results on the complex zeros of solutions to ODEs see also the papers [12, 13, 31, 34], and references given therein. Certainly, we could not survey the whole subject here and refer the reader to the listed publications.

It should be noted that in the above cited works mainly the asymptotic distributions of zeros are investigated. At the same time, bounds for the zeros of solutions are very important in various applications. But to the best of our knowledge, they have been investigated considerably less than the asymptotic distributions. In the paper [18] the author has established bounds for the sums of the zeros of solutions for the second order homogeneous equations with polynomial coefficients. In the interesting paper [9], some results from [18] have been extended to the equation $u^{(m)} = P(z)u$, where P is a polynomial and $m > 2$. In the papers [19, 22] and [21] the main result from [18] have been extended to the second order ODEs with non-polynomial coefficients, to ODEs having singular points and to non-homogeneous second order ODEs, respectively. In addition, in the paper [20] the author has derived a bound for the products of the zeros of solutions to ODEs with polynomial coefficients.

It should be also noted that to the best of our knowledge, perturbations of the zeros of solutions were almost not investigated in the available literature. Here we can mention only the paper [19] on perturbations of the zeros of solutions to second order differential equations with polynomial coefficients.

The present paper reflects some results from the just pointed papers of the author. Besides, the proofs are considerably simplified. Our main tool is the recent results for the zeros of entire functions established in [14, 15] (see also [17]).

A few words about the contents. The paper consists of 13 sections.

Section 2 contains solution estimates for non-homogeneous ODEs.

In Section 3 we recall the basic properties of singular values of compact operators, which we need for the proofs of our main results.

In Sections 4 and 5 we obtain bounds for the sums and products of zeros of finite order entire functions via Taylor coefficients and via orders of the functions. As it was above mentioned, bounds for the zeros of entire functions are our main tool.

In Sections 6 we derive bounds for the sums and products of zeros of solutions to the equation $u'' = P(z)u$ ($z \in \mathbf{C}$), where $P(z)$ is a polynomial. In Section 7 we discuss some applications of these bounds.

In Sections 8, 9, and 10 we obtain perturbation results for the zeros of entire functions, which are used in Section 12 to obtain perturbation bounds for the zeros of ODEs.

In Section 13 we present an example which illustrates the results obtained in Section 12.

2 Solution Estimates for ODEs

Consider the equation

$$\frac{d^2u}{dz^2} = Q(z)u + F(z) \quad (z \in \mathbf{C}, u(0) = u_0 \in \mathbf{C}, u'(0) = u_1 \in \mathbf{C}), \tag{1}$$

where $Q(z)$ and $F(z)$ are entire functions. A solution of (1) is a twice continuously differentiable function $u(z)$ defined for all $z \in \mathbf{C}$ and satisfying the given initial conditions. Since the equation is linear, the existence and uniqueness of solutions is well known, cf. [32]. About the recent results on solution estimates for ordinary differential equations see for instance the books [1, 35] and references given therein. For an entire function f and a positive number r put $M_f(r) = \sup_{|z|\le r} |f(z)|$.

Lemma 1 *A solution $u(z)$ of Equation (1) satisfies the inequality*

$$M_u(r) \le \left(|u_0| + r|u_1| + \int_0^r (r-s)M_F(s)ds\right) \cosh\left(r\sqrt{M_Q(r)}\right)$$

$$(r \ge 0, \cosh x = (e^x + e^{-x})/2, x \ge 0).$$

Proof For a fixed $t \in [0, 2\pi)$ and $z = re^{it}$ from (1) we have

$$\frac{1}{e^{2it}} \frac{d^2u(re^{it})}{dr^2} = Q(re^{it})u(re^{it}) + F(re^{it}).$$

Integrating twice this equation in r, we obtain

$$u(re^{it}) = e^{2it}(u_0 + ru_1) + e^{2it} \int_0^r (r-s)[Q(se^{it})u(se^{it}) + F(se^{it})]ds.$$

Hence,

$$M_u(r) \le |u_0| + r|u_1| + \int_0^r (r-s)(M_Q(s)M_u(s) + M_F(s))ds$$

$$= \int_0^r (r-s)M_Q(s)M_u(s)ds + H(r),$$

where

$$H(r) = |u_0| + r|u_1| + \int_0^r (r-s)M_F(s)ds.$$

Due to the comparison lemma [10, Lemma III.2.1], we have $M_u(r) \le v(r)$, where $v(r)$ is a solution of the equation

$$v(r) = H(r) + \int_0^r (r-s)M_Q(s)v(s)ds = H(r) + (Vv)(r).$$

Here V is the Volterra operator defined by

$$(Vv)(r) = \int_0^r (r-s)M_Q(s)v(s)ds = \int_0^r \int_0^{r_1} M_Q(r_2)v(r_2)dr_2dr_1,$$

and therefore,

$$v = \sum_{k=0}^{\infty} V^k H. \tag{2}$$

But for any positive nondecreasing function $h(r)$ we have

$$(Vh)(r) = \int_0^r \int_0^{r_1} M_Q(r_2)h(r_2)dr_2dr_1 \le h(r)M_Q(r) \int_0^r \int_0^{r_1} dr_2dr_1 = h(r)M_Q(r)r^2/2.$$

Similarly,

$$(V^2h)(r) = \int_0^r \int_0^{r_1} M_Q(r_2) \int_0^{r_2} \int_0^{r_3} M_Q(r_4)h(r_4)dr_4\, dr_3\, dr_2\, dr_1$$

$$\le h(r)M_Q^2(r) \int_0^r \int_0^{r_1} \int_0^{r_2} \int_0^{r_3} dr_4\, dr_3\, dr_2\, dr_1 = h(r)M_Q^2(r)r^4/4!$$

Continuous this process we obtain

$$(V^m h)(r) \le h(r)M_Q^m(r)\frac{r^{2m}}{(2m)!} \quad (m = 1, 2, \ldots).$$

Thus from (2) it follows

$$M_u(r) \le v(r) \le H(r) \sum_{k=0}^{\infty} \frac{M_Q^k(r) r^{2k}}{(2k)!}.$$

But

$$\sum_{k=0}^{\infty} \frac{M_Q^k(r) r^{2k}}{(2k)!} = \cosh(r\sqrt{M_Q(r)}).$$

This implies the required result. Q.E.D.

Since $\cosh(x) \le e^x$, $x \ge 0$, we obtain the following corollary.

Corollary 1 *A solution $u(z)$ of Equation (1) satisfies the inequality*

$$M_u(r) \le \left(|u_0| + r|u_1| + \int_0^r (r-s) M_F(s) ds \right) \exp\left(r\sqrt{M_Q(r)} \right) \quad (r \ge 0).$$

Now let $Q(z) = P(z)$, where

$$P(z) = \sum_{k=0}^{n} c_k z^k \quad (c_n \ne 0; c_k \in \mathbf{C}, k = 0, \ldots, n < \infty).$$

Consider the equation

$$u'' = P(z)u + F(z), \quad u(0) = u_0, u'(0) = u_1 \quad (u_0, u_1 \in \mathbf{C}). \tag{3}$$

In the considered case $M_Q(r) \le p(r)$, where

$$p(r) = \sum_{k=0}^{n} |c_k| r^k.$$

According to Corollary 1, a solution $u(z)$ of (3) satisfies the inequality

$$M_u(r) \le (|u_0| + r|u_1|) \exp\left(r\sqrt{p(r)} \right) \quad (r \ge 0). \tag{4}$$

Recall the Young inequality

$$ab \le \frac{a^m}{m} + \frac{a^l}{l} \quad (\frac{1}{m} + \frac{1}{l} = 1; \ m > 1; a, b > 0).$$

Then with $m = (n+2)/(k+2)$ we have

$$r^k \le \frac{kr^{n+2}}{n+2} + 1 - \frac{k}{n+2} = \frac{(k+2)r^{n+2}}{n+2} + \frac{n+2-(k+2)}{n+2}.$$

Hence

$$r^2 p(r) = \sum_{k=0}^{n} |c_k| r^{k+2} \le \sum_{k=0}^{n} |c_k| \left(\frac{(k+2)r^{n+2}}{n+2} + \frac{n-k}{n+2} \right) = b_0 r^{n+2} + b_1,$$

where

$$b_0 = \frac{1}{n+2} \sum_{k=0}^{n} |c_k|(k+2)$$

and

$$b_1 = \frac{1}{n+2} \sum_{k=0}^{n} |c_k|(n-k).$$

Since

$$\sqrt{a+b} \le \sqrt{a} + \sqrt{b} \quad (a, b \ge 0),$$

we get

$$r\sqrt{p(r)} = \sqrt{r^2 p(r)} \le \sqrt{b_0} r^{n/2+1} + \sqrt{b_1}.$$

Thus

$$\exp[r\sqrt{M_P(r)}] \le \exp[r\sqrt{p(r)}] \le \eta_P \exp[\gamma_P r^{n/2+1}],$$

where

$$\gamma_P = \sqrt{\frac{1}{n+2} \sum_{k=0}^{n} |c_k|(k+2)} \text{ and } \eta_P = \exp\left[\sqrt{\frac{1}{n+2} \sum_{k=0}^{n} |c_k|(n-k)} \right].$$

Now (4) implies

Corollary 2 *A solution $u(z)$ of Equation (3) satisfies the inequality*

$$M_u(r) \le \eta_P \left(|u_0| + r|u_1| + \int_0^r (r-s) M_F(s) ds \right) \exp[\gamma_P r^{n/2+1}] \quad (r \ge 0).$$

3 Singular Values of Compact Operators

Let $\mathscr{H}$ be a complex separable Hilbert space, cf. [24].

For a compact operator A acting in $\mathscr{H}$, by $\lambda_k(A)$ $(k = 1, 2, \ldots)$ we denote the eigenvalues of A taken with their multiplicities and ordered in the non-increasing way of their absolute values; $s_k(A)$ $(k = 1, 2, \ldots)$ are the singular numbers (i.e. the eigenvalues of $(A^*A)^{1/2}$), taken with their multiplicities and ordered in the non-increasing way. Here A^* is the operator adjoint to A. In the sequel we need the following well-known results, cf. [24, Section II.4.2], [23, Section IV.4].

Lemma 2 *Let A and B be compact operators in $\mathscr{H}$. Then*

$$\sum_{k=1}^{j} s_k(A+B) \le \sum_{k=1}^{j} (s_k(A) + s_k(B)),$$

$$\sum_{k=1}^{j} s_k^p(AB) \le \sum_{k=1}^{j} s_k^p(A) s_k^p(B) \quad (p \ge 1)$$

and

$$\prod_{k=1}^{j} s_k(AB) \le \prod_{k=1}^{j} s_k(A) s_k(B) \quad (j = 1, 2, \ldots).$$

In addition, if C and D are bounded linear operators in $\mathscr{H}$, then

$$s_k(CAD) \le \|C\| \|D\| s_k(A) \quad (k \ge 1),$$

where $\|C\|$ means the operator norm of C.

According to Corollary 2.2 from the book [23] the operators A and A^* have the same singular values.

Recall that A is said to be normal if $AA^* = A^*A$.

Lemma 3 (Weyl's Inequalities) *The inequalities*

$$\prod_{j=1}^{k} |\lambda_j(A)| \le \prod_{j=1}^{k} s_j(A)$$

and

$$\sum_{j=1}^{k} |\lambda_j(A)| \le \sum_{j=1}^{k} s_j(A) \quad (k = 1, 2, \ldots)$$

are true. They become equalities if and only if A is normal.

For the proof see Theorem IV.3.1 and Corollary IV.3.4 from [23], or Section II.3.1 from [24, Section II.4.2].

4 Bounds for Zeros of Entire Functions via Taylor Coefficients

Let us consider the entire function

$$h(z) = \sum_{k=0}^{\infty} \frac{a_k z^k}{(k!)^\alpha} \quad (0 < \alpha \le 1,\ z \in \mathbf{C},\ a_0 = 1, a_k \in \mathbf{C}, k \ge 1). \tag{5}$$

Enumerate the zeros $z_k(h)$ of h with the multiplicities in non-decreasing order of their absolute values: $|z_k(h| \le |z_{k+1}(h)|$ $(k = 1, 2, \ldots)$ and assume that

$$\theta(h) := [\sum_{k=1}^{\infty} |a_k|^2]^{1/2} < \infty. \tag{6}$$

The aim of this section is to prove the following theorem.

Theorem 1 *Let h be defined by (5) and condition (6) hold. Then*

$$\sum_{k=1}^{j} \frac{1}{|z_k(h)|} \le \theta(h) + \sum_{k=1}^{j} \frac{1}{(k+1)^\alpha} \ (j = 1, 2, \ldots)$$

and

$$\prod_{k=1}^{j} \frac{1}{|z_k(h)|} \le (\theta(h) + \frac{1}{2^\alpha}) \prod_{k=2}^{j} \frac{1}{(k+1)^\alpha} \ (j = 2, 3, \ldots).$$

To prove this theorem introduce the polynomial

$$f_n(z) = \sum_{k=0}^{n} \frac{a_k z^{n-k}}{(k!)^\alpha} \quad (0 < \alpha \le 1,\ z \in \mathbf{C},\ a_0 = 1, a_k \in \mathbf{C}, k > 1)$$

and the $n \times n$-matrix

$$F_n = \begin{pmatrix} -a_1 & -a_2 & \ldots & -a_{n-1} & -a_n \\ 1/(2^\alpha) & 0 & \ldots & 0 & 0 \\ 0 & 1/(3^\alpha) & \ldots & 0 & 0 \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & 1/(n^\alpha) & 0 \end{pmatrix}.$$

Let $z_k(f_n)$ $(k = 1, \ldots, n)$ be the zeros of f_n with their multiplicities enumerated in non-increasing order of their absolute values. and $\lambda_k(F_n)$ be the eigenvalues of F_n taken with the multiplicities enumerated in the non-increasing order of their absolute values.

Lemma 4 *One has* $\lambda_k(F_n) = z_k(f_n)$ $(k = 1, \ldots, n)$.

Proof Clearly, f_n is the characteristic polynomial of the matrix

$$B = \begin{pmatrix} -a_1 & -\frac{a_2}{2^\alpha} & \ldots & -\frac{a_{n-1}}{((n-1)!)^\alpha} & -\frac{a_n}{(n!)^\alpha} \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Following [17, Lemma 5.2.1, p. 117], put

$$m_k = \frac{1}{k^\alpha} \text{ and } \psi_k = \frac{1}{(k!)^\alpha} = m_1 m_2 \cdots m_k \quad (k = 1, \ldots, n).$$

Then

$$F_n = \begin{pmatrix} -a_1 & -a_2 & \ldots & -a_{n-1} & -a_n \\ m_2 & 0 & \ldots & 0 & 0 \\ 0 & m_3 & \ldots & 0 & 0 \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & m_n & 0 \end{pmatrix}$$

and

$$B = \begin{pmatrix} -a_1 & -a_2\psi_2 & \ldots & -a_{n-1}\psi_{n-1} & -a_n\psi_n \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Let μ be an eigenvalue of B, i.e. for the eigenvector $(x_k)_{k=1}^n \in \mathbf{C}^n$, we have

$$-a_1x_1 - a_2\psi_2x_2 - \ldots - a_{n-1}\psi_{n-1}x_{n-1} - a_n\psi_n x_n = \mu x_1,$$

$$x_k = \mu x_{k+1} \quad (k = 1, \ldots, n-1).$$

Put $x_k = y_k/\psi_k$. Since $\psi_1 = 1$, we obtain

$$-a_1y_1 - a_2y_2 \ldots - a_{n-1}y_{n-1} - a_n y_n = \mu y_1$$

and

$$\frac{y_k}{\psi_k} = \mu \frac{y_{k+1}}{\psi_{k+1}} \quad (k = 1, \dots, n-1).$$

Or

$$m_{k+1} y_k = \frac{y_k \psi_{k+1}}{\psi_k} = \mu y_{k+1} \quad (k = 1, \dots, n-1).$$

These equalities are equivalent to the equality $F_n y = \mu y$ with $y = (y_k)$. In other words $TBT^{-1} = F_n$, where $T = \operatorname{diag}(1, \psi_2, \dots, \psi_n)$ and therefore

$$T^{-1} = \operatorname{diag}(1, \frac{1}{\psi_2}, \dots, \frac{1}{\psi_n}).$$

This proves the lemma. Q.E.D.

Put

$$h_n(z) = z^n f_n(1/z) = \sum_{k=0}^{n} \frac{a_k z^k}{(k!)^\alpha}.$$

Then

$$z_k(h_n) = \frac{1}{z_k(f_n)} = \frac{1}{\lambda_k(F_n)} \quad (k = 1, \dots, n). \tag{7}$$

Here $z_k(h_n)$ are the zeros of h_n with their multiplicities enumerated in non-decreasing order of their absolute values.

Furthermore, note that $F_n = M + C$, where

$$M = \begin{pmatrix} -a_1 & -a_2 & \dots & -a_{n-1} & -a_n \\ 0 & 0 & \dots & 0 & 0 \\ . & . & \dots & . & . \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \text{ and } C = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1/2^\alpha & 0 & \dots & 0 & 0 \\ 0 & 1/3^\alpha & \dots & 0 & 0 \\ . & . & \dots & . & . \\ 0 & 0 & \dots & 1/n^\alpha & 0 \end{pmatrix}.$$

Therefore, with

$$\theta(h_n) = [\sum_{k=1}^{n} |a_k|^2]^{1/2},$$

we have

$$MM^* = \begin{pmatrix} \theta^2(h_n) & 0 \dots 0 & 0 \\ 0 & 0 \dots 0 & 0 \\ . & . \dots . & . \\ 0 & 0 \dots 0 & 0 \end{pmatrix} \text{ and } CC^* = \begin{pmatrix} 0 & 0 & \dots 0 & 0 \\ 0 & 1/2^{2\alpha} & \dots 0 & 0 \\ 0 & 0 & \dots 0 & 0 \\ . & . & \dots . & . \\ 0 & 0 & \dots 0 & 1/n^{2\alpha} \end{pmatrix}.$$

Hence, the singular values of M are $s_1(M) = \theta(h_n)$ and $s_k(M) = 0$ for $k > 1$. In addition,

$$s_k(C) = \frac{1}{(k+1)^\alpha} \quad (k = 1, \dots, n-1), s_n(C) = 0.$$

Thus, applying Corollary 2.2 from [24], we arrive at the following result.

Lemma 5 *The singular values $s_k(F_n)$ of F_n satisfy the inequalities*

$$s_1(F_n) \le \theta(h_n) + \frac{1}{2^\alpha} \text{ and } s_k(F_n) \le \frac{1}{(k+1)^\alpha} \text{ for } k = 2, \dots, n.$$

Hence due to Weyl's inequalities (see the previous section), we obtain

$$\sum_{k=1}^{j} |\lambda_k(F_n)| \le \theta(h_n) + \sum_{k=1}^{j} \frac{1}{(k+1)^\alpha} \quad (j = 1, \dots, n)$$

and

$$\prod_{k=1}^{j} |\lambda_k(F_n)| \le (\theta(h_n) + 1/2^\alpha) \prod_{k=2}^{j} \frac{1}{(k+1)^\alpha} \quad (j = 2, \dots, n).$$

Now (7) implies

$$\sum_{k=1}^{j} \frac{1}{|z_k(h_n)|} \le \theta(h_n) + \sum_{k=1}^{j} \frac{1}{(k+1)^\alpha} \quad (j = 1, \dots, n) \tag{8}$$

and

$$\prod_{k=1}^{j} \frac{1}{|z_k(h_n)|} \le (\theta(h_n) + 1/2^\alpha) \prod_{k=2}^{j} \frac{1}{(k+1)^\alpha} \quad (j = 2, \dots, n). \tag{9}$$

Proof of Theorem 1 In each compact domain $\Omega \in \mathbf{C}$, we have $h_n(z) \to h(z)$ $(n \to \infty)$ uniformly in Ω. Due to the Hurwitz theorem [33, p. 5] $z_k(h_n) \to z_k(h)$ for $z_k(h) \in \Omega$. Now (8) and (9) prove the theorem. Q.E.D.

5 Bounds for Sums and Products of Zeros of an Entire Function via Its Order

Lemma 6 *For an entire function $f(z)$ let the inequality*

$$M_f(r) \le \exp[Br^\rho] \ \ (B = const > 0; \ \rho \ge 1, r > 0) \tag{10}$$

be fulfilled. Then its Taylor coefficients f_j $(j = 1, 2, \ldots)$ satisfy the inequality

$$|f_j| \le \frac{(eB\rho)^{j/\rho}}{(j!)^{1/\rho}} \ \ (j \ge 1).$$

Proof By the well-known inequality for the coefficients of a power series

$$|f_j| \le \frac{M_f(r)}{r^j} \le \frac{e^{Br^\rho}}{r^j}.$$

Employing the usual method for finding extrema it is easy to see that the function in the right-hand side of this inequality takes its smallest value in the range $r > 0$ for $r = (\frac{j}{B\rho})^{1/\rho}$ and therefore

$$|f_j| \le (\frac{eB\rho}{j})^{j/\rho}.$$

Since $j^j \ge j!$, we obtain

$$|f_j| \le (eB\rho)^{j/\rho} \frac{1}{(j^j)^{1/\rho}} \le \frac{(eB\rho)^{j/\rho}}{(j!)^{1/\rho}},$$

as claimed. Q.E.D.

If an entire function $v(z)$ satisfies the inequality

$$M_v(r) \le r^m \exp[Br^\rho] \ \ (m = 1, 2, \ldots),$$

then we can write $v(z) = f(z)z^m$ with $f(z)$ satisfying (10). Then due Lemma 6 the Taylor coefficients at zero v_j of v satisfy the relations

$$|v_j| = |f_{j-m}| \le \frac{(eB\rho)^{(j-m)/\rho}}{((j-m)!)^{1/\rho}} \quad (j=m, m+1, \ldots) \text{ and } v_j = 0 \quad (j=0, 1, \ldots, m-1).$$

Hence we get

Corollary 3 *Let an entire function $v(z)$ be subject to the inequality*

$$M_v(r) \le \exp[B_v r^\rho](D_0 + D_1 r + D_2 r^2 + \ldots + D_m r^m)$$

$$(r > 0; D_j = \text{const} \ge 0, j = 1, \ldots, m < \infty). \tag{11}$$

Then its Taylor coefficients at zero v_j satisfy the inequalities

$$|v_j| \le \sum_{k=0}^{m} D_k \frac{(eB_v\rho)^{(j-k)/\rho}}{((j-k)!)^{1/\rho}} \ (j = 0, 1, \ldots).$$

Note that

$$\frac{1}{(j-k)!} = 0 \text{ if } k > j \text{ and } 0! = 1.$$

Put

$$d_j = v_j (j)^{1/\rho}, \xi = \xi_v := (eB_v\rho)^{1/\rho}$$

and

$$l_j := (j!)^{1/\rho} \sum_{k=0}^{m} D_k \frac{\xi_v^{j-k}}{((j-k)!)^{1/\rho}} \ (j = 1, \ldots).$$

Then due to Corollary 3 $|d_j| \le l_j$. Assume that

$$eB_v\rho < 1. \tag{12}$$

Then $\xi_v < 1$ and for $j < m$ we have

$$l_j \le (j!)^{1/\rho} \sum_{k=0}^{j} D_k \xi^{j-k} \le C_0 (j!)^{1/\rho},$$

where

$$C_0 := \sum_{k=0}^{m} D_k.$$

If $j \geq m$, then

$$l_j \leq (j!)^{1/\rho} \sum_{k=0}^{m} D_k \frac{\xi^{j-k}}{((j-k)!)^{1/\rho}} \leq C_0 j^m \xi^{j-m}.$$

Thus

$$\psi_1(v) := \sum_{j=1}^{\infty} l_j = \sum_{j=1}^{\infty} (j!)^{1/\rho} \sum_{k=0}^{m} D_k \frac{\xi^{j-k}}{((j-k)!)^{1/\rho}}$$

$$\leq C_0 (\sum_{j=0}^{m-1} (j!)^{1/\rho} + \sum_{j=m}^{\infty} \xi^{j-m} j^m) < \infty.$$

The following quantity plays an essential role hereafter:

$$\psi(v) := (\sum_{j=1}^{\infty} l_j^2)^{1/2} = \left[\sum_{j=1}^{\infty} \left((j!)^{1/\rho} \sum_{k=0}^{m} D_k \frac{\xi^{j-k}}{((j-k)!)^{1/\rho}} \right)^2 \right]^{1/2}.$$

Since $\psi(v) \leq \psi_1(v)$, we have $\psi(v) < \infty$. We thus have proved the following lemma.

Lemma 7 *Let an entire function $v(z)$ be subject to inequalities (11) and (12). Then it is representable as*

$$v(z) = \sum_{k=0}^{\infty} \frac{d_k z^k}{(k!)^{1/\rho}},$$

and

$$\theta(v) := \left[\sum_{j=1}^{\infty} |d_j|^2 \right]^{1/2} \leq \psi(v) < \infty.$$

Making use of Theorem 1 and the latter lemma we get our next result.

Theorem 2 *Let v be an entire function satisfying inequalities (11) and (12), and let $v(0) = 1$. Then*

$$\sum_{k=1}^{j} \frac{1}{|z_k(v)|} \leq \psi(v) + \sum_{k=1}^{j} \frac{1}{(k+1)^{1/\rho}} \ (j = 1, 2, \ldots)$$

and

$$\prod_{k=1}^{j} \frac{1}{|z_k(v)|} \le (\psi(v) + \frac{1}{2^{1/\rho}}) \prod_{k=2}^{j} \frac{1}{(k+1)^{1/\rho}} \ (j = 2, 3, \ldots).$$

Remark 1 Condition (12) is not very restrictive. If $eB\rho \ge 1$, then we can apply the substitution $z = wc$ with

$$c^{\rho} < \frac{1}{eB\rho}.$$

Indeed, put $v_c(w) = v(cw)$. Then condition (11) implies.

$$M_{v_c}(r) \le (D_0 + D_1 cr + \ldots + D_m cr^m) \exp[Bc^{\rho} r^{\rho}] \ (r > 0).$$

Condition (12) takes the form

$$eBc^{\rho}\rho < 1.$$

So if

$$c^{\rho} < \frac{1}{eB\rho},$$

then Theorem 2 can be applied to the function $v_c(z)$. Besides, $z_k(v) = cz_k(v_c)$.

6 Sums and Products of Zeros of Solutions to Homogeneous ODEs with Polynomial Coefficients

Again consider the equation

$$u''(z) = P(z)u(z) \ \ (u(0) = 1, u'(0) = u_1 \in \mathbf{C}, z \in \mathbf{C}), \tag{13}$$

where

$$P(z) = \sum_{k=0}^{n} c_k z^k \ \ (c_n \ne 0; c_k \in \mathbf{C}, k = 0, \ldots, n).$$

As is well-known [32], a solution $u(z)$ of (13) is an entire function whose order is no more than $n/2 + 1$, and the zeros $z_k(u)$ of $u(z)$ are simple. Enumerate the zeros of u in the nondecreasing order of their absolute values: $|z_k(u)| \le |z_{k+1}(u)|$ $(k = 1, 2, \ldots)$.

According to Corollary 2 $u(t)$ satisfies inequality (11) with

$$B = \gamma_P, \rho = n/2 + 1, D_0 = \eta_P, D_1 = \eta_P |u_1|, D_k = 0 \ \ (k \geq 2).$$

Recall that

$$\gamma_P = \sqrt{\frac{1}{n+2} \sum_{k=0}^{n} |c_k|(k+2)} \text{ and } \eta_P = \exp\left[\sqrt{\frac{1}{n+2} \sum_{k=0}^{n} |c_k|(n-k)}\right].$$

Condition (12) takes the form

$$e\gamma_P(n + 1/2) < 1. \tag{14}$$

Hence,

$$\xi_P := (e\gamma_P(n + 1/2))^{2/(2n+1)} < 1.$$

Simple calculations show that

$$\psi(u) = \eta_P [\sum_{j=1}^{\infty} \xi_P^{2j} \left(1 + |u_1| \frac{j^{2/(n+2)}}{\xi_P}\right)^2]^{1/2} < \infty.$$

Now Theorem 2 implies

Theorem 3 *If condition (14) holds, then the zeros of a solution $u(z)$ to Equation (13) satisfy the inequalities*

$$\sum_{k=1}^{j} \frac{1}{|z_k(u)|} \leq \psi(u) + \sum_{k=1}^{j} \frac{1}{(k+1)^{2/(n+2)}} \ (j = 1, 2, \ldots) \tag{15}$$

and

$$\prod_{k=1}^{j} \frac{1}{|z_k(u)|} \leq (\psi(u) + \frac{1}{2^{2/(n+2)}}) \prod_{k=2}^{j} \frac{1}{(k+1)^{2/(n+2)}} \ (j = 2, 3, \ldots).$$

Remark 2 Condition (14) is not very restrictive. If it does not hold: $e\gamma_P(n+1/2) \geq 1$, then one can apply the substitution $z = wb$ with

$$0 < b = \text{const} < \frac{1}{e\gamma_P(n/2+1)}. \tag{16}$$

Indeed, substituting $z = wb$ into (13), with $u_b(w) = u(bw)$ we have

$$\frac{1}{b^2} \frac{d^2 u_b(w)}{dw^2} = P(bw) u_b(w).$$

Or

$$\frac{d^2u_b(w)}{dw^2} = P_b(w)u_b(w), \tag{17}$$

where

$$P_b(w) = b^2P(bw) = \sum_{k=0}^{n} c_k b^{k+2} w^k.$$

If $e\gamma_P(n+1/2) \geq 1$, then due to (16) we have $b < 1$ and therefore

$$\gamma_{P_b} = \sqrt{\frac{1}{n+2}\sum_{k=0}^{n} b^{k+2}|c_k|(k+2)} \leq b\sqrt{\frac{1}{n+2}\sum_{k=0}^{n}|c_k|(k+2)} = \gamma_P b.$$

According to (16) condition (14) for Equation (17) is fulfilled:

$$e\gamma_{P_b}(n/2+1) < eb\gamma_P(n+1/2) < 1.$$

Therefore we can apply Theorem 3. Besides $z_k(u) = bz_k(u_b)$ $(k = 1, 2, \ldots)$.

Example 1 To estimate the sharpness of Theorem 3, consider the equation

$$u'' + a^2u = 0, u(0) = 1, u'(0) = 0 \quad (a = \text{const} > 0).$$

Then $u(z) = \cos(az)$ and its zeros are

$$\frac{\pi}{a}(m+1/2) \quad (m = 0, \pm1, \pm2, \ldots).$$

So

$$z_{2k}(cos(az)) = \frac{\pi}{a}(k-1/2), z_{2k-1} = \frac{\pi}{a}(-k+1/2) \quad (k = 1, 2, \ldots)$$

and

$$\sum_{k=1}^{2j} \frac{1}{|z_k(cos(az))|} = \frac{2a}{\pi}\sum_{k=1}^{j}\frac{1}{k-1/2} \quad (j = 1, 2, \ldots). \tag{18}$$

In the considered case $n = 0$, $\gamma_P = a$ and $\eta_P = 1$. Thus,

$$e\gamma_P(n+1/2) = \frac{ea}{2}.$$

So condition (14) holds, provided

$$a < \frac{2}{e}.$$

This condition yields $\xi_P := ea/2 < 1$ and $\psi(u) = \psi(cos(az))$, where

$$\psi(cos(az)) = [\sum_{k=1}^{\infty} \xi_P^{2k}]^{1/2} = [\sum_{k=1}^{\infty} (ea/2)^{2k}]^{1/2} = \frac{1}{[(2/ea)^2 - 1]^{1/2}}.$$

Now Theorem 3 implies

$$\sum_{k=1}^{2j} \frac{1}{|z_k(cos(az))|} \le \psi(cos(az)) + \sum_{k=1}^{2j} \frac{1}{k+1} \ (j = 1, 2, \ldots) \tag{19}$$

and

$$\prod_{k=1}^{2j} \frac{1}{|z_k(cos(az))|} \le (\psi(cos(az)) + \frac{1}{2}) \prod_{k=2}^{2j} \frac{1}{(k+1)} \ (j = 2, 3, \ldots).$$

We can see that (19) and (18) are asymptotically equivalent.

7 Applications of Theorem 3

Again $u(z)$ is a solution of Equation (13). Recall that $u(0) = 1$. Assume that $n \ge 1$ and for the brevity put

$$\alpha = \frac{2}{n+2}.$$

Since $|z_k(u)| \le |z_{k+1}(u)|$, Theorem 3 implies that

$$\frac{j}{|z_j(u)|} \le \psi(u) + \sum_{k=1}^{j} \frac{1}{(k+1)^{\alpha}} \quad (j = 1, 2, \ldots),$$

provided condition (14) holds. But

$$\sum_{k=1}^{j} (k+1)^{-\alpha} \le \int_1^{j+1} \frac{dx}{x^{\alpha}} = \frac{(1+j)^{1-\alpha} - 1}{1-\alpha} \quad (0 < \alpha < 1).$$

Thus,

$$\frac{j}{|z_j(u)|} \le \psi(u) + \frac{(1+j)^{1-\alpha} - 1}{1-\alpha}$$

and therefore,

$$|z_j(u)| \ge \chi_j(u), \tag{20}$$

where

$$\chi_j(u) = \frac{j}{\psi(u) + \frac{(1+j)^{1-\alpha}-1}{1-\alpha}}.$$

If $|z_j(u)| \ge a$ $(a > 0)$, then $u(z)$ has in $\Omega(a) := \{z \in \mathbf{C} : |z| < a\}$ no more than $j-1$ zeros. Denote by $\nu(f,a)$ the number of the zeros of an entire function f inside $\Omega(a)$, i.e. $\nu(f,a)$ is the counting function of the zeros of f. In particular, due to (20) $\nu(u,a) = 0$ for any positive

$$a < \chi_1(u) = \frac{1}{\psi(u) + \frac{2^{1-\alpha}-1}{1-\alpha}}.$$

We thus arrive at

Corollary 4 *Let condition (14) hold and $n \ge 1$. Then the counting function of the zeros of a solution $u(z)$ of (13) satisfies the inequality $\nu(u,a) \le j-1$ for any positive $a \le \chi_j(u)$ $(j = 1, 2, \ldots)$.*

Moreover the circle $\{z \in \mathbf{C} : |z| < \chi_1(u)\}$ is a zero-free domain of $u(z)$.

To consider additional applications of Theorem 3 recall some inequalities for convex functions. The following result is classical, cf. [24, Lemma II.3.4], [23, p. 53].

Lemma 8 *Let $\phi(x)$ $(-\infty \le x \le \infty)$ be a convex continuous function, such that*

$$\phi(-\infty) = \lim_{x \to -\infty} \phi(x) = 0,$$

and a_j, b_j $(j = 1, 2, \ldots, l \le \infty)$ be two non-increasing sequences of real numbers, such that

$$\sum_{k=1}^{j} a_k \le \sum_{k=1}^{j} b_k \quad (j = 1, 2, \ldots, l).$$

Then

$$\sum_{k=1}^{j} \phi(a_k) \le \sum_{k=1}^{j} \phi(b_k) \quad (j = 1, 2, \ldots, l).$$

The next result is also well known, cf. [24, Chapter II], [23, p. 53].

Lemma 9 *Let a scalar-valued function $\Phi(t_1, t_2, \ldots, t_j)$ with an integer j be defined on the domain*

$$-\infty < t_j \le t_{j-1} \ldots \le t_2 \le t_1 < \infty$$

and have continuous partial derivatives, satisfying the condition

$$\frac{\partial \Phi}{\partial t_1} > \frac{\partial \Phi}{\partial t_2} > \ldots > \frac{\partial \Phi}{\partial t_j} > 0 \text{ for } t_1 > t_2 > \ldots > t_j,$$

and $a_k, b_k \quad (k = 1, 2, \ldots, j)$ be two non-increasing sequences of real numbers satisfying the condition

$$\sum_{k=1}^{m} a_k \le \sum_{k=1}^{m} b_k \quad (m = 1, 2, \ldots, j).$$

Then $\Phi(a_1, \ldots, a_j) \le \Phi(b_1, \ldots, b_j)$.

Furthermore, put

$$\vartheta_1 = \psi(u) + \frac{1}{2^\alpha} \text{ and } \vartheta_k = \frac{1}{(k+1)^\alpha} \quad (k = 2, 3, \ldots).$$

Inequality (15) and Lemma 8 yield

Corollary 5 *Let $\phi(t) \quad (0 \le t < \infty)$ be a continuous convex function, such that $\phi(0) = 0$. Then*

$$\sum_{k=1}^{j} \phi\Big(\frac{1}{|z_k(u)|}\Big) \le \sum_{k=1}^{j} \phi(\vartheta_k) \quad (j = 1, 2, \ldots).$$

In particular, for any $p \ge 1$ and $j = 2, 3, \ldots$, we have

$$\sum_{k=1}^{j} \frac{1}{|z_k(u)|^p} \le \sum_{k=1}^{j} \vartheta_k^p$$

and therefore, if $p > 1/\alpha = n/2 + 1$, then

$$\sum_{k=1}^{\infty} \frac{1}{|z_k(u)|^p} \le \vartheta_1^p + \zeta(p\alpha) - 1 - \frac{1}{2^{p\alpha}} < \infty,$$

where $\zeta(z)$ is the zeta Riemann function:

$$\zeta(z) = \sum_{k=1}^{\infty} \frac{1}{k^z} \quad (\text{Re } z > 1).$$

In addition, making use of (15) and Lemma 9, we arrive at

Corollary 6 *Let $\Phi(t_1, t_2, \ldots, t_j)$ satisfy the hypothesis of Lemma 9. Then*

$$\Phi\left(\frac{1}{|z_1(u)|}, \ldots, \frac{1}{|z_j(u)|}\right) \le \Phi(\vartheta_1, \ldots, \vartheta_j).$$

In particular, let $\{m_k\}_{k=1}^{\infty}$ be a decreasing sequence of positive numbers with $m_1 = 1$. Then the previous corollary yields the inequality

$$\sum_{k=1}^{j} \frac{m_k}{|z_k(u)|} \le \psi(u) + \sum_{k=1}^{j} \frac{m_k}{(k+1)^{\alpha}} \quad (j = 1, 2, \ldots).$$

8 A Perturbation Bound for the Zeros of Entire Functions in Terms of Taylor Coefficients

Definition 1 Let $z_j(h)$ and $z_j(\tilde{h})$ $(j = 1, 2, \ldots)$ be the zeros of entire functions h and $\tilde{h}$, respectively, enumerated with the multiplicities in the non-decreasing order of their absolute values. Then the quantity

$$\mathrm{rv}_h(\tilde{h}) = \sup_j \inf_k \left| \frac{1}{z_k(h)} - \frac{1}{z_j(\tilde{h})} \right|$$

will be called the relative variation of the zeros of $\tilde{h}$ with respect to the zeros of h.

In this section we consider entire functions of the form

$$h(z) = \sum_{k=0}^{\infty} \frac{a_k z^k}{(k!)^{\alpha}} \text{ and } \tilde{h}(z) = \sum_{k=0}^{\infty} \frac{\tilde{a}_k z^k}{(k!)^{\alpha}} \quad (a_0 = \tilde{a}_0 = 1,\ 0 < \alpha \le 1) \tag{21}$$

with complex coefficients $a_k, \tilde{a}_k$ $(k = 1, 2, \ldots)$, assuming that

$$\theta(h) = \Big(\sum_{k=1}^{\infty} |a_k|^2\Big)^{1/2} < \infty \text{ and } \theta(\tilde{h}) = \Big(\sum_{k=1}^{\infty} |\tilde{a}_k|^2\Big)^{1/2} < \infty. \tag{22}$$

To investigate the relative variation of the zeros take an integer p satisfying the inequality

$$p > \frac{1}{2\alpha} \tag{23}$$

and put

$$\varpi_p(h) := 2\left[(\theta(h) + \frac{1}{2^\alpha})^{2p} + \zeta(2\alpha p) - 1 - \frac{1}{2^{2\alpha p}}\right]^{1/2p}.$$

Rcall that $\zeta(z)$ is the Riemann zeta function. Finally, denote

$$q := \left[\sum_{k=1}^{\infty} |a_k - \tilde{a}_k|^2\right]^{1/2}$$

and

$$\chi_p(h, y) := \sum_{k=0}^{p-1} \frac{\varpi_p^k(h)}{y^{k+1}} \exp\left[\frac{1}{2} + \frac{\varpi_p^{2p}(h)}{2y^{2p}}\right] \quad (y > 0).$$

Theorem 4 *Let h and $\tilde{h}$ be defined by (21) and conditions (22) be fulfilled. Then for any integer p satisfying inequality (23) we have*

$$\mathrm{rv}_h(\tilde{h}) \le y_p(q, h),$$

where $y_p(q, h)$ is the unique positive root of the equation

$$q\chi_p(h, y) = 1.$$

The proof of this theorem is presented in the next section. Since $\chi_p(h, .)$ is a monotonically decreasing function, the latter theorem implies

Corollary 7 *Let h and $\tilde{h}$ be defined by (21), and conditions (22) and (23) be fulfilled. Then all the zeros of $\tilde{h}$ lie in the union of the sets $W_k(p, h)$ $(k = 1, 2, 3, \ldots)$, where*

$$W_k(p, h) := \left\{\lambda \in \mathbf{C} : q\chi_p\left(h, |\frac{1}{z_k(h)} - \frac{1}{\lambda}|\right) \ge 1\right\}.$$

In particular, if h has $l < \infty$ zeros (we do not take into account the zero limits at infinity), then all the zeros of $\tilde{h}$ lie in the set

$$\cup_{k=0}^{l} W_k(p, h),$$

where

$$W_0(p,h) = \{\lambda \in \mathbf{C} : q\chi_p(h, \frac{1}{|\lambda|}) \geq 1\}.$$

Note that

$$\chi_p\left(h, \frac{1}{|\lambda|}\right) = \sum_{k=0}^{p-1} \varpi_p^k(h)|\lambda|^{k+1} \ \exp\left[\frac{1}{2}(1 + \varpi_p^{2p}(h)|\lambda|^{2p})\right].$$

Furthermore, we need the following lemma.

Lemma 10 *The unique positive root z_a of the equation*

$$\sum_{j=0}^{p-1} \frac{1}{y^{j+1}} \ exp\ [\frac{1}{2}(1 + \frac{1}{y^{2p}})] = a \ \ (a = const > 0)$$

satisfies the inequality $z_a \leq \delta_p(a)$, where

$$\delta_p(a) := \begin{cases} pe/a & if\, a \leq pe, \\ [\ln\,(a/p)]^{-1/2p} & if\, a > pe \end{cases}.$$

For the proof see [16, Lemma 8.3.3] or Lemma 1.6.4 from [17]. Substitute the equality $y = x\varpi_p(h)$ into the equation $q\chi_p(h, y) = 1$, and apply the latter lemma. Then, we have the inequality

$$y_p(q,h) \leq \delta_p(q,h), \tag{24}$$

where

$$\delta_p(q,h) := \begin{cases} epq & \text{if } \varpi_p(h) \leq epq, \\ \varpi_p(h)\ [\ln\,(\varpi_p(h)/qp)]^{-1/2p} & \text{if } \varpi_p(h) > epq. \end{cases}$$

Now Theorem 4 yields the inequality

$$\mathrm{rv}_f(\tilde{h}) \leq \delta_p(q,h).$$

If h has an infinite set of zeros, then according to Theorem 4 for any zero $z(\tilde{h})$ of $\tilde{h}$, there is a zero $z(h)$ of h, such that

$$|z(\tilde{h}) - z(h)| \leq y_p(q,h)|z(\tilde{h})z(h)|$$

Hence,

$$|z(h)| \le |z(\tilde{h})|(1 + y_p(q,h)|z(h)|)$$

and therefore, for any zero $z(\tilde{h})$ of $\tilde{h}$, there is a zero $z(h)$ of h, such that

$$|z(\tilde{h})| \ge \frac{|z(h)|}{y_p(q,h)|z(h)| + 1}.$$

Now (24) implies

$$|z(\tilde{h})| \ge \frac{|z(h)|}{\delta_p(q,h)|z(h)| + 1}.$$

9 Proof of Theorem 4

For an integer $n \ge 2$, consider the polynomials

$$h_n(z) = \sum_{k=0}^{n} \frac{a_k z^k}{(k!)^\alpha} \text{ and } \tilde{h}_n(z) = \sum_{k=0}^{n} \frac{\tilde{a}_k z^k}{(k!)^\alpha}.$$

Put

$$\theta(h_n) := \left[\sum_{k=1}^{n} |a_k|^2\right]^{1/2}, \varpi_p(h_n) := 2\left[(\theta(h_n) + \frac{1}{2^{p\alpha}})^{2p} + \sum_{k=2}^{n} \frac{1}{(k+1)^{2p\alpha}}\right]^{1/2p},$$

$$\chi_p(h_n, y) := \sum_{k=0}^{p-1} \frac{\varpi_p^k(h_n)}{y^{k+1}} \exp\left[\frac{1}{2} + \frac{\varpi_p^{2p}(h_n)}{2y^{2p}}\right] \quad (y > 0),$$

and

$$q_n := \left[\sum_{k=1}^{n} |a_k - \tilde{a}_k|^2\right]^{1/2}.$$

Let

$$f_n(\lambda) = \lambda^n h_n(1/\lambda) = \sum_{k=0}^{n} \frac{a_k}{k^\alpha} \lambda^{n-k} \text{ and } \tilde{f}_n(\lambda) = \lambda^n \tilde{h}_n(1/\lambda) = \sum_{k=0}^{n} \frac{\tilde{a}_k}{k^\alpha} \lambda^{n-k}.$$

Lemma 11 *For any zero $z(\tilde{f}_n)$ of $\tilde{f}_n$, there is a zero $z(f_n)$ of f_n, such that*

$$|z(f_n) - z(\tilde{f}_n)| \le y_p(q_n, f_n),$$

where $y_p(q_n, f_n)$ s the unique (positive) root of the equation

$$q_n \chi_p(h_n, y) = 1.$$

Proof Introduce the matrices

$$F_n = \begin{pmatrix} -a_1 & -a_2 & \dots & -a_{n-1} & -a_n \\ \frac{1}{2^\alpha} & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{3^\alpha} & \dots & 0 & 0 \\ . & . & \dots & . & . \\ 0 & 0 & \dots & \frac{1}{n^\alpha} & 0 \end{pmatrix}$$

and

$$\tilde{F}_n = \begin{pmatrix} -\tilde{a}_1 & -\tilde{a}_2 & \dots & -\tilde{a}_{n-1} & -\tilde{a}_n \\ \frac{1}{2^\alpha} & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{3^\alpha} & \dots & 0 & 0 \\ . & . & \dots & . & . \\ 0 & 0 & \dots & \frac{1}{n^\alpha} & 0 \end{pmatrix}.$$

Thanks to Lemma 4, we have

$$\lambda_k(F_n) = z_k(f_n), \ \lambda_k(\tilde{F}_n) = z_k(\tilde{f}_n) \quad (k = 1, 2, \dots, n),$$

where $\lambda_k(A), k = 1, \dots, n$ are the eigenvalues of an $n \times n$ matrix A with their multiplicities. Besides,

$$\tilde{F}_n - F_n = \begin{pmatrix} \tilde{a}_1 - a_1 & \tilde{a}_2 - a_2 & \dots & \tilde{a}_{n-1} - a_{n-1} & \tilde{a}_n - a_n \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ . & . & \dots & . & . \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

and

$$q_n = \|F_n - \tilde{F}_n\|,$$

where $\|.\|$ is the spectral norm, i.e. the operator norm with respect to the Euclidean vector norm. Making use of [17, Theorem 2.12.4], we can assert that for any $\lambda_j(\tilde{F}_n)$ there is a $\lambda_k(F_n)$, such that

$$|\lambda_j(\tilde{F}_n) - \lambda_k(F_n)| \le y_p(q_n), \tag{25}$$

where $y_p(q_n)$ is the unique (positive) root of the equation

$$q_n \sum_{k=0}^{p-1} \frac{(2N_{2p}(F_n))^k}{y^{k+1}} \exp\left[\frac{1}{2} + \frac{(2N_{2p}(F_n))^{2p}}{2y^{2p}}\right] = 1.$$

Here

$$N_{2p}(F_n) = [\sum_{k=1}^{n} s_k^{2p}(F_n)]^{1/2p}$$

is the Schatten-von Neumann norm of F_n. Due to Lemma 5,

$$s_1(F_n) \le \theta(h_n) + \frac{1}{2^\alpha}$$

and

$$s_k(F_n) \le \frac{1}{(k+1)^\alpha}$$

for $k > 1$. Thus,

$$N_{2p}^{2p}(F_n) \le (\theta(h_n) + \frac{1}{2^\alpha})^{2p} + \sum_{k=2}^{n} \frac{1}{(k+1)^{2p\alpha}} = \varpi_p^{2p}(h_n).$$

Now (25) implies the required result. Q.E.D.

Proof of Theorem 4 Since

$$z_k(h_n) = \frac{1}{z_k(f_n)} \text{ and } z_k(\tilde{h}_n) = \frac{1}{z_k(f_n)},$$

taking into account that $\varpi_p(h_n) \le \varpi_p(h)$, and the roots continuously depend on coefficients, we have the required result, letting in the previous lemma $n \to \infty$. Q.E.D.

10 A Perturbation Bound for the Zeros of an Entire Function via Its Order

Let $h(z)$ and $\tilde{h}(z)$ be entire functions with $h(0) = \tilde{h}(0) = 1$. Put $v(z) = h(z) - \tilde{h}(z)$ and assume that

$$M_h(r) \le \exp[B_h r^\rho](C_0 + C_1 r + C_2 r^2 + \ldots + C_s)$$

$$(r > 0;\ B_h, C_j = \text{const} \ge 0,\ j = 0, 1, \ldots, s < \infty), \tag{26}$$

and

$$M_v(r) \le \exp[B_v r^\rho](D_1 r + D_2 r^2 + \ldots + D_m r^m)$$

$$(r > 0;\, B_v,\, D_j = \text{const} \ge 0,\, j = 1, \ldots, m < \infty). \tag{27}$$

By Corollary 3 the Taylor coefficients at zero v_j of v satisfy the inequalities

$$|v_j| \le \sum_{k=1}^{m} D_k \frac{\xi_v^{j-k}}{((j-k)!)^{1/\rho}} \ (j = 1, \ldots),$$

where $\xi_v = (eB_v\rho)^{1/\rho}$. Similarly, the Taylor coefficients at zero h_j of h satisfy the inequalities

$$|h_j| \le \sum_{k=0}^{s} C_k \frac{\xi_h^{j-k}}{((j-k)!)^{1/\rho}} \ (j = 1, 2 \ldots),$$

where $\xi_h := (eB_h\rho)^{1/\rho}$. Recall that $\frac{1}{(j-k)!} = 0$ if $k > j$ and $0! = 1$. If

$$B_h \le B_v \text{ and } eB_v\rho < 1, \tag{28}$$

then $\xi_v < 1$ and $\xi_h < 1$. Besides,

$$\psi(v) = \left[\sum_{j=1}^{\infty} (j!)^{2/\rho} \left(\sum_{k=1}^{m} D_k \frac{\xi_v^{j-k}}{((j-k)!)^{1/\rho}}\right)^2\right]^{1/2} < \infty$$

and

$$\psi(h) = \left[\sum_{j=1}^{\infty} (j!)^{2/\rho} \left(\sum_{k=0}^{s} C_k \frac{\xi_h^{j-k}}{((j-k)!)^{1/\rho}}\right)^2\right]^{1/2} < \infty$$

(see Section 5). Due to Lemma 7 $v(z)$ is representable as

$$v(z) = \sum_{k=0}^{\infty} \frac{d_k z^k}{(k!)^{1/\rho}}, \tag{29}$$

with

$$\theta^2(v) = \sum_{j=1}^{\infty} |d_j|^2 \le \psi^2(v)$$

and $h(z)$ is representable as

$$h(z) = \sum_{k=0}^{\infty} \frac{a_k z^k}{(k!)^{1/\rho}},$$

with

$$\theta^2(h) = \sum_{j=1}^{\infty} |a_j|^2 \le \psi^2(h).$$

Since $v(z) = h(z) - \tilde{h}(z)$, the Taylor coefficients $\tilde{h}_k$ of $\tilde{h}$ are equal to $h_k - v_k$ and due to (29) we can write

$$\tilde{h}(z) = \sum_{k=0}^{\infty} \frac{\tilde{a}_k z^k}{(k!)^{1/\rho}},$$

with

$$\theta^2(\tilde{h}) = \sum_{j=1}^{\infty} |\tilde{a}_j|^2 < \infty.$$

Besides, in (29) we have $d_j = a_j - \tilde{a}_j$. So

$$q = \left[\sum_{k=1}^{\infty} |a_k - \tilde{a}_k|^2\right]^{1/2} \le \psi(v).$$

Furthermore, according to (23) for an integer p satisfying the inequality

$$p > \rho/2 \tag{30}$$

we have

$$\varpi_p(h) = 2\left[(\theta(h) + \frac{1}{2^{1/\rho}})^{2p} + \zeta(2p/\rho) - 1 - \frac{1}{2^{2p/\rho}}\right]^{1/2p} \le \pi_p(h),$$

where

$$\pi_p(h) := 2\left[(\psi(h) + \frac{1}{2^{1/\rho}})^{2p} + \zeta(2p/\rho) - 1 - \frac{1}{2^{2p/\rho}}\right]^{1/2p}.$$

Hence $\chi_p(h, y) \le \tau_p(h, y)$, where

$$\tau_p(h, y) = \sum_{k=0}^{p-1} \frac{\pi_p^k(h)}{y^{k+1}} \exp\left[\frac{1}{2} + \frac{\pi_p^{2p}(h)}{2y^{2p}}\right] \quad (y > 0).$$

Now Lemma 10 implies

Theorem 5 *Let entire functions h and $\tilde{h}$ satisfy the conditions $h(0) = \tilde{h}(0) = 1$, (26), (27) and (28). Then for any integer p satisfying inequality (30) we have $\mathrm{rv}_h(\tilde{h}) \le x_p(v, h)$, where $x_p(v, h)$ is the unique positive root of the equation*

$$\psi(v)\tau_p(h, y) = 1.$$

Since $\tau_p(h, .)$ is a monotonically decreasing function, the latter theorem implies that all the zeros of $\tilde{h}$ lie in the union of the sets $\hat{W}_k(h, \psi(v))$ $(k = 1, 2, 3, \ldots)$, where

$$\hat{W}_k := \left\{\lambda \in \mathbf{C} : \psi(v)\tau_p\left(h, |\frac{1}{z_k(h)} - \frac{1}{\lambda}|\right) \ge 1\right\}.$$

In particular if h has a finite number l of zeros, then all the zeros of $\tilde{h}$ lie in the set

$$\cup_{k=0}^{l} \hat{W}_k(h, \psi(v)),$$

where

$$\hat{W}_0(h, \psi(v)) = \{\lambda \in \mathbf{C} : \psi(v)\tau_p(h, \frac{1}{|\lambda|}) \ge 1\}.$$

Note that

$$\tau_p\left(h, \frac{1}{|\lambda|}\right) = \sum_{k=0}^{p-1} \pi_p^k(h)|\lambda|^{k+1} \exp\left[\frac{1}{2}(1 + \pi_p^{2p}(h)|\lambda|^{2p})\right].$$

Furthermore, according to (24) we have the inequality

$$x_p(v, h) \le \omega_p(v, h), \tag{31}$$

where

$$\omega_p(v, h) := \begin{cases} ep\psi(v) & \text{if } \pi_p(h) \le ep\psi(v), \\ \pi_p(h)[\ln(\frac{\pi_p(h)}{\psi(v)p})]^{-1/2p} & \text{if } \pi_p(h) > ep\psi(v) \end{cases}.$$

Now Theorem 5 yields the inequality $\mathrm{rv}_f(\tilde{h}) \le \omega_p(v, h)$.

If h has an infinite set of zeros, then according to Theorem 5 and (31), for any zero $z(\tilde{h})$ of $\tilde{h}$, there is a zero $z(h)$ of h, such that

$$|z(\tilde{h}) - z(h)| \le x_p(v,h)|z(\tilde{h})z(h)| \le \omega_p(v,h)|z(\tilde{h})z(h)|.$$

These relations imply the inequalities

$$|z(h)| - |z(\tilde{h})| \le x_p(v,h)|z(\tilde{h})z(h)| \le \omega_p(v,h)|z(\tilde{h})z(h)|.$$

Hence,

$$|z(\tilde{h})| \ge \frac{|z(h)|}{x_p(v,h)|z(h)| + 1} \ge \frac{|z(h)|}{\omega_p(v,h)|z(h)| + 1}.$$

11 An Estimate for the Difference of Two Solutions

Consider the equations

$$u'' = P(z)u, \tag{32}$$

and

$$\tilde{u}'' = \tilde{P}(z)\tilde{u} \quad (z \in \mathbf{C}), \tag{33}$$

where

$$P(z) = \sum_{k=0}^{n} c_k z^k \text{ and } \tilde{P}(z) = \sum_{k=0}^{n} \tilde{c}_k z^k \quad (c_n \ne 0; \tilde{c}_n \ne 0)$$

are polynomials with complex coefficients.

In the sequel it is assumed that

$$\tilde{u}(0) = u(0) = 1, \ \tilde{u}'(0) = u'(0) = u_1 \quad (u_1 \in \mathbf{C}). \tag{34}$$

Recall that numbers η_P and γ_P are defined in Section 2.

Lemma 12 *Let $u(z)$ and $\tilde{u}(z)$ be the solutions of (32) and (33), respectively. Let the initial conditions (34) hold. Then*

$$|u(z) - \tilde{u}(z)| \le \sum_{k=0}^{n} \frac{|c_k - \tilde{c}_k| r^{k+2}}{(k+1)(k+2)} (1 + r|u_1|)\eta_P \eta_{\tilde{P}} \exp[(\gamma_P + \gamma_{\tilde{P}})r^{n/2+1}]$$

$$(|z| \le r, r > 0).$$

Proof Put $w(z) = u(z) - \tilde{u}(z)$. Then

$$w'' = P(z)w + (P(z) - \tilde{P}(z))\tilde{u}.$$

We have

$$|P(z) - \tilde{P}(z)| \leq \hat{v}(r) \quad (|z| = r, r > 0),$$

where

$$\hat{v}(r) := \sum_{k=0}^{n} |c_k - \tilde{c}_k| r^k.$$

Due to Corollary 2

$$M_{\tilde{u}}(r) \leq \eta_{\tilde{P}}(1 + r|u_1|) \exp[\gamma_{\tilde{P}} r^{n/2+1}].$$

Thus,

$$|(P(z) - \tilde{P}(z))\tilde{u}(z)| \leq \beta(r) \quad (|z| \leq r),$$

where

$$\beta(r) := \hat{v}(r)\eta_{\tilde{P}}(1 + r|u_1|) \exp[\gamma_{\tilde{P}} r^{n/2+1}].$$

Since $w(0) = w'(0) = 0$, due to Corollary 1,

$$M_w(r) \leq \eta_P \exp[\gamma_P r^{n/2+1}] \int_0^r (r - s)\beta(s)ds$$

$$\leq (1 + r|u_1|)\eta_P \eta_{\tilde{P}} \exp[(\gamma_P + \gamma_{\tilde{P}})r^{n/2+1}] \int_0^r (r - s)\hat{v}(s)ds.$$

But

$$\int_0^r (r - s)\hat{v}(s)ds \leq \sum_{k=0}^{n} |c_k - \tilde{c}_k| \int_0^r (r - s)s^k ds = \sum_{k=0}^{n} |c_k - \tilde{c}_k| \frac{r^{k+2}}{(k + 1)(k + 2)}.$$

Thus $w(z)$ satisfies the inequality

$$M_w(r) \leq \sum_{k=0}^{n} |c_k - \tilde{c}_k| \frac{r^{k+2}}{(k + 1)(k + 2)} (1 + r|u_1|)\eta_P \eta_{\tilde{P}} \exp[(\gamma_P + \gamma_{\tilde{P}})r^{n/2+1}],$$

as claimed. Q.E.D.

Put $b_0 = \eta_P \eta_{\tilde{P}}$. Then due to the latter lemma we can write

$$|w(z)| = |u(z) - \tilde{u}(z)| \le \sum_{k=2}^{n+3} g_k r^k \ \exp[(\gamma_P + \gamma_{\tilde{P}}) r^{n/2+1}] \ \ (|z| \le r, r > 0), \qquad (35)$$

where

$$g_2 = \frac{b_0}{2}|c_0 - \tilde{c}_0|, \, g_3 = b_0 \left(\frac{|c_1 - \tilde{c}_1|}{6} + |u_1| \frac{|c_0 - \tilde{c}_0|}{2} \right),$$

$$g_4 = b_0 \left(\frac{|c_2 - \tilde{c}_2|}{12} + |u_1| \frac{|c_1 - \tilde{c}_1|}{6} \right), \ \ldots, \, g_{n+1} = b_0 \left(\frac{|c_n - \tilde{c}_n|}{(n+1)(n+2)} + |u_1| \frac{|c_{n-1} - \tilde{c}_{n-1}|}{(n-1)n} \right),$$

$$g_{n+3} = \frac{b_0}{(n+1)(n+2)} |u_1| |c_n - \tilde{c}_n|.$$

We thus arrive at our next result.

Corollary 8 *Let $u(z)$ and $\tilde{u}(z)$ be the solutions of (32) and (33), respectively. Let the initial condition (34) hold. Then inequality (35) is valid.*

12 Perturbations of the Zeros of Solutions to ODEs

As above $z_k(u)$ and $z_k(\tilde{u})$ $(k = 1, 2, \ldots)$ are the zeros of the solutions u and $\tilde{u}$ to (32) and (33), respectively, enumerated in the non-decreasing order of their absolute values. To apply Theorem 5 to perturbations of the zeros of solutions to the considered equations, note that due to Corollary 2,

$$M_u(r) \le \eta_P (1 + r|u_1|) \exp[\gamma_P r^{n/2+1}].$$

So taking in Theorem 5 $h(z) = u(z)$ and $v(z) = w(z)$, where $w(z) = u(z) - \tilde{u}(z)$, we have

$$B_u = \gamma_P, \, \rho = n/2 + 1 \text{ and } \xi_u = \xi_P,$$

where

$$\xi_P := (e\gamma_P (n/2 + 1))^{2/(n+2)}.$$

Besides, $C_0 = \eta_P$, $C_1 = \eta_P |u_1|$ and $C_k = 0$ overwise. Similarly, due to (35) we have

$$B_w = \gamma_P + \gamma_{\tilde{P}} \text{ and } \xi_w = \xi(P, \tilde{P}),$$

where

$$\xi(P, \tilde{P}) := (e(\gamma_P + \gamma_{\tilde{P}})(n/2+1))^{2/(n+2)}.$$

In addition, $D_k = g_k$ $(k = 2, \ldots, n+3)$ and $D_k = 0$ overwise. If

$$e(\gamma_P + \gamma_{\tilde{P}})(n/2+1) < 1, \tag{36}$$

then $\xi(P, \tilde{P}) < 1$ and $\xi_P < 1$, and therefore, $\psi(w) = \Delta$, where

$$\Delta := \left[\sum_{j=1}^{\infty}(j!)^{4/(n+2)}\left(\sum_{k=2}^{n+1} g_k \frac{\xi^{j-k}(P, \tilde{P})}{((j-k)!)^{4/(n+2)}}\right)^2\right]^{1/2} < \infty$$

and

$$\psi(u) = \eta_P\left[\sum_{j=1}^{\infty}(j!)^{4/(n+2)}\left(\frac{\xi_P^j}{(j!)^{4/(n+2)}} + |u_1|\frac{\xi_P^{j-1}}{(j-1)!)^{4/(n+2)}}\right)^2\right]^{1/2} < \infty.$$

Furthermore, for an integer p satisfying the inequality

$$p > \frac{n+2}{4}, \tag{37}$$

we can write

$$\pi_p(u) = 2\left[(\psi(u) + \frac{1}{2^{2/(n+2)}})^{2p} + \zeta(4p/(n+2)) - 1 - \frac{1}{2^{4p/(n+2)}}\right]^{1/2p}$$

and

$$\tau_p(u, y) = \sum_{k=0}^{p-1}\frac{\pi_p^k(u)}{y^{k+1}}\exp\left[\frac{1}{2} + \frac{\pi_p^{2p}(u)}{2y^{2p}}\right] \quad (y > 0).$$

Now Theorem 5 implies

Theorem 6 *Let $u(z)$ and $\tilde{u}(z)$ be the solutions to Equations (32) and (33), respectively. Let conditions (34) and (36) hold. Then for an integer p satisfying the inequality (37) one has*

$$\mathrm{rv}_u(\tilde{u}) \le x_p(\Delta, u),$$

where $x_p(\Delta, u)$ is the unique positive root of the equation

$$\Delta \cdot \tau_p(u, y) = 1.$$

According to Remark 2 condition (36) is not very restrictive. If it does not hold, i.e. if $e(\gamma_P + \gamma_{\tilde{P}})(n + 1/2) \geq 1$, then one can apply the substitution $z = w\hat{b}$ with

$$0 < \hat{b} = \text{const } < \frac{1}{e(\gamma_P + \gamma_{\tilde{P}})(n/2 + 1)}.$$

Furthermore, according to (24) we have the inequality

$$x_p(\Delta, u) \leq \omega_p(\Delta, w), \tag{38}$$

where

$$\omega_p(\Delta, w) := \begin{cases} ep\Delta & \text{if } \pi_p(u) \leq ep\Delta, \\ \pi_p(u)\ [\ln\ (\pi_p(u)/(p\Delta)]^{-1/2p} & \text{if } \pi_p(u) > ep\Delta \end{cases}.$$

Now Theorem 6 yields the inequality

$$\text{rv}_u(\tilde{u}) \leq \omega_p(u, w).$$

Since $\tau_p(u, .)$ is a monotonically decreasing function, the latter theorem implies that all the zeros of $\tilde{u}$ lie in the union of the sets $\hat{W}_k(u, \Delta)$ $(k = 1, 2, 3, \ldots)$, where

$$\hat{W}_k(u, \Delta) := \left\{\lambda \in \mathbf{C} : \Delta \cdot \tau_p\left(u, |\frac{1}{z_k(u)} - \frac{1}{\lambda}|\right) \geq 1\right\}.$$

In particular, if u has $l < \infty$ zeros, then all the zeros of $\tilde{u}$ lie in the set

$$\cup_{k=0}^{l} \hat{W}_k(u, \Delta),$$

where

$$\hat{W}_0(u, \Delta) = \{\lambda \in \mathbf{C} : \Delta \cdot \tau_p(u, \frac{1}{|\lambda|}) \geq 1\}.$$

Note that

$$\tau_p\left(u, \frac{1}{|\lambda|}\right) = \sum_{k=0}^{p-1} \pi_p^k(u)|\lambda|^{k+1}\ \exp\left[\frac{1}{2}(1 + \pi_p^{2p}(u)|\lambda|^{2p})\right].$$

13 Example to Theorem 6

To illustrate Theorem 6, consider the equations

$$u'' + a^2 u = 0, u(0) = 1, u'(0) = 0 \ \ (a = \text{const} \ > 0). \tag{39}$$

and

$$\tilde{u}'' + \tilde{a}^2 \tilde{u} = 0, \tilde{u}(0) = 1, \tilde{u}'(0) = 0 \ \ (\tilde{a} = \text{const} \ > 0). \tag{40}$$

Then $u(z) = \cos(az)$, $\tilde{u}(z) = \cos(\tilde{a}z)$ and their zeros are

$$\frac{\pi}{a}(m + 1/2) \text{ and } \frac{\pi}{\tilde{a}}(m + 1/2) \ \ (m = 0, \pm 1, \pm 2, \ldots),$$

respectively. Consequently,

$$\text{rv}_u(\tilde{u}) = \sup_j \inf_k \left| \frac{1}{z_k(h)} - \frac{1}{z_j(\tilde{h})} \right|$$

$$= \sup_{j=0,\pm 1,\pm 2,\ldots} \inf_{k=0,\pm 1,\pm 2,\ldots} \left| \frac{a}{\pi(k + 1/2)} - \frac{\tilde{a}}{\pi(j + 1/2)} \right|$$

$$= \frac{|\tilde{a} - a|}{\pi} \sup_{j=0,\pm 1,\pm 2,\ldots} \frac{1}{j + 1/2}.$$

Hence,

$$\text{rv}_u(\tilde{u}) = 2\frac{|\tilde{a} - a|}{\pi}. \tag{41}$$

In the considered case $n = 0, \gamma_P = a, \gamma_{\tilde{P}} = \tilde{a}$, and $\eta_{\tilde{P}} = \eta_P = 1$. In addition, $\xi(P, \tilde{P}) = e(a + \tilde{a})$, and $\xi_P = ea$. So condition (36) holds, provided

$$e(a + \tilde{a}) < 1. \tag{42}$$

This condition yields $\psi(u) = \psi(cos(az)) = \psi_1(a)$, where

$$\psi_1(a)) = [\sum_{k=1}^{\infty} \xi_P^{2k}]^{1/2} = [\sum_{k=1}^{\infty} (ae)^{2k}]^{1/2} = \frac{1}{[1/(ae)^2 - 1]^{1/2}}.$$

Moreover, $g_2 = |a - \tilde{a}|/2$, $g_k = 0$ for $k \geq 3$, and

$$\Delta = g_2[\sum_{k=2}^{\infty}(k(k-1)\xi^{k-2}(P,\tilde{P}))^2]^{1/2} \le g_2\sum_{k=2}^{\infty}k(k-1)\xi^{k-2}(P,\tilde{P}).$$

Since

$$\sum_{k=2}^{\infty}x^{k-2}k(k-1) = \frac{d^2}{dx^2}\sum_{k=0}^{\infty}x^k = \frac{d^2}{dx^2}(1-x)^{-1} = \frac{6}{(1-x)^3} \quad (0<x<1),$$

we have $\Delta \le \Delta_1$, where

$$\Delta_1 = \frac{6g_2}{(1-\xi(P,\tilde{P}))^3} = \frac{3|a-\tilde{a}|}{(1-(a+\tilde{a})e)^3}.$$

With $p=1$ the inequality (37) holds and $\pi_1(u) = \hat{\pi}(a)$, where

$$\hat{\pi}(a) := 2\left[(\psi_1(a)+\frac{1}{2})^2 + \zeta(2) - 1 - \frac{1}{2^2}\right]^{1/2},$$

and $\tau_1(u,y) = \hat{\tau}(a,y)$, where

$$\hat{\tau}(a,y) := \frac{1}{y}\exp\left[\frac{1}{2} + \frac{\hat{\pi}^2(a)}{2y^2}\right] \quad (y>0).$$

Since $\cos 0 = 1$ and $\sin 0 = 0$, making use of Theorem 6, we can assert the following result: *if condition (42) holds, then for solutions of (39) and (40) we have* $\mathrm{rv}_u(\tilde{u}) \le x(\Delta,a)$, *where* $x(\Delta,a)$ *is the unique positive root of the equation*

$$\Delta_1\cdot\hat{\tau}(a,y) = 1.$$

By (38)

$$x_1(\Delta,q) \le \omega_1(\Delta_1),$$

where

$$\omega_1(\Delta_1) := \begin{cases} e\Delta_1 & \text{if } \hat{\pi}(a) \le e\Delta_1, \\ \hat{\pi}(a)\,[\ln(\hat{\pi}(a)/\Delta_1)]^{-1/2} & \text{if } \hat{\pi}(a) > e\Delta_1. \end{cases}$$

So if $\hat{\pi}(a) \le e\Delta_1$, then

$$\mathrm{rv}_u(\tilde{u}) \le e\Delta_1 = \frac{3e|a-\tilde{a}|}{(1-e(a+\tilde{a}))^3}.$$

For sufficiently small $\tilde{a}$ and a this inequality is "close" to (41).

References

1. D. Andrica, Th.M. Rassias (eds.), *Differential and Integral Inequalities*. Springer Optimization and Its Applications, vol. 151 (Springer, Switzerland, 2019)
2. N. Anghel, Stieltjes-Calogero-Gil' relations associated to entire functions of finite order. J. Math. Phys. **51**(5), 251–262 (2010)
3. N. Anghel, Entire functions of finite order as solutions to certain complex differential equations. Proc. Am. Math. Soc. **140**, 2319–2332 (2012)
4. S. Bank, A note on the zeros of solutions $w'' - P(z)w = 0$ where P is a polynomial. Appl. Anal. **25**, 29–41 (1987)
5. S. Bank, A note on the location of complex zeros of solutions of linear differential equations. Complex Variables Theory Appl. **12**, 159–167 (1989)
6. S. Bank, On the complex zeros of solutions of linear differential equations. Ann. Mat. Pura Appl. **161**, 83–112 (1992)
7. F. Brűuggemann, On the zeros of fundamental systems of linear differential equations with polynomial coefficients. Complex Variables Theory Appl. **15**, 159–166 (1990)
8. F. Brűuggemann, On the solutions of linear differential equations with real zeros; proof of a conjecture of Hellerstein and Rossi. Proc. Am. Math. Soc. **113**, 371–379 (1991)
9. T.B. Cao, K. Liu, H.-Y. Xu, Bounds for the sums of zeros of solutions of $u(m) = P(z)u$ where P is a polynomial. Electron. J. Qual. Theory Differ. Equ. **60**, 10 (2011)
10. J.L. Daleckii, M.G. Krein, *Stability of Solutions of Differential Equations in Banach Space* (American Mathematical Society, Providence, 1971)
11. A. Eremenko, S. Merenkov, Nevanlinna functions with real zeros. Ill. J. Math. **49**(4), 1093–1110 (2005)
12. L. Gao, On the growth of solutions of higher-order algebraic differential equations. Acta Math. Scientia (B) **22**(4), 459–465 (2002)
13. M. Gaudenzi, On the number of zeros of solutions of a linear differential equation. J. Math. Anal. Appl. **221**(1), 306–325 (1998)
14. M.I. Gil', Approximations of zeros of entire functions by zeros of polynomials. J. Approx. Theory **106**, 66–76 (2000)
15. M.I. Gil', Perturbations of zeros of a class of entire functions. Complex Var. **42**, 97–106 (2000)
16. M.I. Gil', *Operator Functions and Localization of Spectra*. Lectures Notes in Mathematics vol. 1830 (Springer, Berlin, 2003)
17. M.I. Gil', *Localization and Perturbation of Zeros of Entire Functions* (CRC Press, Taylor and Francis Group, New York, 2010)
18. M.I. Gil', Bounds for zeros of solutions of second order differential equations with polynomial coefficients. Results Math. **59**, 115–124 (2011)
19. M.I. Gil', Sums of zeros of solutions to second order ODE with non-polynomial coefficients. Electron. J. Diff. Equ. **2012**(107), 1–8 (2012)
20. M.I. Gil', Bounds for products of zeros of solutions to nonhomogeneous ODE with polynomial coefficients. Int. J. Differ. Equ. **2015**, 690519 (2015)
21. M.I. Gil', Sums of zeros of solutions to non-homogeneous ODE with polynomial coefficients. J. Math. Anal. Appl. **421**(2), 1917–1924 (2015)
22. M.I. Gil', Inequalities for zeros of solutions to second order ODE with one singular point. Differ. Equ. Appl. **8**(1), 69–76 (2016)
23. I.C. Gohberg, S. Goldberg, N. Krupnik, *Traces and Determinants of Linear Operators* (Birkhäuser, Basel, 2000)
24. I.C. Gohberg, M.G. Krein, *Introduction to the Theory of Linear Nonselfadjoint Operators*. Trans. Mathem. Monographs, vol. 18 (American Mathematical Society, Providence, 1969)
25. G.G. Gundersen, On the real zeros of solutions of $f'' + A(z)f = 0$ where $A(z)$ is entire. Ann. Acad. Sci. Fenn. Ser. A.I. Math **11**, 275–294 (1986)
26. S. Hellerstein, J. Rossi, Zeros of meromorphic solutions of second order linear differential equations. Math. Z. **192**, 603–612 (1986)

27. S. Hellerstein, J. Rossi, On the distribution of zeros of solutions of second-order differential equations. Complex Var. Theory Appl. **13**, 99–109 (1989)
28. S. Hellerstein, J. Rossi, Schwarzian derivatives and zeros of solutions to second order linear differential equations. Proc. Am. Math. Soc. **113**, 741–746 (1991)
29. E. Hille, *Lectures on Ordinary Differential Equations* (Addison-Wesley, Ontario, 1969)
30. C.Z. Huang, Real zeros of solutions of second order linear differential equations. Kodai Math. J. **14**, 113–122 (1991)
31. C.-H. Lin, Y. Sibuya, T. Tabara, Zeros of solutions of a second order linear differential equation with polynomial coefficients. Funkc. Ekvacioj **36**(2), 375–384 (1993)
32. I. Laine, *Nevanlinna Theory and Complex Differential Equations* (Walter de Gruyter Berlin, 1993)
33. M. Marden, *Geometry of Polynomials* (American Mathematical Society, Providence, 1985)
34. G.M. Muminov, On the zeros of solutions of the differential equation $\omega^{(2m)} + p(z)\omega = 0$. Demonstr. Math. **35**(1), 41–48 (2002)
35. P.M. Pardalos, T.M. Rassias (eds.), *Mathematics Without Boundaries*. Surveys in Interdisciplinary Research, vol. VIII (Springer, New York, 2014)
36. J. Rossi, The Tsuji characteristic and real zeros of second order ordinary differential equations. J. Lond. Math. Soc. **36**(2), 490–500 (1987)
37. J. Tu, Z.X. Chen, Zeros of solutions of certain second order linear differential equation. J. Math. Anal. Appl. **332**, 1279–291 (2007)

Dynamics of a Higher-Order Ginzburg–Landau-Type Equation

Theodoros P. Horikis, Nikos I. Karachalios, and Dimitrios J. Frantzeskakis

Abstract We study possible dynamical scenarios associated with a higher-order Ginzburg–Landau-type equation. In particular, first we discuss and prove the existence of a limit set (attractor), capturing the long-time dynamics of the system. Then, we examine conditions for finite-time collapse of the solutions of the model at hand, and find that the collapse dynamics is chiefly controlled by the linear/nonlinear gain/loss strengths. Finally, considering the model as a perturbed nonlinear Schrödinger equation, we employ perturbation theory for solitons to analyze the influence of gain/loss and other higher-order effects on the dynamics of bright and dark solitons.

1 Introduction

In this work, our aim is to study the dynamics of a higher-order Ginzburg–Landau type equation. In particular, the model under consideration has the form of a higher-order nonlinear Schrödinger (NLS) equation incorporating gain and loss. The origin of our motivation is the following dimensionless higher-order NLS equation:

$$\partial_t u + \frac{\mathrm{i}s}{2}\partial_x^2 u - \mathrm{i}|u|^2 u = \beta\partial_x^3 u + \mu\partial_x(|u|^2 u) + (\nu - \mathrm{i}\sigma_R)\, u\partial_x(|u|^2), \tag{1}$$

T. P. Horikis (✉)
Department of Mathematics, University of Ioannina, Ioannina, Greece
e-mail: horikis@uoi.gr

N. I. Karachalios
Department of Mathematics, University of Thessaly, Lamia, Greece
e-mail: karan@uth.gr

D. J. Frantzeskakis
Department of Physics, University of Athens, Athens, Greece
e-mail: dfrantz@phys.uoa.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_9

where $u(x,t)$ is a complex field, β, μ, ν and σ_R are real constants, while $s = \pm 1$ stands for normal ($s = +1$) or anomalous ($s = -1$) group-velocity dispersion. Note that Equation (1) can be viewed as a perturbed NLS equation, with the perturbation—in case of small values of relevant coefficients—appearing in the right-hand side (see, e.g., Refs. [1–3]).

Variants of Equation (1) appear in a variety of physical contexts, where they are derived at higher-order approximations of perturbation theory [the lowest-order nonlinear model is simply the NLS equation in the left-hand side of Equation (1)]. The most prominent example is probably that of nonlinear optics [1–3]. In this case, t and x denote propagation distance and retarded time (in a reference frame moving with the group velocity), respectively, while $u(x,t)$ is the complex electric field envelope.

For $\beta = \mu = \nu = \sigma_R = 0$, Equation (1) reduces to the unperturbed equation, i.e., the completely integrable NLS [4], which supports bright soliton solutions (for $s = -1$) [5], or dark soliton solutions (for $s = +1$) [6]. As concerns the origin of the higher-order terms, we mention the following. While the unperturbed NLS equation is sufficient to describe optical pulse propagation, for ultra-short pulses, third-order dispersion and self-steepening (characterized by coefficients β, μ and ν, respectively) become important and have to be incorporated in the model. Similar situations also occur in other contexts and, thus, corresponding versions of Equation (2) have been derived and used, e.g., in the context of nonlinear metamaterials [7–9], but also in the problem of water waves in finite depth [10–12]. Moreover, in the context of optics, and for relatively long propagation distances, higher-order nonlinear dissipative effects, such as the stimulated Raman scattering (SRS) effect, of strength $\sigma_R > 0$, are also important [1–3].

In addition to the above mentioned effects, our aim is to investigate the dynamics in the framework of Equation (1), but also incorporating linear or nonlinear gain and loss. This way, we are going to analyze the following model:

$$\partial_t u + \frac{\mathrm{i}s}{2}\partial_x^2 u - \mathrm{i}|u|^2 u = \gamma u + \delta|u|^2 u + \mu\partial_x(|u|^2 u) + \beta\partial_x^3 u + (\nu - \mathrm{i}\sigma_R)\, u\partial_x(|u|^2), \tag{2}$$

which also incorporates dissipative effects, such as linear loss (for $\gamma < 0$) [or gain (for $\gamma > 0$)]. These effects are physically relevant in the context of nonlinear optics [1–3, 13]: indeed, nonlinear gain ($\delta > 0$) [or loss ($\delta < 0$)] may be used to counterbalance the effects from the linear loss/gain mechanisms, which may potentially lead to the stabilization of optical solitons—see, e.g., Refs. [14, 15]. Notice that it is the presence of gain/loss that renders Equation (2) a higher-order cubically nonlinear Ginzburg–Landau-type equation (see recent studies [16–18] on such models), featuring zero diffusion.

In this work, we will discuss various possible dynamical scenarios associated with Equation (2). In particular, the organization of the presentation and main results of this work can be described as follows.

In Section 2, first we show that the incorporation of gain and loss terms in the model gives rise to the existence of an attractor, capturing the long-time dynamics of the system. A rigorous proof is provided, based on the interpretation of the energy balance equation and properties of the functional (phase) space in which the problem defines an infinite-dimensional flow. It will also be discussed that although the gain/loss effects are pivotal for the dissipative nature of the infinite-dimensional flow that will be defined below, the structure of the attractor is basically determined by the other higher-order effects. In the same Section (Section 2), we also examine conditions for finite-time collapse of the solutions of the model. In particular, upon using energy arguments, we find that the collapse dynamics is chiefly controlled by the linear/nonlinear gain/loss strengths. We also identify a critical value of the linear gain, separating the possible decay of solutions to the trivial zero-state, from collapse.

In addition, considering the higher-order Ginzburg–Landau-type equation as a perturbed NLS equation, in Section 3 we study the dynamics of bright and dark solitons under the influence of the higher-order effects. The analysis is based on various perturbative techniques, relying on general aspects of the perturbation theory for bright and dark solitons. Specifically, we adopt the so-called adiabatic approximation, according to which the soliton form does not chance due to the (small) perturbation, but its characteristics (center, amplitude, velocity, etc.) become unknown functions of time. We derive relevant evolution equations for the soliton characteristics and describe the pertinent soliton dynamics. We also briefly discuss still another method to analyze soliton dynamics, namely a multiscale expansion technique that asymptotically reduces the model to a Korteweg-de Vries–Burgers (KdV-B) equation. This way, we discuss various other nonlinear wave structures that can be supported by the higher-order effects, namely anti-dark solitons, as well as shock waves and rarefaction waves.

2 Limit Set and Collapse

2.1 Existence of the Limit Set

Let us consider the case $s = -1$, and supplement Equation (2) with periodic boundary conditions for u and its spatial derivatives up to the-second order, namely:

$$\begin{aligned} u(x+2L,t) &= u(x,t), \quad \text{and} \\ \partial_x^j(x+2L,t) &= \partial_x^j(x,t), \quad j=1,2, \end{aligned} \tag{3}$$

$\forall\ (x,t) \in \mathbb{R} \times [0,T]$, for some $T > 0$, where $L > 0$ is given. The initial condition

$$u(x,0) = u_0(x), \quad \forall x \in \mathbb{R}, \tag{4}$$

also satisfies the periodicity conditions (3).

As shown in Ref. [19], all possible regimes except $\gamma > 0, \delta < 0$, are associated with finite-time collapse or decay. Furthermore, a critical value γ^* can be identified in the regime $\gamma < 0, \delta > 0$, which separates finite-time collapse from the decay of solutions. On the other hand, for $\gamma > 0, \delta < 0$, below we prove the existence of a limit set (attractor) $\omega(\mathscr{B})$, attracting all bounded orbits initiating from arbitrary, appropriately smooth initial data u_0 (considered elements of a bounded set $\mathscr{B}$ in a suitable Sobolev space). Notice that, as shown numerically in Ref. [20], the attractor $\omega(\mathscr{B})$ captures the full route, ranging from Poincaré–Bendixson limit-cycle dynamics to quasiperiodic or chaotic dynamics.

The starting point of our proof is the power balance equation:

$$\frac{d}{dt}\int_{-L}^{L}|u|^2dx = 2\gamma\int_{-L}^{L}|u|^2dx + 2\delta\int_{-L}^{L}|u|^4dx, \tag{5}$$

satisfied by any local solution $u \in C([0,T], H^k_{per}(\Omega))$, which initiates from sufficiently smooth initial data $u_0 \in H^k_{per}(\Omega)$, for fixed $k \geq 3$. Here, $H^k_{per}(\Omega)$ denotes the Sobolev spaces of periodic functions H^k_{per} [21], in the fundamental interval $\Omega = [-L, L]$. Analysis of (5), results in the asymptotic estimate:

$$\limsup_{t\to\infty}\frac{1}{2L}\int_{-L}^{L}|u(x,t)|^2dx \leq -\frac{\gamma}{\delta}, \tag{6}$$

implying that local in time solutions $u \in C([0,T], H^k_{per}(\Omega))$ are uniformly bounded in $L^2(\Omega)$. This allows for the definition of the extended dynamical system:

$$\varphi(t,u_0) : H^k_{per}(\Omega)) \to L^2(\Omega), \quad \varphi(t,u_0) = u,$$

whose orbits are bounded $\forall t \geq 0$. Moreover, from the above asymptotic estimate, we derive the following: if $L^2(\Omega)$ is endowed with the equivalent averaged norm

$$||u||^2_\alpha = \frac{1}{2L}\int_{-L}^{L}|u|^2dx,$$

then its ball:

$$\mathscr{B}_\alpha(0,\rho) = \left\{u \in L^2(\Omega) \ : \ ||u||^2_\alpha \leq \rho^2, \quad \rho^2 > -\frac{\gamma}{\delta}\right\},$$

attracts all bounded sets $\mathscr{B} \in H^k_{per}(\Omega)$. That is, there exists $T^* > 0$, such that $\varphi(t,\mathscr{B}) \subset \mathscr{B}_\alpha$, for all $t \geq T^*$. Thus, we may define for any bounded set $\mathscr{B} \in H^k_{per}(\Omega))$, $k \geq 3$, its ω-limit set in $L^2(\Omega)$,

$$\omega(\mathscr{B}) = \bigcap_{s\geq 0} \overline{\bigcup_{t\geq s} \varphi(t, \mathscr{B})}.$$

The closures are taken with respect to the weak topology of $L^2(\Omega)$. Then, the standard (embedding) properties of Sobolev spaces imply that the attractor $\omega(\mathscr{B})$ is at least weakly compact in $L^2(\Omega)$, or relatively compact in the dual space $H^{-1}_{per}(\Omega)$.

In the direct numerical simulations of Ref. [20], it was found that apart from the gain/loss parameters γ and δ, the other higher-order effects play also important role on the dynamics. In particular, the competition between the third-order dispersion (characterized by the coefficient β) and SRS effect (characterized by the coefficient σ_R gives rise to rich dynamics (briefly mentioned above): indeed, the dynamics ranges from Poincaré–Bendixson-type scenarios, in the sense that bounded solutions may converge either to distinct equilibria via orbital connections or to space-time periodic solutions, to the emergence of almost periodic and chaotic behavior. A main result is that third-order dispersion has a dominant role in the development of such complex dynamics, since it can be chiefly responsible (even in the absence of other higher-order effects) for the existence of periodic, quasiperiodic, and chaotic spatiotemporal structures.

We conclude by illustrating some representative results illustrating the richness of these dynamics.

Figure 1 depicts the birth of a space time periodic solution emerging from the modulation instability of the continuous wave (cw) steady-state solution of amplitude $|\phi_b|^2 = -\frac{\gamma}{\delta}$ for the choice of parameters $\beta = 0.55$, $\sigma_R = 0.01$, $\gamma = 1.5$, $\delta = -1$, $\sigma_R = 0.3$, $\mu = \nu = 0.01$. The initial condition is a single-mode cw of the form

$$u_0(x) = \epsilon\, e^{-i\frac{K\pi x}{L}},\ K > 0. \tag{7}$$

with $K = 5$ and $\epsilon = 0.01$. This is one of the examples showing the Poincaré-Bendixson type dynamics when the instability of a steady state gives rise to the birth of a limit-cycle. The results visualise the asymptotic behavior in the $2D$-finite dimensional subspace

$$\mathscr{P}_2=\{(X, Y) \in \mathbb{R}^2\ :\ (X(t), Y(t))=(|u(x_1, t)|^2, |u(x_2, t)|^2),\ \ x_1,\ x_2 \in \Omega,\ t \geq 0\},$$

for some arbitrarily chosen fixed spatial coordinates $x_1,\ x_2$. In this subspace, the steady-state ϕ_b is marked by the point $\mathbf{A} = (|\phi_b|^2, |\phi_b|^2) = \left(-\frac{\gamma}{\delta}, -\frac{\gamma}{\delta}\right)$. The $3D$-counterpart is defined as

$$\begin{aligned}\mathscr{P}_3=\{(X, Y, Z) \in \mathbb{R}^3\ :\ &(X(t), Y(t), Z(t))=(|u(x_1, t)|^2, |u(x_2, t)|^2, |u(x_3, t)|^2),\\ &x_1,\ x_2,\ x_3 \in \Omega,\ t \geq 0\}.\end{aligned} \tag{8}$$

The emergence of limit-cycles characterizing the global attractor persists up to certain thresholds for the parameter β.

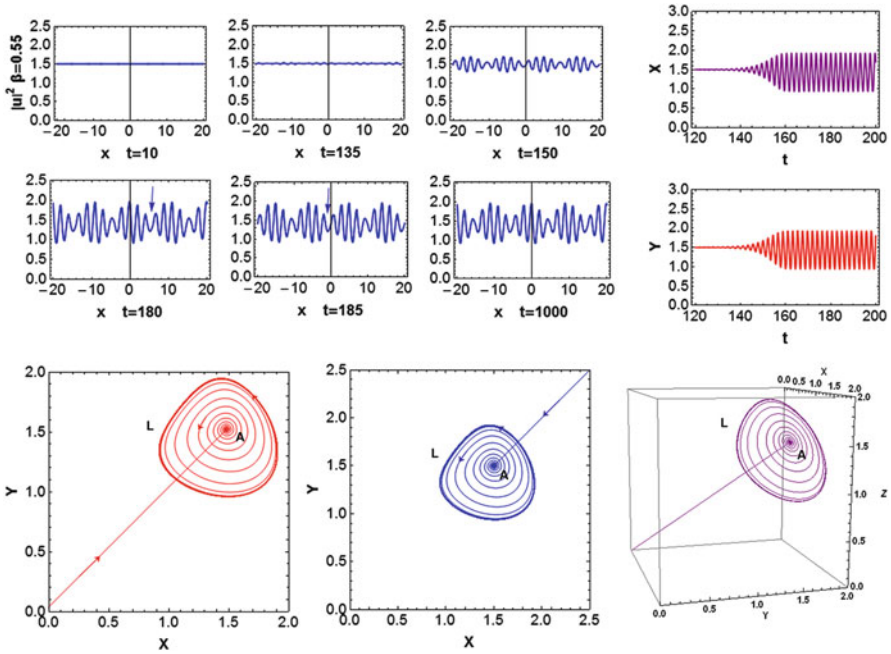


Fig. 1 (Color Online) Upper left panel: The birth of a space-time periodic solution from the instability of the cw-steady state of amplitude $|\phi_b|^2 = -\gamma/\delta$, for cw-initial data (7) of $K = 5$ and $\epsilon = 0.01$. Parameters $\beta = 0.55$, $\sigma_R = 0.01$, $\gamma = 1.5$, $\delta = -1$, $\sigma_R = 0.3$, $\mu = \nu = 0.01$. Upper right panel: Integral curves $X(t) = |u(x_1, t)|^2$ (upper fig.-purple curve) and $Y(t) = |u(x_2, t)|^2$ (bottom fig.-red curve), for the spatial coordinates $x_1 = 0$ and $x_2 = 5$ respectively: Convergence to a periodic solution, for the set of parameters of the upper left panel. Bottom left panel: The space-time periodic solution of the left upper panel, as a limit cycle in the 2D-phase space $\mathscr{P}_2$ for the spatial coordinates $x_1 = 0$ and $x_2 = 5$. Bottom middle panel: Convergence to the limit cycle of the bottom left panel for the cw-initial condition of $K = 5$ and $\epsilon = \sqrt{3}$. Bottom right panel: The space-time periodic solution of the upper right panel as a limit cycle on the 3D-phase space $\mathscr{P}_3$ (defined by (8)), for the choice of spatial coordinates $x_1 = 0$, $x_2 = 5$ and $x_3 = 10$

On the other hand, even when the steady state ϕ_b is asymptotically stable, the convergence may include highly non-trivial transient dynamics. Figure 2 depict an example of the evolution of the initial condition

$$u_0(x) = \epsilon \, \mathrm{sech} x. \tag{9}$$

Such initial data correspond to the profile of a "bright soliton" as an initial state. The example is for $\epsilon = 1$ and parameters $\sigma_R = \mu = \nu = \beta = 0.01$. The gain/loss strengths are $\gamma = 1.5$, $\delta = -1$. We observe formation of a "shock-wave" transitioning to an unstable periodic solution and then, the formation of a decaying travelling pulse, prior to the ultimate convergence to the asymptotically stable state ϕ_b.

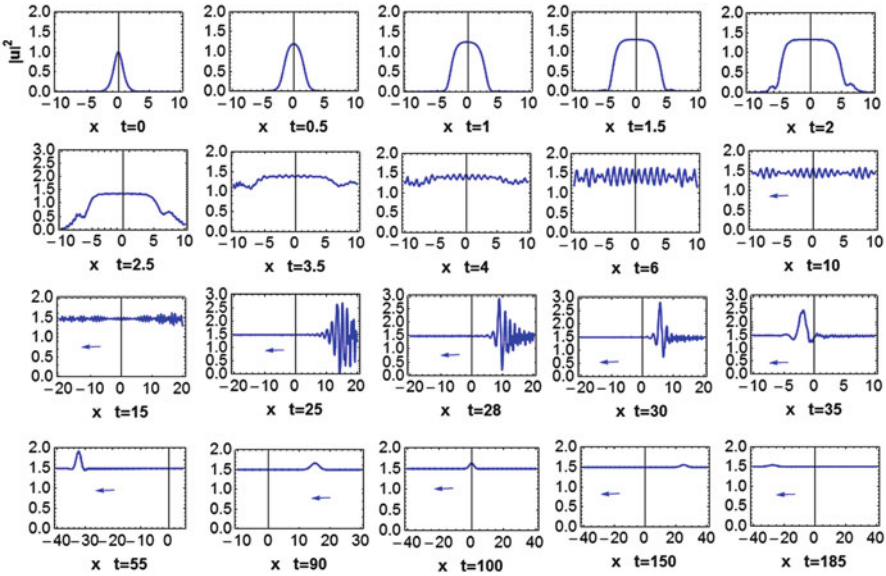


Fig. 2 (Color Online) Snapshots of the evolution of the density $|u|^2$ for initial data (9) with $\epsilon = 1$, when $\sigma_R = \mu = \nu = \beta = 0.01$ and $\gamma = 1.5$, $\delta = -1$. Formation of a decaying "bright" traveling solitary pulse, prior to the ultimate convergence to the steady state ϕ_b. The array indicates the direction of the travelling pulses

2.2 Conditions for Collapse

The question of collapse concerns sufficiently smooth (weak) solutions of Equation (2). The existence of such solutions, is guaranteed by a local existence result associated with the initial-boundary value problem (2)–(4). In particular, the methods which are used in order to prove such a local existence result in the Sobolev spaces of periodic functions H^k_{per} [22, 23], are based on the lines of approach of [24–27]. The application of these methods to establish local existence for the problem (2)–(4), although involving lengthy computations, is now considered as standard. Thus, we refrain from showing the details here, and we just state it in:

Theorem 1 *Let $u_0 \in H^k_{per}(\Omega)$ for any integer $k \geq 2$, and $\beta, \gamma, \delta, \mu, \nu, \sigma_R \in \mathbb{R}$. Then there exists $T > 0$, such that the problem (2)–(4), has a unique solution*

$$u \in C([0, T], H^k_{per}(\Omega)) \quad \text{and} \quad u_t \in C([0, T], H^{k-3}_{per}(\Omega)).$$

Moreover, the solution $u \in H^k_{per}(\Omega)$ depends continuously on the initial data $u_0 \in H^k_{per}(\Omega)$, i.e., the solution operator

$$\mathscr{S}(t) : H^k_{per}(\Omega) \mapsto H^k_{per}(\Omega), \quad t \in [0, T], \tag{10}$$

$$u_0 \mapsto \mathscr{S}(t)u_0 = u,$$

is continuous.

Here, for the sake of completeness, let us recall the definition of $H^k_{per}(\Omega)$:

$$H^k_{per}(\Omega) = \{u : \Omega \to \mathbb{C}, \quad u \text{ and } \frac{\partial^j u}{\partial x^j} \in L^2(\Omega), \quad j = 1, 2, \ldots, k;$$

$$u(x), \quad \text{and} \quad \frac{\partial^j u}{\partial x^j}(x) \text{ for } j = 1, 2, \ldots, k-1, \text{ are } 2L\text{- periodic}\}. \tag{11}$$

Since our analytical energy arguments for examining collapse require sufficiently smoothness of local-in-time solutions, we shall implement Theorem 1 by assuming that $k = 3$, at least. As it follows from the definition of the Sobolev spaces (2.2), this assumption means that the initial condition $u_0(x)$, $x \in \Omega$, and its spatial (weak) derivatives, at least up to the 2nd-order, are $2L$-periodic. Then, it turns out from Theorem 1, that the unique, local-in-time solution $u(x, t)$ of (2) satisfies the periodic boundary conditions (3) for $t \geq 0$, and is sufficiently (weakly) smooth. Such periodicity and smoothness properties of the local-in-time solutions are sufficient for our purposes (see Theorem 2 below).

Next, we adopt the method of deriving a differential inequality for the functional

$$M(t) = \frac{e^{-2\gamma t}}{2L} \int_{-L}^{L} |u(x, t)|^2 dx, \tag{12}$$

and then, showing that its solution diverges in finite-time under appropriate assumptions on its initial value at time $t = 0$; see [22, 23, 28–30] and references therein. Note that the choice of this functional is not arbitrary; in fact, it is a direct consequence of the conservation laws of the NLS model. For a discrete counterpart of this argument applied in discrete Ginzburg–Landau-type equations, we refer to [31]. For applications of these types of arguments in the study of escape dynamics for Klein–Gordon chains, we refer to [32].

Theorem 2 *For $u_0 \in H^k_{per}(\Omega)$, $k \geq 3$, let $\mathscr{S}(t)u_0 = u \in C([0, T_{\max}), H^k_{per}(\Omega))$ be the local- in- time solution of the problem (2)–(4), with $[0, T_{\max})$ be its maximal interval of existence. Assume that the parameter $\delta > 0$ and that the initial condition $u_0(x)$ is such that $M(0) > 0$. Then, $T_{\max}$ is finite, in the following cases:*

$$(i) \qquad T_{\max} \leq \frac{1}{2\gamma} \log\left[1 + \frac{\gamma}{\delta M(0)}\right], \tag{13}$$

$$\text{for } \gamma \neq 0, \quad \text{and} \quad \gamma > -\delta M(0). \tag{14}$$

$$(ii) \qquad T_{\max} \leq \frac{1}{2\delta M(0)}, \quad \text{for } \gamma = 0. \tag{15}$$

Proof For any $T < T_{\max}$, since $k \geq 3$, due to the continuous embedding [22]:

$$C([0,T], H^k_{per}(\Omega)) \subset C([0,T], L^2(\Omega)),$$

the solution $\mathscr{S}(t)u_0 = u \in C([0,T], L^2(\Omega))$. Furthermore, it follows from Theorem 1, that $u_t \in C([0,T], L^2(\Omega))$. Then, by differentiating $M(t)$ with respect to time, we find that:

$$\frac{dM(t)}{dt} = -\gamma \frac{e^{-2\gamma t}}{L} \int_{-L}^{L} |u|^2 dx + \frac{e^{-2\gamma t}}{L} \mathrm{Re} \int_{-L}^{L} u_t \overline{u} dx. \tag{16}$$

In the second term on the right-hand side of (16), we substitute u_t by the right-hand side of Equation (2). Then, after some computations, Equation (16) results in:

$$\frac{dM(t)}{dt} = \delta \frac{e^{-2\gamma t}}{L} \int_{-L}^{L} |u|^4 dx. \tag{17}$$

Next, by the Cauchy-Schwarz inequality, we have

$$\int_{-L}^{L} |u|^2 dx \leq \sqrt{2L} \left(\int_{-L}^{L} |u|^4 dx \right)^{1/2}. \tag{18}$$

Therefore, for the functional $M(t)$ defined in (12), we get the inequality

$$M(t) \leq \frac{e^{-2\gamma t}}{\sqrt{2L}} \left(\int_{-L}^{L} |u|^4 dx \right)^{1/2}, \tag{19}$$

which in turns, implies the estimate

$$M(t)^2 \leq \frac{e^{-4\gamma t}}{2L} \int_{-L}^{L} |u|^4 dx, \tag{20}$$

for all $t \in [0, T_{\max})$. On the other hand, from (17) we have that

$$\int_{-L}^{L} |u|^4 dx = e^{2\gamma t} \frac{L}{\delta} \frac{dM(t)}{dt},$$

and hence, we may rewrite (20) as

$$[M(t)]^2 \leq \frac{e^{-2\gamma t}}{2\delta} \frac{dM(t)}{dt}, \quad \text{or} \quad \frac{\frac{dM(t)}{dt}}{[M(t)]^2} \geq 2\delta e^{2\gamma t}. \tag{21}$$

Since

$$\frac{d}{dt}\left[\frac{1}{M(t)}\right] = -\frac{\frac{dM(t)}{dt}}{[M(t)]^2},$$

we get from (21) the differential inequality

$$\frac{d}{dt}\left[\frac{1}{M(t)}\right] \leq -2\delta e^{2\gamma t}. \tag{22}$$

Integration of (22) with respect to time, implies that

$$\frac{1}{M(t)} \leq \frac{1}{M(0)} - 2\delta \int_0^t e^{2\gamma s} ds,$$

and since $M(t) > 0$, we see that $M(0) > 0$ satisfies the inequality

$$2\delta \int_0^t e^{2\gamma s} ds \leq \frac{1}{M(0)}. \tag{23}$$

From (23), we shall distinguish between the following cases for the damping parameter γ:

- We assume that the damping parameter $\gamma \neq 0$. In this case, (23) implies that

$$\frac{2\delta}{2\gamma}\left(e^{2\gamma t} - 1\right) \leq \frac{1}{M(0)}, \quad \text{or} \quad e^{2\gamma t} \leq 1 + \frac{\gamma}{\delta M(0)}.$$

  Thus, assuming that

$$\frac{\gamma}{\delta M(0)} > -1,$$

  we derive that the maximal time of existence is finite, since

$$t \leq \frac{1}{2\gamma} \log\left[1 + \frac{\gamma}{\delta M(0)}\right].$$

  The inequality above, proves the estimate of the collapse time (13) under assumption (14), that is, case (i) of the Theorem.
- Assume that the damping parameter is $\gamma = 0$. Then, Equation (23) implies that

$$2\delta t \leq \frac{1}{M(0)},$$

or

$$t \leq \frac{1}{2\delta M\,(0)}.$$

This inequality proves the estimate of the collapse time (15) in the absence of damping, that is, case (ii) of the Theorem. □

From condition (14) on the definition of the analytical upper bound of the blow-up time

$$T_{\max}[\gamma, \delta, M(0)] = \frac{1}{2\gamma} \log\left[1 + \frac{\gamma}{\delta M\,(0)}\right], \tag{24}$$

given in (13), we define a critical value of the linear gain/loss parameter as

$$\gamma^* = -\delta M(0). \tag{25}$$

We observe that

$$\lim_{\gamma \to \gamma^*} T_{\max}[\gamma, \delta, M(0)] = +\infty, \tag{26}$$

while $T_{\max}[\gamma, \delta, M(0)]$ is finite if

$$\gamma > \gamma^*, \tag{27}$$

according to the condition (14). Then, (26) suggests that when $\delta > 0$, the critical value γ^* may act as a critical point separating the two dynamical behaviors: blow-up in finite-time for $\gamma > \gamma^*$ and global existence for $\gamma < \gamma^*$.

We also remark that the analytical upper bound for the blow-up time (15) in the case $\gamma = 0$,

$$T_{\max}[\delta, M(0)] = \frac{1}{2\delta M\,(0)}, \tag{28}$$

is actually the limit of the analytical upper bound (24) for $\gamma > 0$ as $\gamma \to 0$, e.g.,

$$\lim_{\gamma \to 0} T_{\max}[\gamma, \delta, M(0)] = T_{\max}[\delta, M(0)]. \tag{29}$$

The analytical estimates for the blow-up time have been proved sharp with respect to their dependence on the several parameters as it was illustrated by the relevant numerical simulations [19].

3 Soliton Dynamics: Perturbative Approach

Below, our aim is to consider the higher-order Ginzburg–Landau equation (2) as a perturbed NLS equation. This can be done, upon rewriting Equation (2) in the following form,

$$iu_t - \frac{s}{2}u_{xx} + |u|^2 u = \varepsilon F[u], \tag{30}$$

where subscripts denote partial derivatives, and the functional perturbation $F[u]$ is given by:

$$F[u] = i\gamma u + i\delta|u|^2 u + i\beta u_{xxx} + i\mu(|u|^2 u)_x + (i\nu + \sigma_R)u(|u|^2)_x. \tag{31}$$

In other, words, we consider the situation where the coefficients of the gain/loss and higher-order terms are small, i.e., of the order of a formal small parameter ε (with $0 < \varepsilon \ll 1$). This problem finds applications in long-haul optical fiber communications, where the terms involved in $F[u]$ can indeed be considered as small perturbations [1].

Based on the fact that, for $\epsilon = 0$, Equation (30) becomes the traditional NLS model that possesses bright or dark solitons for $s = -1$ and $s = +1$ respectively, we will study separately these two cases, and investigate how the perturbation (31) alters the soliton dynamics. Our analysis relies on various perturbation techniques that have been developed in the past, both for bright [33–35] and dark [36–38], including the perturbed inverse scattering method, the variational approach (or Lagrangian method), the Lie transform method, and others (see also [1–3] and references therein). Among these techniques, a particularly convenient method to study the soliton dynamics is the so-called adiabatic approach. According to this, an adiabatic relation is the balance between nonlinearity and dispersion, so that (amplitude)×(width)=const. In other words, it is assumed that—in the presence of the perturbations—the functional form of the soliton remains unchanged, but the soliton parameters change (slowly) as the soliton evolves. Thus, the soliton parameters are treated as unknown functions of t, and their evolution is determined by the evolution of the conserved quantities (integrals of motion) of the unperturbed NLS. Particularly relevant such conserved quantities include the energy:

$$E = \int_{-\infty}^{-\infty} |u|^2 \mathrm{d}x, \tag{32}$$

the momentum,

$$P = \frac{i}{2}\int_{-\infty}^{-\infty} (u\bar{u}_x - \bar{u}u_x)\,\mathrm{d}x, \tag{33}$$

where overbar denotes complex conjugate, and the Hamiltonian:

$$H = \frac{1}{2}\int_{-\infty}^{-\infty} \left(s|u_x|^2 + |u|^4\right) \mathrm{d}x. \tag{34}$$

In addition, for our considerations below, it is also useful to introduce still another conserved quantity, namely the central position of the soliton(s)–alias "soliton center"—given by:

$$R_{\mathrm{bs}} = \int_{-\infty}^{+\infty} x|u|^2 \mathrm{d}x, \quad R_{\mathrm{ds}} = \int_{-\infty}^{+\infty} x\left(u_\infty^2 - |u|^2\right) \mathrm{d}x, \tag{35}$$

for the bright and dark solitons respectively.

3.1 Perturbation Theory for Bright Solitons ($s = -1$)

We start with the case of $s = -1$, i.e., the case of bright solitons. First, using Equation (30) and its complex conjugate, it is straightforward to derive the following equations for the evolution of the NLS conserved quantities under the action of the perturbation:

$$\frac{dE}{dt} = \varepsilon \int_{-\infty}^{+\infty} \left(\bar{u}F + u\bar{F}\right) \mathrm{d}x, \tag{36}$$

$$\frac{dP}{dt} = \varepsilon i \int_{-\infty}^{+\infty} \left(\bar{u}_x F - u_x \bar{F}\right) \mathrm{d}x, \tag{37}$$

$$\frac{dH}{dt} = 2\varepsilon \int_{-\infty}^{+\infty} \left[\left(\frac{1}{2}\bar{u}_{xx} + |u|^2\bar{u}\right) F + \left(\frac{1}{2}u_{xx} + |u|^2 u\right) \bar{F}\right] \mathrm{d}x. \tag{38}$$

For sufficiently small perturbation, the form of the bright soliton solution $u_{\mathrm{bs}}(x,t)$ may be assumed to have the following rather general form, where all its parameters are allowed to vary in t as

$$u_{\mathrm{bs}}(x,t) = \eta(t)\,\mathrm{sech}[\eta(t)(x - x_0(t))] \exp\left[-i\kappa(t)x + i\phi(t)\right], \tag{39}$$

where the soliton's amplitude (and inverse width) η, its central position x_0, the wavenumber κ, and phase ϕ are unknown functions of t that have to be determined. Notice that, in the absence of the perturbation, x_0 and ϕ are constants, given by:

$$\frac{dx_0}{dt} = -\kappa, \quad \frac{d\phi}{dt} = \frac{1}{2}\left(\eta^2 - \kappa^2\right). \tag{40}$$

Substituting the soliton (39) into Equations (37)–(38) [and (35)], we obtain a set of four ordinary differential equations (ODEs) for the four unknown soliton parameters:

$$\frac{d\eta}{dt} = -\mathrm{Im}\left\{\int_{-\infty}^{\infty} F[u]\bar{u}\mathrm{d}x\right\}, \tag{41}$$

$$\frac{d\kappa}{dt} = \mathrm{Re}\left\{\int_{-\infty}^{\infty} F[u]\tanh[\eta(x-x_0)]\bar{u}\mathrm{d}x\right\}, \tag{42}$$

$$\frac{dt_0}{dt} = -\kappa - \frac{1}{\eta^2}\mathrm{Im}\left\{\int_{-\infty}^{\infty} F[u](x-x_0)\bar{u}\mathrm{d}x\right\}, \tag{43}$$

$$\frac{d\phi}{dt} = \frac{1}{2}(\eta^2-\kappa^2) + x_0\frac{d\kappa}{dt} - \mathrm{Re}\left\{\int_{-\infty}^{\infty} F[u]\left[\frac{1}{\eta} - (x-x_0)\tanh[\eta(x-x_0)]\right]\bar{u}\mathrm{d}x\right\}, \tag{44}$$

where Re and Im stand for the real and imaginary parts, respectively.

Before analyzing the full problem, where the perturbation $F[u]$ is given by Equation (31), it is relevant to consider at first a simple example. In particular, let the linear loss/gain term be a small perturbation, i.e., $F[u] = i\gamma u$, with $\gamma \ll 1$, and assume that $\delta = \beta = \nu = \sigma_R = 0$. Then, substituting this form of $F[u]$ into Equations (41)–(44), and performing the integrations, it is found that the soliton wavenumber κ and the central position x_0 remain unaffected of the perturbation, while the soliton amplitude η and phase ϕ evolve, due to the presence of the loss/gain, as follows:

$$\eta(t) = \exp(2\gamma t), \quad \phi(t) = \phi(0) - \frac{1}{8\gamma}\left[1-\exp(4\gamma t)\right]. \tag{45}$$

To obtain the above result, it was assumed that $\eta(0) = 1$ and $\kappa(0) = x_0(0) = 0$ (hence $\kappa(t) = x_0(t) = 0\ \forall t$). Thus, in the presence of loss, $\gamma < 0$ (or gain, $\gamma > 0$) the soliton amplitude decreases (or increases), while its width increases (or decreases), i.e., the soliton broadens (or is compressed) during its evolution.

We now return to the full problem, and study the effect of the perturbation (31) on the dynamics of bright solitons. Following the same procedure, i.e., substituting Equation (31) into Equations (41)–(44), and performing the integrations, we find that the soliton parameters evolve according to the following system:

$$\frac{d\eta}{dt} = \frac{2}{3}\eta(3\gamma + 2\delta\eta^2), \tag{46}$$

$$\frac{d\kappa}{dt} = -\frac{8}{15}\sigma_R\eta^4, \tag{47}$$

$$\frac{dt_0}{dt} = -\kappa + \frac{1}{3}(3\beta - 3\mu - 2\nu)\eta + 3\beta\kappa^2\eta^2, \tag{48}$$

$$\frac{d\phi}{dt} = -\kappa\left[(\mu-3\beta)\eta^2 + \beta(\eta^2-2)q^2\right] + \frac{1}{2}(\eta^2-\kappa^2) - \frac{8}{15}\sigma_R t_0\eta^4. \tag{49}$$

Although the result in this case is more complicated, it is still possible to arrive at a simple analytical result. Indeed, first observe that Equation (46) can be solved analytically to provide the functional form of $\eta(z)$, which is found to be:

$$\eta^2(t) = \frac{3C\gamma e^{4\gamma t}}{1 - 2C\delta e^{4\gamma t}}, \quad C = \frac{\eta^2(0)}{3\gamma + 2\delta\eta^2(0)}. \tag{50}$$

Then, the wavenumber $\kappa(t)$ can be obtained from Equation (47) by simply integrating the above expression for η. Finally, having found $\eta(t)$ and $\kappa(t)$, integration of Equations (48) and (49) yield, respectively, the functional forms of $x_0(z)$ and $\phi(t)$.

3.2 *Perturbation Theory for Dark Solitons ($s = +1$)*

In this section, we consider the case $s = +1$, and provide analytical results based on the perturbation theory for dark solitons devised in Ref. [38]. We start by noting that, for $\varepsilon = 0$, the unperturbed defocusing NLS Equation (30) possesses the following single dark soliton solution:

$$u_{\rm ds}(x, t) = [A + iB\tanh(BX)]]e^{i\sigma_0}, \tag{51}$$

where $X = x - X_0$, with $X_0 = x - \int_0^t A(s)ds - x_0$ being the dark soliton center, $A^2 + B^2 = u_\infty^2$, and $\Delta\phi_0 = 2\tan^{-1}(B/A)$ is the phase jump across the soliton. The latter, is equal to π for stationary, so-called "black" solitons with $A = 0$ (moving solitons with $A \neq 0$ are termed "grey") [39, 40]; finally, A and B depict the velocity and depth of the dark soliton, respectively, while x_0 and σ_0 are real parameters. Notice that u_∞ represents the boundary condition at infinity, i.e., $u_\infty = u(x \to \infty)$, and sets the amplitude of the soliton background. The dark soliton (51) is, therefore, comprised of a background of constant density, and a density dip that propagates on top, accompanied by a phase jump across the minimum density.

The effect of the perturbation of Equation (31) on the dark soliton dynamics will now be studied upon assuming that the soliton parameters are slowly varying functions of t. As shown in Ref. [38], dissipative terms—similar to the ones considered here—give rise to a *shelf*, which develops around the soliton; the shelf has a non-trivial contribution to the integrals employed in order to find expressions for the soliton parameters. Thus, this perturbative approach is better suited here, compared to ones merely relying on the adiabatic approximation [36, 37], as they do not take into account this important contribution.

Our analysis starts with the dynamics of the soliton background. Assuming that $u(x \to \infty) = u_0(t)$, we derive from Equation (30 the equation:

$$iu_{0t} - |u_0|^2 u_0 = i\gamma u_0 + i\delta|u_0|^2 u_0. \tag{52}$$

Then, employing the polar decomposition $u_0 = u_\infty(t)\exp(i\theta(t))$, we obtain:

$$u'_\infty = (\gamma + \delta u_\infty^2) u_\infty, \quad \theta' = u_\infty^2, \tag{53}$$

where primes denote differentiation with respect to t. The role of the term of strength δ is now more obvious: a nontrivial equilibrium (constant solution), exists iff $\gamma\delta < 0$ which is $u_\infty^2 = -\gamma/\delta$. Note the relevance of the solution u_∞^2 with the upper bound in the estimate (6). It is also the density of the cw steady-state solution ϕ_b (see Fig. 1). We focus here on these solutions, i.e., solutions that tend to stabilize the soliton, by keeping its parameters constant. The evolution of the rest of the soliton parameters [see (51)] can be found by means of a multiscale boundary layer theory [38]; the resulting evolution equations are:

$$2BA_t = \mathrm{Re}\left\{\int_{-\infty}^{\infty} F[u_{\mathrm{ds}}](\bar{u}_{\mathrm{ds}})_t \, dx\right\}, \tag{54}$$

$$Bx_{0t} = \mathrm{Im}\left\{\int_{-\infty}^{\infty} x(F[u_\infty]u_\infty - F[u_{\mathrm{ds}}]\bar{u}_{\mathrm{ds}}) \, dx\right\}, \tag{55}$$

$$u_\infty \sigma_{0t} = \mathrm{Im}\left\{\int_{-\infty}^{\infty} (F[u_\infty]u_\infty - F[u_{\mathrm{ds}}]\bar{u}_{\mathrm{ds}}) \, dx\right\} + \mathrm{Re}\left\{F[u_\infty]\right\}, \tag{56}$$

$$BB_t = u_\infty u_{\infty t} - AA_t, \tag{57}$$

$$u_\infty^2 \Delta\phi_{0t} = 2AB_t - 2BA_t, \tag{58}$$

$$q_1^\pm = \frac{1}{2}\frac{\sigma_{0t} \pm \Delta\phi_{0t}}{u_\infty \mp A}, \quad \phi_{1t}^\pm = \mp 2q_1^\pm. \tag{59}$$

Here, we should mention that $q_1^\pm$ and $\phi_1^\pm$ in Equations (59) represent the asympotics of the shelf, induced by the perturbation $F[u]$, as $x \to \pm\infty$ respectively; in fact, they are higher-order corrections to the soliton, so that the shelf amplitude is $u_\infty + q_1^\pm$ and its speed u_∞. Integrating the above equations, and using Equation (53), finally yields:

$$u'_\infty = (\gamma + \delta u_\infty^2) u_\infty, \tag{60}$$

$$A' = \frac{4}{15}\sigma_R A^4 + \frac{2}{3}\delta A^3 - \frac{8}{15}\sigma_R u_\infty^2 A^2 + \left(\gamma + \frac{\delta}{3}u_\infty^2\right) A + \frac{4}{15}\sigma_R u_\infty^4, \tag{61}$$

$$x'_0 = \left(2\beta - \mu - \frac{2\nu}{3}\right) A^2 - \left(2\beta + 2\mu + \frac{4\nu}{3}\right) u_\infty^2, \tag{62}$$

$$\sigma'_0 = \frac{B_z}{u_\infty} - \frac{2B}{3u_\infty}\left(3\gamma + 4u_\infty^2\delta + 2\delta A^2\right). \tag{63}$$

These equations show that the evolution of the soliton center, described by the equation $X'_0 = A + x'_0$, is affected by all parameters of Equation (30) [directly or indirectly from $A(t)$]. On the other hand, the rest of the soliton characteristics, i.e., the background, the dip and the shelf, only depend on γ, δ and σ_R. This implies

that soliton stabilization can be targeted accordingly. Indeed, stable fixed points of this system correspond to stable solitons traveling on top of a constant background with a constant speed. It is possible to identify two such solitons, namely a grey and a black one, supported in the presence ($\sigma_R \neq 0$) and in the absence ($\sigma_R = 0$) of the SRS effect, respectively. In both cases, the background assumes the same form: this can be obtained by means of Equation (60), which depicts the nontrivial stationary solution $u_\infty^2 = -\gamma/\delta$ for $\gamma\delta < 0$, i.e., for linear loss and nonlinear gain, or vice versa.

We start with the case $\sigma_R \neq 0$. Substituting the above mentioned constant background in Equation (61), and seeking stationary solutions for the soliton velocity, we arrive at a 4th-order algebraic equation for A. Solving this equation, we find that there exists only one root, namely $A = (4\delta\sigma_R)^{-1}(-5\delta^2 + \sqrt{25\delta^4 - 16\gamma\delta\sigma_R^2})$, which does not violate the relationship $A^2 + B^2 = u_\infty^2$. Thus, a stable soliton exists for:

$$u_\infty^2 = -\frac{\gamma}{\delta}, \quad A = \frac{-5\delta^2 + \sqrt{25\delta^4 - 16\gamma\delta\sigma_R^2}}{4\delta\sigma_R}. \tag{64}$$

Note that since $\gamma\delta < 0$ the quantity under the square root is always positive.

In general, the solution of Equation (60) with $u_\infty(0) = u_0$ is:

$$u_\infty^2(t) = \frac{\gamma u_0^2 e^{2\gamma t}}{\gamma + \delta u_0^2 - \delta u_0^2 e^{2\gamma t}}, \tag{65}$$

which suggests that there is a (finite) time for which the background exhibits blow-up, as it was discussed in Theorem 2. Indeed, the denominator becomes zero when

$$t = t_* \equiv \frac{1}{2\gamma} \ln\left(1 + \frac{\gamma}{\delta u_0^2}\right). \tag{66}$$

The unexpected feature here is that the addition of the term $i\delta|u|^2u$ which compensates the effect of the linear loss term $i\gamma u$ may result in blow-up of the background in finite time, even when the other soliton parameters remain finite. In addition, Equation (65) indicates that an equilibrium can also be reached in finite time when the denominator is a multiple of the numerator. Nevertheless, while under this requirement the background will be stabilized, this does not necessarily guarantees the stabilization of the other soliton parameters.

Next, we consider the case of $\sigma_R = 0$. In this case, Equations (60) and (61) lead to the following equations for the background and soliton velocity:

$$u_\infty^2 = -\frac{\gamma}{\delta}, \quad A' = \frac{2}{3}(\delta A^2 - \gamma)A, \tag{67}$$

Obviously, the above equation for the velocity depicts a stationary solution $A = 0$ (recall that $\gamma\delta < 0$), that corresponds to a black soliton. Hence, when SRS is absent (which would give a frequency downshift causing the soliton to move), a stable black soliton can exist.

An important comment is in order here. While for these specific choice of u_∞ and A the soliton gets stabilized, this does not mean that the shelf is no longer present. In fact, the shelf is always present in the perturbed NLS, even though its amplitude is small, since it appears as a higher-order correction in the perturbation theory [38]. Thus, the shelf does not affect the soliton propagation but it does, however, affect soliton interactions (see Ref. [41] for a relevant study, but in the framework of another dissipative NLS model). Notice, also, that the shelf can be suppressed with counter effect the destabilization of the soliton.

Finally, we briefly consider the case where gain/loss terms are absent, i.e., $\gamma = \delta = 0$. In this particular case, the dark soliton dynamics is merely driven by the SRS effect. Indeed, now the evolution of the background and soliton velocity is described by the following equations:

$$u'_\infty = 0, \quad A' = \frac{4}{15}\sigma_R(A^2 - u_\infty^2)^2, \tag{68}$$

which recover the results obtained in Refs. [36, 37]. The soliton dynamics in this case can be understood as follows. Since $A^2 \neq u_\infty^2$, the right-hand-side of the second equation is always positive and, thus, the soliton becomes shallower and faster, i.e., $B \to 0$ and $A \to u_\infty$, so that the condition $A^2 + B^2 = u_\infty^2$ is satisfied. Thus, the dark soliton eventually decays to the stationary background. It is therefore clear that no stable dark soliton (in the sense of the existence of stationary soliton parameters) exists in this case.

3.3 *Solitons and Shock Waves in an Effective KdV-Burgers Picture*

Finally, for completeness, it is relevant to briefly mention the following. Apart from the direct perturbation theory for solitons, there exists still another method to analyze the dynamics of dark solitons in the framework of Equation (30). Indeed, as shown in Ref. [42] for the special case of $\gamma = \delta = 0$, it is possible to employ a multiscale expansion method and asymptotically reduce the higher-order NLS equation to a Korteweg-de Vries–Burgers (KdV-B) equation. This can be done upon seeking solutions of the form:

$$u(x,t) = [u_\infty + U(x,t)]\exp[iu_\infty^2 t + i\phi(x,t)], \tag{69}$$

where $U(x,t)$ and $\phi(x,t)$ are unknown real functions (to be determined) representing an amplitude and a phase modulation of the background wave $u_b = u_\infty \exp(iu_\infty^2 t)$. Then, it is assumed that these functions are presented in the form of formal asymptotic series,

$$U = \epsilon^2 U_1 + \epsilon^4 U_2 + \cdots, \quad \phi = \epsilon\phi_1 + \epsilon^3\phi_2 + \cdots, \tag{70}$$

where the unknown functions U_j and ϕ_j ($j = 1, 2, \ldots$) depend on the slow variables $X = \epsilon(x - \upsilon t)$ and $T = \epsilon^3 t$, with υ being an unknown velocity, and ϵ being a formal small parameter. Substituting Equations (69)–(70) into Equation (30), we obtain the following results. First, to the lowest-order of approximation in ϵ of the perturbation technique, we derive the unknown velocity υ and an equation connecting the functions ϕ_{1T} and U_1. Second, to the next order of approximation, we derive the following KdV-B equation:

$$U_{1T} + c_1 U_1 U_{1X} + c_2 U_{1XXX} = c_3 U_{1XX}, \tag{71}$$

where, U_1 obviously represents the soliton amplitude. The coefficients of the underlying KdV equation, c_1 and c_2, depend on the coefficients of the pNLS, β and ν, as well as on the background amplitude u_∞, while the diffusion coefficient c_3 depends on the SRS parameter, σ_R.

Importantly, the relevant asymptotic reduction to the KdV-B equation can be performed for both normal and anomalous dispersion cases, i.e., for both $s = \pm 1$. Of course, in the case $s = -1$ it is known [1–3] that the soliton's background plane wave is prone to the modulational instability (MI), but this long-wavelength instability may be suppressed: indeed, in applications, one expects periodic or other boundary conditions in the x-direction, meaning that the admitted wavenumbers are quantized, hence they are limited from below by a minimum wavenumber, $k_{\min}$, which corresponds to the transverse size of the system. In such a case, if $k_{\min} > K_{\max}$ (where K is the perturbation wavenumber characterizing the MI, and $K_{\max}$ defines the width of the instability band, $0 \leq K \leq K_{\max}$), no quantized wavenumber can get into the instability band and, hence, the MI is eliminated.

The effective KdV-B description of the soliton dynamics offers a number of interesting results. First, in the absence of the SRS effect ($\sigma_R = 0$), dark solitons small-amplitude dark solitary wave solutions can exist for both the normal and anomalous dispersion regimes. This result is in sharp contrast with the conventional form and certain perturbed versions of the NLS equation, where dark solitons solely exist for the normal dispersion regime ($s = +1$). In addition, in this latter regime, there exists another type of solution, namely an *anti-dark soliton*, having the form of a hump, rather than a dip, on top of the background plane wave. Notice that the transformation from the dark to the anti-dark soliton is possible (see details in Ref. [42]).

When the SRS effect is present ($\sigma_R \neq 0$), the soliton dynamics is governed by a KdV–B equation. In this case, the evolution of solitons can be studied by means of the perturbation theory for solitons [33, 35]. The results that can been obtained in this case show that the solitons experience a decrease in their amplitudes and/or their velocities, depending on the direction of propagation and the dispersion region ($s = -1$ or $s = +1$). In particular, right-going solitons experience a decrease in both their amplitudes and their velocities, while the evolution of left-going solitons depends on

s: for $s = -1$, they increase their amplitudes and decrease their velocities, while for $s = +1$, they decrease their amplitudes and increase their velocities—in accordance with the results presented in the previous section.

Still another nonlinear wave structure that can be predicted to occur in this setting, is the one of a traveling shock wave [43]. In the effective KdV-B picture, the existence of such a structure is not surprising, because the KdV-B equation possesses stable traveling shock wave solutions. The latter, are obviously supported by the SRS effect (recall that if $\sigma_R = 0$ then $c_3 = 0$ and the diffusion term in Equation (71) vanishes, as was also found by means of other methods in other studies [44–46]. Notice that, as before, shock wave type structures are possible for both normal and anomalous dispersion cases. In particular, in the case of the normal dispersion ($s = +1$), the structure has the usual shock wave profile, while in the case of the anomalous dispersion ($s = -1$ it has the form of a rarefaction wave.

Finally, based on the analysis of the shock wave structure of the KdV-B equation, one may deduce the relevant profiles in the context of the perturbed NLS equation. Thus, the structure of the front of the shock solutions may be monotonic, in the nonlinearity-dominated regime, or oscillatory in the dispersion-dominated regime. In fact, since the former regime is only accessible for $s = -1$ [43] the front of the rarefaction wave is monotonic. On the other hand, the profile of the front of the shock wave supported for $s = +1$, may be either monotonic in the nonlinearity-dominated regime (resembling the regular stationary solutions of the Burgers equation), or oscillating. It is interesting to mention that the oscillations in the kink front can be studied in the framework of the perturbation theory for solitons of the KdV equation, treating the diffusion term as a small perturbation [33, 35]. This way, it can be deduced that the oscillations of the shock front can be considered as a succession of KdV solitons, a fact that completes the connection between the soliton and shock wave solutions of the perturbed NLS model.

References

1. A. Hasegawa, Y. Kodama, *Solitons in Optical Communications* (Oxford University Press, Oxford, 1996)
2. G.P. Agrawal, *Nonlinear Fiber Optics* (Academic Press, London, 2012)
3. Yu.S. Kivshar, G.P. Agrawal, *Optical Solitons: From Fibers to Photonic Crystals* (Academic Press, London, 2003)
4. M.J. Ablowitz, H. Segur, *Solitons and Inverse Scattering Transform* (SIAM, 1981)
5. V.E. Zakharov, A.B. Shabat, Sov. Phys. JETP. **34**, 62 (1972)
6. V.E. Zakharov, A. B. Shabat, Sov. Phys. JETP. **37**, 823 (1973)
7. M. Scalora, M.S. Syrchin, N. Akozbek, E.Y. Poliakov, G. D'Aguanno, N. Mattiucci, M.J. Bloemer, A.M. Zheltikov, Phys. Rev. Lett. **95**, 013902 (2005)
8. S. Wen, Y. Xiang, X. Dai, Z. Tang, W. Su, D. Fan, Phys. Rev. A **75**, 033815 (2007)
9. N.L. Tsitsas, N. Rompotis, I. Kourakis, P.G. Kevrekidis, D.J. Frantzeskakis, Phys. Rev. E **79**, 037601 (2009)
10. R.S. Johnson, Proc. R. Soc. Lond. A **357**, 131 (1977)
11. Yu.V. Sedletsky, J. Exp. Theor. Phys. **97**, 180 (2003)

12. A.V. Slunyaev, J. Exp. Theor. Phys. **101**, 926 (2005)
13. N.N. Akhmediev, A. Ankiewicz, *Solitons. Nonlinear Pulses and Beams* (Chapman and Hall, London, 1997)
14. H. Ikeda, M. Matsumoto, A. Hasegawa, Opt. Lett. **20**, 1113 (1995)
15. T.P. Horikis, D.J. Frantzeskakis, Opt. Lett. **38**, 5098 (2013)
16. S.C.V. Latas, M.F.S. Ferreira, M.V. Facão, Appl. Phys. B **104**, 131 (2011)
17. S.C.V. Latas, M.F.S. Ferreira, Opt. Lett. **37**, 3897 (2012)
18. I.M. Uzunov, T.N. Arabadzhiev, Z.D. Georgiev, Opt. Fiber Tech. **24**, 15 (2015)
19. V. Achilleos, S. Diamantidis, D.J. Frantzeskakis, T.P. Horikis, N.I. Karachalios, P.G. Kevrekidis, Physica D **316**, 57 (2016)
20. V. Achilleos, A.R. Bishop, S. Diamantidis, D.J. Frantzeskakis, T.P. Horikis, N.I. Karachalios, P.G. Kevrekidis, Phys. Rev. E **94**, 012210 (2016)
21. R. Temam, *Infinite-Dimensional Dynamical Systems in Mechanics and Physics* (Springer, Berlin, 1997)
22. T. Cazenave, *Semilinear Schrödinger Equations*. Courant Lecture Notes, vol. 10 (American Mathematical Society, Providence, 2003)
23. C. Sulem, P.L. Sulem, *The Nonlinear Schrödinger Equation. Self-Focusing and Wave Collapse* (Springer, Berlin, 1999)
24. T. Kato, *Quasilinear Equations of Evolution with Applications to Partial Differential Equations*. Lecture Notes in Mathematics, vol. 448 (Springer, New York, 1975), pp. pp. 25–70
25. T. Kato, *Abstract Differential Equations and Nonlinear Mixed Problems* (Lezione Fermiane Pisa, 1985)
26. T. Kato, C.Y. Lai, J. Funct. Anal. **56**, 15 (1984)
27. T. Kato, *Nonlinear Schrödinger Equations. Schrödinger Operators*. Sønderborg, 1988, 218–263, Lecture Notes in Phys., vol. 345 (Springer, Berlin, 1989)
28. J.M. Ball, Q. J. Math. Oxf. **28**, 473 (1977)
29. T. Ozawa, Y. Yamazaki, Nonlinearity **16**, 2029 (2003)
30. M. Taylor, *Partial Differential Equations III*. Applied Mathematical Sciences, vol. 117 (Springer, New York, 1996)
31. N.I. Karachalios, H. Nistazakis, A. Yannacopoulos, Discrete Cont. Dyn. Syst. A **19**, 711 (2007)
32. V. Achilleos, A. Álvarez, J. Cuevas, D.J. Frantzeskakis, N.I. Karachalios, P.G. Kevrekidis, B. Sánchez-Rey, Phys. D **244**, 1 (2013)
33. V.I. Karpman, E.M. Maslov, Sov. Phys. JETP **46**, 281 (1977)
34. D.J. Kaup, A.C. Newell, Proc. R. Soc. Lond. Ser. A **361**, 413 (1978)
35. Yu.S. Kivshar, B.A. Malomed, Rev. Mod. Phys. **61**, 761 (1989)
36. I.M. Uzunov, V.S. Gerdjikov, Phys. Rev. A **47**, 1582 (1993)
37. Yu.S. Kivshar, X. Yang, Phys. Rev. E **49**, 1657 (1994)
38. M.J. Ablowitz, S.D. Nixon, T.P. Horikis, D.J. Frantzeskakis, Proc. R. Soc. A **467**, 2597(2011)
39. Yu.S. Kivshar, B. Luther-Davies, Phys. Rep. **298**, 81 (1998)
40. D.J. Frantzeskakis, J. Phys. A: Math. Theor. **43**, 213001 (2010)
41. M.J. Ablowitz, S.D. Nixon, T.P. Horikis, D.J. Frantzeskakis, J. Phys. A: Math. Theor. **46**, 095201 (2013)
42. D.J. Frantzeskakis, J. Phys. A: Math. Gen. **29**, 3631 (1996)
43. D.J. Frantzeskakis, J. Opt. Soc. Am. B **9**, 2359 (1997)
44. Yu.S. Kivshar, Phys. Rev. A **42**, 1757 (1990)
45. G.P. Agrawal, C. Headley III, Phys. Rev. A **46**, 1573 (1992)
46. Yu. S. Kivshar and B. A. Malomed, Opt. Lett. **18**, 485 (1993).

The Role of Differential Equations in Applied Statistics

Christos P. Kitsos and C. S. A. Nisiotis

Abstract The target of this paper is to discuss, investigate and present how the differential equations are applied in Statistics. The stochastic orientation of Statistics creates problems to adopt the differential equations as an individual tool, but Applied Statistics is using the differential equations either through applications from other fields, like Chemistry or as a tool to explain "variation" in stochastic processes.

1 Introduction

Differential equations (de), [30] among others, support a number of sciences. Quincy Wright in his early work introduced (de) in political science, on a study of War. In such cases, there is a covered uncertainty. In others, like electric circuits, see Appendix 1, there is a solid background for (de), and not any uncertainty. In Astronomy, were (de) are extensively used, this uncertainty is measured through a Probability model, the Gaussian, [5].

Consider two points in the Universe and their distance l. Due to the expansion of the Universe the distance l it is not a constant function, but it depends on time t. Therefore considering a scale factor $s(t)$ acting as normalizing factor, the distance $D = D(t)$ is eventually of the form

$$D(t) = s(t)l(t)$$

C. P. Kitsos (✉)
Department of Informatics, University of West Attica, Aigaleo, Greece
e-mail: xkitsos@uniwa.gr

C. S. A. Nisiotis
Department of Public and Community Health, University of West Attica, Aigaleo, Greece
e-mail: csnisiotis@uniwa.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_10

From $D(t)$ we can obtain the velocity $V(t)$ and the acceleration $\alpha(t)$ as

$$V(t) = \dot{D}(t) = \dot{s}(t)l(t) + s(t)\dot{l}(t)$$
$$\alpha(t) = \dot{V}(t) = \ddot{s}l + 2\dot{s}\dot{l} + s\ddot{l}$$

where:

$$\dot{s}(t)l(t) = \text{isknownasvelocityduetotheexpansion}$$
$$s(t)\dot{l}(t) = \text{isknownaspeculiarvelocity, usuallydenotedby}u$$
$$2\dot{a}\dot{l}(t)s\ddot{l} = \text{isknownaspeculiaracceleration,}$$

Considering the gravitational potential G the equation of motion is [10],

$$\frac{d^2D}{dt^2} = \nabla_D G$$

If we set:

$$\Psi(l,t) := G(l,t) + \frac{1}{2}s\ddot{s}l^2$$

It can be proven, [10, 46], that:

$$\frac{\partial u}{\partial t} + \frac{\dot{s}}{s}u = -\frac{1}{s}\nabla\Psi$$

Let $p = p(x)$ be the density of the Universe at the position x, and $\bar{p}$ the mean density of the Universe. Thus, if we define:

$$\delta = \delta(x) = \frac{p(x) - \bar{p}}{\bar{p}}$$

the probability distribution of δ is Normal with mean zero and variance σ_δ^2 known i.e. $\delta \sim N(0, \sigma_\delta^2)$ or:

$$P(\delta) = \frac{1}{\sqrt{2\pi}\sigma_\delta}\exp\left(-\frac{1}{2}\frac{\delta^2}{\sigma_\delta^2}\right)$$

The above relation provides evidence that Astronomy is adopting Statistics on the way that Statistics is adopting (de). As Statistics is working with descriptive data, not continuous, the difference equations are adopted widely, but we shall not refer to this characteristic form of equations. In Section 4 we discuss how probability is adopted the (de) to measure the "rate of change" in a random walk.

Now, although the optimal experimental design theory was early originated [41], it was applied in Chemistry only by the pioneering work [15], who worked with the dilution assessment problem, while [25] investigates the problem for the sequential point of view. In principle, the essential problem with the non-linear experimental design problems is, that there is not a global statistical framework, for all the adopted models. There is a solid statistical theoretical background for any nonlinear function, but not for the particular nonlinear equation, which is arisen though a differential equation approach, in Chemical Kinetics, see the early work in [11–13], either from Chemistry or Biology. So each model should be particularly investigated, in order to calculate the optimal design points, which are functions of the unknown parameters! The proposed solution to overpass the unknown parameter repentance, [25, 26, 28] is to adopt the sequential principle of design.

The differential equations play a vital role to Statistical application in Chemical Kinetics, typical example being the Michaelis–Menten model, [29]. They are applied to create the model, which as it is based on observations the continuation is destroyed. Therefore the observations obtained are coming from the model plus a stochastic error. The usual assumption is the normality of the errors, with mean zero and variance constant $\sigma^2 > 0$. Here comes another problem in statistical non-linear case: the variance depends on the input variables and the parameters, we want to estimate. Although the D-optimality criterion appeared to have an aesthetic appeal, as it requests the minimization of the variance, in Chemical Kinetics, see [11] another problem appears: as nonlinear models does not provide, in Statistical terms, the appropriate information, the Taylor expansion has been applied, see the early work [6, 25], among others. Therefore what is consider in this paper is:

- The differential equation (de) that approaches the phenomenon
- Form the model in general Statistical model
- Consider the non Linear Statistical model
- Provide a general framework to obtain the optimal design.

The use and abuse of regression is also discussed as well as the use of differential equations in Statistics. Therefore the (de) are applied, in principle, in three characteristic lines of thoughts:

1. To provide a theoretical framework and a solid background to physical problems, see Appendix 1.
2. Although a theoretical background is developed through (de) principles there is an underlying uncertainty covered by Statistics.

2 Theoretical Framework for the Nonlinear Design

Consider the non-empty set $U \subseteq \mathscr{R}^k$ in which the k predictor variables or covariates or explanatory variables or independent variables are $u = (u_1, u_2, \cdots, u_k)$. We

assume that their values are within the known as experimental region or design space U. A typical example of an input variable from Chemical Kinetics is time.

The parameter space $\Theta \subseteq \mathscr{R}^P$ is the set where the p-term parameter vector $\theta = (\theta_1, \theta_2, \cdots, \theta_P)$ takes values. Where the sequential procedure of design is adopted [26, 31] Θ is assumed compact. When the response vector y is supposed to take any value in the response space, Ψ we also suppose that a *regression model*, in principle a nonlinear function f, that links u and θ, and consists of the deterministic portion $f(u, \theta)$ and the stochastic portion, e, known as error, linked, eventually, through the relation

$$y = f(u, \theta) + , \text{with} \quad E(y) = \eta = f(u, \theta) \tag{1}$$

where $E(y)$ means the expected value of y.

For the independent identically distributed errors we suppose that are normally distributed with mean 0 and variance $\sigma^2 > 0$. The function $f(u, \theta)$ (which describes the chemical kinetic model in our scenario) is assumed to be continuous with the second order derivatives of $f(\cdot)$ with respect to θ existing at and near the true value of the parameter, see [16]. For model (1) we introduce the quantity [3, 4]:

$$S_n(\theta) = \sum (y_i - f(u_i, \theta))^2 = \| \ y - f(u, \theta) \ \|_2 \tag{2}$$

where $\| \ \|_2^{\cdot}$ is the l_2-norm. An estimate $\hat{\theta}$ will be called the *least squares estimate* (LSE) if $S_n(\hat{\theta}) = \min\{S_n(\theta); \theta \in \Theta\}$.

In the non-linear regression problems the variance $\sigma^2 = \sigma^2(u, \theta)$. That is σ^2 depends on the design point and the parameter vector. In the linear case it is independent of the parameter θ.

The concept of the *average-per-observation information matrix* (apoim) is important for the nonlinear design problem, as described [29]. It is defined for ξ_n, the n-point design measure [16] (practically: the portion of observations at the optimal design space), to be, for the discrete case, equal to

$$M(\theta, \xi_n) = n^{-1} \sum I(\theta, u_i) \tag{3}$$

while for the continuous case is

$$M(\theta, \xi) = \int_U I(\theta, u) \xi(du). \tag{4}$$

The following partition of matrix M is usually considered

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \tag{5}$$

with $M_{11} \in \text{Mat}(s, s)$, $M_{12} \in \text{Mat}(s, p-s)$, $M_{22} \in \text{Mat}(p-s, p-s)$, $1 \leq s < p$ and $\text{Mat}(n, p)$ the set of $n \times p$ matrices, see also [20, 25].

We now briefly discussed how the Optimal Experimental Design Theory, for the Non-Linear model proceeds, see [16]. The average per observation information matrix is considered, were θ takes its true value. Then we can define the following operator J_Q applied to (apoim) M, Kitsos (1986), through a considered known matrix Q:

$$J_Q(M) = QM^{-}(\theta, \xi)Q^T,$$

M^{-} a generalized inverse of M and $Q^T \in Mat(p, s)$.

Given the above notation a real valued convex, decreasing function, ω say, on the set of nonnegative definite matrices, say $NMat(s, s)$, is applied to J_Q is what we consider as an optimality criterion function.

The design measure ξ^* is called ω-optimal if and only if:

$$\omega\left\{J_Q\left[M(\theta, \xi^*)\right]\right\} = \min\left\{\omega\left\{QM^{-}(\theta, \xi)Q^T\right\}, \xi \in \Xi\right\}. \tag{6}$$

For the special case for ω such as $\omega(\cdot) = \log\left[\det\left(QM^{-}Q^T\right)\right]$, and for $Q = I \in Mat(p, p)$ *the identity matrix* we are referring to D-optimal design, which is adopted in this paper. Different values of the convex function and the matrix Q define other optimality criteria (When the trace of the covariance matrix is minimized we are referred to A-optimality),which are beyond the target of this paper. Now, two theoretical results (DR1 and DR2) are useful to reduce the number of parameters required to calculate D-optimal designs:

DR1: Let $f(x, \theta)$ be any nonlinear model of the form:

$$f(x, \theta) = L(x_1, \beta_1) + NL(x_2, \beta_2)$$

where $\theta = \left(\theta_1, \theta_2, \ldots, \theta_s; \theta_{s+1}, \ldots, \theta_p\right) = (\beta_1; \beta_2)$, $NL(x_2, \beta_2)$ is the non-linear part, $L(x, \beta_1) = \beta_1^T x_1$, β_1^T the transpose of β_1 and $\dim\beta_1 = \dim x_1 = s$.

Then the D-optimal design depends only on β_2 (the Non-Linearly involved parameters) as well as the D_s-optimal design for the vector β_1.

DR2: Hill [20] defined the partially nonlinear model for the k parameters, $k < p$, to be the one for which $\nabla f(u, \theta) = B(\theta)H(u, \beta)$, where $B(\theta)$ is a matrix depending on $\theta = (\theta_1, \theta_2, \ldots, \theta_p$, β is the vector of the k parameters which appear in a nonlinear way and H is an appropriate matrix. In such a case the design depends only on the parameters of β.

We try to have results reducing the number of the involved parameters, as former information is needed for the involved parameters. So the less the involved parameters, the less the required information, especially when DR1 holds.

3 Growth Curves

In Statistics the nonlinear models are mainly produced by a linear system of differential equations, [7–9, 12, 19, 21, 22, 27, 31, 32, 40, 47], . See also the pioneer work [37], and [35, 45] . In principle compartment models are based on a division of the system into compartments. Then it is assumed that the "rates of flow" of whatever is under investigation, with typical example the drugs between compartments follow the first order kinetics, with the relevant (de) and eventually the rate of transfer is proportional to the concentration.

The target is to find a growth curve, as a mathematical model, to describe the plant dry weight data over a pre-defined season. A number of growth curves (gc) have been developed in various fields. The main approach remains the same, either they serve a chemical oriented purpose, [32, 33] (cgc) or a plant gc, (pgc).

We pay attention to pgc, in the presence of its dry weight W, which varies with time t, $W = W(t)$. Moreover growth exists at the expense of a substate S. The rate of the growth reaction is linearly proportional to S and W, so with λ being a constant it holds:

$$\frac{dIV}{dt} = \lambda SW \tag{7}$$

When W be maximum i.e. $W = W_m$ say, then $S = 0$ and in then the equality can be reduced to:

$$\frac{dW}{dt} = \lambda(W_m - W)W \tag{8}$$

The solution of (8) is, with $W_0 = W(t = 0)$ be the starting observation:

$$W \equiv W_0 \exp(W_m \lambda t) \tag{9}$$

In Statistical terms, as there were introduced in the above section, model (9) is considered as $\eta = \theta_0 \exp(\theta_1 t)$, with the parameters definition to be clear, see [33]. The Michaelis–Menten (M-M) model is a typical example, see [29]

$$E + S \underset{k_1}{\overset{k_2}{\rightleftharpoons}} ES \rightarrow E + P$$

with $E =$ enzyme, $S =$ substrate and $P =$ product. Eventually the model is:

$$u = \frac{u_{max}[s]}{k + [s]} \tag{10}$$

with $u =$ speed of the steady-state reaction $[\cdot]$ declares concentrations:

$$k = \frac{k_2 + k_3}{k_1}$$

Equation (10) declare that:

$$\left.\frac{du}{d[s]}\right|_{[s]=0} = \frac{u_{max}}{k} \tag{11}$$

Adopting the general notation in Statistical terms introduced in Section 2 model (12) can be considered as:

$$\eta = \frac{\theta_1 u}{\theta_2 + u} \tag{12}$$

where η is the rate of the enzyme reaction, θ_1 corresponds to the maximum rate reaction and θ_2 is the half saturation constant, with u being the concentration of substrate, see also [34].

Moreover under DR1 proposition considering $(\theta = \theta_1, \theta_2; \theta_3, \theta_4) = (\beta_1, \beta_2)$ with

$$L(u, \beta_1) = \theta_1 + \theta_2 u$$

$$NL(u, \beta_2) = \frac{\theta_3 u}{\theta_4 + u}$$

then an extending (M-M) model is:

$$f(u, \theta) = \theta_1 + \theta_2 u + \frac{\theta_3 u}{\theta_4 + u} \qquad (extM - M)$$

has exactly the same D-optimal design as

$$n_N = \frac{\theta_3 u}{\theta_4 + u}$$

Again under DR1 the design does not depend on θ_3 so prior information for (ext M-M) only for θ_4 is needed, while for the design approach see [33].

The Gompertz growth equation for given dry matter is D, by two plant research first-order (de) (1st de), with S being the senescence parameter:

$$\frac{dD}{dt} = \lambda D, \qquad \frac{d\lambda}{dt} = -\lambda S \tag{13}$$

or by the one: $\frac{dD}{dt} = \lambda_0 D e^{-St}$ with $\lambda_0 =$ specific growth at $t = 0$.

The Semi-Empirical model, [45], is the model which describes the lightflux densities within the canopy of an isolated plant.

Given a point $M = M(r)$, with r being the radius describing the position of the point M when the ray of light traverse. This de provides, by integration:

$$I = I_0 \exp\left[-\int_0^s kF(r)ds\right] \tag{14}$$

with $I_0 = I(s = 0)$. If k and F does not vary along the chosen path it is known that we can have:

$$I = I_0 e^{-kFS} \tag{15}$$

The estimating photosynthesis rate P for the presence of light variability is another example of adopting a mathematical approach. Let I be the flux density falling a leaf then it can be easily proved that when I changes:

$$\Delta P = \frac{\partial P}{\partial I}\Delta I + \frac{1}{2}\frac{\partial^2 P}{\partial I^2}(\Delta I)^2 \tag{16}$$

If we let the CO_2 density to be C and α, β constants then it is:

$$P = \frac{(\alpha I)(\beta C)}{\alpha I + \beta C} \tag{17}$$

Considering (9) for N fluctuations with N being large and taking into account the first order terms are canceled about the mean value variation:

$$\Delta P = \frac{1}{2} I^2 \frac{\partial^2 P}{\partial I}(CV)^2 \tag{18}$$

with (CV) being the coefficient of variation. From (17) and (18) one can obtain:

$$\frac{\Delta P}{P} = -\frac{(\alpha I)(\beta C)}{(\alpha I + \beta C)^2}(CV)^2 \tag{19}$$

Then the maximum values can be evaluated for the absolute correction P, especially for and the relative correction $\Delta P/P$ as a function of the coefficient of variation:

$$\left(\frac{\Delta P}{P}\right)_{max} = -\frac{1}{4}(CV)^2 \qquad \text{at } \alpha S = \beta C \tag{20}$$

Let us consider the light flux density I from a steady value I_1 say to a new one I_2, say. The effect on photosynthesis rate of charging is under consideration.

Let us assume that the centre is $t = 0$, where this charge occurs fast. Then for $t < 0$ let $I = I_1$ and for $t > 0$ let $I = I_2$, while the density of CO_2 is C. In such a case the system in photosynthesizing at steady rate P_1 and P_2 is respectively:

$$P_i = \frac{aI_i\beta C}{aI_i + \beta C} \qquad i = 1, 2$$

The time constant is in this case:

$$\tau = \frac{k}{aI\beta C}$$

Then as it holds:

$$\frac{dP}{dt} = \frac{1}{\tau}(P_2 - P)$$

one can obtain:

$$\frac{dP}{P_2 - P} = \frac{dt}{\tau}$$

Therefore the final model can be:

$$P = P_2 - (P_2 - P_1)e^{-t/\tau}$$

Thus from the general framework of Section 2 it can be written following the notation on a nonlinear Statistical model as:

$$\eta = \theta_0 - \theta_1 \exp(-\theta_2 u) \tag{21}$$

Notice that under DR1, DR2 the design depends only on θ_2.

Notice that Equation (21) hinds what is behind, as Statistics are involving of solving the stochastic problem

$$\eta = f(u, \theta) + error$$

This is an important point for our scenario: The use of differential equation is only to form, eventually equations like the (21) one. This is exactly what we present briefly in this paper in this section.

For an irreversible, unimolecular homogeneous reaction of the type $A \xrightarrow{k} B$ under isothermal conditions, the rate equation is described by: $\frac{d[A]}{dt} = -k[A]$, where $[A]$, as usually, is the concentration or partial pressure of reagent A, t is the reaction time and k the reaction rate constant[21]. Integrating, with a given initial concentration of A at time $t = 0$, $[A]_0$,

$$\frac{[A]}{[A]_0} = X_A = \exp(-kt) = exp\left[-At \exp\left(-\frac{\Delta E}{RT}\right)\right] \tag{22}$$

where the fraction $\frac{[A]}{[A]_0}$ resembles the molar fraction of A, X_A, k is the rate constant, which is a function of the absolute temperature T. Then (22) arises.

As A and $\frac{\Delta E}{R}$ are highly correlated, it is desirable to introduce k_1, the rate of the reaction at some specific reference temperature T_0.

Thus $k_1 = A \exp\left(-\frac{\Delta E}{RT}\right)$, so that Equation (21) is reduced to:

$$X_A = \exp\left\{-k_1 t_1 \exp\left[-\frac{\Delta E}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right]\right\}. \tag{23}$$

Rewriting Equation (21) in terms of the rate $\theta_1 (= k1)$, and letting (under the general framework of a nonlinear model):

$$\theta_2 = \frac{\Delta E}{R}, \qquad t_2 = \frac{1}{T} - \frac{1}{T_0},$$

we obtain the suitable model for the developed experimental design framework as in Equation (1), with $\theta \subseteq \mathscr{R}^2$, see also [33]:

$$\eta = \exp\left[-\theta_1 t_1 \exp\left(-\theta_2 t_2\right)\right], \quad u = (t_1, t_2) \in U \subseteq \mathscr{R}^+ x \Delta, \quad \Delta = [380, 450]. \tag{24}$$

This is the first order decay law for η, the molar fraction of A (X_A).

If it is assumed that $f(\infty, \theta) - f(0, \theta) = 1$ then the model is reduced to

$$\eta = \theta_1 - \exp(-\theta_2 u). \tag{25}$$

Under DR1 model (24) depends only on $theta_2$, therefore prior information only for θ_2 is needed.

Recall to verify that the design depends only on θ_2.

The reaction network of two irreversible first-order reactions in series proceeds according to the scheme $A \xrightarrow{k_1} B \xrightarrow{k_2} C$: a raw material A reacts to form a partial product B which, in turn, decomposes to give a substance C.

We shall refer bellow to the law of mass action is the proposition that the rate of the chemical reaction is directly proportional to the product of the activities or concentrations of the reactants. This law forms (at least) a differential equation and not only explains abut also predicts behaviors of solutions in dynamic equilibrium.

Using the Guldberg-Waage form, of the reaction rates, to describe the network, provide for constant volume the simultaneous equations:

$$\left.\begin{aligned} &\frac{d[A]}{dt} = -k_1[A], \qquad \frac{d[B]}{dt} = k_1[A] - k_2[B], \qquad \frac{d[C]}{dt} = k_2[B], \\ &[A]_0 + [B]_0 + [C]_0 = [A] + [B] + [C] \end{aligned}\right\} \tag{26}$$

Integration of the differential equation for the concentration of A, $[A]$, with the initial conditions $[A] = [A]_0$ at $t = 0$ yields $[A] = [A]_0 \exp(-k_1 t)$. Substituting

into the differential equation for the concentration of B, $[B]$, gives:

$$\frac{d[B]}{dt} + k_2[B] = k_1[A]_0 \exp(-k_1 t). \tag{27}$$

With the initial conditions $[B]_0 = [C]_0 = 0$ at $t = 0$ the solution of (27) for the concentration of B, adopting the integrating factor method, is eventually

$$[B] = \frac{k_1[A]_0}{k_2 - k_1}\left[\exp(-k_1 t) - \exp(-k_2 t)\right]. \tag{28}$$

The experimental design model produced between the response η (concentration of B, $[B]$) and time t, will be in terms of Equation (1):

$$\eta = \frac{\theta_1}{\theta_1 - \theta_2}\left[\exp(-\theta_2 u) - \exp(-\theta_1 u)\right], \qquad u \in U = \mathscr{R}^+, \quad \Theta \subseteq \mathscr{R}^2 \tag{29}$$

where θ_1 and θ_2 correspond to k_1 and k_2, respectively, both functions of the temperature according to the Arrhenius law, [41, 43]. The optimal design for the model that appears in (29) is considered with $(t, T) \in U \subseteq \mathscr{R}^+ x \Delta, \quad \Delta = [380, 450]$. For the experimental design of this model, also known as compartmental, see [2]. Kitsos and Kolovos [33] worked on a number of Chemical Models and calculated their D-optimal statistical design points, see following Section 2.

4 Differential Equations in Probability

Differential equations can be traced as a useful tool in Probability Theory, [5, 14, 23] . In the sequence we present in a compact form differential equations are essential in Applied Statistical Theory.

Although the difference equations play an important role in Probability theory differential equations are also very useful. The well known Chapman–Kolmogorov differential equations are the most applied either the forward one or the backwards type. In this section we are focused on the following type differential equations applied in Probability Theory.

4.1 Brownian Motion

The English botanist Brown, in 1827, introduced the physical phenomenon, known hear after the Brownian motion. In physics, was introduced in 1905 by Einstein. For us in this paper Brownian motion is an example of a continuous time, continues state space, Markov process $X(t) = X_t, t \in T$ with characteristics:

B1: A process with independent increments. It is reflected then that the changes of X_t over non-over-lapping time periods are independent random variables (irv.).
B2: The probability distribution of $X(t_2) - X(t_1)$, $t_2 > t_1$ depends only on $t_2 - t_1$
B3: $P[X(r) - X(s) \leq x] = \frac{1}{\sqrt{2\pi(r-s)}} \int_{-\infty}^{x} \exp\left[-\frac{u^2}{2} A(r-s)du\right]$, $r > s$ with A being a positive constant.

Let assume that $X_0 = 0$. Note that $E(X(t)) = 0$, $Var(X(t)) = \sigma^2(X_t) = At$. At time t_0 the particle process the position $X(t_0) = X_0$. The conditional probability density of $X(t + t_0)$ given that $X(t_0) = x_0$ is denoted by $\phi(x, b|x_0)$, see also Appendix 2.

It is important that Einstein proved that $\phi(x, b|x_0)$ satisfies the partial differential equation, known as the diffusion equation:

$$\frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2}$$

with D being the diffusion coefficient. For the Diffusion equation the requests for the density function $p(x, t|x_0)$ act as boundary conditions and provide a unique solution. Moreover:

$$D = \frac{2RT}{Nf}$$

where R is the gas constant, T the temperature, N is Avogadro's number and f is a coefficient of friction. With the appropriate scales $D = \frac{1}{2}$ and then:

$$\phi(x, t|x_0) = \frac{1}{\sqrt{2\pi t}} \exp\left[-\frac{1}{2t}(x - x_0)^2\right]$$

the well known Gaussian distribution.

Therefore it comes easily that a Brownian motion is a stochastic process $\{X(t), t \geq 0\}$ with:

Br1. $\forall t, s \qquad X(t + s) - X(s) \sim N(0, \sigma^2 t)$
Br2. Consider time intervals $[t_1, t_2] \cap [t_3, t_4] = \emptyset$, $t_L \leq t_{L+1}$, $L = 1, 2, 3$ then $X(t_{L+1}) - X(t_L) \sim N(0, \sigma^2 t)$ $L = 1, 2, 3$ and are independent random variables. This is true for r disjoint time intervals.
Br3. $X(0) = 0$, $X(t)$ is continuous as $t = 0$

The $X^*(t) = X(t)/\sigma \sim N(0, 1)$ is the standard Brownian motion. Given the interval $I = (t_0, t_1)$, $X(0) = 0$, the probability that $X(t)$ has at least one zero within I is

$$\gamma = \frac{2}{\pi} \arccos \sqrt{\frac{t_0}{t_1}}$$

Another very important stochastic process in Probability Theory is the Wiener process. The Wiener process, W_t, is a continuous-time process based on:

W1 : $W_0 = 0$
W2 : W_t is almost surely continuous
W3 : W_t has independent increments
W4 : $W_t - W_s \sim N(0, t-s), \quad 0 \leq s < t$

Through the Wiener process the stochastic differential equation (s-de):

$$dX_t = \mu_t dt + \sigma_t dW_t, \qquad \mu_t = \mu(X_t, t), \quad \sigma_t = \sigma(X_t, t)$$

and μ is the drift and the diffusion coefficient:

$$D = D(X_t, t) = \frac{1}{2}\sigma^2(X_t, t)$$

The Fokker–Plank equation for the probability density $p(x, t)$ of a given variable (rv) X_t is:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}\left[\mu(x,t)p(x,t)\right] + \frac{\partial^2}{\partial x^2}\left[D(x,t)p(x,t)\right]$$

When the drift is zero and the diffusion constant we are referred to Brownian motion discussed above.

4.2 Pure Birth Process

Suppose we have a system $E = \{E_1, E_2, \ldots, E_j, E_{j+1}, \ldots, E_n\}$ and that from state E_j you can move only to E_{j+1}. Moreover in state E_n at time t, the probability of a jump at a sort time interval $I = (t, t+h)$ equals $\lambda_n h + O(h)$.

The probability of more than one jump with I is $O(h)$.

Let $P_n(t)$ be the probability that at time t the system is at stage E_n. Then,

$$\frac{P_n(t+h) - P_n(t)}{h} = -\lambda P_n(t) + P_{n-1}(t) + \frac{O(h)}{h} \tag{30}$$

with $h \to 0$ the above is reduced to:

$$P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t) \tag{31}$$

with:

$$P_0'(t) = -\lambda P_0(t) \tag{32}$$

The Poisson process satisfies the above (de) and it holds

$$P_k(0) = 1, \qquad P_n(0) = 0, \qquad n \neq k$$

The Poisson process acts as the solution of (31) with form

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \qquad n \geq 1.$$

4.3 The Birth-and-Death Process

The assumption about the system $E = \{E_1, E_2, \ldots, E_{j-1}, E_j, E_{j+1}\}$ is now changing, that the system changes only through transitions from states to their nearest ones (i.e., can move backwards). In principle from E_n moves either to E_{n+1} or E_{n-1}, while E_0 give rise only to E_1 movement. Consider the time interval $I = (t, t + h)$ the probability that the transition is from E_n to E_{n+1} equals to $\lambda_n h + O(h)$, while the probability that from E_n there is a movement to E_{n-1} equals to $\mu_n h + O(h)$. The probability that within I move that one change takes place is $O(h)$.

If we let $\nu_n = \lambda_n + \mu_n$, due to the independence of the events, it can be eventually proved:

$$P_n(t + h) = P_n(t)[1 - \nu_n h] + \lambda_{n-1} h P_{n-1}(t) + \mu_{n+1} h P_{n+1}(t) + O(h) \tag{33}$$

Thus, from the above equation we obtain (divining by h, $h \to 0$)

This theory is extended to more variables under the light of physical applications and will not discussed here. We believe that the birth / death process is very important in applications.

$$\begin{aligned} P_n'(t) &= -\nu_n P_n(t) + \lambda_{n-1} P_{n-1}(t) + \mu_{n+1} P_{n+1}(t), \qquad n \geq 1 \\ P_0'(t) &= -\lambda_0 P_0(t) + \mu_1 P_1(t) \end{aligned} \tag{34}$$

with initial conditions:

$$P_1(0) = 1, \qquad P_n(0) = 0$$

If the coefficients are bounded there is a unique solution, under the regularity condition $\sum P_n(t) = 1$, while for $\sum P_n(t) < 1$ there exist infinitely many solutions.

The solution of (34) can be obtained by induction. Let:

$$\pi_0 = 1, \qquad \pi_j = \frac{\lambda_0 \lambda_1 \cdots \lambda_{j-1}}{\mu_1 \mu_2 \cdots \mu_j} \qquad j \geq 1$$

Then from (34) we obtain: $P_1 = \lambda_0/\mu_1 \qquad P_1 = \pi_1 P_0$

Assuming that: $P_k = \pi_k P_0 \qquad k = 1, \ldots j$ we have eventually, [23] $P_{j+1} = \pi_{j+1} P_0$

Trivially the consequence P_j, $j = 1, 2, \ldots$ defines a distribution, if $\sum P_j = 1$. Provided that $\sum \pi_k < \infty$ it holds:

$$P_j = \frac{\pi_j}{\sum \pi_k} \qquad j = 0, 1, 2, \ldots$$

It can be proven that the limit

$$\lim_{k \to \infty} P_n(t) = P_n \tag{35}$$

exist and is independent of the initial conditions. Due to (35) for an ordinary Markov chain it holds:

$$P_{n,n+1} = \frac{\lambda_n}{\nu_n}, \qquad P_{n,n-1} = \frac{\mu_n}{\nu_n}$$

Let $a_i, i = 1, 2, \ldots, n$ be the probability of absorption into state 0 from the initial state 1. Then we obtain that:

$$a_i = \frac{\lambda_i}{\nu_i} a_{i+1} + \frac{\mu_i}{\nu_i} a_{i-1} \qquad i \geq 1$$

where $a_0 = 0$. The above relation can be considered through the "embedded random walk" associated to a given *b-d* process.

As far as applications concerns see the early work [14] and [23].

The values of birth and death process give rise to a number of applications. The most well known is coming from the queuing theory, where the Kendall's notation is described by three factors: if "Arrivals" denotes the time between arrivals in queue, "Service" is the time service distribution and "channels" is the number of service channels open at the nobe, the queue is denoted by Assistance/Service/Channel. If the arrival process and the service time distribution is Markovian (or Memoryless) the M/M/1 system is a typical one. In principle the Arrival process is a Poisson process and the Service time distribution is an exponential distribution, see the pioneering work [24].

The birth and death rates are constant, say $\lambda_i = \lambda$, $\mu_i = \mu$ with the average rate of arrivals to be λ, and the average service time to be $1/\mu$.

The corresponding (de) for the evaluation of the probability that M/M/1 system is at state k at the given time t with $\nu = \lambda + \mu$ are:

$$P_0'(t) = \mu P_1(t) - \lambda P_0(t)$$
$$P_k'(t) = \lambda P_{k-1}(t) + \mu P_{k+1}(t) - \nu P_k(t)$$

when more than one channel exists a M/M/C queue exists with $(C) = S$ servers and an infinite buffer, the birth and death process is characterized by $\lambda_i = \lambda$ as above and:

$$\mu_i = \mu, \text{for} i \leq S-1, \mu_i = S_\mu, \text{for} i \geq S$$

We introduce the notation:

$$\nu_k = \lambda + k\mu, \quad \nu_s = \lambda + S_\mu$$

Then the corresponding (de) are:

$$P_0'(t) = \mu P_1(t) - \lambda P_0(t)$$
$$P_k'(t) = \lambda P_{k-1}(t) + (k+1)\mu P_{k+1}(t) - \nu_k P_k(t) \text{for} k = 1, 2, \ldots, S-1$$
$$P_k'(t) = \lambda P_{k-1}(t) + S\mu P_{k+1}(t) - \nu_s P_k(t), k \geq S$$

For different queues different (de) are needed, so (de) are essential in Probability Theory.

5 Discussion

The Design theory is widely applied, [36, 39, 48] to a number of different oriented experiments. At the same time different fields of Mathematics are considering Statistics, [1], among others. The target of this paper is to discuss how differential equations (de) are applied to Statistics. The back-bone of Statistics is the optimal Design theory. Although the problem of (de) seems to have with Statistics as intersection the null set this is not the true situation. As the Ca problem as faced [31] seems to have no relation with [38] this is not true: to solve the low dose problems, as Robins-Monro iterations, it is similar to solve an non-linear equation with Newton-Raphson method.

Same story with the solution of the partial (de), [42] and the optimal Design Problem, [16, 44]. Still the hidden theory is based on (de), so does a number of Growth Models, facing by Statistics. As it was proven in Section 3 the Growth Curves in most of the Chemical, Biological cases are based on non-linear problems coming from (de). For the Linear case of Growth Curves see [18] were no (de) are needed.

In Probability theory the use of (de) is clear to birth and death stochastic process-esare based on random walks, see Appendix 2. In Physics or Electrical Engineering (de) are applied extensively, see Appendix 1, and their use defines a solid and compact theoretical model, while in Statistics there exists a stochastic orientation

of the use: either the stochastic error in models or the Probabilistic development in random works. The Browian motion, eventually joints three sciences: Math, Stat and Physics.

There are three lines of though for facing a de: the theoretical inside, see [42] the applied—how we can solve a de—see [30] and the Numerical Analysis point of view, see [17].

It is true that no such a deep development of de exists in Statistics. But we would say, that as Statistics serves all Sciences, so do and de. We tried to cover how this so widely used lines of thought are, eventually, communicate.

Appendix 1

Let us consider the main types of fundamental electrical circuits, so essential to build a differential equations approach to electrical circuits. To avoid any confusion the imaginary unit is denoted by $j = (0, 1)$ and the current is denoted by $i = i(t)$.

1. *An RL—circuit.*
   The voltage source, where it is assumed that it is AC, and thus $E = E_0 \sin \omega t$. Then based on the 2nd law of Kirchoff:

$$L\frac{di}{dt} + R_i = E_0 \sin \omega t \qquad t \geq 0$$

   The (general) solution is:

$$i(t) = i_0 e^{-\frac{R}{L}t} + E_0 \left(R^2 + L^2\omega^2\right)^{-1/2} \sin\left(\omega t - Arctan\frac{\omega L}{R}\right)$$
$$+E_0\omega \left(R^2 + L^2\omega^2\right)^{-1} e^{-\frac{R}{L}t}$$

2. *An RC—Circuit.*
   Then it holds:

$$R_i + \frac{1}{C}\int_0^t i dt_1 = E(t)$$

   And if $E(t) = \text{constant} = E_0$ then:

$$R_i + \frac{1}{C}\int_0^t i dt_1 = E_0$$

   And the solution is:

$$i = i(t) = \frac{E_0}{R} e^{-t/RC},$$

3. *RC—circuit* with $E = E_0 \cos(\omega t)$
   Then it holds:

$$R_i + \frac{1}{C}\int_0^t i\,dt_1 = E_0 \cos(\omega t)$$

   And if we let:

$$u = \int_0^t i\,dt_1$$

   We obtain that:

$$R\frac{du}{dt} + \frac{u}{C} = E_0 \cos(\omega t), \qquad u(0) = 0$$

   Thus:

$$u = te^{-t/RC}\int_0^t \frac{E_0}{R} e^{-t_1/RC}\cos(t)du$$
$$= \frac{E_0}{R}\left[\frac{(1/RC)\left(\cos(\omega t) - e^{-t/RC}\right) + \omega\sin(\omega t)}{(1/RC)^2 + \omega^2}\right] = u_1 + u_2$$

   With:

$$u_1 = \frac{-E_0/R\left(e^{-t/RC}/RC\right)}{(1/RC)^2 + \omega^2}, \quad u_2 = \frac{E_0}{R}\frac{(1/RC)\cos(\omega t) + \omega\sin(\omega t)}{(1/RC)^2 + \omega^2}$$

   Eventually we can evaluate that:

$$i_1 = \frac{du_1}{dt} = \frac{-E_0/R\left(e^{-t/RC}/RC\right)^2}{(1/RC)^2 + \omega^2} = \frac{E_0}{R}e^{-t/RC}/(1 + \omega^2(RC)^2)$$
$$i_2 = \frac{du_1}{dt} = \frac{E_0}{R}\left[\frac{-\omega RC\sin(\omega t) + \omega^2(RC)^2\cos(\omega t)}{1 + (RC)^2\omega^2}\right]$$
$$\cong \frac{E_0}{R}\left[\frac{-\omega RC\sin(\omega t)}{1}\right] = -\omega C E_0 \sin(\omega t)$$

   Notice that with $t \ll i_1$, $t = RC$ say then: $i_1 \cong \frac{E_0}{R}\frac{1}{e}$.

4. *RL—circuit*
   Assuming that $E = E_0$ the (de) concerning this circuit is:

$$L\frac{di}{dt} + Ri = E_0, \qquad L(0) = 0$$

With solution: $i = \frac{E_0}{R} - \frac{E_0}{R} e^{-\frac{R}{L}t}$

5. *LC—circuit* with $E = E_0$.
   The corresponding (de) is:

$$L\frac{di}{dt} + \frac{1}{C}\int_0^t i dt_1 = E(t) \text{and the solution is :}$$

$$i = E_0\sqrt{C/L}\sin\left(t/\sqrt{LC}\right) \text{with} i(0) = 0 \text{and} L_i'(0) = E_0$$

6. *LC—circuit* with $E = E_0\cos(\omega t)$.
   The corresponding equation is: $L\frac{di}{dt} + \frac{1}{C}\int_0^t i dt = E_0\cos(\omega t)$.
   With solution eventually: $i = \frac{e^{j\omega t}}{(-L\omega^2 + j\omega R + 1/C)}$
   As it is known that it holds:

$$\sin(\omega t) = \frac{e^{j\omega t} - e^{-j\omega t}}{2j}$$

   with the appropriate calculations the general solution eventually is:

$$i = -\frac{E_0\omega\left(1/C - L\omega^2\right)\sin(\omega t + E_0 R\omega^2\cos(\omega t)}{\left(-L\omega^2 + 1/C\right)^2 + \omega^2 R^2}$$

   There is an extensive approach which is beyond the target of this appendix.

Appendix 2: Introduction to Heat Equation

Consider a random walk, where we assume that the probability to move to the closest up, down, backwards, forwards points are equal to 1/4. Let $P(x, y; t)$ be the probability that a particle at time t is at $(x, y) \in R^2$ point. Then we can see that at time $t + 1$ holds:

$$P(x, y; t+1) = \frac{1}{4}\{P(x-1, y; t) + P(x, y-1; t) + P(x+1, y, t) + P(x, y+1; t)\}$$

Thus for the difference:

$$P(x, y; t+1) - P(x, y; t) = \frac{1}{4}\Big\{P(x+1, y; t) - 2P(x, y; t) + P(x-1, y, t) + P(x, y+1; t) - 2P(x, y; t) + P(x, y-1; t)\Big\}$$

This difference equation approximated by the two-dimensional heat equation:

$$\frac{\partial P}{\partial t} = C\left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2}\right)$$

If we assume that the random—walk takes place in a limited domain D, usually of the form $D = (l, u) \times [L, U] \subseteq R^2$. More over it is assumed that the particle is absorbed when it reaches the boundary. Let (x_0, y_0) be the boundary points and $W = W(x_b, y_b)$ the associated "profit that is paid out".

If we denote by $P(x, y; x_0, y_0)$ the probability that the particle start from the (interior) point (x, y) to be absorbed at the boundary (x_b, y_b) the expected "profit" is:

$$u(x, y) = \sum_b P(x, y; x_b, y_b) W(x_b, y_b)$$

and satisfies the difference equation

$$u(x, y) = \frac{1}{4}\{u(x+1, y) + u(x-1, y) + u(x, y+1) + u(x, y-1)\}$$

with $u(x_b, y_b) = W(x_b, y_b)$. This equation is a well-known approximation of the Laplace equation:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

Recall that

$$\nabla^2 u = \begin{cases} 0 & \text{Laplaceequation} \\ C(x, y) & \text{Poissonequation} \end{cases}$$

Through the Laplace equation three different well-known problems, associated with the boundary are defined: Dirichlet's problem, Neuman's problem, Robbin–Churchill's problem for elliptic equations.

References

1. S.I. Amari, *Differential Geometrical Methods in Statistics* (Springer, Berlin, 2012)
2. A.C. Atkinson, K. Chaloner, A.M. Herzberg, J. Juritz, Optimum experimental designs for properties of a compartmental model. Biometrics **49**(2), 325–337 (1993)
3. Y. Bard, *Nonlinear Parameter Estimation* (Academic Press, New York, 1974)
4. M.D. Bates, G.D. Watts, *Nonlinear Regression Analysis and Its Applications* (Wiley, New York, 1988)

5. B.R. Bhat, *Modern Probability Theory* (Wiley, New Delhi, 1985)
6. G.E.P. Box, H.L. Lucas, Design of experiments in Nonlinear situation. Biometrika **49**, 77–90 (1959)
7. G.E.P. Box, W.G. Hunter, The experimental study of physical mechanisms, Technometrics **7**, 23–42 (1965)
8. S. Brunauer, P.H. Emmett, E. Teller, Adsorption of gases in multimolecular layers. J. Am. Chem. Soc. **60**(2), 309–319 (1938)
9. N.L. Carr, Kinetics of catalytic isomerisation of n-pentane. Indus. Eng. Chem. **52**(5), 391–396 1960
10. M. Davis, P.J.E. Peebles, On the integration of the BBGKY equations for the development of strongly nonlinear clustering in an expanding universe. Astrophys. J. Suppl. Ser. **34**, 425–50 (1977)
11. M.E. Davis, R.J. Davis, *Fundamentals of Chemical Reaction Engineering*, 1st edn. (McGraw Hill, New York, 2003), 368 pp.
12. C.T. De Wit, Resource use efficiency in agriculture. Agric. Syst. **40**(1–3), 125–151 (1992)
13. J. Downie, K.A. Shelstad, W.F. Graydon, Kinetics of the vapour-phase oxidation of toluene over a vanadium catalyst. Can. J. Chem. Eng. 39(5), 201–204 (1961)
14. W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1968)
15. R.A. Fisher, On the Mathematical Foundations of Theoretical Statistics. Philos. Trans. R. Soc. Lond. Ser. A, **222**, 309–368 (1922)
16. I. Ford, C.P. Kitsos, D.M. Titterington, Recent advances in nonlinear experimental design. Technometrics **31**(1), 49–60 (1989)
17. C.E. Fröberg, *Numerical Mathematics: Theory and Computer Applications* (Benjamin-Cummings, San Francisco, 1985)
18. A.F. Graybill, *Theory and Application of the Linear Model* (Duxbury Press, 1976)
19. K. Harmsen, A modified Mitscherlich equation for rainfed crop production in semi-arid areas: 1. Theory. Neth. J. Agric. Sci. **48**(3), 237–250 (2000)
20. P.D.H. Hill, D-optimal Designs for partially Nonlinear Regression Models. Technometrics **22**, 275–276 (1980)
21. L.H. Hosten, G.F. Froment, Isomerization of n-pentane. Ind. Eng. Chem. Process. Des. Dev. **10**(2), 280–287 (1971)
22. I.S. Jaswal, R.F. Mann, J.A. Juusola, J. Downie, The vapour-phase oxidation of benzene over a vanadium pentoxide catalyst. Can. J. Chem. Eng. **47**(3), 284–287 (1969)
23. S. Karlin, M.H. Taylor, *A First Course in Stochastic Processes* (Academic Press, New York, 1975)
24. D.G. Kendall, Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. Ann. Math. Stat. **1**, 338–54 (1953)
25. C.P. Kitsos, Design and inference for Non-linear Problems. Unpublished Ph.D. thesis, Univ. of Glascow
26. C.P. Kitsos, Fully-sequential procedures in nonlinear design problems. Comput. Stat. Data Anal. **8**(1), 13–19 (1989)
27. C.P. Kitsos, On the support points of D-optimal non-linear experimental design for kinetics, in *MODA4: Advances in Model-Oriented Data Analysis*, ed. by C.P. Kitsos, W.G. Muller (Physica Verlag, Heidelberg, 1995), pp. 71–76
28. C.P. Kitsos, Nonlinear: optimal—sequential experiment designs and applications, in *Proceedings of the 51st International Conference on Industrial Statistics: Aims and Computational Aspects*, ed. by C.P. Kitsos, L. Edler, Athens, Greece, August 16–17, 1997 (Physica-Verlag, Heidelberg, 1997), pp. 151–163
29. C.P. Kitsos, Design aspects for the Michaelis Menten model. Biometrical Lett. **38**(1), 53–66 (2001)
30. C.P. Kitsos, *Technological Mathematics and Statistics*, vols. I, II (New Technologies Pub., Greece, 2009)
31. C.P. Kitsos, Sequential Approaches for Ca tolerance models. Biometrie und Medizinische Informatik, Greifswalder Seminarbeichte, vol. 18 (Shaker Verlag, 2011), pp. 87–98

32. C.P. Kitsos, K.G. Kolovos, An Optimal Calibration Design for pH Meters. Instrumentation Sci. Tech. **38**(6), 436–447 (2010)
33. C.P. Kitsos, K.G. Kolovos, A compilation of the D-optimal designs in chemical kinetics. Chem. Eng. Commun. **200**(2), 185–204 (2013)
34. S. Lopez, J. France, W.J. Gerrits, M.S. Dhanoa, D.J. Humphries, J. Dijkstra, A generalized Michaelis-Menten equation for the analysis of growth. J. Anim. Sci. **78**(7), 1816–1828 (2000)
35. P. Mars, D.W. van Krevelen, Oxidations carried out by means of vanadium oxided catalysts. Chem. Eng. Sci. **3**, 41–59 (1954)
36. I. Martinez, I. Ortiz, C. Rodriguez, Optimal design for weighted rational models. Appl. Math. Lett. **2009**, **22**(12), 1892–1895
37. E.A. Mitscherlich, The law of the minimum and the law of diminishing soil productivity (In German). Landwirtschafliche Jahrbuecher **38**, 537–552 (1909)
38. J. Ortega, W. Rheinbolt, *Iterative Solution of Equations in Several Variables* (Academic Press, New York 1970)
39. L.J. Rodriguez-Aragon, J. Lopez-Fidalgo, Optimal designs for the Arrhenius equation. Chemom. Intell. Lab. Syst. **77**(1–2), 131–138 (2005)
40. G.A.F. Seber, C.J. Wild, *Nonlinear Regression* (Wiley, New York, 1989), 768 pp.
41. K. Smith, On the standard deviation of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. Biometrika **12**(1–2), 1–85 (1918)
42. G.D. Smith, *Solution of Partial Differential Equations* (Oxford University Press, London, 1965)
43. J.M. Smith, Chemical Engineering Kinetics (McGraw Hill, Singapore, 1981)
44. D. Steinberg, W. Hunter, Experimental design: review and comment. Technometrics **26**(2), 71–97 (1984)
45. H.M.J. Thorney, *Mathematical Models in Plant Physiology* (Academic Press, London, 1976)
46. G.B. Van Albada, Formation and evolution of clusters of galaxies (Errata: 15 330). Bull. Bull. Astron. Inst. Neth. **15**, 165 (1960)
47. Q. Wright, *A Study of War. 1942*, 2 vols. (Univ. of Chicago Press, Chicago, 1942)
48. V. Zarikas, V. Gikas, C.P. Kitsos, Evaluation of the optimal design Cosinor model for enhancing the potential of robotic theodolite kinematic observation. Measurement **43**(10), 1416–1424 (2010)

Geometric Derivation and Analysis of Multi-Symplectic Numerical Schemes for Differential Equations

Odysseas Kosmas, Dimitrios Papadopoulos, and Dimitrios Vlachos

Abstract In the current work we present a class of numerical techniques for the solution of multi-symplectic PDEs arising at various physical problems. We first consider the advantages of discrete variational principles and how to use them in order to create multi-symplectic integrators. We then consider the nonstandard finite difference framework from which these integrators derive. The latter is now expressed at the appropriate discrete jet bundle, using triangle and square discretization. The preservation of the discrete multi-symplectic structure by the numerical schemes is shown for several one- and two- dimensional test cases, like the linear wave equation and the nonlinear Klein–Gordon equation.

1 Introduction and Motivation

In general, symplectic integrators are robust, efficient and accurate in preserving the long time behavior of the solutions of Hamiltonian ordinary differential equations (ODEs) [1]. The basic feature of a symplectic integrator is that the numerical performance is designed to preserve a physical observable property, i.e., the symplectic form at each time step. Recently, it was shown that many conservative partial differential equations (PDEs) allow for description similar to the symplectic structure of Hamiltonian ODEs, called the multi-symplectic formulation (see, e.g., Refs. [2–5]). For example, in Ref. [2] authors develop the multi-symplectic

O. Kosmas
Modelling and Simulation Centre, MACE, University of Manchester, Manchester, UK
e-mail: odysseas.kosmas@manchester.ac.uk

D. Papadopoulos
Delta Pi Systems Ltd., Thessaloniki, Greece
e-mail: dimitris@delta-pi-systems.eu

D. Vlachos (✉)
Department of Informatics & Telecommunications, University of Peloponnese, Tripolis, Greece
e-mail: dvlachos@uop.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_11

structure of Hamiltonian PDEs from a Lagrangian formulation, using the variational principle. The wave equation and its multisymplectic structure have been studied by [6–8] from the Hamiltonian viewpoint.

On the other hand, in the past decades, nonstandard finite difference schemes have been well established by Mickens [9–11] to compensate the weaknesses that may be caused by standard finite difference methods as, such as the numerical instabilities. Regarding the positivity, the boundedness, and the monotonicity of solutions, nonstandard finite difference schemes have a better performance than standard ones, due to their flexibility to construct a nonstandard finite difference method. The latter can preserve certain properties and structures, which are obeyed by the original equations.

In the present paper, following our previous work [12] we pay special attention to the geometric structure of multisymplectic integrators through the use of nonstandard finite difference schemes for variational partial differential equations (PDEs). The considered approach comes as a first step towards developing a Veselov type discretization for PDEs in variational form, e.g. [2, 4, 5] and combines it with nonstandard nonstandard finite difference schemes of Mickens [9–11]. The resulting multisymplectic-momentum integrators have very good energy performance in the level of the conservation of a nearby Hamiltonian, under appropriate circumstances, up to exponentially small error [2].

In the following Section 2 we present a short overview of the standard numerical techniques relying on variational integrator schemes and their special case of exponential variational integrators in Section 3. Afterwards, nonstandard finite difference properties are employed for the derivation of nonstandard variational integrators by using a triangle discretization of the spacetime (Section 4.1). Then, in Sections 5 and 6, we demonstrate concrete applications of the proposed integrators, for the numerical solution of the linear wave equation, the Laplace equation and the Poisson equation. In Section 7, we perform dispersion analysis and convergence experiments to further illustrate the numerical properties of the method. Finally, in Section 8, we summarize the main conclusions coming out of our study.

2 Review of Variational Integrators

The discrete Euler–Lagrange equations can be derived in correspondence to the steps of derivation of the Euler–Lagrange equations in the continuous formulation of Lagrangian dynamics [3]. Denoting the tangent bundle of the configuration manifold Q by TQ, the continuous Lagrangian $L : TQ \to \mathbb{R}$ can be defined. In the discrete setting, considering approximate configurations $q_k \approx q(t_k)$ and $q_{k+1} \approx q(t_{k+1})$ at the time nodes t_k, t_{k+1}, with $h = t_{k+1} - t_k$ being the fixed time step, a discrete Lagrangian $L_d : Q \times Q \to \mathbb{R}$ is defined to approximate the action integral along the curve segment between q_k and q_{k+1}, i.e.,

$$L_d(q_k, q_{k+1}) \approx \int_{t_k}^{t_{k+1}} L(q(t), \dot{q}(t))dt. \tag{1}$$

Defining the discrete trajectory $\gamma_d = (q_0, \ldots, q_N)$, $N \in \mathbb{N}$, one can obtain the action sum

$$S_d(\gamma_d) = \sum_{k=1}^{N-1} L_d(q_k, q_{k+1}). \tag{2}$$

The discrete Hamilton's principle states that a motion γ_d of the discrete mechanical system extremizes the action sum, i.e., $\delta S_d = 0$. Through differentiation and rearrangement of the terms, holding the end points q_0 and q_N fixed, the discrete Euler–Lagrange equations are obtained [3]

$$D_2L_d(q_{k-1}, q_k) + D_1L_d(q_k, q_{k+1}) = 0, \qquad k = 1, \ldots, N-1, \tag{3}$$

where the notation D_iL_d indicates derivative with respect to the i-th argument of L_d, see also [3, 12–16].

The definition of the discrete conjugate momentum at time steps k and $k+1$ reads

$$p_k = -D_1L_d(q_k, q_{k+1}), \quad p_{k+1} = D_2L_d(q_k, q_{k+1}), \qquad k = 0, \ldots, N-1. \tag{4}$$

The above equations, also known as position-momentum form of a variational integrator, can be used when an initial condition (q_0, p_0) is known, to obtain (q_1, p_1).

To construct high order methods, we approximate the action integral along the curve segment between q_k and q_{k+1} using a discrete Lagrangian that depends only on the end points. We obtain expressions for configurations q_k^j and velocities $\dot{q}_k^j$ for $j = 0, \ldots, S-1$, $S \in \mathbb{N}$ at time $t_k^j \in [t_k, t_{k+1}]$ by expressing $t_k^j = t_k + C_k^j h$ for $C_k^j \in [0, 1]$ such that $C_k^0 = 0$, $C_k^{S-1} = 1$ using

$$q_k^j = g_1(t_k^j)q_k + g_2(t_k^j)q_{k+1}, \qquad \dot{q}_k^j = \dot{g}_1(t_k^j)q_k + \dot{g}_2(t_k^j)q_{k+1}, \tag{5}$$

where $h \in \mathbb{R}$ is the time step. We choose functions

$$g_1(t_k^j) = \sin\left(u - \frac{t_k^j - t_k}{h}u\right)(\sin u)^{-1}, \quad g_2(t_k^j) = \sin\left(\frac{t_k^j - t_k}{h}u\right)(\sin u)^{-1}, \tag{6}$$

to represent the oscillatory behavior of the solution, see [17, 18]. For continuity, $g_1(t_{k+1}) = g_2(t_k) = 0$ and $g_1(t_k) = g_2(t_{k+1}) = 1$ is required.

For any different choice of interpolation used, we define the discrete Lagrangian by the weighted sum

$$L_d(q_k, q_{k+1}) = h \sum_{j=0}^{S-1} w^j L(q(t_k^j), \dot{q}(t_k^j)), \tag{7}$$

where it can be easily proved that for maximal algebraic order

$$\sum_{j=0}^{S-1} w^j (C_k^j)^m = \frac{1}{m+1}, \tag{8}$$

where $m = 0, 1, \ldots, S-1$ and $k = 0, 1, \ldots, N-1$ see [17, 18].

Applying the above interpolation technique with the trigonometric expressions of (6), following the phase lag analysis of [13, 14, 17, 18], the parameter u can be chosen as $u = \omega h$. For problems that include a constant and known domain frequency ω (such as the harmonic oscillator) the parameter u can be easily computed. For the solution of orbital problems of the general N-body problem, where no unique frequency is given, a new parameter u must be defined by estimating the frequency of the motion of any moving point mass [16, 19–21].

3 Exponential Integrators

We now consider the Hamiltonian systems

$$\ddot{q} + \Omega q = g(q), \qquad g(q) = -\nabla U(q), \tag{9}$$

where Ω is a diagonal matrix (will contain diagonal entries ω with large modulus) and $U(q)$ is a smooth potential function. We are interested in the long time behavior of numerical solutions when ωh is not small.

Since $q_{n+1} - 2\cos(h\omega)q_n + q_{n-1} = 0$ is en exact discretisation of (9) we can consider the numerical scheme

$$q_{n+1} - 2\cos(h\omega)q_n + q_{n-1} = h^2 \psi(\omega h) g(\phi(\omega h) q_n), \tag{10}$$

where the functions $\psi(\omega h)$ and $\phi(\omega h)$ are even, real-valued functions satisfying $\psi(0) = \phi(0) = 1$, see [1]. The resulting methods using the latter numerical scheme are known as exponential integrators (for some examples of those integrators see the appendix).

3.1 Exponential High Order Variational Integrators

If we now use the phase fitted variational integrator for the system (9) the result of the discrete Euler–Lagrange equations (3) will be

$$q_{n+1} + \Lambda(u, \omega, h, S)q_n + q_{n-1} = h^2\Psi(\omega h)g(\Phi(\omega h)q_n), \tag{11}$$

where

$$\Lambda(u, \omega, h, S) = \frac{\sum_{j=0}^{S-1} w^j \left[\dot{g}_1(t_k^j)^2 + \dot{g}_2(t_k^j)^2 - \omega^2 \left(g_1(t_k^j)^2 + g_2(t_k^j)^2 \right) \right]}{\sum_{j=0}^{S-1} w^j \left[\dot{g}_1(t_k^j)\dot{g}_2(t_k^j) - \omega^2 g_1(t_k^j) g_2(t_k^j) \right]}. \tag{12}$$

Using the above expressions, to obtain exponential variational integrators that use expressions for configurations q_k^j and velocities $\dot{q}_k^j$ taken from (5), we get

$$\Lambda(u, \omega, h, S) = -2\cos(\omega h). \tag{13}$$

In [16] we have proved (using the phase lag analysis of [22]) that exponentially fitted methods using phase fitted variational integrators can be derived when (13) holds. So phase fitted variational integrators using trigonometric interpolation can be considered as exponential integrators, i.e. when using phase fitted variational integrators, keeping the phase lag zero the resulting methods are exponentially fitted methods (exponential integrators). Those methods has been tested on several numerical results in [16].

3.2 Frequency Estimation for Mass Points Motion in Three Dimensions

In our previous work [16], we constructed adaptive time step variational integrators using phase fitting techniques and estimated the required frequency through the use of a harmonic oscillator with given frequency ω. Here, in solving the general N-body problem by using a constant time step, a new frequency estimation is necessary in order to find for each body i) the frequency at an initial time t_0 and ii) the frequency at time t_k for $k = 1, \ldots, N-1$.

It is now clear that, by applying the trigonometric interpolation (6), the parameter u can be chosen as $u = \omega h$. For problems for which the domain of frequency ω is fixed and known (such as the harmonic oscillator) the parameter u can be easily computed. For the solution of orbital problems involved in the general N-body

problem, where no unique frequency is determined, the parameter u must be defined by estimating the frequency of the motion of any moving material point.

Towards this purpose, we consider the general case of N masses moving in three dimensions. If $q_i(t)$ $(i = 1, \ldots, N)$ denotes the trajectory of the i-th particle, its curvature can be computed from the known expression

$$k_i(t) = \frac{|\dot{q}_i(t) \times \ddot{q}_i(t)|}{|\dot{q}_i(t)|^3}, \tag{14}$$

where $\dot{q}_i(t)$ the velocity of the i-th mass with magnitude $|\dot{q}_i(t)|$ at a point $q_i(t)$. After a short time h, the angular displacement of that mass is $h|\dot{q}_i(t) \times \ddot{q}_i(t)|/|\dot{q}_i(t)|^2$, which for each mass's actual frequency gives the expression

$$\omega_i(t) = \frac{|\dot{q}_i(t) \times \ddot{q}_i(t)|}{|\dot{q}_i(t)|^2}. \tag{15}$$

From (14) and (15) the well known relation $\omega_i(t) = k_i(t)|\dot{q}_i(t)|$ holds (see also [16]).

For the specific case of many-particle physical problems, that can be described via a Lagrangian of the form $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q)$, where $M(q)$ represents a symmetric positive definite mass matrix and V is a potential function, the continuous Euler–Lagrange equations are $M(q)\ddot{q} = -\nabla V(q)$. In this case, the expression for frequency estimation (15), referred to the i-th body at time t_k, $k = 1, \ldots, N-1$, takes the form

$$\omega_i(t_k) = h^{-1} \frac{\left|M^{-1}(q_k)p_k \times \left(M^{-1}(q_k)p_k - M^{-1}(q_{k-1})p_{k-1}\right)\right|}{\left|M^{-1}(q_k)p_k\right|^2}, \tag{16}$$

where the quantities on the right hand side are the mass matrix, the configuration and the momentum of the i-th body. Since the frequency $\omega_i(t_k)$ must be also known at an initial time instant t_0 (in which the initial positions are $\bar{q}_0$ and initial momenta are $\bar{p}_0$), using the continuous Euler–Lagrange equation at t_0 we obtain

$$\omega_i(t_0) = \frac{\left|M^{-1}(\bar{q}_0)\bar{p}_0 \times \left(-M^{-1}(\bar{q}_0)\nabla V(\bar{q}_0)\right)\right|}{\left|M^{-1}(\bar{q}_0)\bar{p}_0\right|^2}. \tag{17}$$

Equations (16) and (17) provide an "estimated frequency" for each mass in the general motion of the N-body problem. This allows us to derive high order variational integrator methods using trigonometric interpolation where the frequency is estimated at every time step of the integration procedure. These methods show better energy behavior, i.e. smaller total energy oscillation than other methods which employ constant frequency, see [14, 16].

Before closing this section, it should be mentioned that the linear stability of our method is comprehensively analyzed in our previous works [14, 16, 19].

4 Triangle and Square Discretization

In order to express the discrete Lagrangian and discrete Hamilton function, we will use the definition of the tangent bundle TQ and cotangent bundle T^*Q as in [2] to fields over the higher-dimensional manifold X. In this way, we also view fields over X as sections of some fiber bundle $B \to X$, with fiber Y, and then consider the first jet bundle J^1B and its dual (J^1B^*) as the appropriate analogs of the tangent and cotangent bundles.

It is then possible to use the generalization of the Veselov discretization [4, 5] to multisymplectic field theory, by discretizing the spacetime X. For simplicity reasons we will restrict ourselves to the discrete analogue of $\dim X = 2$. Thus, we take $X = \mathbb{Z} \times \mathbb{Z} = (i, j)$ and the fiber bundle Y to be $X \times F$ for some smooth manifold F [2, 12].

4.1 Triangle Discretization

Assume that we have a uniform quadrangular mesh in the base space, with mesh lengths Δx and Δt. The nodes in this mesh are denoted by $(i, j), \in \mathbb{Z} \times \mathbb{Z}$, corresponding to the points $(x_i, t_j) := (i\Delta x, j\Delta t) \in \mathbb{R}^2$. We denote the value of the field u at the node (i, j) by u^i_j. We label the triangle at (i, j) with three ordered triple $((i, j), (i+1, j), (i, j+1))$ as $\bigtriangleup_{ij}$, and we define $X_{\bigtriangleup}$ to be the set of all such triangles, see Figure 1.

Then, the discrete jet bundle is defined as follows [2]

$$J^1_{\bigtriangleup} Y := \{(u^j_i, u^j_{i+1}, u^{j+1}_i) \in \mathbb{R}^3 : ((i, j), (i+1, j), (i, j+1)) \in X_{\bigtriangleup}\}, \tag{18}$$

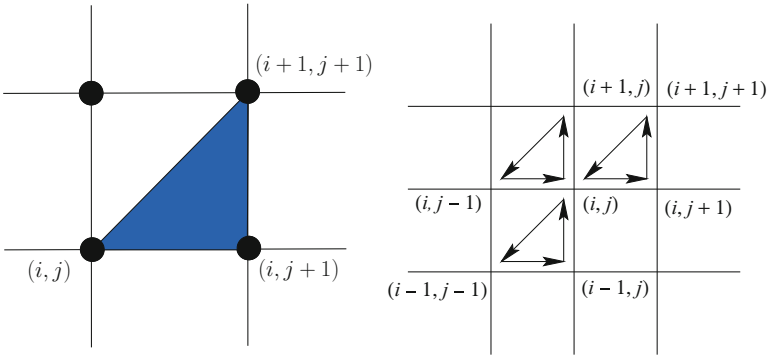


Fig. 1 The triangles which touch (i, j)

which is equal to $X_{\Delta} \times \mathbb{R}^3$. The field u can be now defined by averaging the fields over all vertices of the triangle (see Figure 1a)

$$u \to \frac{u_i^j + u_i^{j+1} + u_{i+1}^{j+1}}{3}, \tag{19}$$

while the derivatives can be expressed using nonstandard finite differences [9–11]

$$\frac{du}{dt} \to \frac{u_i^{j+1} - u_i^j}{\phi(\Delta t)}, \qquad \frac{du}{dx} \to \frac{u_{i+1}^{j+1} - u_i^{j+1}}{\psi(\Delta x)}, \tag{20}$$

with [9, 10]

$$\phi(\Delta t) = 2\sin\left(\frac{\Delta t}{2}\right), \quad \psi(\Delta x) = 2\sin\left(\frac{\Delta x}{2}\right). \tag{21}$$

Using the latter expressions, we can obtain the discrete Lagrangian at any triangle, which depends on the edges of the triangle, i.e., $L_d(u_i^j, u_i^{j+1}, u_{i+1}^{j+1})$, while the discrete Euler–Lagrange field equations are

$$D_1 L_d(u_i^j, u_i^{j+1}, u_{i+1}^{j+1}) + D_2 L_d(u_i^{j-1}, u_i^j, u_{i+1}^j) + D_3 L_d(u_{i-1}^{j-1}, u_{i-1}^j, u_i^j) = 0, \tag{22}$$

see Figure 1 (right).

4.2 Square Discretization

For the cases where square discretization is used, and if we also denote a square at (i, j) with four ordered quaternion $((i, j), (i+1, j), (i+1, j+1), (i, j+1))$ by $\square_j^i$, we can consider $X_{\square}$ to be the set of all such squares, see Figure 1. Then, the discrete jet bundle is defined as (for more details see [2] and references therein)

$$J^1_{\square} Y := \{(u_i^j, u_{i+1}^j, u_{i+1}^{j+1}, u_i^{j+1}) \in \mathbb{R}^4 : ((i, j), (i+1, j), (i+1, j+1), (i, j+1)) \in X_{\square}\}, \tag{23}$$

which is equal to $X_{\square} \times \mathbb{R}^4$.

By averaging the fields over all vertices of the square, the field u can be now obtained as (see Figure 2 (left))

$$u \to \frac{u_i^j + u_{i+1}^j + u_i^{j+1} + u_{i+1}^{j+1}}{4}. \tag{24}$$

As above, the expressions for the derivatives can be taken from [9–11] for the discrete Lagrangian, which now depends on the edges of the square, i.e.,

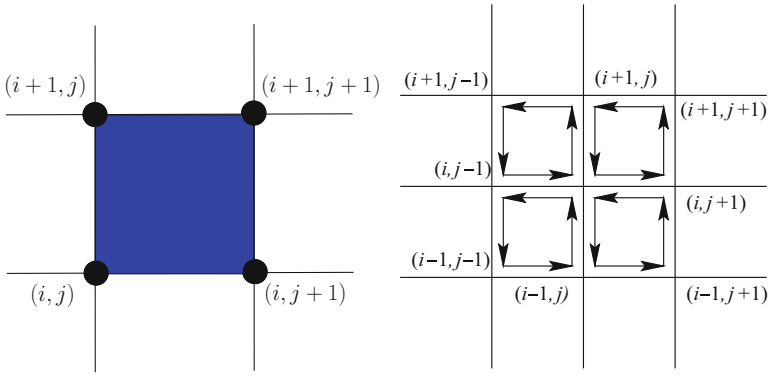


Fig. 2 The triangles which touch (i, j)

$L_d(u_i^j, u_{i+1}^j, u_{i+1}^{j+1}, u_i^{j+1})$. As a result, the discrete Euler–Lagrange field equations are

$$D_1 L_d(u_i^j, u_i^{j+1}, u_{i+1}^{j+1}, u_{i+1}^j) + D_2 L_d(u_i^{j-1}, u_i^j, u_{i+1}^j, u_{i+1}^{j-1}) + D_3 L_d(u_{i-1}^{j-1}, u_{i-1}^j, u_i^j, u_i^{j-1}) + D_4 L_d(u_{i-1}^j, u_{i-1}^{j+1}, u_i^{j+1}, u_i^j) = 0, \tag{25}$$

see Figure 2 (right).

5 Numerical Examples Using Triangle Discretization

To illustrate the proposed method, we consider the basic PDEs of three physical problems, i.e., the linear wave equation, the Laplace equation, and the Poisson equation (see [2] and [23, 24]). In the following subsections, for representation requirements, quadrilaterals have been by interpolating the solution on triangles.

5.1 Linear Wave Equation

The linear wave equation contains second order partial derivatives of the wavefunction $u(x, t)$ with respect to time and space, respectively, as (see e.g. [23, 24])

$$\frac{\partial^2 u}{\partial t^2} + c\frac{\partial^2 u}{\partial x^2} = 0. \tag{26}$$

This equation may be considered for the description of the wave function, i.e., the amplitude of oscillation, that is created from a one-dimensional medium (e.g. a

string extended in the x-direction). For the special case that the velocity of the wave, representing by the parameter c, is chosen as $c = -1$, the corresponding Lagrangian is [12]

$$L(u, u_t, u_x) = \frac{1}{2}u_t^2 - \frac{1}{2}u_x^2, \tag{27}$$

where the derivatives are $\partial u/\partial t = u_t$ and $\partial u/\partial x = u_x$.

If we use triangle discretization, described in Section 4.1, we end up with discrete Lagrangian

$$L_d\left(u_i^j, u_i^{j+1}, u_{i+1}^{j+1}\right) = \frac{1}{2}\Delta t \Delta x \left[\frac{1}{2}\left(\frac{u_i^{j+1} - u_i^j}{\phi(\Delta t)}\right)^2 - \frac{1}{2}\left(\frac{u_{i+1}^{j+1} - u_i^{j+1}}{\psi(\Delta x)}\right)^2\right], \tag{28}$$

where Δt and Δx are the mesh lengths for time and space, respectively. Applying the above discrete Lagrangian to the discrete Euler–Lagrange field equations (22), we get

$$\frac{u_i^{j+1} - 2u_i^j + u_i^{j-1}}{(\phi(\Delta t))^2} - \frac{u_{i+1}^j - 2u_i^j + u_{i-1}^j}{(\psi(\Delta x))^2} = 0. \tag{29}$$

The latter expression represents the variational integrator for the linear wave Equation (26), resulting through the use of the proposed nonstandard finite difference schemes.

In Figure 3 the solution $u(x, t)$ of (29) is shown in a 3-D diagram. We have chosen as initial conditions $0 < x < 1, u(x, 0) = 0.5[1-\cos(2\pi x)], u_t(x, 0) = 0.1$ and as boundary conditions $u(0, t) = u(1, t)$, $u_x(0, t) = u_x(1, t)$, the latter being periodic. The grid discretization has been taken to be $\Delta t = 0.01$ and $\Delta x = 0.01$. As seen, the time evolution of the solution $u(x = const., t)$ is a continuous function, while the periodicity is preserved.

5.2 *Laplace Equation*

As another physical example, we have chosen the Laplace equation over a 2-D scalar field $u(x, y)$. It is written as

$$u_{xx} + u_{yy} = 0. \tag{30}$$

The function $u(x, y)$ may describe a potential in a 2-D medium or a potential inside a 3-D medium, which does not depend on the third coordinate z. Thus, the 2-dimensional second order PDE (30) governs a variety of equilibrium physical

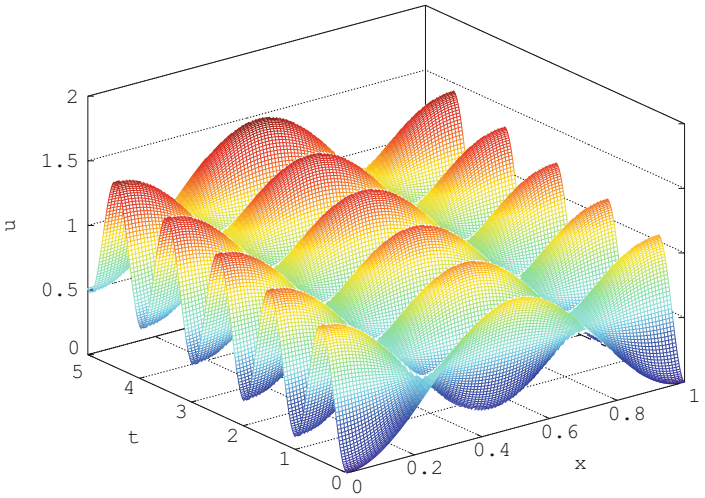


Fig. 3 The waveforms of linear wave equation (26)

phenomena such as temperature distribution in solids, electric field in electrostatics, inviscid and irrotational two-dimensional flow (potential flow), groundwater flow, etc.

The corresponding continuous Lagrangian of (30) takes the form

$$L(u, u_x, u_y) = \frac{1}{2}u_x^2 + \frac{1}{2}u_y^2. \tag{31}$$

By applying the triangle discretization of Section 4.1, the discrete Lagrangian can be written as

$$L_d\left(u_i^j, u_i^{j+1}, u_{i+1}^{j+1}\right) = \frac{1}{2}\Delta x \Delta y \left[\frac{1}{2}\left(\frac{u_i^{j+1} - u_i^j}{\phi(\Delta x)}\right)^2 + \frac{1}{2}\left(\frac{u_{i+1}^{j+1} - u_i^{j+1}}{\psi(\Delta y)}\right)^2\right]. \tag{32}$$

From the latter Lagrangian, working in a similar manner to that followed in Section 4.2, results the integrator from the proposed nonstandard finite difference schemes

$$\frac{u_i^{j+1} - 2u_i^j + u_i^{j-1}}{(\phi(\Delta x))^2} + \frac{u_{i+1}^j - 2u_i^j + u_{i-1}^j}{(\psi(\Delta y))^2} = 0. \tag{33}$$

The solution of the above equation, when considering the boundary conditions $u(x, 0) = 0, u(x, 1) = 1$ and $u(0, y) = u(1, y) = 0$, is plotted in Figure 4. The grid discretization has been chosen to be $\Delta x = 0.02$ and $\Delta y = 0.02$.

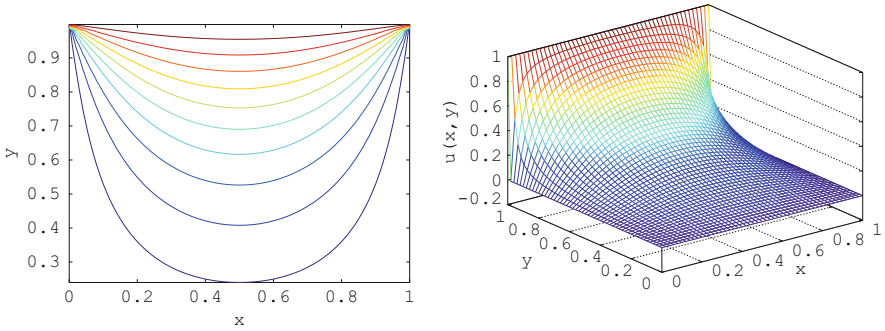


Fig. 4 Contour plot (left) and three-dimensional surface plot (right) of the solution of Laplace equation with boundary conditions $u(x, 0) = 0$, $u(x, 1) = 1$, $u(0, y) = u(1, y) = 0$, and discretization: $\Delta x = 0.02$, $\Delta y = 0.02$

5.3 Poisson Equation

As a final application to illustrate the advantages of the proposed variational integrator relying on nonstandard finite difference schemes, we examine the Poisson equation, which is an elliptic PDE of the form

$$-u_{xx} - u_{yy} = f(x, y). \tag{34}$$

Obviously, this equation in physical applications presents an additional complexity compared to the Laplace equation (30). Now the right hand side is a non-zero function $f(x, y)$, which may be considered as a source (or a load) function defined on some two-dimensional domain denoted by $\Omega \subset \mathbb{R}^2$ (it could also be a general non-linear function $f(u, x, y)$). A solution u satisfying (34) will also satisfy specific conditions on the bounderies of the domain Ω. For example, for the element $\partial\Omega$ the general condition holds

$$\alpha u + \beta \frac{\partial u}{\partial n} = g \qquad \text{on} \quad \partial\Omega, \tag{35}$$

where $\partial u/\partial n$ denotes the directional derivative in the direction normal to the boundary $\partial\Omega$ and α and β are constants [23, 24].

As it is well known, the system of (34) and (35) is referred to as a boundary value problem for the Poisson equation. If the constant β in Equation (35) is zero, then the boundary condition is of Dirichlet type, and the boundary value problem is referred to as the Dirichlet problem for the Poisson equation. Alternatively, if the constant α is zero, then we correspondingly have a Neumann boundary condition, and the problem is referred to as a Neumann problem. A third possibility exists when the Dirichlet conditions hold on a part of the boundary $\partial\Omega_D$, and Neumann conditions

hold on the remainder $\partial\Omega \setminus \partial\Omega_D$ (or indeed mixed conditions where α and β are both nonzero), see [23, 24] and references therein.

Equation (34) can also be obtained by starting from the Lagrangian

$$L(u, u_x, u_y) = \frac{1}{2}u_x^2 + \frac{1}{2}u_y^2 - fu. \tag{36}$$

The triangle discretization of Section 4.1 in the Poisson problem defines the discrete Lagrangian

$$L_d\left(u_i^j, u_i^{j+1}, u_{i+1}^{j+1}\right) = \frac{1}{2}\Delta x\Delta y\left[\frac{1}{2}\left(\frac{u_i^{j+1}-u_i^j}{\phi(\Delta x)}\right)^2 + \frac{1}{2}\left(\frac{u_{i+1}^{j+1}-u_i^{j+1}}{\psi(\Delta y)}\right)^2\right] - \frac{f_i^j u_i^j + f_i^{j+1}u_i^{j+1} + f_{i+1}^{j+1}u_{i+1}^{j+1}}{3}. \tag{37}$$

By inserting the latter discrete Lagrangian into the discrete Euler–Lagrange field equations (22) and elaborating as done in [2], the resulting integrator from the proposed nonstandard finite difference schemes is

$$-\frac{u_i^{j+1}-2u_i^j+u_i^{j-1}}{(\phi(\Delta x))^2} - \frac{u_{i+1}^j-2u_i^j+u_{i-1}^j}{(\psi(\Delta y))^2} = f_i^j + \partial f_i^j/\partial u_i^j. \tag{38}$$

As a special case we chose the source term $f(x, y) \equiv 1$, so $\partial f_i^j/\partial u_i^j = 0$ in (38), and the boundary conditions $u(0, y) = u(1, y) = 0$ and $u(x, 0) = u(x, 1) = 0$. Figure 5 shows the numerical results obtained with the discretization $\Delta x = 0.02$ and $\Delta y = 0.02$.

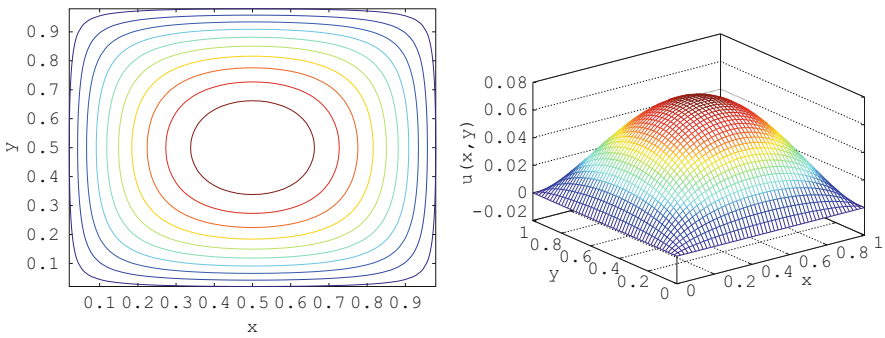


Fig. 5 Contour plot (left) and three-dimensional surface plot (right) of the solution of Poisson equation, using the variational integrator with nonstandard finite difference schemes. The source term was chosen $f(x, y) \equiv 1$, while the boundary conditions $u(0, y) = u(1, y) = 0$, $u(x, 0) = u(x, 1) = 0$ for discretization $\Delta x = 0.02$, $\Delta y = 0.02$

6 Numerical Examples Using Square Discretization

To illustrate the behavior of the proposed method, we will consider the Klein–Gordon equation, which plays a significant role in many scientific applications such as solid state physics, nonlinear optics and quantum field theory, see for example [25].

6.1 Klein–Gordon

For the general case, the initial-value problem of the one-dimensional nonlinear Klein–Gordon equation is given by

$$u_{tt} + \alpha u_{xx} + g(u) = f(x,t), \tag{39}$$

where $u = u(x,t)$ represents the wave displacement at position x and time t, α is a known constant and $g(u)$ is the nonlinear force which, in the physical applications has also other forms [25].

Here we will consider the special case that $\alpha = -1$, $g(u) = u^3 - u$ and $f(x,t) = 0$ resulting

$$u_{tt} = u_{xx} - u^3 + u. \tag{40}$$

The above equation can be described using the Lagrangian

$$L(u, u_t, u_x) = \frac{1}{2}u_t^2 - \frac{1}{2}u_x^2 - \frac{1}{4}u^4 - \frac{1}{2}u^2.$$

Following Section 4.2 we can obtain the discrete Lagrangian that now uses square discretization as

$$\begin{aligned} L_d(u_i^j, u_{i+1}^j, u_{i+1}^{j+1}, u_i^{j+1}) = {} & \frac{\Delta t \Delta x}{2}\left(\frac{u_i^{j+1} - u_i^j}{2\phi(\Delta t)} + \frac{u_{i+1}^{j+1} - u_{i+1}^j}{2\phi(\Delta t)}\right)^2 - \\ & \frac{\Delta t \Delta x}{2}\left(\frac{u_{i+1}^{j+1} - u_i^{j+1}}{2\psi(\Delta x)} + \frac{u_{i+1}^j - u_i^j}{2\psi(\Delta x)}\right)^2 - \\ & - \frac{\Delta t \Delta x}{4} u_i^j u_{i+1}^j u_{i+1}^{j+1} u_i^{j+1} + \\ & \frac{\Delta t \Delta x}{2}\left(\frac{u_i^j u_{i+1}^j + u_i^j u_{i+1}^{j+1} + u_i^j u_i^{j+1} + u_{i+1}^j u_{i+1}^{j+1} + u_{i+1}^{j+1} u_i^{j+1} + u_{i+1}^{j+1} u_i^{j+1}}{6}\right), \end{aligned}$$

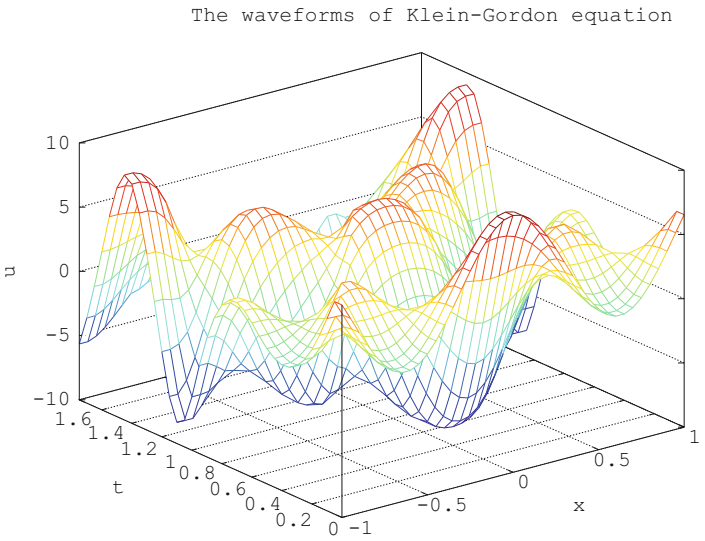


Fig. 6 Numerical solution of the Klein–Gordon equation (40) using square discretization of Section 4.2

which we will consider for the discrete Euler–Lagrange equations (25) in order to derive the resulting integrator from the proposed nonstandard finite difference schemes.

Figure 6 shows the numerical results obtained with the discretization $\Delta t = 0.05$ and $\Delta x = 0.05$. To that we have used initial conditions $u(x, 0) = A(1 + \cos(\frac{2\pi x}{L}))$ where $A = 5$ and $u_t(x, 0) = 0$, while the boundary conditions were $u(-1, t) = u(1, t)$ and $u_x(-1, t) = u_x(1, t)$.

7 Analysis of the Proposed Schemes

A dispersion analysis and mesh convergence experiments are performed in this section in order to show the numerical properties of the proposed method.

7.1 Dispersion Analysis

We will now turn our study to the dispersion-dissipation properties of the derived numerical schemes and compare them with the ones of [2]. To that end, similar to [26], we consider the discrete analog of the Fourier mode

$$u_i^j = \hat{u} e^{\mathbf{i}(ik\Delta x + j\omega\Delta t)}, \tag{41}$$

where $\mathbf{i}^2 = 1$. Using $\bar{k} = k\Delta x$ and $\bar{\omega} = k\Delta t$, the latter equation results in

$$u_i^j = \hat{u} e^{\mathbf{i}(i\bar{k} + j\bar{\omega})}. \tag{42}$$

Following the above, the multi-symplectic scheme of [2], also known as leap frog algorithm, for the case of the linear wave (26) gives

$$\frac{u_i^{j+1} - 2u_i^j + u_i^{j-1}}{(\Delta t)^2} - \frac{u_{i+1}^j - 2u_i^j + u_{i-1}^j}{(\Delta x)^2} = 0. \tag{43}$$

When substituting (42) in the latter equation, we get the discrete dispersion relationship

$$\frac{e^{\mathbf{i}\bar{k}}}{(\Delta t)^2}\left[e^{2\mathbf{i}\bar{\omega}} - 2e^{\mathbf{i}\bar{\omega}} + 1\right] - \frac{e^{\mathbf{i}\bar{\omega}}}{(\Delta x)^2}\left[e^{2\mathbf{i}\bar{k}} - 2e^{\mathbf{i}\bar{k}} + 1\right] = 0. \tag{44}$$

As a second example we consider the second order implicit Runge–Kutta scheme described in [27, 28] and [29]. This scheme, also known as implicit Crank-Nicolson, is a symplectic time discretization of order two, which for the case of (26) gives

$$\begin{aligned} &4\left(u_i^{j+2} - 2u_i^{j+1} + u_i^j\right) - \lambda^2\left(u_{i-1}^{j+2} - 2u_i^{j+2} + u_{i+1}^{j+2}\right) \\ &-2\lambda^2\left(u_{i-1}^{j+1} - 2u_i^{j+1} + u_{i+1}^{j+1}\right) - \lambda^2\left(u_{i-1}^j - 2u_i^j + u_{i+1}^j\right) = 0, \end{aligned} \tag{45}$$

where

$$\lambda^2 = \left(\frac{\Delta t}{\Delta x}\right)^2. \tag{46}$$

Substituting to the above integrator the form (42) we obtain the discrete dispersion relationship

$$\frac{4e^{\mathbf{i}\bar{k}}}{(\Delta t)^2}\left[e^{2\mathbf{i}\bar{\omega}} - 2e^{\mathbf{i}\bar{\omega}} + 1\right] - \frac{\left[e^{2\mathbf{i}\bar{k}} - 2e^{\mathbf{i}\bar{k}} + 1\right]}{(\Delta x)^2}\left[e^{2\mathbf{i}\bar{\omega}} - 2e^{\mathbf{i}\bar{\omega}} + 1\right] = 0. \tag{47}$$

For the case of the linear wave Equation (26) the integrator with the proposed technique, i.e., (29) for u_i^j of (42) gives

$$(\cos\bar{\omega} - 1)(1 - \cos\Delta x) - \left(\cos\bar{k} - 1\right)(1 - \cos\Delta t) = 0. \tag{48}$$

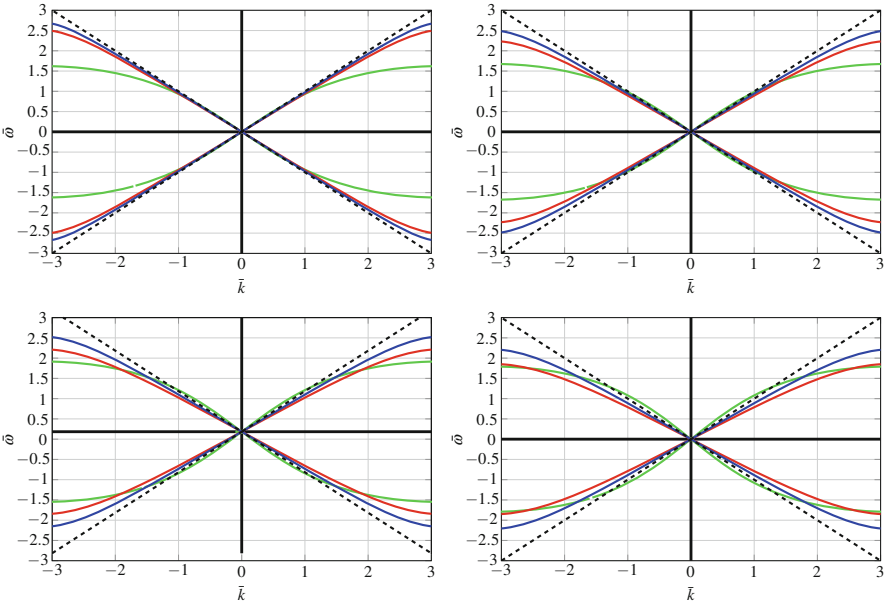


Fig. 7 Dispersion curves for the linear wave equation with the proposed method (red), the leap frog scheme of [2] (blue), the implicit Runge–Kutta (green) and the analytic one (dashed black) for $\lambda = \{0.95, 0.9, 0.85, 0.8\}$

For now we will restrict ourselves only to $\lambda \leq 1$, but due to symmetry, all other cases can be easily obtained. Figure 7 shows the discrete dispersion relationships for $\lambda = \{0.95, 0.9, 0.85, 0.8\}$. Specifically, to each sub plot we can see the dispersion curve of the leap frog scheme, i.e., Equation (44), with blue line, the red line corresponds to the proposed method, described by (48), while the green line is the one for the implicit Runge–Kutta scheme, Equation (47). For all the choices of λ tested the behaviour of the method using nonstandard finite difference schemes is close to the excellent behaviour of the leap frog scheme, and much better than the implicit Runge–Kutta scheme.

7.2 *Convergence Experiments*

In order to show the grid independence of the solution, following the finite element convention, the l^{∞}-norm error is calculated between the solutions on two successive grids according to

$$e_h = \max_i \{|u_i^f - u_i^c|, \ldots, |u_{n_{\text{el}}}^f - u_{n_{\text{el}}}^c|\}, \tag{49}$$

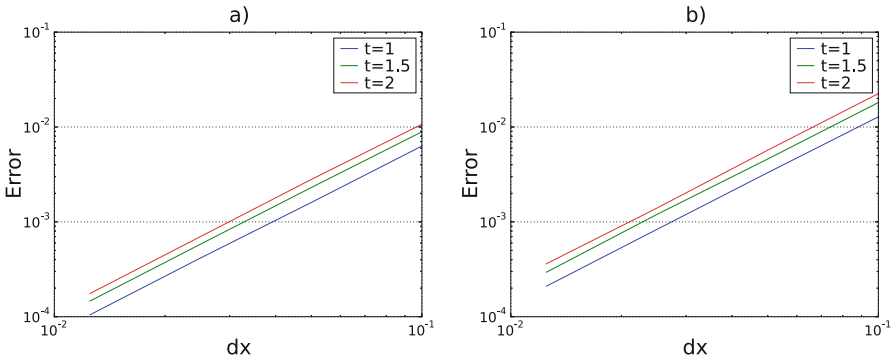


Fig. 8 Error of numerical solution as a function of grid size Δx for different time steps: (**a**) triangle discretization (**b**) square discretization

where u_i^f is the solution on the fine grid, u_i^c the solution on a coarse grid interpolated on the fine one. Here, n_{el} are the total number of elements, where the elements of the mesh are either triangles or squares. A sample convergence of the calculations for the Klein–Gordon case is shown in Figure 8 in a logarithmic plot for triangle and square discretizations and for different time steps. It can be easily seen that by decreasing the space discretization the error is also decreased linearly in the log scale.

8 Summary and Conclusions

The derivation of advantageous multisymplectic numerical methods, relying on nonstandard finite difference schemes, is investigated. The numerical solution of the linear wave equation, the 2-D Laplace equation, and the 2-D Poisson equation, which are addressed in this study, show a good energy behavior and the preservation of the discrete multisymplectic structure of the proposed numerical schemes. Moreover, we showed with the help of dispersion analysis and mesh convergence experiments the numerical properties of the proposed method.

Future applications may include the field equation of incompressible fluid dynamics, like that of Cotter et al [30] and Pavlov et al [31], which could be of interest in investigating the properties of 3-D media. For partial differential equations arising in the field of fluid dynamics, dissipative terms should be taken into consideration. These dissipative perturbations necessitate application of techniques similar to [32, 33] but in the case of PDEs. Furthermore, a possible application in complex geometries, as they appear in real world problems, would necessitate the extension of this methodology to non-uniform grids.

The variational method presented in this work can be applied in a variety of physical problems, ranging from magnetic field simulations in NMR [34] to

inverse problems that arise in geophysics [35] and others. Future work may include comparison with other numerical of PDEs, such as the finite element method or the finite volume method.

Acknowledgments Dr. Odysseas Kosmas wishes to acknowledge the support of EPSRC via grand EP/N026136/1 "Geometric Mechanics of Solids".

Appendix

By denoting $\text{sinc}(\xi) = \sin(\xi)/\xi$, special cases of the exponential integrators described using (10) can be obtained, i.e.

- Gautschi type exponential integrators [36] for

$$\psi(\Omega h) = \text{sinc}^2\left(\frac{\Omega h}{2}\right), \qquad \phi(\Omega h) = 1$$

- Deuflhard type exponential integrators [37] for

$$\psi(\Omega h) = \text{sinc}(\Omega h), \qquad \phi(\Omega h) = 1$$

- García-Archilla et all. type exponential integrators [38] for

$$\psi(\Omega h) = \text{sinc}^2(\Omega h), \qquad \phi(\Omega h) = \text{sinc}(\Omega h)$$

Finally, in [1] a way to write the Störmer-Verlet algorithm as an exponential integrators is presenting.

References

1. E. Hairer C. Lubich, G. Wanner, Geometric numerical integration illustrated by the Störmer-Verlet method. Acta Numer. **12**, 399 (2003)
2. J.E. Marsden, G.W. Patrick, S. Shkoller, Multisymplectic geometry, variational integrators, and nonlinear PDEs. Commun. Math. Phys. **199**, 351 (1998)
3. J.E. Marsden, M. West, Discrete mechanics and variational integrators. Acta Numer. **10**, 357 (2001)
4. A.P. Veselov, Integrable discrete-time systems and difference operators. Funkts. Anal. Prilozhen. **22**, 1 (1988)
5. A.P. Veselov, Integrable Lagrangian correspondences and the factorization of matrix polynomials. Funkts. Anal. Prilozhen. **25**, 38 (1991)
6. T.J. Bridges, Multi-symplectic structures and wave propagation. Math. Proc. Camb. Philos. Soc. **121**, 1 (1997)
7. T.J. Bridges, S. Reich Multi-symplectic integrators: numerical schemes for Hamiltonian PDEs that conserve symplecticity. Phys. Lett. A. **284**, 4–5 (2001)

8. T.J. Bridges, S. Reich, Numerical methods for Hamiltonian PDEs. J. Phys. **39**, 19 (2006)
9. R.E. Mickens, *Applications of Nonstandard Finite Difference Schemes* (World Scientific Publishing, Singapore, 2000)
10. R.E. Mickens, Nonstandard finite difference schemes for differential equations. J. Differ. Equ. Appl. **8**, 823 (2002)
11. R.E. Mickens, Dynamic consistency: a fundamental principle for constructing nonstandard finite difference schemes for differential equations. J. Differ. Equ. Appl. **11**, 645 (2005)
12. O.T. Kosmas, D. Papadopoulos, Multisymplectic structure of numerical methods derived using nonstandard finite difference schemes. J. Phys. Conf. Ser. **490** (2014)
13. O.T. Kosmas, Charged particle in an electromagnetic field using variational integrators. ICNAAM Numer. Anal. Appl. Math. **1389**, 1927 (2011)
14. O.T. Kosmas, S. Leyendecker, Analysis of higher order phase fitted variational integrators. Adv. Comput. Math. **42**, 605 (2016)
15. O.T. Kosmas D.S. Vlachos, Local path fitting: a new approach to variational integrators. J. Comput. Appl. Math. **236**, 2632 (2012)
16. O.T. Kosmas, S. Leyendecker, Variational integrators for orbital problems using frequency estimation. Adv. Comput. Math. **45**, 1–21 (2019)
17. O.T. Kosmas, D.S. Vlachos, Phase-fitted discrete Lagrangian integrators. Comput. Phys. Commun. **181**, 562–568 (2010)
18. O.T. Kosmas, S. Leyendecker, Phase lag analysis of variational integrators using interpolation techniques. PAMM Proc. Appl. Math. Mech. **12**, 677–678 (2012)
19. O.T. Kosmas, S. Leyendecker, Stability analysis of high order phase fitted variational integrators, in *Proceedings of WCCM XI—ECCM V—ECFD VI*, vol. 1389 (2014), pp. 865–866
20. O.T. Kosmas, S. Leyendecker, Family of high order exponential variational integrators for split potential systems. J. Phys. Conf. Ser. **574**, 012002 (2015)
21. O.T. Kosmas, D.S. Vlachos, A space-time geodesic approach for phase fitted variational integrators. J. Phys. Conf. Ser. **738**, 012133 (2016)
22. L. Brusca, L. Nigro, A one-step method for direct integration of structural dynamic equations. Int. J. Numer. Methods Eng. **15**, 685–699 (1980)
23. L.C. Evans, *Partial Differential Equations* (American Mathematical Society, Providence, 1998)
24. V.I. Arnold, *Lectures on Partial Differential Equations* (Springer, Berlin, 2000)
25. H. Han, Z. Zhang, Split local absorbing conditions for one-dimensional nonlinear Klein–Gordon equation on unbounded domain. J. Comput. Phys. **227**, 8992 (2008)
26. J.W. Thomas, *Numerical Partial Differential Equations*, Finite Difference Methods, vol. 1 (Springer, New York, 1995)
27. J.M. Sanz-Serna, M.P. Calvo, *Numerical Hamiltonian Problems* (Chapman & Hall, London, 1994)
28. J.M. Sanz-Serna, *Solving Numerically Hamiltonian Systems*. Proceedings of the International Congress of Mathematicians (Birkhäuser, Basel, 1995)
29. S. Reich, Multi-symplectic Runge–Kutta collocation methods for Hamiltonian wave equations. J. Comput. Phys. **157**, 473 (2000)
30. C.J. Cotter, D.D. Holm, P.E. Hydon, Multisymplectic formulation of fluid dynamics using the inverse map. Proc. R. Soc. A **463**, 2671 (2007)
31. D. Pavlov, P. Mullen, Y. Tong, E. Kanso, J.E. Marsden, M. Desbrun, Structure-preserving discretization of incompressible fluids. Phys. D: Nonlinear Phenom. **240**, 443 (2011)
32. E. Hairer, C. Lubich, Invariant tori of dissipatively perturbed Hamiltonian systems under symplectic discretization. Appl. Numer. Math **29**, 57–71 (1999)
33. D. Stoffer, On the qualitative behaviour of symplectic integrators. III: perturbed integrable systems. J. Math. Anal. Appl. **217**, 521–545 (1998)
34. D. Papadopoulos, M.A. Voda, S. Stapf, F. Casanova, M. Behr, B. Blümich, Magnetic field simulations in support of interdiffusion quantification with NMR. Chem. Eng. Sci. **63**, 4694 (2008)
35. D. Papadopoulos, M. Herty, V. Rath, M. Behr, Identification of uncertainties in the shape of geophysical objects with level sets and the adjoint method. Comput. Geosci. **15**, 737 (2011)

36. W. Gautschi, Numerical integration of ordinary differential equations based on trigonometric polynomials. Numer. Math. **3**, 1 (1961)
37. P. Deuflhard, A study of extrapolation methods based on multistep schemes without parasitic solutions. Z. Angew. Math. Phys. **30**, 2 (1979)
38. B. García-Archilla, M.J. Sanz-Serna, R.D. Skeel, Long-time-step methods for oscillatory differential equations. SIAM J. Sci. Comput. **20**, 3 (1999)

Non-radial Solutions of a Supercritical Equation in Expanding Domains: The Limit Case

Nikos Labropoulos

Abstract In this article, we introduce a new method to prove the existence of an infinite sequence of distinct non-radial but symmetric nodal (i.e. sign changing) solutions for supercritical nonlinear elliptic problems defined in the whole Euclidean space. By 'symmetric' we mean that both the domain and the solution remain invariant under the action of a compact subgroup G of the isometry group $O(n)$, without finite subgroup. The key ingredient of the method is a process through which an open symmetric domain of the n-dimensional space can be extended in an appropriate manner to 'fill' eventually the entire space 'almost everywhere', remaining symmetric, and giving a sequence of domains where in each of them subsequently we solve an appropriate auxiliary problem. Passing to the limit we obtain the solution of the problem as a limit of the sequence formed by the solutions of the corresponding to the domains sequence of equations.

The base model problem of interest is stated bellow:

$$\text{(P)} \quad \begin{cases} \Delta_p u = |u|^a u, \quad u \in C^2(\mathbb{R}^n), \quad n \geqslant 3 \\ 1 < p < n-k, \quad 0 \leq a \leq p^*(k) = \frac{(n-k)p}{n-k-p}, \end{cases}$$

where p^*_G is the critical exponent of the embedding

$$H^{1,p}_{0,G}(\Omega) \hookrightarrow L^{p^*_G}(\Omega)$$

(being the critical of the supercritical one) and k is the minimum orbit dimension in G. However, we will focus on the critical of the supercritical case $a = p^*(k)$, since on the one hand it is the most important and on the other hand it covers all the other cases.

N. Labropoulos (✉)
Department of Mathematics, National Technical University of Athens, Athens, Greece
e-mail: nal@math.ntua.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_12

By $H^{1,p}_{0,G}(\Omega)$ is denoted the closure of the subspace $C^{\infty}_{0,G}(\Omega)$ consisting of all G-invariant functions in $C^{\infty}_{0}(\Omega)$.

1 Introduction

In this article, the main objective is to prove the existence of non-radial nodal (sign-changing) solutions of the above problem (P), in the case where the exponent a is the critical of supercritical exponent, since the rest of the cases have been studied. Thus, the problem (P) is set out in detail as follows:

$$\text{(P)} \quad \begin{cases} \Delta_p u = |u|^{p^*(k)-2}u \quad \text{in } \mathbb{R}^n, \quad n \geq 3 \\ 1 < p < n-k, \quad p^*(k) = \frac{(n-k)p}{n-k-p}, \end{cases}$$

Here, G is a group of symmetries that acts on the domains and the functions defined on them together, k is the minimum dimension orbit of all orbits of G,

$$\Delta_p u = -\text{div}\left(|\nabla u|^{p-2}\nabla u\right), \ 1 < p \neq 2$$

is the p-Laplacian operator (note that if $p = 2$, is the Laplace–Beltrami operator) and $p^*(k)$ is the critical exponent of the Sobolev embedding

$$H^{1,p}_{G}(\Omega) \hookrightarrow L^p(\Omega).$$

By $H^{1,p}_{G}(\Omega)$ is denoted the subspace of all G-invariant functions in $H^{1,p}(\Omega)$.

In problem (P) the solutions obtained are such that

$$\int\limits_{\mathbb{R}^n} |\nabla u|^p dx \to \infty.$$

We study both the cases, $p = 2$ and $p \in (1, 2) \cup (2, n-k)$, however, to avoid any confusion we note that throughout the article we denote by

$$\Delta_p u = -\text{div}\left(|\nabla u|^{p-2}\nabla u\right), \ 1 < p < n-k$$

the p-Laplacian as well as the Laplace–Beltrami operator but when we refer to other articles the Laplace–Beltrami operator is denoted as in the referred articles, i.e. without the minus conversion.

For the problem (P), we prove the existence and find both the type and the number of the solutions to the problem (P). For this aim we use *the method of expanding domains* which was successfully introduced for the first time in [42]. In

that article this method was used firstly to prove the existence of a solution and secondly to determine the type and the number of the solutions to critical nonlinear elliptic problem:

$$(\mathrm{P}_0) \qquad -\Delta u = |u|^{\frac{4}{n-2}}u, \quad u \in C^2(\mathbb{R}^n), \quad n \geq 3.$$

Concerning the method itself it seems to have particular value because it can be used and in other types of partial differential equations.

Both cases, i.e. $p = 2$ and $p \neq 2$, are extremely interesting and that is why for several decades now many researchers have been paying attention to them.

Problem (P_0) consists a special case of (P) for $p = 2$ and it owns its origin in many astrophysical and physical contexts and more precisely in the Lane-Emden-Fowler problem,

$$(\mathrm{P}_0') \qquad \begin{cases} -\Delta u = u^q \\ u > 0 \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega, \end{cases}$$

where Ω is a domain with smooth boundary in $\mathbb{R}^N$ and $p > 1$. But its greatest interest lies in its relation to the Yamabe problem (see in [5, 57, 64, 68]) and for a complete and detailed study we refer to [6], nevertheless it has an autonomous presence holding an important place among the most famous nonlinear partial differential equations). We refer, also, to the classical papers [20, 30, 47], which are some of the large number of very good papers that are devoted to the study of this problem.

Gidas, Ni, and Nirenberg, in their celebrated paper [30], proved symmetry and some related properties of positive solutions of a larger class of second order elliptic equations. Concerning the equation

$$-\Delta u = |u|^{\frac{4}{n-2}}u, \quad u \in C^2(\mathbb{R}^n), \quad n \geq 3,$$

they proved that any positive solution of this, which has finite energy, namely

$$\int_{\mathbb{R}^n} |\nabla u|^2 dx < +\infty,$$

is necessarily of the form

$$u(x) = \left(\frac{\lambda\sqrt{n(n-2)}}{\lambda^2 + |x - x_0|^2}\right)^{\frac{n-2}{2}}, \quad \lambda > 0, x_0 \in \mathbb{R}^n.$$

These solutions yield the well-known one-instanton solutions in a regular gauge of the Yang–Mills equation. In addition, since this equation is invariant under the conformal transformations of $\mathbb{R}^n$, if u(x) is a solution, then

$$\lambda^{\frac{n-2}{2}} u(\frac{x - x_0}{\lambda}), \ \forall \ \lambda > 0 \text{ and } x_0 \in \mathbb{R}^n$$

is also a solution. Moreover, all solutions obtained in this way have the same energy and we will say that these solutions are equivalent. In particular, all these solutions are equivalent.

Ding in his also celebrated article [20] using Ambrosetti and Rabinowitz analysis (see in [2]) proved that this problem has infinite distinct solutions $u_k \in C^2(\mathbb{R}^n)$, $k = 1, 2, \cdots$, which changes sign and such that

$$\lim_{k\to\infty} \int_{\mathbb{R}^n} |\nabla u_k|^2 dx \to \infty.$$

Ding showed that it is possible to solve the equation in the whole Euclidean space, reduced the problem to an equivalent problem on $\mathbb{S}^n$, the Euclidean n-sphere throughout a conformal deformation. However, this method cannot be used in the case of the p-Laplacian operator, because this operator is not a conformal invariant operator.

Mazzeo and Smale in their also celebrated article [47] proved that if Ω is an open set in $\mathbb{R}^n$ and u is a positive C^2 function on Ω such that the metric $g = u^{\frac{4}{n-2}} g_0$ on Ω has scalar curvature $R(g) = n(n - l)$, then u must satisfy the equation

$$\Delta u + \frac{n(n-2)}{4} u^{\frac{n+2}{n-2}} = 0, \ u > 0$$

on Ω, where g_0 is the Euclidean metric on $\mathbb{R}^n$.

Caffarelli, Gidas, and Spruck in their classical paper [11] studied non-negative smooth solutions of the conformally invariant equation

$$-\Delta u = u^{\frac{n+2}{n-2}}, \quad u \geq 0, \quad n \geq 3,$$

in a punctured ball $B_1(0)\backslash\{0\} \subseteq \mathbb{R}^n$, with an isolated singularity at the origin. In this paper, the authors introduced a heuristic idea of asymptotic symmetry technique which can roughly be described as follows: After an inversion, the function u becomes defined in the complement of B_1, is strictly positive of ∂B_1, and in some sense 'goes to zero' at infinity. If the function u can be extended to B_1 as a super solution of our problem, then the reflection process at infinity can start and move all the way to ∂B_1. This would imply asymptotic radial symmetry at infinity. With this comprehensive report on this issue we would like, on the one hand, to emphasize the important contribution of this great article of Caffarelli, Gidas, and Spruck on the study on the direction of finding the radial solutions of our problem and on the other hand, we wish to make clear that in our procedural paper we do not care about the radial solutions but we do care about the existence of non-radial solutions.

Schoen in [57] built solutions of (P) with prescribed isolated singularities. Schoen, also, in [58], have used the geometrical meaning of problem (P) in order to derive, through ideas of conformal geometry, the existence of weak solutions having a singular set whose Hausdorff dimension is less than or equal to $\frac{n-2}{2}$. Let us notice that in this paper the authors explain how to build solutions of (P) with a singular set whose Hausdorff dimension is not necessarily an integer.

Bartsch and Schneider in [8] proved that for $N > 2m$ the equation

$$(-\Delta)^m = |u|^{\frac{4m}{N-2m}} u$$

on $\mathbb{R}^N$ has a sequence of nodal, finite energy solutions which is unbounded in $\mathscr{D}^{m,2}(\mathbb{R}^N)$, the completion of $\mathscr{D}(\mathbb{R}^N)$ with respect to the scalar product:

$$(u, \upsilon) = \begin{cases} \int_{\mathbb{R}^N} \Delta^{\frac{m}{2}} u \cdot \Delta^{\frac{m}{2}} \upsilon, & m \text{ even} \\ \int_{\mathbb{R}^N} \nabla\Delta^{\frac{m-1}{2}} u \cdot \nabla\Delta^{\frac{m-1}{2}} \upsilon, & m \text{ odd.} \end{cases}$$

This result generalizes the result of Ding for $m = 1$, and provides interesting information concerning the number and the kind of the solutions of the equation.

Wang in [66] studied the following nonlinear Neumann elliptic problem:

$$(\mathrm{P_N}) \quad \begin{cases} -\Delta u = u^{\frac{N+2}{N-2}}, & u > 0 \text{ in } \mathbb{R}^N \backslash \Omega, \\ u(x) \to 0 & \text{as } |x| \to +\infty, \\ \frac{\partial u}{\partial n} = 0 & \text{on } \partial\Omega, \end{cases}$$

where n denotes interior unit normal vector and Ω is a smooth bounded domain in $\mathbb{R}^N$, $N \geq 4$. In this paper, it is proved that if $N \geq 4$, (Wang believes that the results will also hold in the case of $N = 3$), and Ω is a smooth and bounded domain then the problem $(\mathrm{P_N})$ has infinity many non-radial positive solutions, whose energy can be made arbitrarily large when Ω is convex, as seen from inside (with some symmetries). We refer to the Wang's problem $(\mathrm{P_N})$ due to its close relationship with our problem and as we will see later, if we choose suitable Ω we can have a result on this problem in almost all the space. In particular, in both problems we have to solve the same non-linear differential equation with critical exponent with boundary conditions Dirichlet and Neumann, respectively. In addition, in both cases the domain Ω presents some symmetries. However, a subsequent process in each case is completely different from that of another. In our case, our goal is to solve the problem *almost in the whole space*, starting from an open symmetric domain Ω of n-dimensional space and we extend the Ω so that it remains symmetrical to fill almost all the space. In the other case is considered the corresponding Neumann problem in $\mathbb{R}^N \backslash \Omega$ where Ω is convex as seen from inside with some symmetries. If we choose appropriate a such Ω with a small volume as much as we can say that the solutions of Wang satisfy the conditions of the problem *almost in the whole space*. Finally, in both problems we take an infinity number of non-radial solutions,

whose energy can be made arbitrary large, however in the first problem we find nodal solutions while in the second are founded positive solutions.

Concerning to the progress of the study of the problem (P) for $p = 2$ a number of important articles are available (cf. [1, 3, 4, 9–11, 20, 21, 23, 27, 29, 30, 32, 37, 42, 43, 45, 47, 48, 52, 57, 58, 63, 66]).

The p-Laplace operator (or p-harmonic operator) occupies a similar position to the standard Laplace operator when it comes to nonlinear phenomena. In fact, many of the things that apply to the usual Laplace operator and consequently to the equations that relate to it also apply to the p-Laplace as well as his equations, except that the Principle of Superposition which is of course lost. A very detailed and complete study is provided by Lindqvist [44]. Also, a Morse theoretic study of a very general class of homogeneous operators that includes the p-Laplacian as a special case is presented by Perera, Agarwal, and O'Regan in [53]. However, the p-Laplacian operator also appears in many areas of physics, such as non-Newtonian fluid flows, turbulent filtration in porous media, plasticity theory, rheology, glaciology, radiation of heat (cf. [24, 35, 49]).

The p-Laplace operator is a particularly interesting and remarkable case and this fact is confirmed not only by the large number of articles dedicated to it but also by the multifaceted study of the problems related to it (cf. [13, 18, 22, 26, 28, 31, 36, 40, 46, 54, 55, 60, 67, 69]).

In the problem (P), considered for any $1 < p < n - k$, a main difficulty comes from the double lack of compactness. By lack of compactness, we mean that the functional that we consider do not satisfy the Palais-Smale condition (cf. [50, 51, 59, 61, 62, 70]), (i.e. there exists a sequence along which the functional remains bounded, its gradient goes to zero, and does not converge). However, for $p \neq 2$, a second difficulty arises from the fact that the p-Laplace operator is not conformal invariant operator so the methods used in the case of the Laplace operator cannot be applied.

Concerning the lack of compactness, the first difficulty comes from the fact that the exponent

$$p^*(k) = \frac{(n-k)p}{n-k-p}$$

is supercritical (in fact the critical of the supercritical), and the second one is some extra difficulty because of the lack of compactness in unbounded domains. But, it is well known (see in [15, 16, 25, 32]) that the symmetry property of the domain allows us to improve the Sobolev embedding in higher L^p spaces and we overcome the obstruction of the exponent. Regarding the problem of lack of compactness in unbounded domains we avoid solving problems in such domains by remaining in bounded domains and then we pass to unbounded with limit procedures. In addition, this ensures us the ability to overcome the problems due to the non-conformality of the p-Laplace operation.

To overcome all the above obstacles we consider the following corresponding problem

$$(\mathrm{P}_\varepsilon) \qquad \begin{cases} \Delta_p u_\varepsilon + \varepsilon\, a(x)|u_\varepsilon|^{p-2}u_\varepsilon = f(x)|u_\varepsilon|^{p^*(k)-2}u_\varepsilon \\ \qquad u_\varepsilon \not\equiv 0 \ \text{ in } \ \Omega_\varepsilon, \ u_\varepsilon = 0 \text{ on } \partial\Omega_\varepsilon \\ \qquad 1 < p < n-k, \ p^*(k) = \frac{(n-k)p}{n-k-p}, \end{cases}$$

where Ω_ε, $\varepsilon > 0$, is an expanding domain in $\mathbb{R}^n$, $n \geq 3$, invariant under the action of a subgroup G of the isometry group $O(n)$ and $a,\ f \in C^\infty(\overline{\Omega}_\varepsilon)$ are two smooth G-invariant functions on $\overline{\Omega}_\varepsilon$.

The problem (P_ε) has been studied by many authors. We refer to [3, 4, 10, 20, 23, 27, 29, 32] and the references therein for a further discussion of both the problem itself and several variants of it. Some special cases have been also studied. For example, no solution can exist if Ω is starshaped, as a consequence of the Pohozaev identity (see in [56]). Furthermore, if Ω is an annulus, there are infinite solutions (see in [43]). Also, a general result of Bahri and Coron guarantees the existence of positive solutions in domains Ω having nontrivial topology (i.e. certain homology groups of Ω are non trivial) (see in [7]). The existence and multiplicity of positive or nodal solutions of critical equations on bounded domains or in some contractible domains have been determined by other authors (see for example in [21, 27, 32, 52, 63]). Some more nonexistence results in this case are available, (see in [1, 4, 12, 37]).

Our proof is via approximation by an infinite sequence of problems defined on a sequence of expanding symmetric bounded domains. Firstly, we solve the problem (P_ε). (see in [14, 42] for the case of the Laplacian and for the case of the p-Laplacian see in [15, 16], for $n = 3$, and for $n \geq 3$, respectively). Then we consider a sequence of problems $(\mathrm{P}_{\varepsilon_j})$, $j = 1, 2, \cdots$, defined in a sequence of expanding domains Ω_{ε_j}, $j = 1, 2, \cdots$, and henceforth, sending $\varepsilon \to 0$, we obtain the solution of the limit problem (P) as the limit of the sequence of the solutions of the problems $(\mathrm{P}_{\varepsilon_j})$. This method is a generalization of the method we used in [42] and thus a uniform treatment of both cases $p = 2$ and $p \in (1, 2) \cup (2, n-k)$ is achieved. In addition, the used method is a different from previous ones and can be used to solve poly-harmonic equations with supercritical exponent and even in the critical of supercritical case, as in our case, providing an alternative way of utilizing the best constants of the appearing Sobolev inequalities. Furthermore, this method enables us to determine the kind and the number of solutions of the problem in both cases, i.e. for $p = 2$ and for $p \in (1, 2) \cup (2, n-k)$.

This article is organized as follows: Section 2 is devoted to notations and in some necessary background material. In Section 3, we introduce our main tool, meaning the process through which an open symmetric domain of n-dimensional space can be extended in an appropriate manner to *'fill'* eventually the entire space *'almost everywhere'*, remaining symmetric, and subsequently we solve the auxiliary problem (P_ε). Section 4 is devoted to some basic definitions and to the proof of the main theorem.

2 Notations and Some Background Material

As referred in the beginning of this article, our main objective is to prove the existence of an infinite sequence of distinct non-radial nodal G-invariant solutions defined in 'almost the whole' Euclidean space for the supercritical nonlinear elliptic problem (P). However, before dealing with problem (P) let us consider the following basic problem which will play an important role in solving the problem (P).

$$(\mathrm{P}_0) \quad \begin{cases} \Delta_p u + a(x)|u|^{p-2}u = f(x)|u|^{q-2}u \\ \quad u \not\equiv 0 \ \text{ in } \ \Omega, \ \ u = 0 \ \text{ on } \partial\Omega, \end{cases}$$

where Ω is a bounded, smooth, domain of $\mathbb{R}^n$, $n \geq 3$.

If we consider

$$p^* = \frac{np}{n-p}$$

it is well known by Sobolev's embedding theorem (cf. [6]) the embedding

$$H_0^{1,p}(\Omega) \hookrightarrow L^p(\Omega)$$

is compact for any $p \in [1, \ p^*)$ but the embedding

$$H_0^{1,p}(\Omega) \hookrightarrow L^{p^*}(\Omega)$$

is only continuous.

We say that the exponent

$$p^* = \frac{np}{n-p}$$

for the Sobolev embedding

$$H_0^{1,p}(\Omega) \hookrightarrow L^p(\Omega)$$

is the *critical exponent* for this embedding and that the problem (P) is *supercritical*, *critical* or *supercritical* if $q-1 < p^*$, $q-1 = p^*$ or $q-1 > p^*$ respectively. If $p > n$ the problem (P) is always sub-critical.

In order to make this article self-contained we will open at this point a parenthesis where we will introduce some useful background material from the geometry. (More details see in [9] or [38]).

Consider a group G acting on a set X. The *orbit* of a point x in X is the set of elements of X to which x can be moved by the elements of G. (Just as gravity moves a planet around in its orbit, the group action moves an element around in its orbit.)

The $G-orbit$ of x is denoted by

$$O_G(x) = \{\tau(x), \tau \in G\},$$

and for any $Y \subseteq X$, we write

$$G(Y) = \{\tau(y) : y \in Y \text{ and } \tau \in G\}.$$

If for some subset $Y \subseteq X$ is valid

$$G(Y) = Y,$$

then, we say that Y is *invariant* under the action of G and in this case we denote it by Y_G.

For every $x \in X$, we define the *stabilizer subgroup* of G with respect to x (also called the *isotropy group*) as the set of all elements in G that fix x:

$$S_G(x) = \{\tau \in G : \tau(x) = x\}.$$

Moreover, if the set X is equipped with a metric, then the *isometry group* of this metric space is the set of all isometries (i.e. distance-preserving maps) from the metric space onto itself, with the function composition as group operation. Its identity element is the identity function (i.e. the isometry group of a two-dimensional sphere is the orthogonal group $O(3)$).

Given (M, g) a Riemannian manifold (complete or not, but connected), we define by $I(M, g)$ its group of isometries. It is well known (see for instance [38]) that $I(M, g)$ is a Lie group with respect to the compact open topology, and that $I(M, g)$ acts differentiably on M. Since (this is actually due to E. Cartan) any closed subgroup of a compact Lie group is a Lie group, we get that any compact subgroup of $I(M, g)$ is a sub-Lie group of $I(M, g)$. It is now classical (see [9] and [19]), that for any $x \in M$, $O_G(x)$ is a smooth compact sub-manifold of M.

We denote by $|O_G(x)|$ the volume of $O_G(x)$ for the Riemannian metric induced on $O_G(x)$. In the special case where $O_G(x)$ has finite cardinal, then,

$$|O_G(x)| = \text{card } O_G(x).$$

Let G be a closed subgroup of $I(M, g)$. Assume that for any $x \in M$,

$$\text{card } O_G(x) = +\infty,$$

and set

$$k = \min_{x \in M} \dim O_G(x).$$

Then $k \geq 1$ (see [32]), and is called *minimum orbit dimension*.

We consider a bounded, smooth domain Ω of $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$, $k \geq 2, n-k \geq 1$ such that

$$\overline{\Omega} \subset \left(\mathbb{R}^k \backslash \{0\}\right) \times \mathbb{R}^{n-k}.$$

Suppose that $\overline{\Omega}$ is invariant under the action of $G_{k,n-k}$, that is

$$\tau(\overline{\Omega}) = \overline{\Omega}, \quad \text{for all } \tau \in G_{k,n-k},$$

where $G_{k,n-k} = O(k) \times Id_{n-k}$ (then denoted by G), is the subgroup of the isometry group $O(n)$ of the type

$$(x_1, x_2) \longrightarrow (\sigma(x_1), x_2), \quad \sigma \in O(k), \quad x_1 \in \mathbb{R}^k, \quad x_2 \in \mathbb{R}^{n-k}.$$

For example, a such Ω in $\mathbb{R}^3$ is the solid torus

$$\overline{T} = \left\{ (x, y, z) \in \mathbb{R}^3 : \left(\sqrt{x^2 + y^2} - R\right)^2 + z^2 \leq r^2, \ R > r > 0 \right\}.$$

Also, as such Ω we can see the part of the n-dimensional ball B_n from which we have removed a part of it in such a way that the rest is invariant under the action of the group G and its cover belongs to $\overline{B_n} \subset \left(\mathbb{R}^k \backslash \{0\}\right) \times \mathbb{R}^{n-k}$. This is because the balls enjoy a large number of symmetries in addition to the radial symmetry.

We define

$$C^\infty_G(\Omega) = \left\{ u \in C^\infty(\Omega) : u \circ \tau = u \, , \, \forall \, \tau \in G \right\},$$

and

$$C^\infty_{0,G}(\Omega) = \left\{ u \in C^\infty_0(\Omega) : u \circ \tau = u \, , \, \forall \, \tau \in G \right\},$$

where $C^\infty(\Omega)$ denotes the space of smooth functions on Ω and where $C^\infty_0(\Omega)$ denotes the space of smooth functions with compact support on Ω.

We define, also, the Sobolev space $H^{1,p}(\Omega)$ as the completion of $C^\infty(\Omega)$ with respect to the norm

$$\|u\|_{H^{1,p}(\Omega)} = \left(\|\nabla u\|^p_{L^p(\Omega)} + \|u\|^p_{L^p(\Omega)} \right)^{1/p}, \ p \geq 1,$$

and the Sobolev space $H^{1,p}_0(\Omega)$ as the closure of $C^\infty_0(\Omega)$ in $H^{1,p}(\Omega)$.

Finally, we denote by $H^{1,p}_G(\Omega)$ and $H^{1,p}_{0,G}(\Omega)$ the subspaces of $H^{1,p}(\Omega)$ and $H^{1,p}(\Omega)$, respectively, of all G-invariant functions defined on Ω.

It is well known that the symmetry property of the domain allows us to improve the Sobolev embedding in higher L^p spaces. More precisely, let us consider a smooth compact n-dimensional, $n \geq 3$, Riemannian manifold (M, g) invariant under the action of an arbitrary compact subgroup G of $\mathrm{Isom}_g(M)$. Let us also assume that

$$\mathrm{Card}\, O_G^x = +\infty$$

for any orbit O_G^x of G and $k \geq 1$. It is well known that the Sobolev embedding

$$H_G^{1,p}(M) \hookrightarrow L^q(M)$$

is compact for any

$$1 \leq q < \frac{(n-k)p}{n-k-p}$$

but if

$$1 \leq q \leq \frac{(n-k)p}{n-k-p}$$

is only continuous (cf. in [14, 16, 20, 25, 33, 34]).

3 Preliminary Results

Let Ω be a domain such that $\overline{\Omega} \subset \left(\mathbb{R}^k \backslash \{0\}\right) \times \mathbb{R}^{n-k}$ and also invariant under the action of the group G defined above. For any small $\varepsilon > 0$ and some $m > 0$ (which will be determined later) we consider the family of expanding domains

$$\Omega_\varepsilon = \varepsilon^{-m}\Omega = \{\varepsilon^{-m}x : x \in \Omega\}$$

Then, it is very simple to be confirmed that Ω_εs inherit the symmetry properties of Ω for any ε.

At this point we need to comment on the term 'almost the whole' space and specify the impact of this term on solutions to the problem. To do this we must describe the process by which we 'fill' the space by properly expanding the domain Ω and then see how the method of solving the problem works. In fact we consider a sequence consisting of Ω_{ε_j}, where the sequence of ε_js (for the time being) is a sequence that tends to 0 in such a way that Ω_{ε_j}s extend continuously and as $\varepsilon \to 0$ they cover "almost everywhere" the entire space. This is because this extension also entails the inside boundary of the Ω_{ε_j}s (i.e. the one that is on the zero side) and of course increases the volume of the orbit with the minimum dimension. This does not

pose a problem for us in the solution because it depends only on the volume of this orbit (apart from the other parameters) (see Theorem 3) so we can extend these Ω_{ε_j}s to the inside as much as we want by extending with zero values the functions defined in them. The outer boundary of Ω_{ε_j}s does not impose any restrictions and this is because any orbits close to it do not play any role since as mentioned above only the orbit with the minimum volume affects the solutions and it is on the opposite side, the side of zero. Finally, the fact that the domain is also expanding does not affect either the Sobolev inequalities associated with the problem or the solutions because we can, for example, normalize the functions u_j i.e. so that their norms are equal to 1.

We consider now the transformation

$$\phi : \Omega \to \Omega_\varepsilon \ : \ X = \varepsilon^{-m} x, \ x \in \Omega, \ X \in \Omega_\varepsilon \tag{1}$$

and for $\ell > 0$ we set

$$u_\varepsilon(X) = \varepsilon^{-\ell} u\left(\varepsilon^m X\right).$$

In particular we obtain

$$|\nabla u| = \varepsilon^{-m} |\nabla u_\varepsilon| \tag{2}$$

and

$$\Delta_p u = -\varepsilon^{-mp} \operatorname{div}\left(|\nabla u_\varepsilon|^{p-2} \nabla u_\varepsilon\right). \tag{3}$$

Note the equality (3) remains valid for $p = 2$, i.e. for $\Delta_2 = \Delta$, the Laplace–Beltrami operator.

In the following, we will suppose that $p \neq 2$, since the case where $p = 2$ was studied in [42].

Applying the transformation (1) in the equation of the problem (P_0), because of (2) and (3), we obtain the following equation

$$\Delta_p u_\varepsilon + \varepsilon^{mp+\ell(2-p)} a(x) |u_\varepsilon|^{p-2} u_\varepsilon = \varepsilon^{mp+\ell(2-q)} f(x) |u_\varepsilon|^{q-2} u_\varepsilon.$$

Since ℓ is an arbitrary positive real, we can choose $\ell = \frac{mp}{q-2}$ and thus we obtain the following equation:

$$\Delta_p u_\varepsilon + \varepsilon^{mp+mp(2-p)/(q-2)} a(x) |u_\varepsilon|^{p-2} u_\varepsilon = f(x) |u_\varepsilon|^{q-2} u_\varepsilon. \tag{4}$$

Finally, replacing the $\varepsilon^{mp+mp(2-p)/(q-2)}$ by ε, we can write the equation (4) in the following form

$$\Delta_p u_\varepsilon + \varepsilon a(x) |u_\varepsilon|^{p-2} u_\varepsilon = f(x) |u_\varepsilon|^{q-2} u_\varepsilon. \tag{5}$$

Let Ω be a smooth bounded domain in $\mathbb{R}^n$, G-invariant and k be the minimum of the dimensions of all orbits of G with infinite cardinal. Let, also, Ω_ε as defined above. A such Ω is the above defined solid torus T and a such Ω_ε, in this case, is an expanding torus T_ε.

Now for any $\varepsilon > 0$ consider the following auxiliary problem:

$$(\mathrm{P}_\varepsilon) \quad \begin{cases} \Delta_p u_\varepsilon + \varepsilon\, a(x)|u_\varepsilon|^{p-2}u_\varepsilon = f(x)|u_\varepsilon|^{p^*(k)-2}u_\varepsilon \\ \qquad u_\varepsilon \not\equiv 0 \ \text{ in } \ \Omega_\varepsilon, \ u_\varepsilon = 0 \text{ on } \partial\Omega_\varepsilon \\ \qquad 1 < p < n-k, \ p^*(k) = \frac{(n-k)p}{n-k-p}, \end{cases}$$

where a, f are two smooth $\mathscr{H}_\sigma^p$-invariant functions (defined bellow).

Before, we solve the problem (P_ε), we must compute the best constant $K_G^p(\Omega_\varepsilon)$ in the following Sobolev inequality, which appears in this problem:

$$\left(\int_{\Omega_\varepsilon} |u|^{p^*(k)}\,dx\right)^{\frac{p}{p^*(k)}} \leqslant \left(K_G^p(\Omega_\varepsilon) + \epsilon\right)\int_{\Omega_\varepsilon} |\nabla u|^p\,dx + B_\epsilon \int_{\Omega_\varepsilon} |u|^p\,dx, \tag{6}$$

where ϵ is a positive constant no matter how small, but it cannot disappear and B_ϵ a positive constant.

In fact we will express the best constant $K_G(\Omega_\varepsilon)$ of inequality (6) as a function of the optimal constant the best constant $K_G^p(\Omega)$ and ε.

Concerning this best constant the following theorem holds:

Theorem 1

$$K_G(\Omega_\varepsilon) = \varepsilon^m K_G(\Omega) = \frac{K\,(n-k,\,p)}{\varepsilon^{-m}\mathscr{V}^{\frac{1}{n-k}}}$$

where $K(n-k,\,p)$ is the best constant in the classical Sobolev inequality of $\mathbb{R}^{n-k}$ and $\mathscr{V}$ denotes the minimum of the volume of the k-dimensional orbits in Ω.

Proof According to the Theorem 2.1 in [16], (see, also, Theorem 3.1 in [15]) we have

$$K_G^p(\Omega_\varepsilon) = \frac{K(n-k,\,p)}{\mathscr{V}_\varepsilon^{\frac{1}{n-k}}}$$

and since

$$\mathscr{V}_\varepsilon = \varepsilon^{-m(n-k)}\mathscr{V}$$

we obtain

$$K_G^p(\Omega_\varepsilon) = \frac{K(n-k,\, p)}{\varepsilon^{-m}\sqrt{}^{\frac{1}{n-k}}}$$

□

Now, for the problem (P_ε), consider the functional

$$J(u_\varepsilon) = \int\limits_{\Omega_\varepsilon} \Big(|\nabla u_\varepsilon|^p + \varepsilon a(x) |u_\varepsilon|^p \Big) dx$$

and suppose that the operator

$$L_p(u_\varepsilon) = \Delta_p u_\varepsilon + \varepsilon\, a(x) |u_\varepsilon|^{p-2} u_\varepsilon$$

is coercive.

Denote

$$\mathscr{H}^p = \left\{ u_\varepsilon \in H_{0,G}^{1,p}(\Omega_\varepsilon) : \int\limits_{\Omega_\varepsilon} f(x) |u_\varepsilon|^q dx = 1 \right\},$$

$$\mu_\varepsilon = \inf J(u_\varepsilon),$$

for all $u_\varepsilon \in \mathscr{H}^p$, and suppose that exists an isometry σ such that $\sigma(\Omega_\varepsilon) = \Omega_\varepsilon$. Moreover we suppose that the functions $a(x)$ and $f(x)$ are invariant under the action of σ, and

$$\mathscr{H}_\sigma^p = \mathscr{H}^p \cap \left\{ u_\varepsilon \in H_{0,G}^{1,p}(\Omega_\varepsilon) : u_\varepsilon \circ \sigma = -u_\varepsilon \right\} \neq \varnothing.$$

Then, we have the following theorems.

Theorem 2 *For $p = 2$ and $n \geq 3$, the problem (P_ε), always, has a non-radial nodal solution u. Moreover, if $f(x) > 0$ for all $x \in \overline{\Omega}_\varepsilon$, (P_0) has an infinity sequence $\{u_{\varepsilon_i}\}$ of non-radial nodal solutions, such that*

$$\lim_{i\to\infty} \int\limits_{\Omega_\varepsilon} (|\nabla u_{\varepsilon_i}|^2 + u_{\varepsilon_i}^2) dx = +\infty.$$

In addition, u and $\{u_{\varepsilon_i}\}_{i=1,2,\ldots}$ are G-invariant and σ-antisymmetrical.

Theorem 3 *Let a and f be two smooth functions, $\mathscr{H}_\sigma^p$-invariant and p, q be two real numbers defined as in (P_ε). Suppose that $\sup_{x\in\Omega_\varepsilon} f(x) > 0$ and the operator L_p is coercive. Then the problem (P_ε) has a non-radial nodal $\mathscr{H}_\sigma^p$-invariant solution, that belongs to $C^{1,\alpha}(\Omega_\varepsilon)$ for some $\alpha \in (0, 1)$, if*

$$\mu_\varepsilon < K_G^p(\Omega_\varepsilon)^{-p} \left(\sup_{x\in\Omega_\varepsilon} f(x) \right)^{-p/q}.$$

The proofs of Theorems 2 and 3 use standard variational methods, under the assumptions of Lemma 3.6 in [16], (cf. [14, 15, 17, 25]).

4 Solution of the Problem (P)

We return to our main problem

$$(\mathrm{P}) \quad \begin{cases} \Delta_p u = |u|^{p^*(k)} u, \quad u \in C^2(\mathbb{R}^n), \quad n \geqslant 3 \\ 1 < p < n-k, \quad p^*(k) = \frac{(n-k)p}{n-k-p}. \end{cases}$$

In the problem (P) direct variational methods are not applicable because of the double lack of compactness. To overcome this problem we will use an approximate method. That is, we consider a sequence of expanding Ω_{ε_j} (where $\varepsilon_j \to 0$ as $j \to \infty$) as well as the sequence of problems

$$(\mathrm{P}_{\varepsilon_j}) \quad \begin{cases} \Delta_p u_{\varepsilon_j} + \varepsilon_j a(x) \left|u_{\varepsilon_j}\right|^{p-2} u_{\varepsilon_j} = f(x) \left|u_{\varepsilon_j}\right|^{p^*(k)-2} u_{\varepsilon_j} \\ u_{\varepsilon_j} \not\equiv 0 \ \text{ in } \ \Omega, \quad u_{\varepsilon_j} = 0 \ \text{ on } \ \partial\Omega \\ 1 < p < n-k, \quad p^*(k) = \frac{(n-k)p}{n-k=p}, \end{cases}$$

where a, f are as in the problem $(\mathrm{P}_{\varepsilon_j})$.

According to the Theorems 2 and 3, every problem $(\mathrm{P}_{\varepsilon_j})$ has a non-radial nodal $\mathscr{H}_\sigma^p$-invariant solution. Thus, a solution to the problem (P) may be then obtained by the limit procedure as $\varepsilon_j \to 0$.

Before we will approximate the solutions in $\mathbb{R}^n$ by solutions in bounded domains $\Omega_{\varepsilon_j} \in \mathbb{R}^n$, we note that, in the generalized setting of the problems in Ω_{ε_j}s, the Dirichlet condition $u_{\varepsilon_j}(x) = 0$ on $\partial\Omega_{\varepsilon_j}$ may actually be included in the condition $u_{\varepsilon_j} \in H_{0,G}^{1,p}(\Omega_{\varepsilon_j})$.

Moreover, since any function $u_{\varepsilon_j} \in H_{0,G}^{1,p}(\Omega_{\varepsilon_j})$ can be extended onto $\mathbb{R}^n$ by

$$\tilde{u}_\varepsilon(x) = \begin{cases} u_{\varepsilon_j}(x), \quad x \in \Omega_{\varepsilon_j} \\ 0, \quad x \in \mathbb{R}^n \backslash \Omega_{\varepsilon_j}, \end{cases}$$

generalized solutions may be defined in Ω_{ε_j}s analogously to the case in $\mathbb{R}^n$.

We need now the following two definitions:

Definition 1 A function $u_{\varepsilon_j} \in H^{1,p}_{0,G}(\Omega_{\varepsilon_j})$ is a *generalized solution* of $(\mathrm{P}_{\varepsilon_j})$ if the function

$$g(x, u_{\varepsilon_j}) = \varepsilon_j a(x) u_{\varepsilon_j} - f(x)|u_{\varepsilon_j}|^{p^*(k)-2}u_{\varepsilon_j}$$

is locally integrable and for all $\varphi \in C_0^\infty(\Omega_{\varepsilon_j})$, the following holds:

$$\int\limits_{\Omega_{\varepsilon_j}} |\nabla u_{\varepsilon_j}|^{p-2}(\nabla u_{\varepsilon_j}, \nabla\varphi)dx + \int\limits_{\Omega_{\varepsilon_j}} f(x, u_{\varepsilon_j})\varphi dx = 0.$$

Definition 2 A function $u_\varepsilon \in C^2(\Omega_\varepsilon) \cap C(\overline{\Omega}_\varepsilon)$ is a *classical solution* to (P_ε) if after substituting it into equation of (P_ε), this equation becomes the identity at each $x \in \Omega_\varepsilon$ and $u_\varepsilon(x) = 0$ provided $x \in \partial\Omega_\varepsilon$.

Provided that all the conditions of the Theorem 3 are satisfied, we apply it to the sequence of the problems $(\mathrm{P}_{\varepsilon_j})$ and denote by $\{u_j\}_{j=1}^\infty$ the sequence of the corresponding solutions.

Under the above considerations to following theorem holds.

Theorem 4 *The problem*

$$\Delta_p u = f(x)|u|^{p^*(k)-2}u \quad \text{in } \mathbb{R}^n, \quad n \geq 3$$

has a generalized non-radial nodal $\mathscr{H}^p_\sigma$-invariant solution u and there is a subsequence of $\{u_j\}$ (again denoted by $\{u_j\}$) such that

$$u_j \rightharpoonup u \quad \text{in} \quad H^{1,p}_{0,G} \quad \text{as} \quad j \to \infty.$$

In addition

$$\lim_{j\to\infty} \int\limits_{\mathbb{R}^n} |\nabla u_j|^p dx = +\infty.$$

Proof The case $p = 2$ is presented in [42], thus, we will prove the case $p \in (1, 2) \cup (2, n - k)$. However, we present a unified proof for both cases. For the proof we borrow ideas from [42] and carried out in 5 steps.

Step 1. According to the above Theorem 3, every problem $(\mathrm{P}_{\varepsilon_j})$ has at least one non-radial nodal G-invariant and σ antisymmetrical solution u_j. Let u_j, $j = 1, 2, \ldots$, an arbitrary sequence of such solutions. Since the problem $(\mathrm{P}_{\varepsilon_j})$ has a nontrivial solution belonging to one of the spaces considered earlier, then for any $\lambda > 0$ the function

$$v_j = \lambda^{\frac{1}{p^*(k)-p}} u_j \in H^{1,p}_{0,G}\left(\Omega_{\varepsilon_j}\right)$$

is a non trivial solution to the problem:

$$(\mathrm{P}^{\lambda}_{\varepsilon_j}) \quad \begin{cases} \Delta_p v_j + \varepsilon_j a(x) \left|v_j\right|^{p-2} v_j = \lambda f(x) \left|v_j\right|^{p^*(k)-2} v_j \\ v_j \not\equiv 0 \ \text{ in } \ \Omega, \ \ v_j = 0 \ \text{ on } \ \partial\Omega \\ 1 < p < n-k, \quad p^*(k) = \frac{(n-k)p}{n-k=p}. \end{cases}$$

In this first step of the proof we prove that there exists a sub-sequence of the sequence of the solutions to the problems $\left(P^{\lambda}_{\varepsilon_j}\right)$ which converges weakly in $H_0^{1,p}(\mathbb{R}^n)$.
For

$$\lambda = \left\|u_j\right\|^{-(p^*(k)-p)}_{H^{1,p}\left(\Omega_{\varepsilon_j}\right)}$$

we obtain that

$$v_j = \frac{u_j}{\left\|u_j\right\|_{H^{1,p}\left(\Omega_{\varepsilon_j}\right)}},$$

which means that the sequence $\{v_{\varepsilon_j}\}$ is bounded in $H^{1,p}(\Omega_{\varepsilon_j})$ for all $j = 1, 2, \ldots$.
Therefore, there is a positive constant C not dependent on j and such that:

$$\|v_j\|_{H^{1,p}(\Omega_{\varepsilon_j})} \leq C, \ \forall \ j = 1, 2, \cdots. \tag{7}$$

Because of the reflexivity of $H_0^{1,p}(\mathbb{R}^n)$ and condition (7) we may choose a sub-sequence of $\{v_j\}$ (again denoted by $\{v_j\}$) such that:

$$v_j \rightharpoonup v \ \text{ in } \ H_0^{1,p}(\mathbb{R}^n) \ \text{ as } \ j \to +\infty. \tag{8}$$

Step 2. In this step we prove that the function v is a nontrivial G-invariant generalized solution of the limit problem obtained from the sequence of problems $(\mathrm{P}^{\lambda}_{\varepsilon_j})$ as $j \to \infty$.
We choose an arbitrary $\varphi \in C_0^{\infty}(\mathbb{R}^n)$. Then, according to the definition of $C_0^{\infty}(\mathbb{R}^n)$, the support of φ is bounded in $\mathbb{R}^n$, which means that there is an Ω_{ε_0} such that supp$\varphi \subset \Omega_{\varepsilon_0}$. Since, by definition, the Ω_{ε_j}s constitute a family of expanding domains, we can choose the Ω_{ε_0} such that $\Omega_{\varepsilon_0} \subset \Omega_{\varepsilon_1}$ and so $\Omega_{\varepsilon_0} \subset \Omega_{\varepsilon_j}$ for all $j = 1, 2, \ldots$.
Let

$$g(x, v_j) = \varepsilon_j a(x) v_j - \lambda f(x) |v_j|^{p^*(k)-2} v_j.$$

Then, because the υ_j is a generalized solution to $(\mathrm{P}^{\lambda}_{\varepsilon_j})$, it holds

$$\int\limits_{\mathbb{R}^n} |\nabla v_j|^{p-2}(\nabla v_j, \nabla\varphi)dx = -\int\limits_{\Omega_{\varepsilon_j}} g(x, v_j)\varphi dx = -\int\limits_{\Omega_{\varepsilon_0}} g(x, v_j)\varphi dx. \tag{9}$$

for all Ω_{ε_j}.
By the weak convergence (8), we obtain the following limit relation for the left-hand side of (9):

$$\lim_{j\to\infty}\int\limits_{\mathbb{R}^n} |\nabla v_j|^{p-2}(\nabla v_j, \nabla\varphi)dx = \int\limits_{\mathbb{R}^n} |\nabla v|^{p-2}(\nabla v, \nabla\varphi)dx. \tag{10}$$

In addition, the critical exponent of the embedding

$$H^{1,p}_G(\Omega_{\varepsilon_0}) \hookrightarrow L^p(\Omega_{\varepsilon_0})$$

is equal to

$$p^*(k) = \frac{(n-k)p}{n-k-k} > \frac{np}{n-p} = p^*.$$

Let some p_0 such that

$$p^* < p_0 < p^*(k).$$

Then the embedding is compact and thus from the Sobolev and Kondrashov theorems together and (9) arises that

$$\upsilon_j \to \upsilon \ \text{ in } \ L^{p_0-1}(\Omega_{\varepsilon_0}), \ \text{ as } \ j \to +\infty. \tag{11}$$

Furthermore, by definition of $a(x)$ and $f(x)$, there exists a positive constant C such that:

$$|g(x,t)| \le C(|t| + |t|^{p_0-1}), \quad p^* < p_0 < p^*(k),$$

for almost all $x \in \Omega_{\varepsilon_j}$, $j = 1, 2, \ldots$ and for all $t \in \mathbb{R}$.
Therefore, by Vainberg-Krasnoselskii Theorem (cf. [39, 65] or [41]) gives that:

$$\varphi g(\cdot, \upsilon_j(\cdot)) \to \varphi g(\cdot, \upsilon(\cdot)) \ \text{ in } \ L^{\frac{p_0}{p^*}}(\Omega_{\varepsilon_0}) \ \text{ as } \ j \to +\infty \tag{12}$$

and the Hölder inequality from (12) follows that:

$$\varphi g(\cdot, \upsilon_j(\cdot)) \to \varphi g(\cdot, \upsilon(\cdot)) \ \text{ in } \ L^{r}(\Omega_{\varepsilon_0}) \ \text{ as } \ j \to +\infty, \tag{13}$$

for all $1 \leq r \leq \frac{p_0}{p^*}$.

By (13) the limit relation from the right hand-side of (10) yields:

$$\lim_{j\to\infty} \int\limits_{\Omega_{\varepsilon_0}} g(x, \upsilon_j)\varphi dx = \int\limits_{\Omega_{\varepsilon_0}} g(x, \upsilon)\varphi dx. \tag{14}$$

Finally, passing to the limit in (9) because of (8) and (14), we obtain:

$$\int\limits_{\mathbb{R}^n} |\nabla \upsilon|^{p-2}(\nabla \upsilon, \nabla \varphi)dx = -\int\limits_{\Omega_{\varepsilon_0}} g(x, \upsilon)\varphi dx = -\int\limits_{\mathbb{R}^n} g(x, \upsilon)\varphi dx,$$

which corresponds to the definition of a weak solution. This is a generalized solution by the force of (9) and since the function f is regular enough it is a classical solution, (see §§ 1.2 and 3.1 in [41]). As convergence in L^p spaces implies a.e. convergence by (11) follows that the function υ will be G-invariant.

Step 3. In this step we prove that the solution υ is non trivial, that is $\upsilon \not\equiv 0$. Suppose, by contradiction, that $\upsilon \equiv 0$. Then, for any $\varepsilon > 0$ we have

$$|\upsilon| < \frac{\varepsilon}{2}. \tag{15}$$

On the other hand, from (13) arises that

$$\upsilon_j \to \upsilon \ \text{ in } \ L^1(\Omega_{\varepsilon_0}),$$

which means that for any $\varepsilon > 0$ there exists a positive integer j_0 such that:

$$|\upsilon_j - \upsilon| < \frac{\varepsilon}{2} \ \text{ for all } \ j > j_0. \tag{16}$$

Therefore, by the standard inequality

$$|\upsilon_j| \leq |\upsilon_j - \upsilon| + |\upsilon|$$

due to (15) and (16) we obtain that:

$$|\upsilon_j| < \varepsilon \ \text{ for any } \ j \geq j_0. \tag{17}$$

We recall now that every solution to the problem (P_{ε_j}) belongs to the set

$$\mathscr{H}_\varepsilon^\sigma = \left\{ u_\varepsilon \in H^{1,p}_{0,G}(\Omega_{\varepsilon_j}) : \ u_{\varepsilon_j} \circ \sigma = -u_{\varepsilon_j} \ \text{ and } \int\limits_{\Omega_{\varepsilon_j}} f(x)|u_{\varepsilon_j}|^{p^*(k)}dx = 1 \right\}.$$

Since every υ_j corresponds to an $u_{\varepsilon_j} \in \mathscr{H}_{\varepsilon}^{\sigma}$, and $\upsilon_{\varepsilon_j} = \lambda^{\frac{1}{p^*(k)-p}} u_{\varepsilon_j}$, by definition, we have the following:

$$1 = \int\limits_{\Omega_{\varepsilon_j}} f(x)\lambda^{-\frac{p^*(k)}{p^*(k)-p}} |\upsilon_j|^{p^*(k)} dx < \int\limits_{\Omega_{\varepsilon_j}} f(x)\lambda^{-\frac{p^*(k)}{p^*(k)-p}} \varepsilon^{p^*(k)} dx,$$

which is false due to (17) as the $\varepsilon > 0$ can be chosen as small as we want.

Step 4. We have proved that the limit problem

$$(\mathrm{P}^{\lambda}) \quad \Delta_p \upsilon = \lambda f(x)|\upsilon|^{p^*(k)-2}\upsilon \quad \text{in } \mathbb{R}^n, \quad n \geq 3$$

has a generalized non-radial nodal G-invariant and σ-anti-symmetrical solution υ, which means that the function $u = \lambda^{\frac{1}{p^*(k)-p}} \upsilon$ is a generalized non-radial nodal G-invariant and σ-anti-symmetrical solution to the limit problem:

$$(\mathrm{P}) \quad \Delta_p u = f(x)|u|^{p^*(k)-2}u \quad \text{in } \mathbb{R}^n, \quad n \geq 3.$$

Step 5. It remains to prove that

$$\lim_{j\to\infty} \int\limits_{\mathbb{R}^n} |\nabla u_j|^p dx = +\infty.$$

The Sobolev inequality (6) after a normalization of the sequence u_js so that $\|u_j\|_{L^{p^*(k)}(\Omega_{\varepsilon_j})=1}$ and provided that the constants B_ϵ are positive give us that

$$1 \leqslant \left(K_G^p\left(\Omega_{\varepsilon_j}\right) + \epsilon\right) \int\limits_{\Omega_{\varepsilon_j}} \left|\nabla u_j\right|^p dx. \tag{18}$$

From (18) after a replacement of the constant $K_G^p(\Omega_{\varepsilon_j})$ from the one calculated in Theorem 1 we obtain the inequality

$$\frac{1}{\varepsilon_j^m V^{-\frac{1}{n-k}} K\left(n-k,\ p\right) + \epsilon} < \int\limits_{\Omega_{\varepsilon_j}} \left|\nabla u_j\right|^p dx. \tag{19}$$

By inequality (19) taking the limits for $j \to \infty$ we have that $\varepsilon_j \to 0$ and then

$$\int\limits_{\mathbb{R}^n} \left|\nabla u_j\right|^p dx \to \infty.$$

This completes the proof of the theorem. □

Corollary 1 *The problem:*

$$(\mathrm{P}) \quad \begin{cases} \Delta_p u = |u|^{p^*(k)} u, \quad u \in C^2(\mathbb{R}^n), \quad n \geqslant 3 \\ 1 < p < n-k, \quad p^*(k) = \frac{(n-k)p}{n-k-p}, \end{cases}$$

has a sequence $\{u_j\}$ *of non-radial nodal* G*-invariant and* σ*-anti-symmetrical solutions, such that:*

$$\lim_{j \to +\infty} \int_{\mathbb{R}^n} |\nabla u_j|^p 0 dx = +\infty.$$

Proof The result is obtained if we put

$$f(x) = \frac{1}{|\Omega_{\varepsilon_j}|} - \varepsilon_j |x|^\alpha, \ \alpha > -n$$

and follows the spirit of the approach in Theorem 4. □

Remark 1 The number of the sequences of non-radial nodal G-invariant and σ-anti-symmetrical solutions to the problem (P), depends on the number of all subgroups of $O(n)$ of which the cardinal of orbits with minimum volume is infinite, that are on the dimension n of the domain.

References

1. K. Adimurthi, S.L. Yadava, Elementary proof of the nonexistence of nodal solutions for the semilinear elliptic equations with critical Sobolev exponent. Nonlinear Anal. **14**(9), 785–787 (1990)
2. A. Ambrosetti, P.H. Rabinowitz, Dual variational methods in critical point theory and applications. J. Funct. Anal. **14**, 349–381 (1973)
3. A. Ambrosetti, M. Struwe, A note on the problem $-\Delta u = \lambda u + |u|^{2^*-2}u$. Manuscripta Math. **54**, 373–379 (1986)
4. F.V. Atkinson, H. Brezis, L.A. Peletier, Nodal solutions of elliptic equations with critical Sobolev exponents. J. Diff. Eq. **85**, 151–170 (1990)
5. T. Aubin, Equations différentielles non linéaires et probléme de Yamabe concernant la courbure scalaire. J. Math. Pures Appl. **55**, 269–296 (1976)
6. T. Aubin, *Some Non Linear Problems in Riemannian Geometry* (Springer, Berlin, 1998)
7. A. Bahri, J.M. Coron, On a nonlinear elliptic equation involving the limiting Sobolev exponent. Comm. Pure Appl. Math. **41**, 253–294 (1988)
8. T. Bartsch, M. Schneider, Multiple solutions of a critical polyharmonic equation. J. Reine Angew. Math. **571**, 131–143 (2004)
9. H. Brezis, Elliptic equations with limiting Sobolev exponent—the impact of topology, in *Procceding of the 50th Anniversary Courant Institute Communications on Pure Applied Mathematics*, vol. 39 (1986), pp. 517–539
10. H. Brezis, L. Nirenberg, Positive solutions of nonlinear elliptic equations involving critical Sobolev exponents. Comm. Pure Appl. Math. **36**(4), 437–477 (1983)

11. L. Caffarelli, B. Gidas, J. Spruck, Asymptotic symmetry and local behavior of semilinear elliptic equations with critical Sobolev growth. Comm. Pure Appl. Math., **42**, 271–297 (1989)
12. A. Carpio Rodriguez, M. Comte, R. Lewandoski, A nonexistence result for nonlinear equation involving critical Sobolev exponent. Ann. Inst. Henri Poincaré Anal. Non Linéaire **9**, 243–261 (1992)
13. Y.-H. Cheng, Eigenvalue problems with p-Laplacian operators. Electron. J. Differ. Equations **2014**(139), 1–11 (2014)
14. A. Cotsiolis, D. Iliopoulos, Equations elliptiques non lineaires a croissance de Sobolev sur-critique. Bull. Sci. Math. **119**, 419–431 (1995)
15. A. Cotsiolis, N. Labropoulos, Dirichlet problem on a solid torus in the critical of supercritical case. Bull. Greek Math. Soc. **53**, 39–57 (2007)
16. A. Cotsiolis, N. Labropoulos, Best constants in Sobolev inequalities on manifolds with boundary in the presence of symmetries and applications. Bull. Sci. Math. **132**, 562–574 (2008)
17. F. Demengel, E. Hebey, On some nonlinear equations involving the p-Laplacian with critical Sobolev growth. Adv. Diff. Equations **3**, 533–574 (1998)
18. P. De N'apoli1, M.C. Mariani, Mountain pass solutions to equations of p-Laplacian type. Nonlinear Anal. **54**, 1205–1219 (2003)
19. J. Dieudonné, Eléments d'Analyse , tome 3. Gauthier-Villars **1974**, 562–574 (2008)
20. W. Ding, On a conformally invariant elliptic equation on $\mathbb{R}^n$. Commun. Math. Phys. **107**, 331–335 (1986)
21. W. Ding, Positive solutions of $\Delta u + u^{\frac{n+2}{n-2}} = 0$ on contactible domains. J. Partial Diff. Equation **2**, 83–88 (1989)
22. H. Ding, J. Zhou, Global existence and blow-up for a mixed pseudo-parabolic p-Laplacian type equation with logarithmic nonlinearity. J. Math. Anal. Appl. **478**, 393–420 (2019)
23. L. Dupaigne, M. Ghergu, V. Rădulescu, Lane-Emden-Fowler equations with convection and singular potential. J. Math. Pures Appl. **87**(6), 563–581 (2007)
24. J. Esteban, J. Vazquez, On the equation of turbulent filtration in one dimensional porous media. Nonlinear Anal. **10**, 1303–1325 (1986)
25. Z. Faget, Best constants in Sobolev inequalities on Riemannian manifolds in the presence of symmetries. Potential Anal. **17**, 105–124 (2002)
26. R. Filippucci, P. Pucci, F. Robert, On a p-Laplace equation with multiple critical nonlinearities. J. Math. Pures Appl. **91**, 156–177 (2009)
27. D. Fortunato, E. Jannelli, Infinity many nodal solutions for some nonlineal elliptic problems in symmetrical domains. Proc. R. Soc. Edinb. **105**(A), 205–213 (1987)
28. M. Fraas, Y. Pinchover, Positive liouville theorems and asymptotic behavior for p-Laplacian type elliptic equations with a fuchsian potential. Confluentes Mathematici **3**(2), 291–323 (2011)
29. M. Ghergu, V. Rădulescu, Ground state solutions for the singular Lane-Emden-Fowler equation with sublinear convection term. J. Math. Anal. Appl. **333**, 265–273 (2007)
30. B. Gidas, W.-M. Ni, L. Nirenberg, Symmetry and related properties via the maximum principle. Commun. Math. Phys. **68**, 209–243 (1979)
31. C.-Y. Guo, M. Kar, Quantitative uniqueness estimates for p-Laplace type equations in the plane. Nonlinear Anal. **143**, 19–44 (2016)
32. E. Hebey, M. Vaugon, Existance and multiplicity of nodal solutions for nonlinear elliptic equations with critical Sobolev growth. J. Funct. Anal. **119**, 298–318 (1994)
33. E. Hebey, M. Vaugon, Meilleures constantes dans le théorème d'inclusion de Sobolev. Ann. Inst. Henri Poincaré, Analyse Non Linéaire **13**, 57–93 (1996)
34. E. Hebey, M. Vaugon, Sobolev spaces in the presence of symmetries. J. Math. Pures Appl. **76**, 859–881 (1997)
35. S. Kamin, J. Vazquez, Fundamental solutions and asymptotic behaviour for the p-Laplacian Equation. Revista Matematica Iberoamericana **4**(2), 339–354 (1988)
36. B. Kawohl, J. Horák, On the geometry of the p-Laplacian operator. Discrete Contin. Dyn. Syst. Ser. S **10**, 799–813 (2017)

37. J. Kim, M. Zhu, Non-existence results about $-\Delta u = u^{\frac{n+2}{n-2}}$ on non-starshaped domains. J. Differ. Equations **225**, 737–753 (2006)
38. S. Kobayashi, Transformation groups in differential geometry, in *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 70 (1972)
39. A. Krasnoselskii, *Topological Methods in the Theory of Nonlinear Integral Equations* (Macmillan, New York and London, 1964)
40. A. Kristály, H. Liseib, C. Vargab, Multiple solutions for p-Laplacian type equations. Nonlinear Anal. **68**, 1375–1381 (2008)
41. I. Kuzin, S. Pohozaev, Entire solutions of semilinear elliptic equations, in *Progress in Nonlinear Differential Equations and their Applications*, vol. 33 (Birkhauser, Basel, 1997)
42. N. Labropoulos, An alternative approach to critical PDEs. Electron. J. Differ. Equations **2017**(150), 1–22 (2017)
43. Y.Y. Li, Existence of many positive solutions of semilinear elliptic equations on annulus. J. Diff. Eq. **83**, 348–367 (1990)
44. P. Lindqvist, *Notes on the p-Laplacian Equation, University of Jyväskyliä—Lecture notes*, 2nd edn. (2017)
45. C. Loewner, L. Nirenberg, Partial differential equations invariant under conformal and projective transformations, in *Contributions to Analysis* (Academic Press, New York, 1974), pp. 245–272
46. I. Ly, The first eigenvalue for the p-Laplacian operator. J. Inequal. Pure Appl. Math. **6**(3/91), 1–12 (2005)
47. R. Mazzeo, N. Smale, Conformally flat metrics of constant positive scalar curvature on subdomains of the sphere. J. Differ. Geom. **34**, 581–621 (1991)
48. L. Nirenberg, Variational and topological methods in nonlinear problems. Bull. Am. Math. Soc. **4**, 267–302 (1981)
49. J. Padial, P. Takáč, L. Tello, An antimaximum principle for a degenerate parabolic problem. Adv. Differ. Equations, **15**(7–8), 601–648 (2010)
50. R.S. Palais, Morse theory on Hilbert manifolds. Topology **2**, 299–340 (1963)
51. R.S. Palais, Lusternik-Schnirelman theory on Banach manifolds. Topology **5**, 115–132 (1966)
52. D. Passaseo, Multiplicity of positive solutions of nonlinear elliptic equations with critical Sobolev exponent in some contractible domains. Manuscripta Math. **65**, 147–165 (1989)
53. K. Perera, R.P. Agarwal, D. O'Regan, Morse Theoretic Aspects of p-Laplacian Type Operators, in *Mathematical Surveys and Monographs*, vol. 161 (American Mathematical Society, Providence, 2010)
54. M. Pereira, R. Silva, Remarks on the p-Laplacian on thin domains. Prog. Nonlin. Differ. Equ. Appl. **86**, 389–403 (2015)
55. Y. Pinchover, K. Tintarev, On positive solutions of p-Laplacian-type equations in Analysis, in *Partial Differential Equations and Applications—The Vladimir Maz'ya Anniversary Volume*. Operator Theory: Advances and Applications, vol. 193, eds. by A. Cialdea et al. (Birkäuser, Basel, 2009), pp. 245–268
56. S.I. Pohozaev, Eigenfunctions of the equation $\Delta u + \lambda f(u) = 0$. Sov. Math. Dokl. **6**, 1408–1411 (1965)
57. R. Schoen, Conformal deformation of a Riemannian metric to constant scalar curvature. J. Diff. Geometry **20**, 479–495 (1984)
58. R. Schoen, The existence of weak solutions with prescribed singular behavior for a conformally invariant scalar equation. Commun. Pure Appl. Math. **41**, 317–392 (1988)
59. J. Schwartz, *Nonlinear Functional Analysis* (Gordon and Breach, New York, 1968)
60. L. Shi, X. Chang, Multiple solutions to p-Laplacian problems with concave nonlinearities. J. Math. Anal. Appl. **363**, 155–160 (2010)
61. W. Smith, Morse theory without critical points. Rocky Mountain J. Math. **19**(2), 531–543 (1989)
62. M. Struwe, *Variational Methods Applications to Nonlinear Partial Differential Equations and Hamiltonian Systems* (Springer, Berlin, 2008)

63. G. Tarantello, Nodal solutions of semilinar elliptic equations with critical exponent. Differ. Integral Equation **5**, 25–42 (1992)
64. N.S. Trüdinger, On imbeddings into Orlicz spaces and some applications. J. Math. Mech. **17**(5), 473–483 (1967)
65. M. Vainberg, *The Variational Method for the Study of Non-linear Operators* (Holden-Day, San Francisco, 1964)
66. L. Wang, Infinitely many solutions for an elliptic problem with critical exponent in exterior domain. J. Part. Differ. Equation **23**(1), 80–104 (2010)
67. E.-M. Wang, Y. Zheng, Regularity of the first eigenvalue of the p-laplacian and Yamabe invariant along geometric flows. Pacific J. Math. **254**(1), 239–255 (2011)
68. H. Yamabe, On a deformation of Riemannian structures on compact manifolds. Osaka. Math. J. **12**, 21–37 (1960)
69. Z. Yang, D. Geng, H. Yan, Three solutions for singular p-Laplacian type equations. Electron. J. Differ. Equations **2008**(61), 1–12 (2008)
70. E. Zeidler, *Nonlinear Analysis and its Applications III: Variational Methods and Optimization* (Springer, Berlin, 1985)

Financial Contagion in Interbank Networks: The Case of Erdős–Rényi Network Model

K. Loukaki, P. Boufounou, and J. Leventides

Abstract In this study, we extend the model developed in Leventides et al. (J Econ Behav Organ 158:500–525, 2019) to include a wide variety of network topologies and provide a better understanding of the relation between network structure, banks' characteristics and interbank contagion. While the focus of this paper is on the various factors that affect interbank contagion such as bank capital ratios, leverage, interconnectedness and homogeneity across banks' sizes, the model lacks flexibility as far as the variability of the networks links is concerned. In order to circumvent this problem, we introduce the Erdős–Rényi probabilistic network model in our study to provide a wider vicinity of scenarios concerning the network structure of the interbank system and study how homogeneity within the interbank network affects the propagation of financial distress from one institution to the other parts of the system through bilateral exposures.

1 Introduction

Meeting the SDGs has currently secured prior importance for both businesses' and nations' socio-economic-environmental transformation to achieving sustainable development. More than 10,000 companies around the world have already signed up to the principles of sustainable business behavior and an adequate number of special toolkits has been developed to assist them towards this transformation. As explicitly stated in [1] "Achieving the Global Goals would create a world that is comprehensively sustainable: socially fair; environmentally secure; economically prosperous; inclusive; and more predictable". According to Oxford Analytica Foundation [2], "companies that see the business case – as well as the moral imperative – for achieving all the Global Goals will take a 'Global Goals lens' to every aspect of their business strategy to change the way they operate and put

K. Loukaki · P. Boufounou · J. Leventides (✉)
Department of Economics, National and Kapodistrian University of Athens, Athens, Greece
e-mail: ylevent@econ.uoa.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_13

more focus on inclusion". Accordingly, Weber [3] underlined that "the banking sector became aware of the opportunity to finance the change to more sustainable development instead of just focusing on risks for their lending business". Alves et al. [4] debated that "EU aid policy evolved over the last fifteen years in accordance to the notion of financial liberalization and to the importance of private initiative and combined with the SDGs increased the promotion of the financial sector as engine of growth and development in the developing countries' and that "the New International Financial architecture assigns new roles for developing nations in the global financial markets". Achieving the Agenda 2030 depends on aligning the entire global chain of the financial and the banking system with sustainability and long-term outcomes therefore delineating of interbank linkages network structure becomes of outmost importance.

Furthermore, in the wake of the aftermaths of global financial crisis of 2007–2009 and the European sovereign debt crisis, there is a lot of attention of systemic risk, interconnectedness and contagious effects. Thus, there is a critical need for a better understanding of the fragility of the financial systems, their inner interconnections, their interaction with real economy and the conditions that can drive them from stability to instability and complete breakdown. In recent years, both academics and regulators has started to study various architectures of the financial system in order to assess certain risks within the system that potentially lead to huge losses for the overall economy.

The global financial system can be represented as a large complex network in which banks, hedge funds and other financial institutions are interconnected to each other through various forms of financial linkages. For example, in the banking sector, banks can be interconnected through direct and indirect links. Direct interconnectedness arises from bilateral transactions; borrowing or lending relationships between banks. A default by one bank, for example, can impose distress on other entities that hold significant liabilities of the defaulting bank. Thus, the failure of a bank can jeopardize the ability of its creditors banks to meet their obligations to their interbank creditors which may lead to a domino effect. There are also indirect ways that banks can be interconnected, since they invest in common securities, namely portfolio overlap. If, for example, a bank holds identical assets with other banks the correlation between their portfolios can cause fire sales in the market during a crisis period depressing thus overall prices in the market, ultimately leading to downward spirals for asset sales and inducing significant losses for all the participants in the market. The complexity of the financial system led many academics to utilize the network theory to study the effects of the interconnectedness and network topology on financial stability. Studying the financial system as a network is one of the methods to investigate the emergence of systemic risk through the connections of banks. In such a network structure every node represents a bank, the connections between banks are represented by edges where edge's weight represents the magnitude of exposure between the two parties and edge directionality allows one to determine who is the creditor and who is the lender. A robust interbank market plays an important role on the stability of the financial system. Through the interbank market, banks which suffer a liquidity shortage can

borrow from banks with liquidity surpluses. The interbank market can stabilize the financial system, by redistributing the funds in a effective way among the banks but at the same time can make the system prone to contagion of financial trouble from one bank to another (linkages).

In this paper, we focus our attention on the direct contagion channel and aim to identify the main drivers that affect interbank contagion. The flourishing literature which ensued in recent years has developed both theoretical models and empirical applications aimed at addressing the various issues concerning systemic risk. Counterfactual simulations on data have been extensively employed to study interbank contagion under different scenarios related to the topology of the interbank network, the size of interbank exposures and the degree of heterogeneity and interconnectedness within the network. In our assessment of the various drivers that affect interbank contagion, we extend the model developed in Leventides et al. [5] to include a wide variety of network topologies and provide a better understanding of the relation between network structure, banks' characteristics and interbank contagion. While the focus of this paper is on the various factors that affect interbank contagion such as bank capital ratios, leverage, interconnectedness and homogeneity across banks' sizes, the model lacks flexibility as far as the variability of the networks links is concerned. In this effort, interbank exposure and capital equity among banks displayed a stochasticness and the ability to construct a wide range of scenarios regarding connective links among banks is limited.

The introduction of the Erdős–Rényi probabilistic network model provides us with a wider vicinity of scenarios concerning the network structure of the interbank system. Under this framework, we build up multiple scenarios of various network structures that include a satisfactory number of cases via Monte Carlo simulations. In every single network that we construct, we investigate the dynamics of cascading defaults from an initial random shock that hits the system. Erdős–Rényi random graph model is one of the earliest theoretical network models and was introduced in the early 1960s by the Hungarian mathematicians Paul Erdős and Alfréd Rényi. In this random graph, each possible link between any two nodes can occur with a certain independent and identical probability, the Erdős and Rényi probability.

The Erdős and Rényi random graph model is a model in which has been extensively applied for the study of contagion in financial networks, e.g. [23, 24]. However, a number of alternatives have been recently developed that differ in the probability law governing the distribution of links between nodes. Using the Erdős–Rényi network structure, the degree distribution or the connectivity among banks can vary with respect to the chosen probability p. Thus, each random network generated with the same parameters N, p looks slightly different. Not only the detailed wiring network graph changes between realizations, but so does the number of links. Random graphs or Erdős–Rényi graphs are useful for modeling, analysis, and solving of structural and algorithmic problems arising in mathematics, theoretical computer science, statistical mechanics, natural sciences, and even in social sciences. However, the utility of an Erdős–Rényi model lies mainly in its mathematical simplicity, not in its realism. Virtually, the comparison with real-

world networks indicates that the random network model does not capture the degree distribution of real networks but it provides a useful baseline for more complicated network models.

The remainder of the paper is organized as follows. The following section discusses briefly the recent literature that has addressed on the topic of interbank contagion. Various aspects of systemic risk and network structures that are either found in real-world data or used in some theoretical studies of interbank contagion are addressed before we introduce the model investigated in Section 3. In Section 4 we describe the variables, considered in our subsequent analysis, provide in full detail the computer experiments conducted and discuss our simulations results. Summary and concluding remarks are drawn in the final section.

2 Related Literature

According to Upper [11], the channels through which a shock spreads can be broken down into two groups: indirect and direct contagion channels. A direct contagion channel results from the direct interbank linkages among banks and can take effect when an idiosyncratic shock travels through the network. This shock can be due to the inability of some banks to meet their financial obligations or due to interbank exposures that are quite large relative to the lender's capital. The possibility of the occurrence and transmission of direct contagion depends mainly on the structure and size of the interbank market. On the other hand, indirect contagion is created by indirect linkages among banks such as identical assets, portfolio returns and overlapping portfolios. If, for example, a bank holds identical assets with other banks, the correlation between their portfolios can cause fire sales in the market during a crisis period, thus depressing overall prices in the market and inducing significant losses for all participants [12]. Distinguishing among the various contagion channels is crucial for understanding financial contagion and the mechanisms through which it spreads and evolves.

There are a number of recent studies that have dealt with the issue of interbank contagion. Memmel and Sachs [13] simulate interbank contagion effects for the German banking sector and find that bank capital ratios, the share of interbank assets in the system and the degree of equality in the distribution of interbank exposures are the most important determinants for financial stability. Georgescu [14] compares the contagion potential of accounting induced regulatory constraints to that of funding constraints in a bank network and concludes that the interplay between illiquidity and solvency can lead to bank failures which are manifested by the vulnerable funding structure of banks during a crisis. Tonzer [15] examines the relationship between cross-border bank linkages and financial stability and show that larger cross-border exposures increase bank risks, however, when bilateral interbank linkages exist there is a shift toward a more stable banking system. Fink et al. [16] model contagion in the German interbank market via the credit quality channel and propose a novel metric which estimates the potential regulatory capital loss to a banking system due to contagion via interbank loans. They show that

contagion effects can be reduced if banks alter their lending and borrowing habits in response to policy interventions.

Our analysis also relates to the role of heterogeneity in the structure of interbank networks and how this characteristic affect systemic risk. Iori et al. [6] use an Erdős–Rényi network model of 400 banks comprising the interbank market in which the lending and borrowing functions are endogenously generated. The authors find that the likelihood of contagion is lowered in case the interconnected institutions are homogeneous, i.e. they have similar characteristics such as size or investment opportunities and thus, no institution becomes significant for either borrowing or lending. The authors also suggest, in line with Allen and Gale [17], that as connectivity increases the system becomes more stable. In a related study, Caccioli et al. [18] study the role of heterogeneity in degree distributions (the number of incoming and outgoing links), balance sheet size and degree correlations between banks. They find that networks with heterogeneous degree distributions are shown to be more resilient to contagion triggered by the failure of a random bank, but more fragile with respect to contagion triggered by the failure of highly connected nodes. The authors also provide evidence that when the average degree of connectivity is low, the probability of contagion due to failure of highly connected banks is higher than that due to the failure of large banks. However, when the average degree of connectivity is high, the opposite holds. Since the second scenario seems to be more realistic (networks with high connectivity), having "too big to fail" banks is more effective in eliminating a shock. Ladley [19] develops a partial equilibrium model of a closed economy in which heterogeneous banks interact with borrowers and depositors through the interbank market. Banks in the model are subject to regulation and the aim of the model is to qualitatively show how regulation and network structure can constrain or enhance the risk of contagion. The results show that for high levels of connectivity the system is more stable when the shock is small, while the contagion effects are amplified in case of larger initial shocks. Chinazzi et al. [20] explore the interplay between heterogeneity, network structure and balance sheet composition in the transmission of contagion. They argue that heterogeneity in connectivity provides additional resiliency to the system when the initial default is random and also show that 'too-connected-to-fail' banks are more dangerous than 'too-big-to-fail' ones and should be the primary concern of policy makers since their failure can trigger systemic breakdowns. Amini et al. [10] focus on bank heterogeneity in terms of the number of banks included in the network and the magnitude of their interconnections with other banks. They conclude that the more heterogeneity is introduced, the less resilient the network becomes. Contrary to these findings, the study of Georg and Poschmann [21] finds no significant evidence that the heterogeneity of the financial system has a negative impact on financial stability.

Finally, as far as the structure of an interbank system is concerned, the most common network structures that are either found in real-world data or used in some theoretical studies of interbank contagion are the Erdős–Rényi random network structure, introduced in Erdős and Rényi (1960), the small-world structure,

introduced in Watts and Strogatz (1998) and the scale-free structure, introduced in Barabasi and Albert (1999).

The Erdős–Rényi network structure, which is applied in our study, can be obtained by connecting any two nodes with a fixed and independent probability p. Thus, in an Erdős–Rényi network structure the degree or the number of links of a node is p(n-1). The expected degree distribution for such networks is Binomial, converging to Poisson for large n. The Erdős and Rényi (1960) random graph model is a model in which has been extensively applied for the study of contagion in financial networks, e.g. in the contributions from Iori et al. [6], Nier et al. [7], Gai and Kapadia [8], May and Arinaminpathy[9] and Amini et al. [10]. A number of alternatives models have been recently developed that differ in the probability law governing the distribution of links between nodes. Nier et al. [7] study the extent to which the resilience of an interbank network depends on a combination of variables characterizing the network topology, banks' characteristics in terms of net worth and interbank exposures, and market concentration. Using Monte Carlo simulation experiments in Erdős–Rényi random graphs, they find that the effect of the degree of connectivity is non-monotonic. Specifically, a small initial increase in connectivity increases the chance of contagion defaults. However, after a certain threshold value, connectivity improves the capacity of a banking system to withstand shocks. In addition, the authors find that the banking system is more resilient to contagious defaults if its banks are better capitalized and this effect is non-linear. Finally, the size of interbank liabilities tends to increase the risk of default cascades, even if banks hold capital against such exposures and more concentrated banking systems are shown to be prone to larger systemic risk. Gai and Kapadia [8] using a network model of a banking system study how the probability and potential impact of contagion is influenced by aggregate and idiosyncratic shocks, network structure and liquidity. The authors agree with Haldane (2009) concerning the "robust-yet-fragile" property that the financial system exhibit. Even when the probability of contagion is very low, its effects can have tremendous consequences to the financial system. Higher connectivity may reduce the probability of default when contagion has not started yet but it may also increase the probability of having large default cascades when contagion begins. May and Arinaminpathy [9] apply an Erdos–Renyi network structure of which they build on the models of Nier et al. [7] and Gai and Kapadia [8] and study the interplay between the characteristics of individual banks and the overall behavior of the network. The authors consider that banks interact through different asset classes and study contagion between those asset classes. May and Arinaminpathy [9] find that increasing the level of connectivity is beneficial only when the initial shock has been caused by a default on interbank loans. However, by contrast, the opposite holds in case of liquidity shocks since they do not experience attenuation and for a given asset class, they tend to grow as more and more banks hold the failing asset. Finally, the authors emphasize the importance of having large capital buffers that will make for greater robustness both of individual banks and of the system as a whole. Finally, Amini et al. [10] test the impact of heterogeneity in an interbank network structure and the relation between resilience and connectivity using three different network models; a scale-

free network with equal and heterogeneous weights and an Erdős–Rényi network with equal weights. The main result of this study is that the most heterogeneity is introduced, the least the resilience of the network.

3 Erdős–Rényi Random Graph Model

The random graph model which is one of the earliest theoretical network models was introduced by Erdős and Rényi (1960). In this random graph, each possible link between any two nodes can occur with a certain independent and identical probability, p. This model is typically denoted G(n, p) and has two parameters: n the number of vertices and p, the probability that each simple edge (i, j) exists, which is constant for each pair nodes.

The adjacency matrix of a random graph is given by

$$\forall i > j, A_{ij} = A_{ji} = \begin{cases} 1, & \text{edge (i, j) exists; prob (p)} \\ 0, \text{edge (i, j) does not exist; prob } & (1 - \text{p}) \end{cases}$$

In other words, each edge is included in the graph with probability p, independent from every other edge. The probability to create randomly a graph with n nodes and m edges is given by $p^m(1-p)^{\binom{n}{2}-m}$. Furthermore, the probability p serves as the parameter of our model and as p increases, the graph is more likely to have more edges.

The restriction of $i > j$ appears because edges are undirected or to put it differently, the adjacency matrix is symmetric across the diagonal, and there are no self loops. In the network there are n (n − 1) possible links to be created, resulting in an expected number of edges in the network equal to pn (n − 1), so that the (expected) average degree is p(n − 1). Thus, the degree distribution of such a graph is given by

$$p(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \tag{1}$$

The mean degree, c, in the G(n,p) graph model is given by

$$c = (n-1)\, p \tag{2}$$

In other words, each vertex has (n-1) possible partners and each of these exist with the same independent probability p. Asymptotically, as $n \to \infty$, the degree distribution of a random graph converges to a Poisson (c) distribution

$$p(k) = \frac{e^{-c}c^{k}}{k!} \tag{3}$$

Due to the above property, the Erdős–Rényi random graph model is sometimes referred as Poisson random graph or random graph. The Erdős and Rényi (1960) graph model results in networks with small diameters and short average path lengths, capturing very well the "small-world" property, observed in many real networks. The clustering coefficient of an Erdős–Rényi graph model is equal to the probability of an edge's existence between two nodes, p. The Erdős and Rényi (1960) random graph model is a model in which has been extensively applied for the study of contagion in financial networks, e.g. in the contributions from Iori et al. [6], Nier et al. [7], Gai and Kapadia [8] and Montagna and Kok (2013).

In an Erdős –Rényi model we begin with n isolated nodes as presented in the first snapshot in Figure 1. Then, with probability p > 0 each pair of nodes is connected by a link. Therefore, in this model the network is determined only by the number

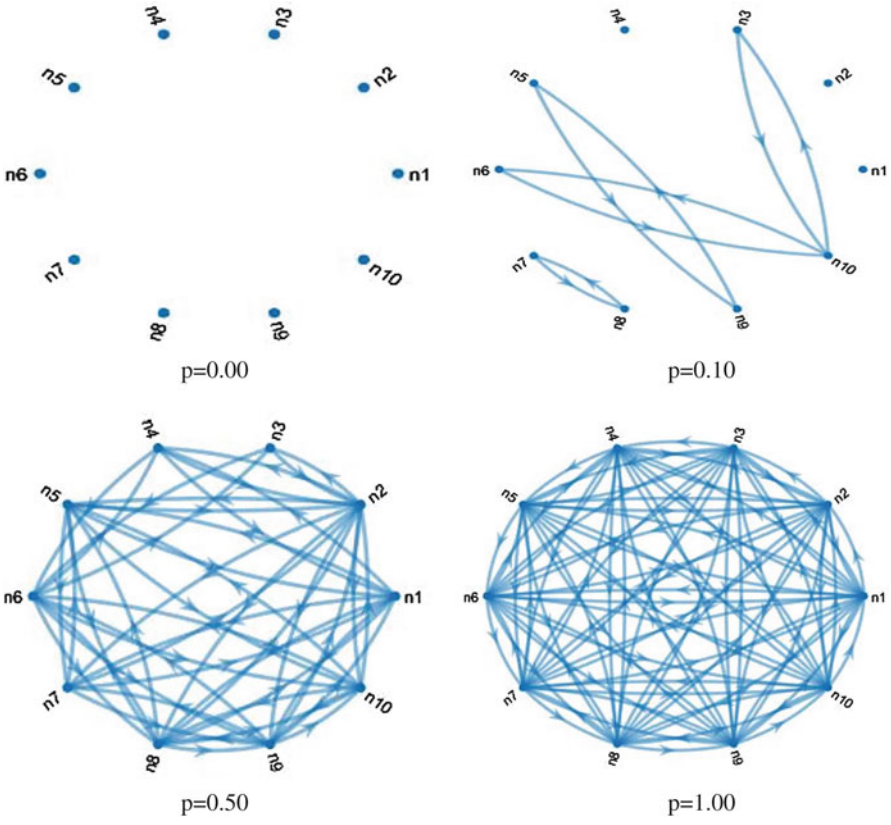


Fig. 1 Erdős–Rényi random networks: Erdős–Rényi random networks with ten nodes and different probabilities of connecting a pair of nodes

of nodes, n, and edges, m, and usually an Erdős–Rényi random graph is written as G(n, m) or G(n, p). In Figure 1 we present some examples of Erdős–Rényi random graphs with the same number of nodes and different linking probabilities. It is easy to understand that if we repeat the process for the same number of nodes and the same probability, we will not necessarily get the same network.

However, a number of alternatives models have been recently developed that differ in the probability law governing the distribution of links between nodes. Since, the Erdős–Rényi probability, p, is assumed to be equal and constant across all pairs of nodes, the resulting network structure does not present marked heterogeneity. Thus, modeling interbank networks using the Erdős–Rényi structure fails to mimic the heterogeneity observed in real interbank network systems.

In order to fully understand the heterogeneity of an Erdős–Rényi random network, we now consider one particular random realization of an Erdős–Rényi random network with 1000 nodes and $p = 0.04$, that is $G(n = 1000, p = 0.04)$ and plot the probability p(k) of finding a node of degree k, versus the degree, we obtain Figure 2, where it can be seen that the maximum of the distribution is about the value $k = (n - 1)\, p = 39$. Obviously, the probability p(k)follows a binomial distribution of the form represented in Equation (1). As we explained above, for large values of n, the degree distribution of a random graph converges to a Poisson (c) distribution. Figure 2 displays the heterogeneity plot for G(1000, 0.04), where two characteristic features of the Erdős–Rényi networks are observed. The first is a typical dispersion of the points around the value $x = 0$, and the second is the very small value of $\rho(G)$, which in this case is 0.0066.

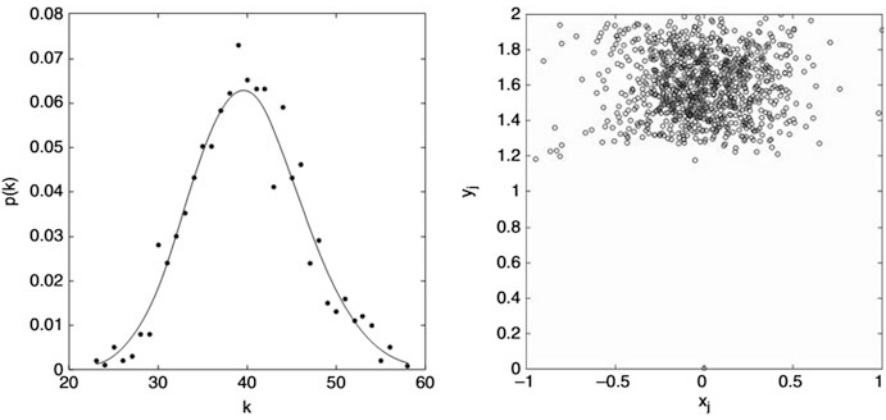


Fig. 2 Heterogeneity of Erdős–Rényi random networks. A typical Poisson degree distribution of an Erdős–Rényi random network with 1000 nodes and $p = 0.04$ (left), and the characteristic heterogeneity plot for the same network. (Source: Estrada [22])

3.1 The Mathematical Description of the Contagion Model

In this section we study the case of an Erdős–Rényi network model in which, as we stated earlier, all nodes have the same probability of being connected to another node in the network. Our model is tailored to simulate default cascades triggered by an exogenous shock in an interbank network as in Leventides et al. [5]. We first introduce the interbank network model, describe the default cascades initiated by a random negative shock on this network and analyze the parameters that affect interbank contagion.

3.2 The Interbank Network

As in Leventides et al. [5], we assume that the banking system contains $i = 1,..., N$ banks. Every bank has its own balance sheet and the accounting equation holds at all times. Total assets are divided in three categories: interbank assets A_i^{IB}, other assets A_i^{OT} and cash reserves C_i. On the liabilities side of the balance sheet we have included: interbank liabilities L_i^{IB}, other liabilities L_i^{OT} and equity capital E_i. A schematic overview of the balance sheet is given in Table 1. Although the proposed balance sheet structure does not capture all elements of a bank balance sheet, it includes all those positions that are relevant to our study.

We introduce a standard notation for our model and we define a simple interbank network as $G = (V, E)$, where V represents the nodes of the graph while E represents the edges. We further consider A, the adjacency matrix of the graph, defined as

$$\forall i > j,\ A_{ij} = A_{ji} = \begin{cases} 1, & \text{edge (i, j) exists} \\ 0, & \text{edge (i, j) does not exist} \end{cases}$$

The uth row or column of A has k_u entries, where k_u is the *degree* of the node u, which is simply the number of nearest neighbours that u has. Denoting by 1 a $|V| \times 1$vector, the column vector of node degrees $\kappa$is given by

Table 1 Stylized Balance sheet structure

Assets A_i	Liabilities L_i
Interbank Assets $\left(A_i^{IB}\right)$	Interbank Liabilities $\left(L_i^{IB}\right)$
Other Assets $\left(A_i^{OT}\right)$	Other Liabilities $\left(L_i^{OT}\right)$
Cash (C_i)	Equity Capital (E_i)

The table presents a stylized balance sheet structure in the interbank network. Total assets are divided in three categories: Interbank assets $\left(A_i^{\mathrm{IB}}\right)$, other assets $\left(A_i^{OT}\right)$, and cash reserves(C_i). Total liabilities include: Interbank liabilities$\left(L_i^{IB}\right)$, other liabilities$\left(L_i^{OT}\right)$, and equity capital (E_i). It is assumed that the accounting equation holds at all times

$$\kappa = \left(1^T \mathrm{A}\right)^T = \mathrm{A}^T 1 \tag{4}$$

We define the *indegree* as the number of links pointing toward a given node, and the *outdegree* as the number of links departing from the corresponding node. Specifically:

$$\kappa^{in} = \left(1^T \mathrm{A}\right)^T = \mathrm{A}^T 1 \tag{5}$$

$$\kappa^{out} = \mathrm{A}1 \tag{6}$$

Thus, our interbank network of credit exposures between n banks can be visualized by a graph $G = (V, E)$ where V represents the set of financial institutions—nodes, and E is the set of the edges linking the banks, that is, the set of ordered couples$(i, j) \in V \times V$ indicating the presence of a loan made by bank i to bank j. The number of nodes defines the size of the interbank network. Every edge (i, j) is weighted by the face value of the interbank claim and the representation of interbank claims is made by a single weighted $N \times N$ matrix X:

$$\mathrm{X} = \begin{bmatrix} 0 & \cdots & x_{1j} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ x_{i1} & \cdots & 0 & \cdots & x_{iN} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nj} & \cdots & 0 \end{bmatrix}$$

where x_{ij} is the credit exposure of bank i vis-à-vis bank j and N is the number of banks in the network. Interbank assets are represented along the rows while columns represent interbank liabilities. Once X is in place, the interbank entries of each bank are given according to the following rules:

(i) $A_i = \sum_{j=1}^{N} x_{ij}$ (horizontal summation), where A_i is the total interbank assets of bank i.

(ii) $L_i = \sum_{i=1}^{N} x_{ij}$ (vertical summation), where L_i is the summation of the total interbank liabilities of bank j.

One can observe that the diagonal line contains zeros due to the fact that banks do not lend to themselves. In this framework, a random network is generated based on two parameters, the size of the network (number of nodes/banks) and the probability $\mathrm{p_{ij}}$ that there is a lending/borrowing link between two nodes/banks. Thus, each possible link between two nodes exists with an independent and identical probability, which is often called the Erdős–Rényi probability.

Although, we have undirected edges in this framework, we cannot really speak of undirected links, since the two directions of the same link are given different weights.

3.3 Shock Propagation and Contagion Dynamics

The failure of a bank can affect other banks through their interbank connections. Below, we describe the mechanism through which an initial shock affecting a bank propagates onto its counterparties along the network. Contrary to the recent literature, the term contagion here translates into total capital losses due to multiple default cascades. The cascade dynamics we use in this study are straightforward to implement and enable us to run a great number of simulations on a variety of different scenarios (Table 2).

The default procedure starts with an exogenous shock being simulated, typically by setting to zero the equity of one randomly chosen bank i and the cascade of defaults proceeds on a timestep-by-timestep basis, assuming zero recovery for shock transmissions. The zero recovery assumption, which is a realistic one in the short run, is often used in the literature to analyze worst case scenarios and

Table 2 OLS regression analysis for Scenario 1 (Heterogeneous banks with homogeneous exposures)

	N = 20	N = 50	N = 80	N = 100
CATIN 1	0.051 (16.198)***	−0.002 (−0.459)	−0.001 (−0.347)	−0.007 (−2.044)**
CATIN2	0.098 (4.195)***	0.004 (0.170)	0.179 (8.073)***	0.104 (5.059)***
LEVIN	0.389 (17.018)***	0.413 (19.043)***	0.260 (12.205)***	0.315 (15.935)***
NOUTGOING	−0.080 (−3.915)***	0.097 (2.773)***	−0.170 (−4.933)***	−0.053 (−1.534)
COUNT	0.602 (138.571)***	0.572 (134.093)***	0.576 (136.735)***	0.540 (124.326)***
VARCAP	−0.088 (−53.348)***	−0.075 (−61.005)***	−0.053 (−53.890)***	−0.054 (−57.165)***
P	−0.101 (−5.089)***	−0.080 (−2.338)**	0.165 (4.885)***	0.107 (3.148)***
Adjusted R^2	0.800	0.763	0.756	0.749

The table presents the regression results for Scenario 1. The dependent variable is CATEND measured as the total loss of capital due to contagion as percentage of total capital in the network. Explanatory variables are, CATIN1, CATIN2, LEVIN, NOUTGOING, COUNT, VARCAP and P, the probability for a link to exist between two nodes. Each cell displays the OLS standardized coefficients along with the corresponding t-statistics (shown in parentheses). The sample comprises of 18,000 realizations (simulated banking crises). *, ** and *** denote significance at the 10, 5 and 1 percent level, respectively

refers to a situation where creditor banks lose all of their interbank assets held against a defaulting bank [8, 20]. A bank's default implies that it is no longer able to meet its interbank liabilities to its counterparties. Since these liabilities constitute other banks' assets, the banks that get into trouble affect simultaneously their counterparties, leading to write-downs in their balance sheets. The interbank asset loss due to failure of bank *i* is subtracted from the bank's *j* capital. Bank *j* will fail if its exposure against bank *i* exceeds its equity. A second round of bank failure occurs if bank creditors cannot withstand the losses realized due to its default and eventually, contagion stops if no additional bank goes bankrupt, otherwise a third round of contagion takes place. An initial shock can be amplified through banks' interconnections and further transmitted to other institutions, such that the overall effect on the system goes largely beyond the original shock. As Upper and Worms (2004) demonstrate, in response to a liquidity shock banks prefer to withdraw their deposits at other banks instead of liquidating their long-term assets, creating further instability and liquidity dry-ups in the financial system.

A general mathematical description of the dynamical system expressing the shock propagation mechanism is presented hereafter. We consider a network consisting of N banks numbered from 1 to N. We define b_i as the capital possessed by bank *i* in the network and

$$b_0 = (b_1, b_2, \ldots, b_N) \tag{7}$$

stands for the initial vector of bank capital. X is defined as a $N \times N$ matrix with entries:

x_{ij}= the credit exposure of bank *i* vis-à-vis bank j in the network

$$x_{i\,i} = b_i \tag{8}$$

We consider the case where some of the banks (one or more) collapse. We wish to study how the crisis travels through the bank network and when exactly it comes to a fixed point. The collapse of banks $i_1, i_2, \ldots, i_k$ (where $k \leq N$), can be described in the following way. Consider the element $x_0 \in Z_N^2 = \{0, 1\}^N$ which has zero entries everywhere except the positions $i_1, i_2, \ldots, i_k$ where x_0 takes on the value 1. Then,

$$b_1 = b_0 - \mathrm{X} \cdot x_0 \tag{9}$$

is the new vector of capital of the N banks. We now take

$$x_1(i) = \begin{cases} 1, b_1(i) \leq 0; \\ 0, b_1(i) > 0. \end{cases} \tag{10}$$

Then $x_1 \in Z_2^N$ and x_1 indicates the banks that have collapsed after the bankruptcy of the first k banks. The vector x_1 takes on the value 1 in the positions i_1, i_2, ..., i_k . If $x_1 \neq x_0$, this indicates that the collapse of the first k banks has adversely affected other banks leading them to bankruptcy. Similarly, from x_1 we take:

$$b_1 = b_0 - \mathrm{X} \cdot x_1 \tag{11}$$

and then

$$x_2(i) = \begin{cases} 1, b_2(i) \leq 0; \\ 0, b_2(i) > 0. \end{cases} \tag{12}$$

The vector x_2 indicates the banks that collapse after the bankruptcy of the banks of x_1. Therefore, we have a map:

$$F : Z_2^{\mathrm{N}} \to Z_2^{\mathrm{N}} \tag{13}$$

$$x \to \mathrm{F}(x) = f\,(b_0 - \mathrm{X} \cdot x) \tag{14}$$

The map F(x) defines a dynamical system $x_{n+1} = \mathrm{F}(x_n)$ which describes the evolution of contagion in the interbank network.

3.4 Monte Carlo simulations

In this section we apply Monte Carlo simulations in four different stages. As in Leventides et al. [5], we introduce randomness in three areas: amount of capital, interbank claims and network structure. The stochasticness introduced in our model provides us with a wide vicinity of scenarios that may come across in real world. Using the Erdős–Rényi network structure, the degree distribution or the connectivity among banks can vary with respect to the chosen probability p. Thus, each random network generated with the same parameters N, p looks slightly different.

The second stage involves estimating the parameters of interest, i.e. the value of the coefficients in the regression model. In the third stage the test statistics of interest are saved, while in the fourth stage we go back to the first stage and repeat N times. The quantity N is the number of replications which should be as large as is feasible. As Monte Carlo is based on random sampling from a given distribution (with results equal to their analytical counterparts asymptotically), setting a small number of replications will yield results that are sensitive to odd combinations of random number draws. Generally speaking, the sampling variation is measured by the standard error estimate, denoted $S_x = \sqrt{\mathrm{var}(x)/N}$, where x denotes the value of

the parameter of interest and var(x) is the variance of the estimates of the quantity of interest over the N replications.

Similar to Leventides et al. [5], we consider four different scenarios, in line with Chinazzi et al. [20], where we let vary the balance sheet composition, the size of the network and the link probability among banks which is held constant for each pair of nodes. The four scenarios tested are as follows:

Scenario 1:	• **Heterogeneous banks with homogeneous exposures**. In this scenario, we construct interbank networks where banks have different equity size and their interbank claims are evenly distributed across the outgoing links
Scenario 2:	• **Heterogeneous banks with heterogeneous exposures**. In this scenario, the interbank networks allow for heterogeneous bank sizes and heterogeneous interbank claims among banks.
Scenario 3:	• **Homogeneous banks with heterogeneous exposures**. In this scenario, we construct interbank networks where banks have the same equity size and unevenly distribute their exposures across creditor banks
Scenario 4:	• **Homogeneous banks with homogeneous exposures**. In this last scenario, we construct interbank networks where banks have the same equity size and interbank claims are evenly distributed across creditor banks

In each case, we do not control the number of outgoing links as in Leventides et al. [5] but for each network that is generated a random probability, which is constant for each pair of nodes, defines the lending/borrowing relation of each bank. The probability p_{ij} is assumed to be equal and constant across all pairs (i,j). For simplicity, we denote the probability, termed as the Erdős–Rényi probability, by p. Since the probability of forming a link is homogeneous, the resulting network structure does not present marked heterogeneity.

We examine banking systems consisting of small banks with low, medium and large interbank exposures, as well as systems of large banks with corresponding exposure levels. We consider a basic model that uses only two components from a bank's balance sheet, that is, equity and interbank loans–in the words of May and Arinaminpathy [9] '*a caricature for banking ecosystems'*. We generate our model in two separate steps. First, we construct a model structure of N nodes representing the banks in our system and randomly choose the probability p of forming a link between each of the $\binom{N}{2}$ possible links.

For all the possible couples of nodes, a link is created with probability p which represent lending/borrowing relationship, while in a second step, we assign each node to a stylized balance sheet structure. Once the banking networks are created, the default propagation dynamics are implemented to examine the effects of an idiosyncratic shock hitting one bank. The effect of a shock is simulated, typically by setting to zero the equity of the affected bank. We estimate the initial loss of capital by letting the first bank default and subsequently record the loss as percentage of the total capital in the system. Consequently, the defaulted bank will be unable to

repay its creditors and the interbank loans that were granted will be written-off, as we have selected to work under a zero recovery assumption. This bad debt will be recorded and expressed as percentage of the total capital in the system. Moreover, the creditors of the defaulted bank will experience a shock on their balance sheets and the recorded losses will be subtracted from their equity.

If at any time the total losses realized by a bank exceed its net worth, the bank is deemed in default and is removed from the network. Note that time steps are modeled as being discrete and there is the possibility that many banks default simultaneously in each timestep. These shocks propagate to their creditors and take effect in the next timestep. When no further failures are observed, the default procedure terminates and various contagion indicators[1] are calculated based on the contagion map as described in Subsection 3.3.

4 Main Findings

This section discusses the main findings of this study. Subsection 4.1 describes in full detail the computer experiments conducted while Subsection 4.2 discusses the simulation results of all four scenarios considered.

4.1 Computer Experiments

Having generated banking systems via an Erdős–Rényi network structure framework and balance sheet allocation, the dynamics of an initial shock affecting a bank within the interbank network can be investigated. Given the complexity of the interbank network outlined above, it is extremely difficult to derive analytical solutions. In order to obtain data to describe the variables that affect contagion, we employ several Monte Carlo simulations. In each realization, we construct an interbank network with $N \in [20, 50, 80, 100]$ nodes under the rewiring process of the Erdős–Rényi methodology. In a second step, we test the four scenarios mentioned before by varying the equity size of banks and the interbank exposure structure across creditor banks. For each scenario tested we check a wide range of link probabilities, such that we can observe dense or sparse interbank network systems. Since the probability of forming a link is homogeneous, the resulting network structure does not present marked heterogeneity.

When homogeneity across bank sizes is considered, all banks are assumed to have the same equity size and thus, each bank is endowed with a balance sheet that

[1] We refer the interested reader to Appendix in Leventides et al. [5] for a formalization of the aforementioned mechanism in a pseudocode which simulates the default cascade in the interbank network.

consists of 100 units of equity. On the other hand, when homogeneity is present with respect to interbank exposures, interbank claims are randomly allocated within the interbank network and are categorized as follows: small loans granted (4 units), medium loans (8 units) and large loans (14 units). With respect to scenarios tested where heterogeneity of bank size is introduced, the amount of equity of each bank is drawn from a uniform distribution in the range: $b_i \in [0, 100]$, whereas when heterogeneity is introduced with respect to interbank claims, credit is allocated in the following ranges: $a_{ij} \in [0, 4]$, $a_{ij} \in [0, 8]$, $a_{ij} \in [0, 14]$.[2] Interbank exposures are set 60% lower than these in Leventides et al. [5]. This is due to the fact that we cannot control the connectivity across banks since the link probability in randomly selected. The interbank exposure decrease was set by trial and error in order not to observe enormous high leveraged systems. In addition, we control the leverage of the system by setting the rule that the maximum leverage ratio of each network system cannot exceed five. Then, balance sheets are assigned to each node, consistent with each specific scenario tested. We randomly choose a single bank in the system to default due to an exogenous shock and the default cascades proceed sequentially, assuming zero recovery. When no further failures are observed results are recorded before another realization begins. We then impose another shock on the second bank in the network and this procedure continues until all banks in the interbank network are hit by an exogenous shock.

For each scenario tested and for each network size we have three cases in which we allow the weight of outgoing links (small, medium and large interbank claims)to vary among banks. Each case gives us 6000 realizations or, to put it differently, 6000 banking crises. We deem that 6000 realizations provide a satisfactory number of runs and robustness to our analysis. Thus, for each scenario tested and each network size we employ $6000 \times 3 = 18{,}000$ realizations using the following variables in each realization:

- Total loss of capital due to contagion as percentage of total capital in the system (**CATEND**)
- Initial loss of capital by defaulting bank *i* as percentage of total capital in the system (**CATIN1**), i.e. bank's *i* depleted equity divided by the total equity in the network
- Loss of capital at the first stage (interbank loans that cannot be repaid) by defaulting bank *i* as percentage of total capital in the system (**CATIN2**), i.e. total amount of loans granted to bank *i* that cannot be repaid divided by the total equity in the network
- Leverage of the interbank network(**LEVIN**), i.e. total interbank exposures as measured by the sum of matrix's A elements, divided by the total capital in the network

[2] Although those ranges have been selected arbitrarily, they are not sensitive to any regression model employed in the following analysis and thus, our regression results will be unaffected from a qualitative point of view if different ranges are used.

- Number of outgoing links of bank *i* (**NOUTGOING**), i.e. the outdegree of a bank i which corresponds to the number of creditors in the network. It is defined as the summation of the ith column of the adjacency matrix A.
- Shock propagation variable (**COUNT**) which measures the number of rounds needed until no further bank defaults
- Variance of capital (equity) (**VARCAP**) used in those scenarios tested where only heterogeneous bank sizes are considered
- Variance of interbank loans (**VARLOANS**) used in those scenarios tested where only heterogeneous interbank loan exposures are considered
- Erdős–Rényi probability p_{ij} **(p)** that there is a lending/borrowing link between two nodes/banks.

Our selection of variables is motivated by economic intuition and by the findings of previous studies on the dynamics of systemic risks [7] and Leventides et al. [5]. Table 3 presents summary statistics on these variables. In order to study the effect the aforementioned variables have on contagion risk, we estimate the following ordinary least squares (OLS) models:

$$CATEND = \beta_1 CATIN1 + \beta_2 CATIN2 + \beta_3 LEVIN + \beta_4 NOUTGOING + \beta_5 COUNT + \beta_6 VARCAP + \beta_7 p \tag{15}$$

$$CATEND = \beta_1 CATIN1 + \beta_2 CATIN2 + \beta_3 LEVIN + \beta_4 NOUTGOING + \beta_5 COUNT + \beta_6 VARCAP + \beta_7 VARLOANS + \beta_8 p \tag{16}$$

$$CATEND = \beta_1 CATIN2 + \beta_2 LEVIN + \beta_3 NOUTGOING + \beta_4 COUNT + \beta_5 VARLOANS + \beta_6 p \tag{17}$$

$$CATEND = \beta_1 CATIN2 + \beta_2 LEVIN + \beta_3 NOUTGOING + \beta_4 COUNT + \beta_5 p \tag{18}$$

The model described in Equation (15) is applied to scenarios involving heterogeneous bank sizes with homogeneous exposures in the network structure, Equation (15) refers to a situation where emphasis is placed on heterogeneous interbank loan exposures combined with heterogeneous bank sizes, Equation (17) takes into account homogeneous banks with heterogeneous exposures while Equation (18) considers only homogeneous bank sizes and interbank claims. The variable CATIN1 has been omitted from Eqs. (17) and (18) due to the fact that banks in the interbank system are homogeneous, i.e. we keep constant the equity of each bank and thus CATIN1 remains stable during our simulation runs. There is an explanation in the next subsection concerning the fact that in our experiments we have selected to

Table 3 Summary statistics

		Heterogeneous banks–homogeneous exposures			Heterogeneous banks–heterogeneous exposures			Homogeneous banks–heterogeneous exposures			Homogeneous banks–homogeneous exposures		
	Variable	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
n = 20 banks	CATEND	0.241	0.067	0.368	0.185	0.063	0.313	0.220	0.050	0.350	0.251	0.050	0.367
	CATIN1	0.050	0.050	0.028	0.050	0.050	0.029	0.050	0.050	0.000	0.050	0.050	0.000
	CATIN2	0.079	0.063	0.062	0.070	0.052	0.067	0.147	0.130	0.108	0.175	0.160	0.124
	COUNT	2.624	1.000	2.342	2.278	1.000	1.999	1.916	1.000	1.893	2.029	1.000	1.876
	LEVIN	1.572	1.273	1.193	1.406	1.081	1.241	2.935	2.679	1.888	3.510	3.510	2.046
	P	0.491	0.486	0.286	0.459	0.438	0.278	0.402	0.371	0.253	0.276	0.214	0.226
	NOUTGOING	9.303	9.000	5.729	8.701	8.000	5.595	7.601	7.000	5.139	5.228	4.000	4.603
	VARCAP	829.371	824.415	171.611	836.673	832.512	172.930	–	–	–	–	–	–
	VARLOANS	–	–	–	38.407	8.230	45.561	47.449	33.195	40.430	–	–	–
n = 50 banks	CATEND	0.364	0.034	0.459	0.352	0.033	0.453	0.177	0.020	0.348	0.215	0.020	0.367
	CATIN1	0.020	0.020	0.011	0.020	0.020	0.011	0.020	0.020	0.000	0.020	0.020	0.000
	CATIN2	0.047	0.044963	0.029	0.046	0.045	0.030	0.068	0.063	0.047	0.068	0.064	0.051
	COUNT	4.078	2.000	3.473	4.026	2.000	3.532	2.209	1.000	2.618	2.451	1.000	2.767
	LEVIN	2.330	2.320	1.365	2.310	2.336	1.347	3.380	3.392	1.996	3.401	3.240	2.002
	NOUTGOING	16.754	14.000	12.604	18.478	14.000	14.272	10.092	8.000	9.003	5.025	4.000	4.562
	P	0.342	0.290	0.250	0.377	0.270	0.286	0.206	0.159	0.176	0.103	0.079	0.084
	VARCAP	834.415	832.740	106.833	830.829	829.254	115.273	–	–	–	–	–	–
	VARLOANS	–	–	–	38.207	8.299	44.715	47.688	33.015	39.929	–	–	–
n = 80 banks	CATEND	0.383	0.022	0.469	0.359	0.021	0.460	0.180	0.012	0.362	0.198	0.012	0.3676
	CATIN1	0.012	0.012	0.007	0.012	0.012	0.007	0.012	0.012	0.000	0.012	0.012	0.000
	CATIN2	0.031	0.030	0.020	0.029	0.028	0.020	0.045	0.042	0.031	0.040	0.035	0.032

(continued)

Table 3 (continued)

		Heterogeneous banks–homogeneous exposures			Heterogeneous banks–heterogeneous exposures			Homogeneous banks–heterogeneous exposures			Homogeneous banks–homogeneous exposures		
	Variable	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
	COUNT	4.638	3.000	4.259	4.662	2.000	4.442	2.359	1.000	2.906	2.455	1.000	2.882
	LEVIN	2.461	2.429	1.464	2.365	2.297	1.430	3.623	3.595	2.099	3.209	3.120	2.065
	NOUTGOING	17.752	15.000	14.422	19.403	14.000	16.368	10.829	8.000	9.556	4.860	3.000	4.777
	P	0.225	0.196	0.178	0.246	0.164	0.203	0.137	0.104	0.115	0.061	0.046	0.055
	VARCAP	820.667	820.990	85.218	822.902	815.162	86.217	–	–	–	–	–	–
	VARLOANS	–	–	–	38.955	8.376	46.160	47.854	33.478	39.566	–	–	–
n = 100 banks	CATEND	0.370	0.017	0.468	0.382	0.018	0.468	0.149	0.010	0.332	0.220	0.010	0.385
	CATIN1	0.010	0.010	0.006	0.010	0.010	0.006	0.010	0.010	0.000	0.010	0.010	0.000
	CATIN2	0.025	0.024	0.015	0.025	0.025	0.017	0.035	0.033	0.024	0.036	0.032	0.026
	COUNT	4.649	3.000	4.261	4.823	3.000	4.291	2.383	1.000	3.212	2.830	1.000	3.469
	LEVIN	2.474	2.458	1.423	2.555	2.629389	1.515	3.539	3.615	2.006	3.600	3.552	1.967
	NOUTGOING	18.732	14.000	15.964	21.815	15.00000	18.283	10.898	8.000	9.488	5.267	4.000	4.565
	P	0.189	0.142	0.157	0.220	0.141615	0.181	0.110	0.079	0.092	0.053	0.041	0.040
	VARCAP	822.626	821.214	80.128	827.379	826.1881	76.812	–	–	–	–	–	–
	VARLOANS	–	–	–	37.755	8.359417	44.330	48.180	33.479	39.959	–	–	–

The mean, median, and standard deviation are depicted for interbank networks consisting of 20, 50, 80, and100 banks, respectively. Four scenarios are included: (a) Heterogeneous banks–homogeneous exposures; (b) Heterogeneous banks-heterogeneous exposures; (c) Homogeneous banks–heterogeneous exposures; (d) Homogeneous banks–homogeneous exposures. The variables are: CATEND, defined as total loss of capital due to contagion as percentage of total capital in thesystem;CATIN1, defined as bank's i depleted equity divided by the total equity in thenetwork;CATIN2, defined as the total amount of loans granted to bank i that cannot be repaid, divided by the total equity in the network; LEVIN, defined as the leverage of the interbank network; NOUTGOING, defined as the number of outgoing links of bank i, which corresponds to the number of its creditors in the network; COUNT, defined as the number of rounds needed until no further bank defaults; VARCAP, defined as the variance of bank capital; VARLOANS, defined as the variance of interbank loans and p, the Erdős–Rényi probability p_{ij} that there is a lending/borrowing link between two nodes/banks

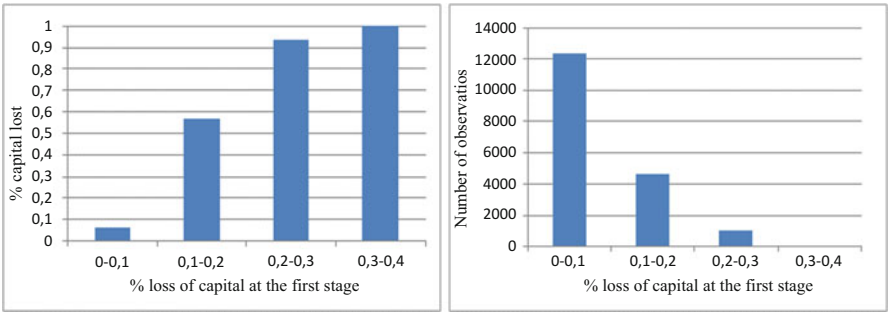
work with standardized variables—both dependent and independent variables—and have not included the intercept term in the regression models as it will be zero. Our concern is to measure effects not in terms of the original units of the dependent variable or the independent variables, but in standard deviation units.[3]

4.2 Simulation Results

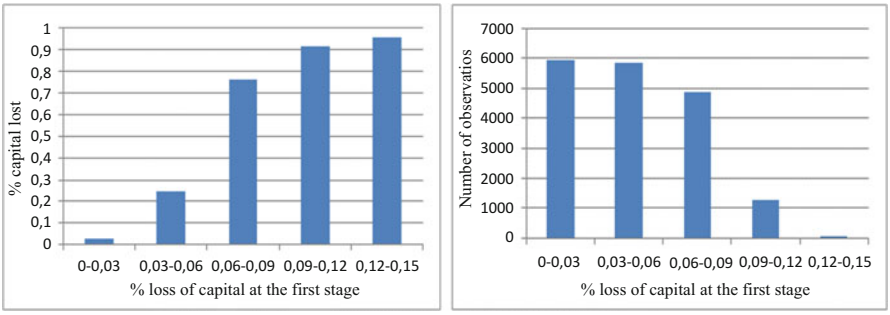
In this section, we discuss the regression results of all four scenarios. Since our variables are measured on different scales, we cannot directly infer which independent variable has the largest effect on the dependent variable. In order to circumvent this problem we standardize our series to have zero mean and unit variance. Table 2 presents the regression results of the **first scenario** using the OLS model described in Equation (15), where heterogeneous banks distribute evenly their interbank claims across the outgoing links of a network consisting of $N = 20$, 50, 80 and 100 banks. Almost all regressor coefficients are found to be statistically significant for all the sizes of the network. We discern only two cases where regressor coefficients are found to be statistically insignificant and has to do with CATIN1 variable and one case that has to do with CATIN2. *R*-squared coefficients take on large values ranging from 74.9 to 80% and highlight the ability of our selected variables to explain financial distress in interbank networks.

The variable CATIN1 captures the initial effect defaulting bank i exerts on the network, whereas the magnitude of interconnectedness across the banks that comprise the interbank network is measured through parameter CATIN2. As we observe from Table 1, variable CATIN1 does not seem to affect much the dependent variable, whereas two regressor coefficients are found to be insignificant. Financial shocks will propagate into the defaulting bank's counterparties along the network, erode their capital and make them more vulnerable to subsequent shocks. The magnitude of the positive relationship between CATIN2 and CATEND – the dependent variable - seems to increase as the size of the interbank network increases with the only exception being the $N = 50$ bank network segment which follows an autonomous path (although statistically insignificant). The increasing magnitude of the above relationship seems to cease as we move from the case of $n = 80$ banks to the case of $n = 100$ banks. This finding implies that as we move from smaller to larger network settings, systemic risk and the likelihood of contagion increases. However, when we move from the case of $n = 80$ banks to the case of $n = 100$ banks the likelihood of contagion seems to decreases. Figure 3 visually illustrates the extent of contagion as a function of the percentage loss of capital due to bank's i default. It is shown that as the network size increases from small to medium sized networks, we observe that capital losses rises, confirming the findings from the

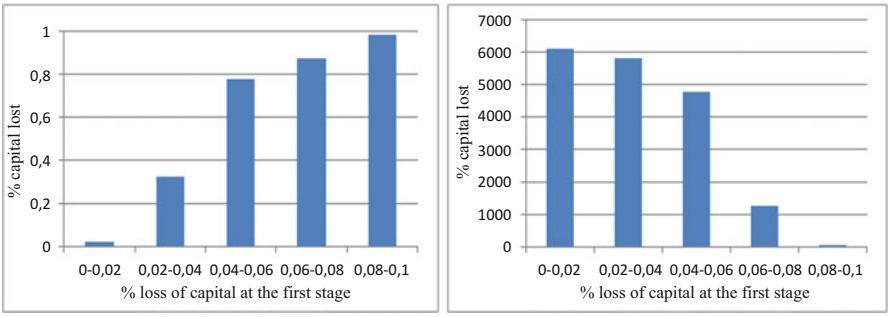
[3]See Wooldridge (2003) for an interesting discussion on standardization and explanation of the absence of the standardized intercept.



(a) N=20 banks

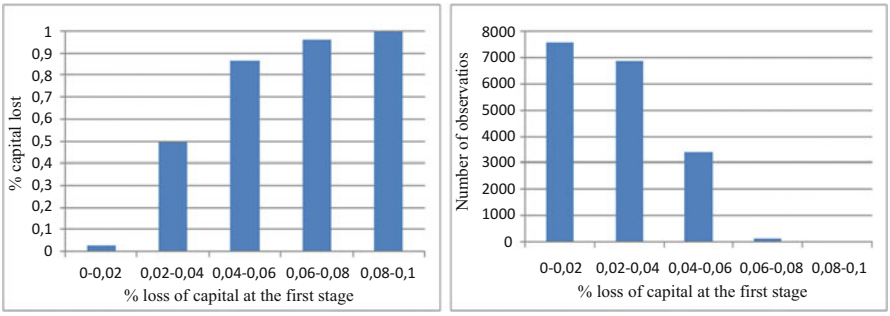


(b) N=50 banks



(c) N=80 banks

Fig. 3 Scenario 1: **Heterogeneous Banks with homogeneous exposures. Extent of contagion** (expressed as the total capital lost from the banking system due to the failure of at least one bank) **as a function of the % initial loss of capital** due to default of the first bank. Panels (**a**–**d**) show the relation between the % initial loss of capital due to default of the first bank and the extent of contagion across interbank networks with different number of banks



(d) N=100 banks

Fig. 3 (continued)

regression model. As we can observe from Figure 3, as we move from the n = 80 interbank network scheme to n = 100 the likelihood of contagion seems to decrease since we have very few cases that cause systemic break downs and defaults.

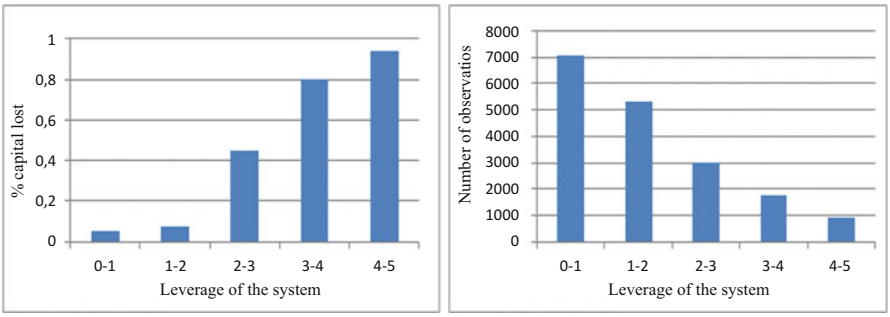
As expected, we also find that there is a positive relationship between the leverage of the network and the capital losses due to contagion, which is depicted by Figure 4. This result is in line with the findings of Nier et al. [7] who provide evidence that systemic risk increases when system-wide leverage increases. Highly leveraged banks in the interbank network are clearly more exposed to the risk of default on interbank loans. The greater the size of default on debt is, the larger the losses are that banks transmit to their neighbors, other things being equal. Thus, highly leveraged banks contribute more to systemic risk as they become a vehicle for transmitting shocks within the network. Moreover, it is shown that the magnitude of the positive relationship between the network's leverage and contagion risk increases as we move from smaller to larger interbank networks (illustrated in Table 2) with the only exception being the n = 80 bank network scheme where the magnitude of the standardized coefficients seems to decrease.

Our results also suggest that connectivity, expressed in our experiments as the outgoing[4] of the first bank that defaults, has a negative effect on interbank contagion with the only exception being the case of n = 50 banks where we can observe a positive relationship between contagious defaults and connectivity.

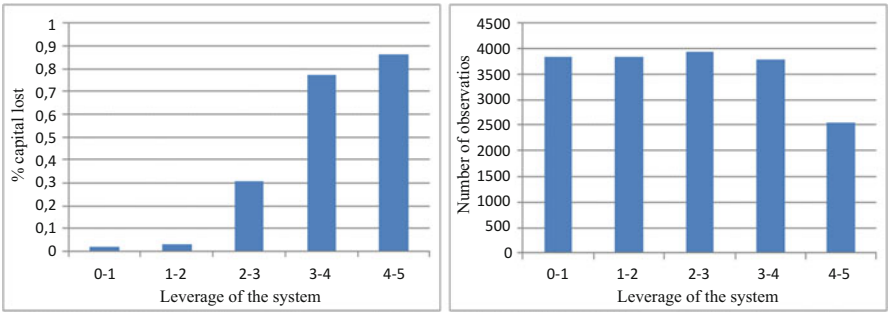
Interestingly, as we move from small networks consisted of twenty banks to networks consisted of 50 banks the effect of connectivity to interbank contagion turns from negative to positive and after then connectivity keeps affect negatively the systemic risk of the network. Thus, as we move from network systems consisted of fifty banks to networks consisted of 100 banks this negative relationship seems to decrease. In relatively small interbank networks, a high level of connectivity

[4]It should be highlighted that in the Erdős–Rényi network structure the outdegree equals the indegree since we have an undirected network structure. However, in our framework, the two directions of the same link are given different weights.

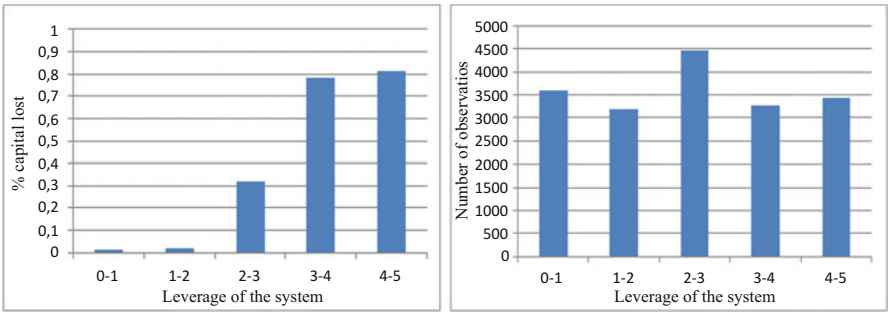
will allow an efficient absorption of shocks, whereas in medium size networks the increased connectivity will spread the shock throughout the system, potentially leading to many default cascades. The link probability, that is assumed to be equal across all pairs, seems to contribute to the resilience of the system for small and medium size networks. However, as we move from medium to large size networks



(a) N=20 banks

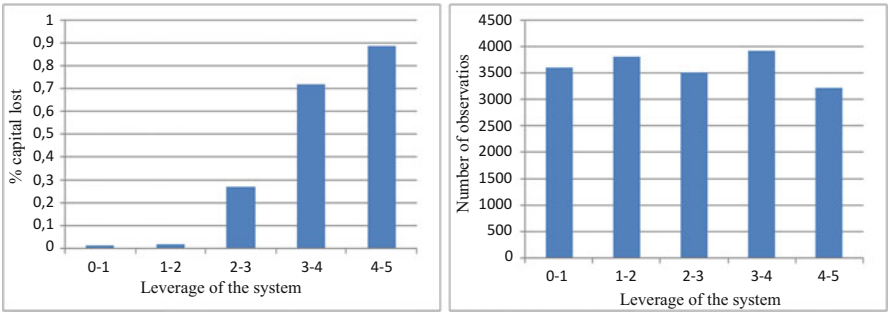


(b) N=50 banks



(c) N=80 banks

Fig. 4 Scenario 1: **Heterogeneous Banks with homogeneous exposures. Extent of contagion** (expressed as the total capital lost from the banking system due to the failure of at least one bank) **as a function of the leverage of the system**. Panels (**a–d**) show the relation between the leverage of the system and the extent of contagion across interbank networks with different number of banks



(d) N=100 banks

Fig. 4 (continued)

this effect turns negative to the resilience of the system as it seems to contribute positively to systemic risk.

Our regression analysis also shows that the COUNT variable which measures the number of rounds until no further bank defaults, has a positive impact on interbank contagion. Heterogeneity expressed as the variance of capital exhibits a negative and statistically significant relationship with interbank contagion, showing that size heterogeneity can have positive effects on the stability of an interbank network.

However, the positive magnitude seems to decrease as we move from small to large interbank networks. An interbank network consisting of banks of various sizes can more easily withstand a negative shock, therefore no institution becomes significant for either borrowing or lending. Furthermore, in such network both smaller and larger banks can act as shock absorbers when an initial shock hits the banking system, making contagion a less likely phenomenon. This finding is in line with the results of Iori et al. [6] concerning bank size heterogeneity.

Table 4 presents the regression results of the **second scenario** using the model described in Equation (16), where banking institutions with heterogeneous bank sizes are linked to one another via heterogeneous interbank claims. The regressor coefficients are statistically significant in almost all cases and the R-squared values are quite high and lie in the vicinity of 75–83%, highlighting the good explanatory power of the model.

CATIN1 does not seem to impact much the dependent variable in all network segments and the regressor coefficients in the relatively large interbank networks becomes statistically insignificant. The magnitude of standardized coefficients is almost the same with the corresponding magnitude of those derived from the first scenario. In other words, an initial shock from defaulting bank i will spill over more easily in the network. Thus, the first bank defaulting has the dynamics to spread the initial shock and contaminate the entire interbank network. CATIN2 has a large positive impact on contagion risk, however, its magnitude fades away as we move from smaller to larger networks. It should also be highlighted that the CATIN2 coefficients are much larger than those recorded in the first scenario in all network

Table 4 OLS regression analysis for Scenario 2 (Heterogeneous banks with heterogeneous exposures)

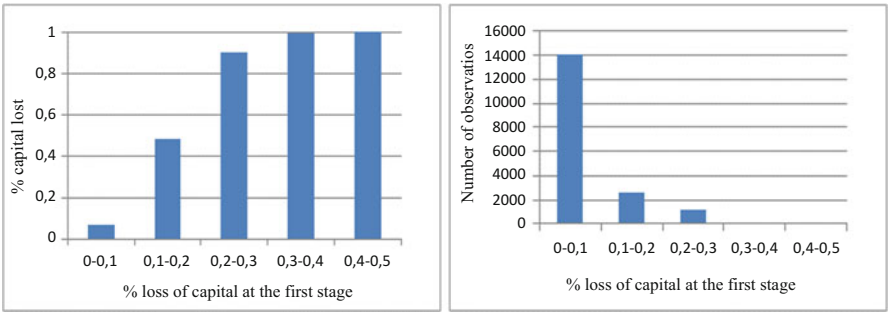
	N = 20	N = 50	N = 80	N = 100
CATIN 1	0.070 (23.660)***	0.007 (2.024)**	0.000 (0.047)	−0.001 (−0.283)
CATIN2	0.201 (19.541)***	0.113 (11.183)***	0.106 (9.669)***	0.071 (6.015)***
LEVIN	0.653 (58.484)***	0.346 (30.847)***	0.321 (26.320)***	0.399 (30.132)***
NOUTGOING	−0.136 (−11.540)***	−0.150 (−6.539)***	−0.052 (−2.253)**	0.038 (1.575)
COUNT	0.456 (111.687)***	0.630 (156.274)***	0.577 (141.939)***	0.573 (131.397)***
VARCAP	−0.032 (−18.897)***	−0.067 (−50.848)***	−0.053 (−52.027)***	−0.041 (−40.597)***
VARLOANS	−0.246 (−45.472)***	−0.091 (−14.882)***	−0.018 (−3.113)***	−0.082 (−12.307)***
P	−0.254 (−21.462)***	0.038 (1.620)	0.064 (2.678)***	−0.110 (−4.311)***
Adjusted R^2	0.830	0.796	0.776	0.751

The table presents the regression results for Scenario 2. The dependent variable is CATEND measured as the total loss of capital due to contagion as percentage of total capital in the network. Explanatory variables are the constant term CATIN1, CATIN2, LEVIN, NOUTGOING, COUNT, VARCAP, VARLOANS and P, the probability for a link to exist between two nodes. Each cell displays the OLS standardized coefficients along with the corresponding *t*-statistics (shown in parentheses). The sample comprises of 18,000 realizations (simulated banking crises). *, ** and *** denote significance at the 10, 5 and 1 percent level, respectively

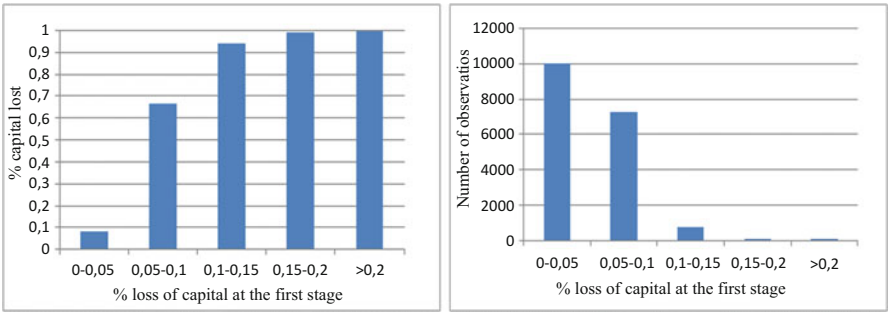
sizes. An initial shock following the default of bank *i* seems to contribute much to a banking crisis scenario within small and medium-sized networks and the size of total capital losses is smaller than that documented in the first scenario. Figure 5 depicts the extent of contagion as a function of the percentage loss of capital due to default of the first bank and confirms the results recorded in Table 5.

The results also show that there still exists a positive relationship between leverage and contagion (illustrated in Figure 6); however, the coefficient estimates are larger in almost all cases than those recorded in the previous scenario. Moreover, the magnitude of the leverage coefficients decreases as the number of banks in the interbank network increases, with the only exception being the 100 bank network segment where one can observe a slight increase compared to the 80 bank network segment.

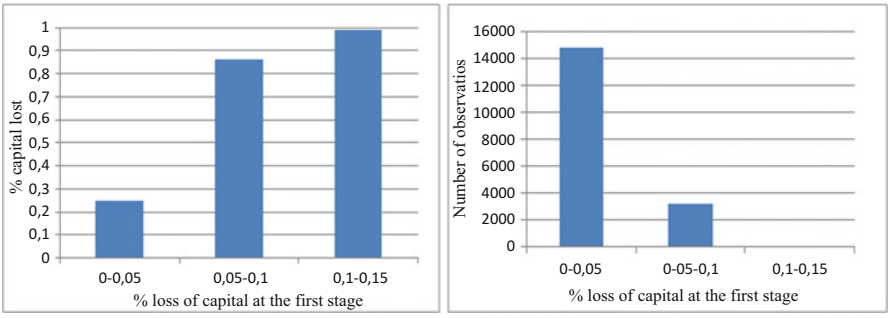
Results on connectivity are qualitatively similar to those of the first scenario, showing that connectivity negatively impacts contagion risk especially in small and medium interbank networks with the only exception being the 100 bank network segment which follows an autonomous path and is positively related to contagion (although statistically insignificant).



(a) N=20 banks

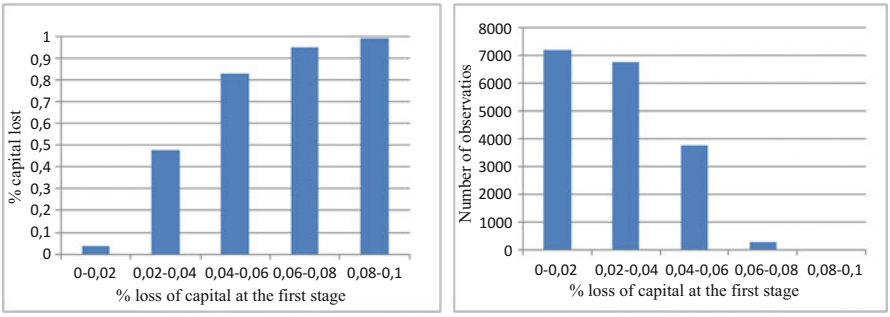


(b) N=50 banks



(c) N=80 banks

Fig. 5 Scenario 2: **Heterogeneous Banks with heterogeneous exposures. Extent of contagion** (expressed as the total capital lost from the banking system due to the failure of at least one bank) **as a function of the % initial loss of capital** due to default of the first bank. Panels (**a**–**d**) show the relation between the % initial loss of capital due to default of the first bank and the extent of contagion across interbank networks with different number of banks



(d) N=100 banks

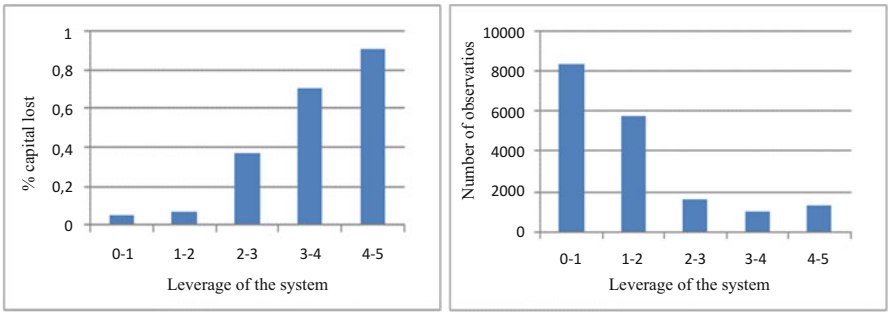
Fig. 5 (continued)

Table 5 OLS regression analysis for Scenario 3 (Homogeneous banks with heterogeneous exposures)

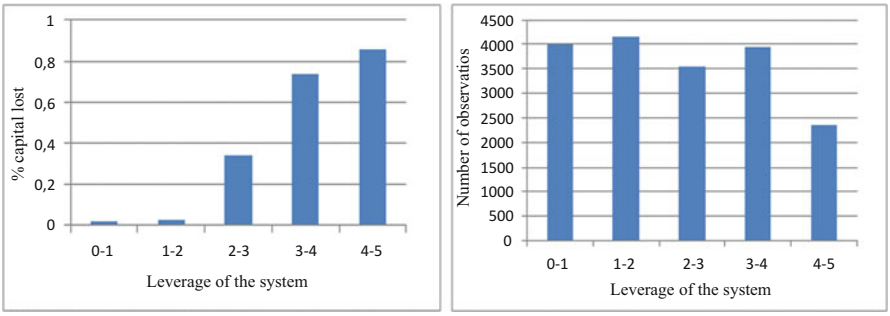
	N = 20	N = 50	N = 80	N = 100
CATIN2	0.196 (25.178)***	0.143 (16.422)***	0.125 (15.232)***	0.088 (9.806)***
LEVIN	0.324 (39.268)***	0.298 (32.475)***	0.275 (31.619)***	0.279 (30.578)***
NOUTGOING	−0.163 (−15.308)***	−0.168 (−10.438)***	−0.126 (−8.707)***	−0.087 (−5.561)***
COUNT	0.736 (191.690)***	0.761 (175.841)***	0.790 (195.383)***	0.793 (186.247)***
VARLOANS	−0.175 (−43.977)***	−0.190 (−44.390)***	−0.180 (−46.723)***	−0.167 (−41.937)***
P	−0.253 (−24.270)***	−0.313 (−19.153)***	−0.322 (−21.833)***	−0.339 (−21.575)***
Adjusted R^2	0.860	0.823	0.845	0.809

The table presents the regression results for Scenario 3. The dependent variable is CATEND measured as the total loss of capital due to contagion as percentage of total capital in the network. Explanatory variables are the constant term CATIN2, LEVIN, NOUTGOING, COUNT, VARLOANS and P, the probability for a link to exist between two nodes. Each cell displays the OLS standardized coefficients along with the corresponding *t*-statistics (shown in parentheses). The sample comprises of 18,000 realizations (simulated banking crises). *, ** and *** denote significance at the 10, 5 and 1 percent level, respectively

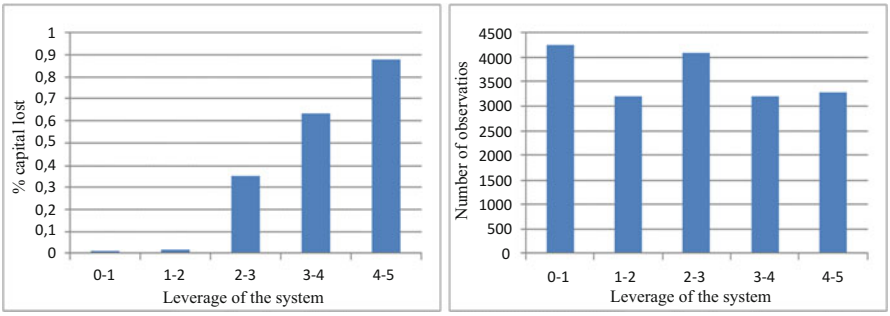
As far as the link probability is concerned, we can observe a different pattern from that of the first scenario. For small and large sized networks, link probability seems to contribute negatively to systemic risk while for medium sized networks there is a positive relationship between link probability and contagion. The number of rounds until no further bank defaults positively impacts contagion risk and contributes the most to total capital losses in the banking system when medium and large interbank networks are formed. Under this scenario, the heterogeneity allowed on both bank sizes and interbank exposures has had a great impact on the resilience



(a) N=20 banks



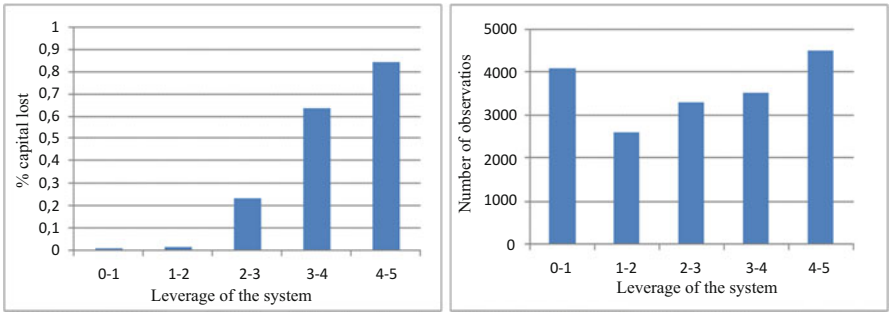
(b) N=50 banks



(c) N=80 banks

Fig. 6 Scenario 2: **Heterogeneous Banks with heterogeneous exposures. Extent of contagion** (expressed as the total capital lost from the banking system due to the failure of at least one bank) **as a function of the leverage of the system**. Panels (**a–d**) show the relation between the leverage of the system and the extent of contagion across interbank networks with different number of banks

of the network system. Heterogeneity impacts negatively on interbank contagion although its intensity decreases as the size of the network increases. Moreover, as we can see from the Table 4 heterogeneity of bank size contributes less to the resilience



(d) N=100 banks

Fig. 6 (continued)

of the interbank network than heterogeneity of interbank exposures when it comes to small and medium sized networks.

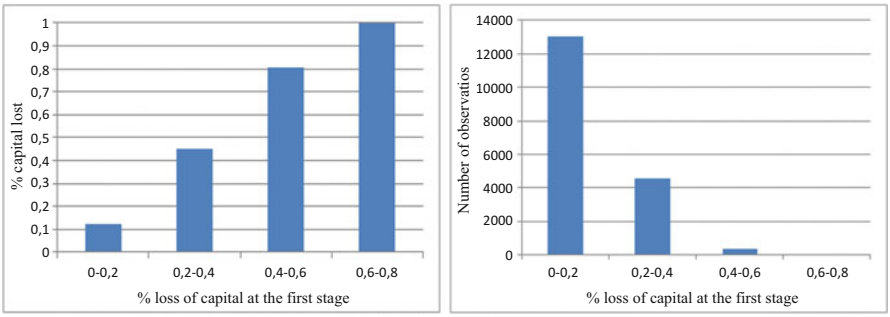
The heterogeneity of interbank exposures acts as a diversification tool and contributes to a smaller extent to an unfolding crisis compared to the scenario where homogeneous banks are interconnected via heterogeneous exposures (shown in Table 4).

Table 5 depicts the results of the **third scenario** as described in Equation (17). In this scenario, we construct network systems where banks have the same equity size and unevenly distribute their exposures across creditor banks. We note that an initial shock fades away with the failure of the first bank and does not spillover to other banks within the network. This is mainly due to our choice of parameters regarding the equity of each bank, the links among banks and the interbank claims among creditor banks. In order to observe default cascades we relax our initial assumptions and lower the equity of each bank in the network system.

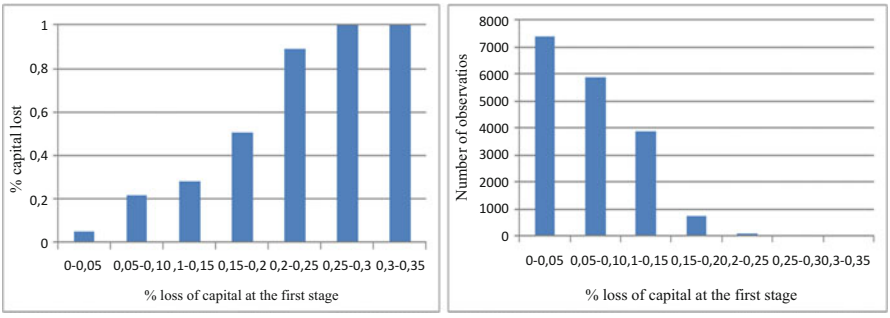
Specifically, each bank is now endowed with a balance sheet that consists of 25 units of equity and interbank claims among creditor banks are distributed in the following ranges: $a_{ij} \in [0, 10]$, $a_{ij} \in [0, 20]$, $a_{ij} \in [0, 35]$. Interbank exposures levels were kept the same as in Leventides et al. (2018). Moreover, we control the leverage of the system by setting the rule that the maximum leverage ratio of each network system cannot exceed seven. Similar to the previous scenarios, the regressor coefficients are statistically significant in all cases and the R-squared values are still large, in fact the largest of all three scenarios tested. Variable CATIN2 has a highly significant positive impact on systemic risk that fades away as the network system gets larger. The same observation holds for the level of connectivity in the banking system i.e. a strong negative impact on contagion risk that dissipates as N increases.

The leverage of the system has a positive impact on systemic risk and its magnitude decreases as the size of the network increases. Figures 7 and 8 illustrate the third scenario as a function of the percentage loss of capital due to default of the first bank in the network and as a function of leverage in the banking system, respectively. As in the previous cases, we find the number of rounds until no further

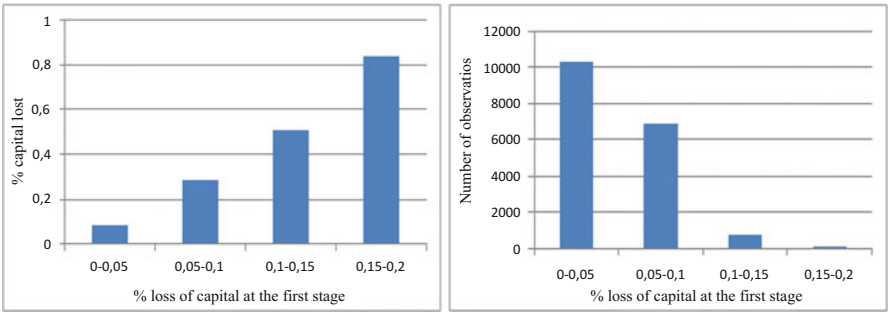
bank defaults to affect contagion risk positively and statistically significantly, and such impact is magnified in relatively larger interbank networks. The heterogeneity of interbank exposures plays a significant role in the stability of the financial network especially in the medium sized interbank networks.



(a) N=20 banks

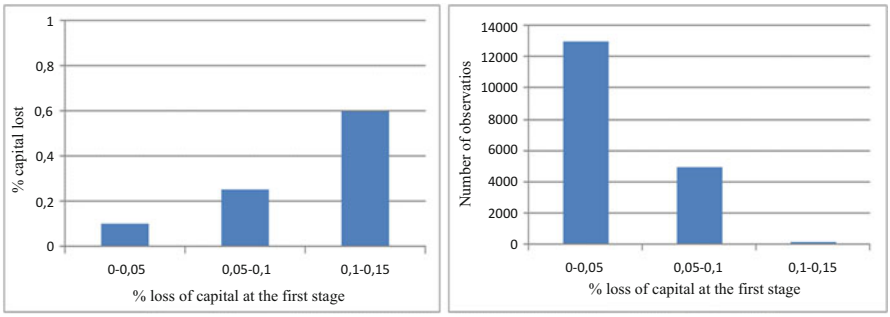


(b) N=50 banks



(c) N=80 banks

Fig. 7 Scenario 3: **Homogeneous banks with heterogeneous exposures** (expressed as the total capital lost from the banking system due to the failure of at least one bank) **as a function of the % initial loss of capital** due to default of the first bank. Panels (**a**–**d**) show the relation between the % initial loss of capital due to default of the first bank and the extent of contagion across interbank networks with different number of banks



(d) N=100 banks

Fig. 7 (continued)

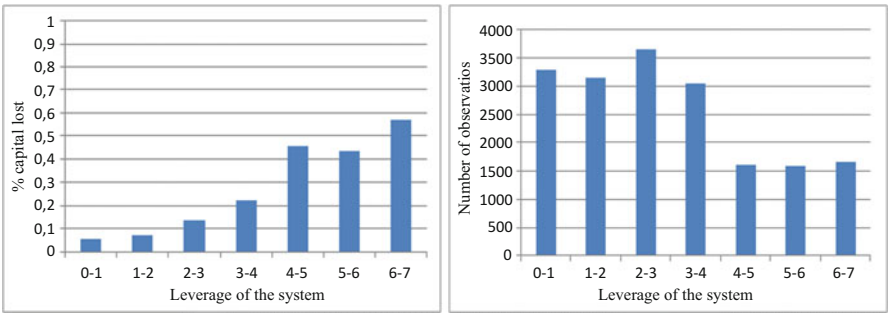
Table 6 OLS regression analysis for Scenario 4 (Homogeneous banks with homogeneous exposures)

	N = 20	N = 50	N = 80	N = 100
CATIN2	0.228 (21.978)***	0.153 (14.098)***	0.137 (12.902)***	0.105 (9.426)***
LEVIN	0.137 (14.890)***	0.268 (28.512)***	0.352 (37.106)***	0.352 (37.707)***
NOUTGOING	−0.257 (−15.906)***	−0.146 (−9.715)***	−0.130 (−8.719)***	−0.095 (−6.262)***
COUNT	0.645 (198.356)***	0.617 (172.925)***	0.568 (150.736)***	0.573 (148.381)***
P	−0.156 (−10.231)***	−0.304 (−21.593)***	−0.378 (−26.723)***	−0.379 (−27.197)***
Adjusted R^2	0.834	0.806	0.817	0.779

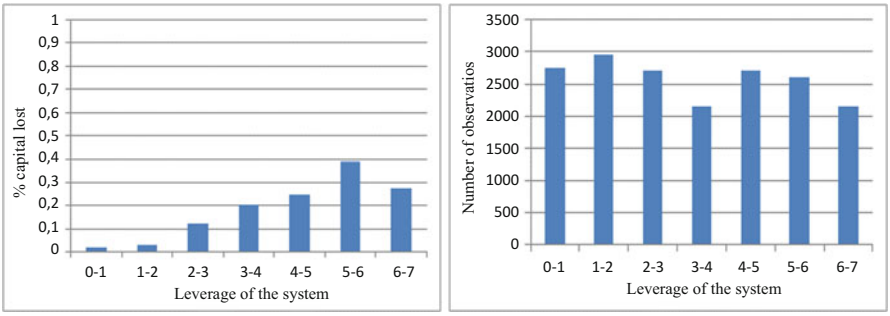
The table presents the regression results for Scenario 4. The dependent variable is CATEND measured as the total loss of capital due to contagion as percentage of total capital in the network. Explanatory variables are the constant term CATIN2, LEVIN, NOUTGOING, COUNT and P, the probability for a link to exist between two nodes.. Each cell displays the OLS standardized coefficients along with the corresponding *t*-statistics (shown in parentheses). The sample comprises of 18,000 realizations (simulated banking crises). *, ** and *** denote significance at the 10, 5 and 1 percent level, respectively

Finally, Table 6 depicts the results of the **fourth scenario** as described in Equation (18). In this scenario, we construct network systems where banks have the same equity size and interbank claims are evenly distributed across creditor banks. We acknowledge the fact that this scenario is a bit unrealistic as banks in real-world interbank networks do not have the same equity size and do not necessarily distribute their interbank claims evenly across their creditors. However, by testing a wide range of link probabilities between any two nodes, we are in a position to effectively examine the effect of different calibrations on systemic risk. Thus, although this scenario can be regarded as a special case with magnifying effects, it provides useful insights on interbank market resiliency during periods of stress.

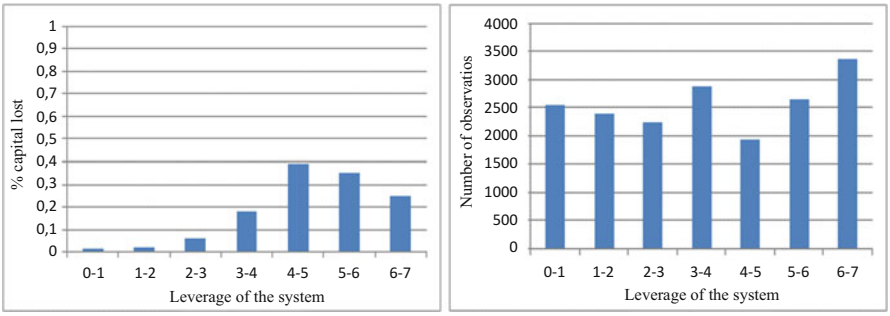
The variable CATIN2 has a strong positive impact on systemic risk that dissipates as the network system gets larger. Simulations show that this scenario yields qualitatively similar results with the previous three scenarios in relation to the leverage of the network, that is, leverage positively and significantly affects



(a) N=20 banks

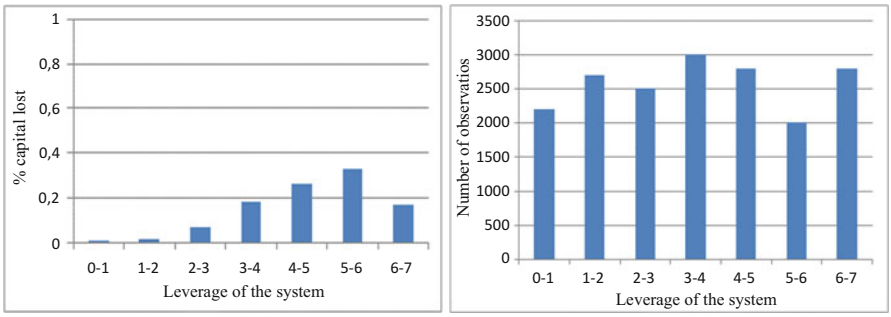


(b) N=50 banks



(c) N=80 banks

Fig. 8 Scenario 3: **Homogeneous banks with heterogeneous exposures. Extent of contagion** (expressed as the total capital lost from the banking system due to the failure of at least one bank) **as a function of the leverage of the system.** Panels (**a–d**) show the relation between the leverage of the system and the extent of contagion across interbank networks with different number of banks



(d) N=100 banks

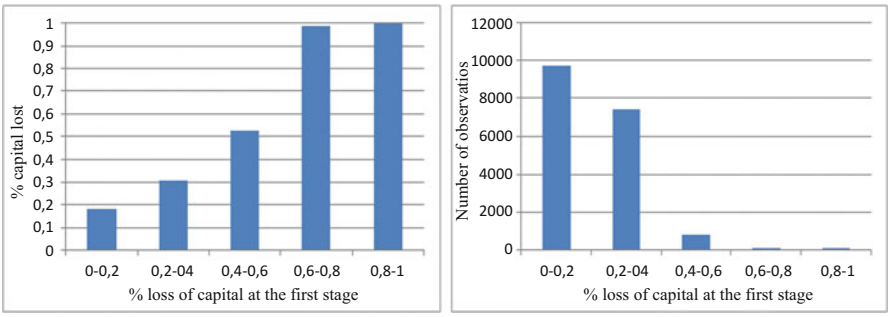
Fig. 8 (continuued)

contagion risk (illustrated in Figure 10). However, in this scenario, we observe that this effect becomes stronger progressively when the number of constituent banks in the network increases. Figure 9 confirms the results recorded in the Table 6 concerning the relationship between the extent of contagion and the percentage loss of capital in the network. For instance, the likelihood of systemic breakdowns seems to decrease as we move from smaller to larger network systems since we have very few cases that cause large capital losses. Connectivity impacts negatively on interbank contagion, although this negative impact dissipates as the number of banks in the interbank networks increases. As expected, the link probability has the same negative impact as connectivity on the interbank contagion. Contrary to the previous findings concerning connectivity, the negative impact of the link probability on interbank contagion seems to scale up as we move from smaller to larger interbank networks (Figure 10).

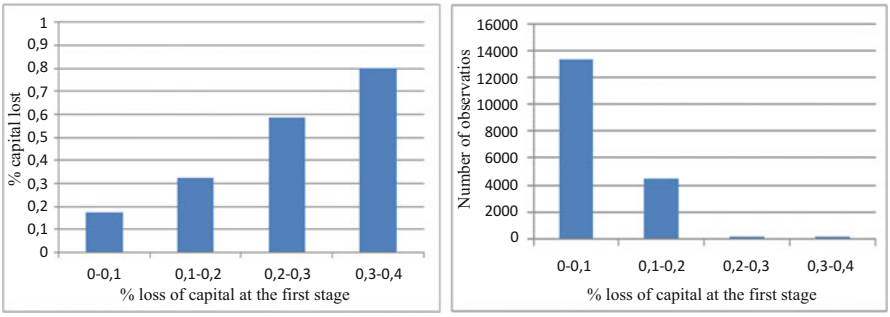
Finally, the number of rounds until no further bank defaults affects contagion risk in a statistically significant manner especially when small interbank networks are considered.

The main intuition behind these results is that increasing connectivity on a homogeneous interbank network can reduce the frequency of contagion in case the first bank that defaults is less leveraged as the interbank network has the dynamics to absorb more easily the shock and thus the initial shock is dissipated at a faster rate. This is the case for small network systems. As the size of the network increase and the system gets more leveraged, the stabilizing force of connectivity weakens and default cascades prevail.

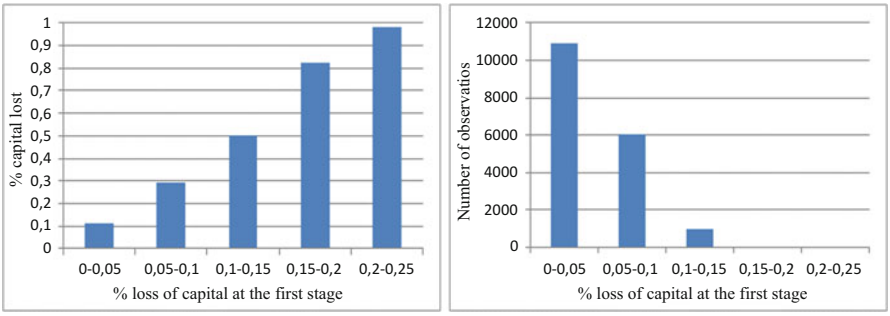
Tables 7, 8, 9, and 10 depict robustness tests on all four scenarios based on random sampling. We have performed second run Monte Carlo simulations in order to examine whether the new results differ from the previous ones, thus checking how random sampling affects our main conclusions. We observe qualitatively similar results in all four cases to those from the first run providing evidence that our findings are stable across different simulation scenarios.



(a) N=20 banks

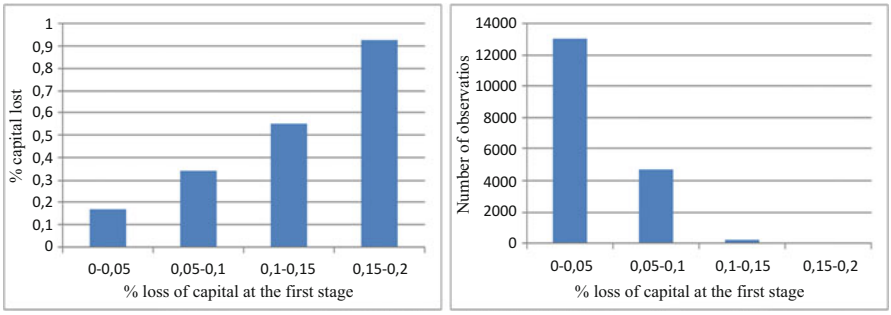


(b) N=50 banks



(c) N=80 banks

Fig. 9 Scenario 4: **Homogeneous banks with homogeneous exposures** (expressed as the total capital lost from the banking system due to the failure of at least one bank) **as a function of the % initial loss of capital** due to default of the first bank. Panels (**a**–**d**) show the relation between the % initial loss of capital due to default of the first bank and the extent of contagion across interbank networks with different number of banks



(d) N=100 banks

Fig. 9 (continued)

5 Conclusions

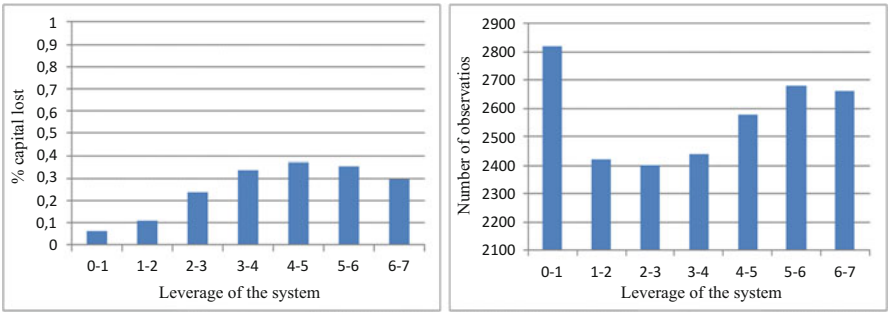
This paper investigates how complexity under a specific network structure, that has been extensively applied for the study of contagion in financial networks, affects interbank contagion. Similar to Leventides et al. [5], we explore the interplay between heterogeneity, balance sheet composition in the spreading of contagion using four basic scenarios, under an Erdős–Rényi network structure using a wide range of link probabilities between any two banks.

Our findings indicate a non-monotonic relation between diversification and interbank contagion across the different sizes of interbank networks for all scenarios tested. While for small or medium interbank networks, connectivity can act as an absorbing barrier, such that interbank systems of these sizes can withstand the initial shock, for large network systems connectivity does not seem to provide an effective shield against capital losses. Our results, for the four scenarios tested, indicate that small and thus more concentrated interbank network systems are more prone to contagion. In these cases, we observe that the size of total capital losses is, in most cases, larger than that documented in medium and large sized systems, which is in line with the findings of Nier et al. [7].

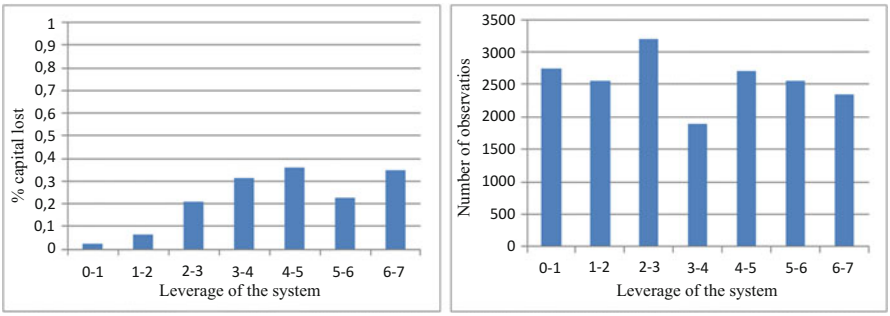
As far as heterogeneity is concerned, this enters in our experiments in the form of interbank claims and bank sizes. Our results clearly suggests that heterogeneity plays a significant role in the stability of the financial system. Similar to Leventides et al. [5], we still find that when heterogeneity is introduced with respect to the size of each bank, the system's shock absorption capacity is enhanced. In addition, when we allow for heterogeneity on interbank exposures in our model, we observe additional resilience to the interbank network as an initial shock dissipates more easily than in the case of homogeneous interbank claims.

Finally, we should also justify the fact that we choose to work under an Erdős–Rényi network structure even if this network framework is not very realistic. In an such a network framework, where the probability of forming a link is homogeneous, the resulting network structure does not present marked heterogeneity. As we

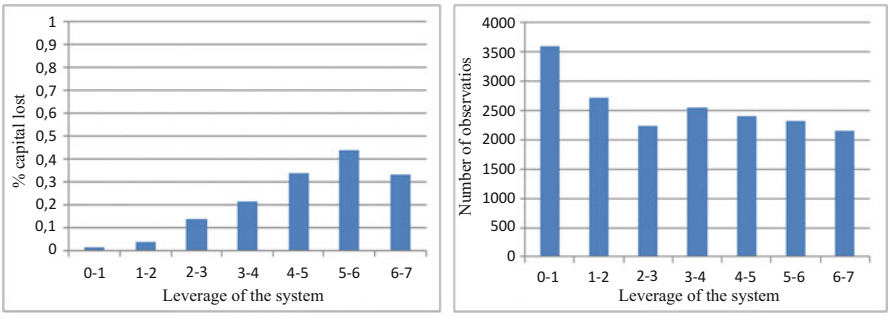
observed from all the four scenarios tested, the initial shock that hits the system seems to propagates into the system jeopardizing thus the stability of the entire system. This strengthens even more our arguments concerning the critical role that heterogeneity plays in the resilience of the financial system.



(a) N=20 banks

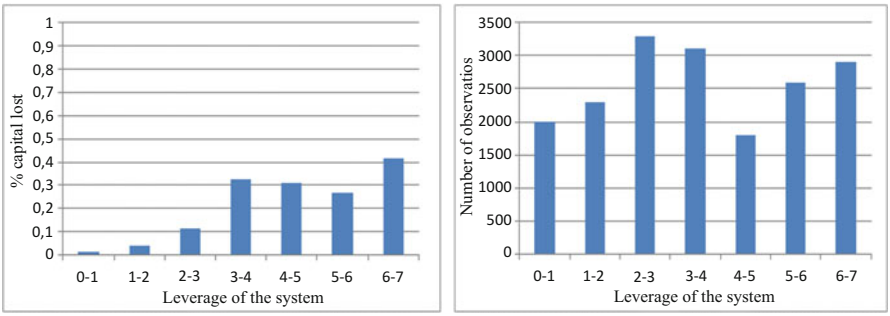


(b) N=50 banks



(c) N=80 banks

Fig. 10 Scenario 4: **Homogeneous banks with homogeneous exposures. Extent of contagion** (expressed as the total capital lost from the banking system due to the failure of at least one bank) **as a function of the leverage of the system**. Panels (**a–d**) show the relation between the leverage of the system and the extent of contagion across interbank networks with different number of banks



(d) N=100 banks

Fig. 10 (continued)

Table 7 Robustness tests: OLS regression analysis for Scenario 1 (Heterogeneous banks with homogeneous exposures)

	N = 20	N = 50	N = 80	N = 100
CATIN 1	0.044 (13.363)***	0.001 (0.389)	−0.004 (−1.135)	0.002 (0.579)
CATIN2	0.259 (10.095)***	0.163 (7.044)***	0.073 (3.301)***	0.050 (2.614)***
LEVIN	0.272 (10.899)***	0.255 (11.408)***	0.412 (19.261)***	0.402 (22.155)***
NOUTGOING	−0.213 (−10.339)**	−0.162 (−4.518)***	0.014 (−4.933)***	0.042 (1.328)
COUNT	0.569 (128.765)***	0.604 (147.789)***	0.523 (126.895)***	0.539 (131.678)***
VARCAP	−0.085 (−50.816)***	−0.075 (−57.790)***	−0.057 (−58.108)***	−0.054 (−64.019)***
P	0.021 (−5.089)***	0.141 (3.998)**	−0.006 (−0.171)	−0.012 (−0.405)
Adjusted R^2	0.786	0.785	0.768	0.789

The table presents the regression results for Scenario1 applied on a second run of Monte Carlo simulations based on random sampling as robustness test. The dependent variable is CATEND measured as the total loss of capital due to contagion as percentage of total capital in the network. Explanatory variables are, CATIN1, CATIN2, LEVIN, NOUTGOING, COUNT, VARCAP and P, the probability for a link to exist between two nodes. Each cell displays the OLS standardized coefficients along with the corresponding *t*-statistics (shown in parentheses). The sample comprises of 18,000 realizations (simulated banking crises). *, ** and *** denote significance at the 10, 5 and 1 percent level, respectively

Table 8 Robustness tests: OLS regression analysis for Scenario 2 (Heterogeneous banks with heterogeneous exposures)

	N = 20	N = 50	N = 80	N = 100
CATIN 1	0.071 (23.366)***	0.001 (0.273)	0.008 (2.368)***	0.006 (1.730)*
CATIN2	0.207 (20.109)***	0.098 (8.941)***	0.101 (9.336)***	0.068 (6.721)***
LEVIN	0.602 (53.999)***	0.469 (38.008)***	0.313 (26.051)***	0.304 (26.494)***
NOUTGOING	−0.154 (−12.833)***	−0.096 (−4.023)***	−0.064 (−2.747)***	0.008 (0.391)
COUNT	0.459 (107.602)***	0.567 (131.004)***	0.590 (144.084)***	0.609 (156.107)***
VARCAP	−0.038 (−21.431)***	−0.067 (−41.713)***	−0.053 (−54.105)***	−0.051 (−40.597)***
VARLOANS	−0.220 (−42.365)***	−0.091 (−24.628)***	−0.018 (−2.107)**	−0.009 (−12.307)***
P	−0.223 (−18.398)***	−0.080 (−3.217)***	0.092 (3.779)***	0.061 (2.751)***
Adjusted R^2	0.817	0.770	0.772	0.800

The table presents the regression results for Scenario2 applied on a second run of Monte Carlo simulations based on random sampling as robustness test. The dependent variable is CATEND measured as the total loss of capital due to contagion as percentage of total capital in the network. Explanatory variables are the constant term CATIN1, CATIN2, LEVIN, NOUTGOING, COUNT, VARCAP, VARLOANS and P, the probability for a link to exist between two nodes. Each cell displays the OLS standardized coefficients along with the corresponding *t*-statistics (shown in parentheses). The sample comprises of 18,000 realizations (simulated banking crises). *, ** and *** denote significance at the 10, 5 and 1 percent level, respectively

Table 9 Robustness tests: OLS regression analysis for Scenario 3 (Homogeneous banks with heterogeneous exposures)

	N = 20	N = 50	N = 80	N = 100
CATIN2	0.200 (25.728)***	0.153 (18.410)***	0.127 (13.078)***	0.157 (21.173)***
LEVIN	0.282 (34.036)***	0.189 (22.195)***	0.329 (32.672)***	0.308 (37.777)***
NOUTGOING	−0.187 (−16.831)***	−0.168 (−11.145)***	−0.114 (−6.923)***	−0.145 (−11.721)***
COUNT	0.745 (190.987)***	0.773 (184.138)***	0.736 (164.477)***	0.765 (190.795)***
VARLOANS	−0.167 (−41.084)***	−0.137 (−33.586)***	−0.164 (−38.890)***	−0.196 (−46.316)***
P	−0.217 (−19.612)***	−0.226 (−14.785)***	−0.371 (−22.288)***	−0.323 (−25.206)***
Adjusted R^2	0.862	0.824	0.789	0.864

The table presents the regression results for Scenario3 applied on a second run of Monte Carlo simulations based on random sampling as robustness test. The dependent variable is CATEND measured as the total loss of capital due to contagion as percentage of total capital in the network. Explanatory variables are the constant term CATIN2, LEVIN, NOUTGOING, COUNT, VARLOANS and P, the probability for a link to exist between two nodes. Each cell displays the OLS standardized coefficients along with the corresponding *t*-statistics (shown in parentheses). The sample comprises of 18,000 realizations (simulated banking crises). *, ** and *** denote significance at the 10, 5 and 1 percent level, respectively

Table 10 Robustness tests: OLS regression analysis for Scenario4 (Homogeneous banks with homogeneous exposures)

	N = 20	N = 50	N = 80	N = 100
CATIN2	0.266 (25.631)***	0.196 (18.103)***	0.153 (13.994)***	0.126 (11.749)***
LEVIN	0.163 (17.098)***	0.247 (25.943)***	0.283 (29.111)***	0.357 (37.852)***
NOUTGOING	−0.306 (−19.180)***	−0.220 (−14.254)***	−0.164 (−10.529)***	−0.118 (−8.020)***
COUNT	0.616 (188.365)***	0.600 (161.885)***	0.609 (160.323)***	0.565 (145.072)***
P	−0.150 (−9.783)***	−0.256 (−17.542)***	−0.309 (−20.730)***	−0.371 (−26.650)***
Adjusted R^2	0.834	0.804	0.790	0.798

The table presents the regression results for Scenario4 applied on a second run of Monte Carlo simulations based on random sampling as robustness test. The dependent variable is CATEND measured as the total loss of capital due to contagion as percentage of total capital in the network. Explanatory variables are the constant term CATIN2, LEVIN, NOUTGOING, COUNT and P, the probability for a link to exist between two nodes.. Each cell displays the OLS standardized coefficients along with the corresponding *t*-statistics (shown in parentheses). The sample comprises of 18,000 realizations (simulated banking crises). *, ** and *** denote significance at the 10, 5 and 1 percent level, respectively

References

1. Business and Sustainable Development Commission Report, Better business better world (2017)
2. Oxford Analytica Foundation, The corporate stewardship compass-Guiding values for sustainable development, Project Report for the Caux Round Table of Moral Capitalism, December 2017
3. O. Weber, The financial sector and the SDGs: Interconnections and future directions, CIGI (Center for International Governance) Innovation Papers, No. 201, November 2018
4. C. Alves, V. Boufounou, A. Dellis, C. Pitelis, Toporowski, J., Synthesis Report: empirical analysis for new ways of global engagement, FESSUD (Financialisation, Economy, Society and Sustainable Development) Working Paper Series, No.163, August 2016
5. J. Leventides, K. Loukaki, V.G. Papavassiliou, Simulating financial contagion dynamics in random interbank networks. J. Econ. Behav. Organ. **158**, 500–525 (2019)
6. G. Iori, S. Jafarey, F. Padilla, Systemic risk on the interbank market. J. Econ. Behav. Organ. **61**, 525–542 (2006)
7. E. Nier, J. Yang, T. Yorulmazer, A. Alentorn, Network models and financial stability. J. Econ. Dyn. Control. **31**, 2033–2060 (2007)
8. P. Gai, S. Kapadia, Contagion in financial networks. Proc. R. Soc. A Math. Phys. Eng. Sci. **466**, 2401–2423 (2010)
9. R.M. May, N. Arinaminpathy, Systemic risk: the dynamics of model banking systems. J. R. Soc. Interface **7**, 823–838 (2010)
10. H. Amini, R. Cont, A. Minca, Resilience to contagion in financial networks. Math. Financ. **26**, 329–365 (2016)
11. C. Upper, Simulation methods to assess the danger of contagion in interbank markets. J. Financ. Stab. **7**, 111–125 (2011)

12. F. Allen, A. Babus, Networks in finance, Working Paper No. 08-07, Wharton Financial Institutions Center, 2008
13. C. Memmel, A. Sachs, Contagion in the interbank market and its determinants. J. Financ. Stab. **9**, 46–54 (2013)
14. O.-M. Georgescu, Contagion in the interbank market: funding versus regulatory constraints. J. Financ. Stab. **18**, 1–18 (2015)
15. L. Tonzer, Cross-border interbank networks, banking risk and contagion. J. Finan. Stabil. **18**, 19–32 (2015)
16. K. Fink, U. Krüger, B. Meller, L.-H. Wong, The credit quality channel: modeling contagion in the interbank market. J. Financ. Stab. **25**, 83–97 (2016)
17. F. Allen, D. Gale, Financial contagion. J. Polit. Econ. **108**, 1–33 (2000)
18. F. Caccioli, T.A. Catanach, J.D. Farmer, Heterogeneity, correlations and financial contagion. Adv. Comp. Syst. **15**(Suppl 02), 1250058 (2012)
19. D. Ladley, Contagion and risk-sharing on the inter-bank market. J. Econ. Dyn. Control. **37**, 1384–1400 (2013)
20. M. Chinazzi, S. Pegoraro, G. Fagiolo, Defuse the bomb: rewiring interbank networks. LEM Working Paper Series, Institute of Economics, Scuola Superiore Sant' Anna (2015)
21. G. Georg, J. Poschmann, Systemic risk in a network model of interbank markets with central bank activity. Jena Economic Research Papers, No. 2010, 033, 2010
22. E. Estrada, *The Structure of Complex Networks: Theory and Applications* (Oxford University Press, 2011)
23. G. Iori, G. de Masi, O. Precup, G. Gabbi, G. Caldarelli, A network analysis of the Italian overnight money market. J. Econ. Dyn. Control. **32**, 259–278 (2008)
24. A. Krause, S. Giansante, Interbank lending and the spread of bank failures: a network model of systemic risk. J. Econ. Behav. Organ. **83**, 583–608 (2012)

Higher Order Strongly m-convex Functions

Muhammad Aslam Noor and Khalida Inayat Noor

Abstract Some new concepts of the m-convex functions, where $m \in (0, 1]$ are introduced and studied. Basic properties of m-convex functions are discussed. New modified Regula Falsi methods are suggested for solving nonlinear equations. Characterizations of the higher order strongly m-convex functions are investigated under suitable conditions. It is shown that the parallelogram laws for Banach spaces can be obtained as applications of higher order strongly m-convex functions. Results obtained in this paper can be viewed as refinement and significant improvement of previously known results.

1 Introduction

Lin and Fukushima [11] introduced the concept of higher order strongly convex functions and used it in the study of mathematical program with equilibrium constraints. These mathematical programs with equilibrium constraints are defined defined by a parametric variational inequality or complementarity system and play an important role in many fields such as engineering design, economic equilibrium, and multilevel game. Mishra and Sharma [12] derived the Hermite–Hadamard type inequalities for higher order strongly convex functions. Characterizations of the higher order strongly convex functions discussed in Lin and Fukushima [11] are not correct. These facts and observations motivated Noor and Noor [16] to consider higher order strongly convex function. Several new characterizations of the higher order strongly convex functions were discussed. Parallelogram laws for uniformly

Nonlinear Analysis, Differential Equations and Applications (Edits: Themistocles M. Rassias), Springer Volume

M. A. Noor (✉) · K. I. Noor
COMSATS University Islamabad, Islamabad, Pakistan

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_14

Banach spaces can be deduced from the definitions of the higher order strongly convex functions, which is itself a novel and interesting property.

Strongly convex functions were introduced and studied by Polyak [20], which play an important part in the optimization theory and related areas. Karmardian [9] used the strongly convex functions to discuss the unique existence of a solution of the nonlinear complementarity problems. Strongly convex functions also played important role in the convergence analysis of the iterative methods for solving variational inequalities and equilibrium problems, see Zu and Marcotte [24]. Nikodem and Pales [14] investigated the characterization of the inner product spaces using the strongly convex functions, which can be viewed as a novel and innovative application. It is also known that the minimum of the strongly convex functions is unique. Qu and Li [21] investigated the exponentially stability of primal-dual gradient dynamics using the concept of strongly convex functions. Awan et al. [3] have derived Hermite–Hadamard type inequalities for various classes of strongly convex functions, which provide upper and lower estimate for the integrand. For more applications and properties of the strongly convex functions, see [1–4, 9–19, 21] and the references therein.

Relevant to the convex set and convex functions, we also have the concept of m-convex sets and m-convex functions, which were introduced by Toader [22]. For the properties and other aspects of the m-convex functions, see [10] and the references therein. Lara et al. [10] introduced and investigated the properties of strongly m-convex functions. We would like to mention that the concepts of strongly m-convex [10] are not correct. To overcome these deficiencies of the higher order strongly convex functions and strongly m-convex functions, we consider some new classes of convex functions, which are called higher order strongly m-convex functions. Using the techniques and ideas of this paper, one can easily obtain the refine and correct versions of higher order strongly convex functions and strongly m-convex functions.

In Section 2, we introduce the new concepts of m-convex functions and discuss their characterizations. It is pointed our that these m-convex functions are distinctly different the concepts of m-functions of Toader [22], which are being investigated in recent years. Higher order strongly m-convex functions are introduced in Section 3. Some results are discussed in Section 4. In Section 5, we discuss some applications of the higher order strongly convex functions. It is shown that the weakly parallelogram laws can be deduced from the definitions, which characterize the uniformly reflex Banach spaces. As special cases, one can obtain various new and refined versions of known results. It is expected that the ideas and techniques of this paper may stimulate further research in this field.

2 m-Convex Functions

Let K be a nonempty closed set in a real Hilbert space H. We denote by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ be the inner product and norm, respectively.

Definition 1 ([11, 19]) A set K in H is said to be a convex set, if

$$u + t(v - u) \in K, \qquad \forall u, v \in K, t \in [0, 1].$$

We denote by $m \in (0, 1]$, unless otherwise specified.

Definition 2 ([23]) The set K in H is said to be m-convex set, if

$$(1 - t)u + tmv \in K, \qquad \forall u, v \in K, t \in [0, 1].$$

Also one can define the concept of m_1-convex set as:

Definition 3 ([23]) The set K in H is said to be m_1-convex set, if

$$(1 - t)m_1u + tv \in K, \qquad \forall u, v \in K, t \in [0, 1], m_1 \in (0, 1].$$

We remark that m-convex set and m_1-convex set are entirely two different generalizations of the so-called convex sets. that is m-convex set $\neq m_1$-convex set.

For example, $[m_1a, b] \neq [a, mb], \quad m, m_1 \in (0, 1]$. Consequently these sets have different properties.

In [10], the authors remarked that $[m_1a, b] = [a, mb]$, which is not true. In passing, we remark that the concept of strongly m-convex functions introduced and discussed in [10] is also wrong. Our results can be viewed as the refinement and improvement of the known results.

In this paper, we only consider the m-convex set K_m, unless otherwise specified.

We now introduce new concepts of m-convex functions.

Definition 4 ([17]) A function F is said to be m-convex function, if

$$F((1 - t)u + tmv) \leq (1 - t)F(u) + tF(mv), \ \forall u, v \in K_m, \ t \in [0, 1], m \in (0, 1].$$

We now consider the m-convex functions on the interval $I = [a, mb], \quad m \in (0, 1]$.

Definition 5 Let $I = [a, mb]$. Then F is m-convex function, if and only if,

$$\begin{vmatrix} 1 & 1 & 1 \\ a & x & mb \\ F(a) & F(x) & F(mb) \end{vmatrix} \geq 0; \quad a \leq x \leq mb.$$

One can easily show that the following are equivalent:

1. F is m-convex function.
2. $F(x) \leq F(a) + \frac{F(mb)-F(a)}{mb-a}(x - a)$.
3. $\frac{F(x)-F(a)}{x-a} \leq \frac{F(mb)-F(a)}{mb-a}$.
4. $(mb - x)F(a) + (a - mb)F(x) + (x - a)F(mb) \geq 0$.
5. $\frac{F(a)}{(mb-a)(a-x)} + \frac{F(x)}{(x-mb)(a-x)} + \frac{F(mb)}{(mb-a)(x-mb)} \leq 0$,

where $x = (1-t)a + tmb \in [a, mb]$.

We now suggest an iterative method for solving the nonlinear equations $F(x) = 0$ in the interval $[a, mb]$, $m \in (0.1]$ and $[ma, b]$, $m \in (0.1]$. From $F(x) = F(a) + \frac{F(mb)-F(a)}{mb-a}(x-a)$, and using the fact that $F(x) = 0$, we have

$$x = \frac{aF(mb) - mbF(a)}{mb - a}, x \in [a, mb], m \in (0, 1].$$

This enable to suggest the following iterative method for solving the nonlinear equations $F(x) = .$

Algorithm 1 For given u_0, u_1, find the approximate solution u_{n+1} by the iterative scheme

$$u_{n+2} = \frac{aF(mu_{n+1}) - mu_{n+1}F(u_n)}{mu_{n+1} - u_n}, \quad m \in (0, 1],$$

which is called the Modified Regula Falsi Method.

In a similar way, we can also suggest the following modified Regula Falsi Method.

Algorithm 2 For given u_0, u_1, find the approximate solution u_{n+1} by the iterative scheme

$$u_{n+2} = \frac{mu_nF(u_{n+1}) - u_{n+1}F(mu_n)}{u_{n+1} - mu_n}, \quad m \in (0, 1].$$

It is worth mentioning that Algorithms 1 and 2 are quite different and provide different results.

Using the technique of Toader [22], we can introduce the following concept of m-convex functions:

Definition 6 ([22]) A function F is said to be m-convex function in the Toader sense, if

$$F((1-t)u + tmv) \leq (1-t)F(u) + mtF(v), \quad \forall u, v \in K_m, \quad t \in [0, 1].$$

We would like to point out that these concepts defined in Definitions 2 and 6 are equivalent, if the functions $F(mv) = mF(v)$, that is, the function F is homogeneous. Consequently all the results proved in this paper can be extended for the m-convex functions in the Toader sense with suitable modifications.

In this section, we consider some basic properties of m-convex functions. Using the technique of Pavic and Ardic [18], we derive the following result.

Theorem 1 *Let $I = [a, mb] \subset R$ be an interval containing the zero and let $m \in (0,]$ be a constant. Let $a, b, c \in I$ be a point such that $a \leq mc \leq b$. Then the m-convex function satisfy the inequality*

$$\int_a^b F(x)dx \leq \frac{mc-a}{2}F(a) + \frac{b-mc}{2}F(b) + \frac{b-a}{2}F(mc). \tag{1}$$

Proof Assume that $a \leq x \leq mc$. Then, from (1), we have

$$\begin{aligned}\int_a^{mc} F(x)dx &\leq \int_a^{mc} \frac{mc-x}{mc-a}F(a)dx + \int_a^{mc} \frac{x-a}{mc-a}F(mc)dx \\ &= \frac{mc-a}{2}\big(f(a)+f(mc)\big). \end{aligned} \tag{2}$$

In a similar way, for $mc \leq x \leq b$, we have

$$\begin{aligned}\int_{mc}^b F(x)dx &\leq \int_{mc}^b \frac{mc-x}{mc-a}F(mc)dx + \int_{mc}^b \frac{x-a}{mc-a}F(b)dx \\ &= \frac{mc-a}{2}\big(F(b)+F(mc)\big). \end{aligned} \tag{3}$$

From (2) and (3), we have

$$\begin{aligned}\int_a^b F(x)dx &= \int_a^{mc} F(x)dx + \int_{mc}^b F(x)dx \\ &= \frac{mc-a}{2}\big(F(a)+F(mc)\big) + \frac{mc-a}{2}\big(F(b)+F(mc)\big) \\ &= \frac{mc-a}{mc-a}F(a) + \frac{b-mc}{2}F(b) + \frac{b-a}{2}F(mc),\end{aligned}$$

the required (1).

Remark 1 For the interval $I = a, b]$ containing the zero, one can choose an point $c \in [a, b]$ in (3), since $mc \in [a, b]$. Using this information,we can obtain the following inequality for the m-convex functions for the case $c = a$ or $b = c$.

$$\int_a^b F(x)dx \leq \frac{mb-a}{2}F(a) + \frac{b-ma}{2}F(b),$$

from which, we can have

$$\int_a^b F(x)dx \leq \frac{b-a}{2}\{F(a)+F(b)\}.$$

Theorem 2 *Let F be a strictly m-convex function. Then any local minimum of F is a global minimum.*

Proof Let the m-convex function F have a local minimum at $u \in K_m$. Assume the contrary, that is, $F(mv) < F(u)$ for some $mv \in K_m$. Since F is m-convex, so

$$F(u + t(mv - u)) < tF(mv) + (1 - t)F(u), \quad \text{for} \quad 0 < t < 1.$$

Thus

$$F(u + t(mv - u)) - F(u) < t[F(mv) - F(u)] < 0,$$

from which it follows that

$$F(u + t(mv - u)) < F(u),$$

for arbitrary small $t > 0$, contradicting the local minimum.

Theorem 3 *If the function F on the m-convex set K_m is m-convex, then the level set $L_\alpha = \{u \in K_m : F(u) \leq \alpha, \quad \alpha \in R\}$ is a m-convex set.*

Proof Let $u, mv \in L_\alpha$. Then $F(u) \leq \alpha$ and $F(mv) \leq \alpha$. Now, $\forall t \in (0, 1), \quad w = u + t(mv - u) \in K_m$, since K_m is a m-convex set. Thus, by the m-convexity of F, we have

$$\begin{aligned} F(u + t(mv - u) &\leq (1 - t)F(u) + tF(mv) \\ &\leq (1 - t)\alpha + t\alpha = \alpha, \end{aligned}$$

from which it follows that $u + t(mv - u) \in L_\alpha$ Hence L_α is a m-convex set.

Theorem 4 *The function F is a m-convex function, if and only if,*

$$epi(F) = \{(u, \alpha) : u \in K_m : F(u) \leq \alpha, \alpha \in R\}$$

is a m-convex set.

Proof Assume that F is a m-convex function. Let $(u, \alpha), \quad (mv, \beta) \in epi(F)$. Then it follows that $F(u) \leq \alpha$ and $(mv) \leq \beta$. Thus, $\forall t \in [0, 1], \quad u, mv \in K_m$, we have

$$\begin{aligned} F(u + t(mv - u)) &\leq (1 - t)F(u) + tF(mv) \\ &\leq (1 - t)\alpha + t\beta, \end{aligned}$$

which implies that

$$(u + t(mv - u), (1 - t)\alpha + t\beta) \in epi(F).$$

Thus $epi(F)$ is a m-convex set. Conversely, let $epi(F)$ be a convex set. Let $u, mv \in K$. Then $(u, F(u)) \in epi(F)$ and $(mv, e^{F(mv)}) \in epi(F)$. Since $epi(F)$ is a m-convex set, we must have

$$(u + t(mv - u), (1 - t)F(u) + tF(mv)) \in epi(F),$$

which implies that

$$F(u + t(mv - u)) \leq (1 - t)F(u) + tF(mv).$$

This shows that F is a m-convex function.

Theorem 5 *The function F is quasi m-convex, if and only if, the level set $L_\alpha = \{u \in K_m, \alpha \in R : e^{F(u)} \leq \alpha\}$ is a m-convex set.*

Proof Let $u, mv \in L_\alpha$. Then $u, mv \in K_m$ and $\max(F(u), F(mv)) \leq \alpha$. Now for $t \in (0, 1)$, $w = u + t(mv - u) \in K_m$, We have to prove that $u + t(mv - u) \in L_\alpha$. By the quasi m-convexity of F, we have

$$F(u + t(mv - u)) \leq \max F(u), F(mv) \leq \alpha,$$

which implies that $u + t(mv - u) \in L_\alpha$, showing that the level set L_α is indeed a m-convex set.

Conversely, assume that L_α is a m-convex set. Then for any $u, mv \in L_\alpha, t \in [0, 1]$, $u + t(mv - u) \in L_\alpha$. Let $u, mv \in L_\alpha$ for

$$\alpha = max(F(u), F(mv)) \quad \text{and} \quad F(mv) \leq F(u).$$

Then from the definition of the level set L_α, it follows that

$$F(u + t(mv - u)) \leq \max F(u), F(mv) \leq \alpha.$$

Thus F is an quasi m-convex function. This completes the proof.

Theorem 6 *Let F be a m-convex function. Let $\mu = \inf_{u \in K_m} F(u)$. Then the set $E = \{u \in K_m : F(u) = \mu\}$ is a m-convex set of K. If F is a exponentially m-convex function, then E is a singleton.*

Proof Let $u, mv \in E$. For $0 < t < 1$, let $w = u + t(mv - u)$. Since F is a m-convex function, then

$$\begin{aligned} F(w) = F(u + t(mv - u)) &\leq (1 - t)F(u) + tF(mv) \\ &= t\mu + (1 - t)\mu = \mu, \end{aligned}$$

which implies that to $w \in E$. and hence E is a m-convex set. For the second part, assume to the contrary that $F(u) = F(mv) = \mu$. Since K is a m-convex set, then for $0 < t < 1, u + t(mv - u) \in K_m$. Further, since F is strictly m-convex function,

$$F(u + t(mv - u)) < (1 - t)F(u) + tF(mv) = (1 - t)\mu + t\mu = \mu.$$

This contradicts the fact that $\mu = \inf_{u \in K_m} F(u)$ and hence the result follows.

Theorem 7 *If F is a m-convex function such that $F(mv) < F(u), \forall u, mv \in K_m$, then F is a strictly quasi m-convex function.*

Proof By the m-convexity of the function F, $\forall u, mv \in K_m, m, t \in [0, 1]$, we have

$$F(u + t(mv - u)) \leq (1 - t)F(u) + tF(mv) < F(u),$$

since $F(mv) < F(u)$, which shows that the function F is strictly quasi m-convex.

We now discuss some properties of the differentiable m-convex functions.

Theorem 8 *Let F be a differentiable function on the m-convex set K_m. Then the function F is m-convex function, if and only if,*

$$F(mv) - F(u) \geq \langle F'(u), mv - u\rangle, \quad \forall mv, u \in K_m. \tag{4}$$

Proof Let F be a y m-convex function. Then

$$F(u + t(mv - u)) \leq (1 - t)F(u) + tF(mv), \qquad \forall u, mv \in K_m,$$

which can be written as

$$F(mv) - F(u) \geq \{\frac{F(u + t(mv - u)) - F(u)}{t}\}.$$

Taking the limit in the above inequality as $t \to 0$, we have

$$F(mv) - F(u) \geq \langle F'(u), mv - u)\rangle,$$

which is (4), the required result.

Conversely, let (4) hold. Then $\forall u, mv \in K_m, \quad t \in [0, 1]$, $v_t = u + t(mv - u) \in K_m$, we have

$$F(mv) - F(v_t) \geq \langle F'(v_t), mv - v_t)\rangle = (1 - t)\langle F'(v_t), mv - u\rangle. \tag{5}$$

In a similar way, we have

$$F(u) - F(v_t) \geq \langle F'(v_t), u - v_t)\rangle = -t\langle F'(v_t), mv - u\rangle. \tag{6}$$

Multiplying (5) by t and (6) by $(1 - t)$ and adding the resultant, we have

$$F(u + t(mv - u) \leq (1 - t)F(u) + tF(mv),$$

showing that F is a m-convex function.

Theorem 8 enables us to introduce the concept of the m-monotone operators, which appears to be new ones.

Definition 7 The differential $F'(.)$ is said to be m-monotone, if

$$\langle F'(u) - F'(mv), u - mv\rangle \geq 0, \quad \forall u, mv \in H.$$

Definition 8 The differential $F'(.)$ is said to be pseudo m-monotone, if

$$\langle F'(u), mv - u\rangle \geq 0, \quad \Rightarrow \langle F'(mv), mv - u\rangle \geq 0, \quad \forall u, mv \in H.$$

From these definitions, it follows that m-monotonicity implies pseudo m-monotonicity, but the converse is not true.

Theorem 9 *Let F be differentiable m-convex function on the mconvex set K_m. Then (19) holds, if and only if, $F'(.)$ satisfies*

$$\langle F'(u) - F'(mv), u - mv\rangle \geq 0, \quad \forall u, mv \in K_m. \tag{7}$$

Proof Let F be a m-convex function on the m-convex set K_m. Then, from Theorem 8, we have

$$F(mv) - F(u) \geq \langle F'(u), mv - u\rangle, \quad \forall u, mv \in K_m. \tag{8}$$

Changing the role of u and mv in (8), we have

$$F(u) - F(mv) \geq \langle F'(mv), u - mv)\rangle, \quad \forall u, mv \in K_m. \tag{9}$$

Adding (7) and (8), we have

$$\langle F'(u) - F'(mv), u - mv\rangle \geq 0,$$

which shows that F' is a m-monotone.

Conversely, from (7), we have

$$\langle F'(mv), u - mv\rangle \leq \langle F'(u), u - mv)\rangle. \tag{10}$$

Since K is a m-convex set, $\forall u, mv \in K_m, \quad t \in [0, 1]\ v_t = u + t(mv - u) \in K_m$.

Taking $v = v_t$ in (10), we have

$$\langle F'(v_t), u - v_t\rangle \leq \langle F'(u), u - v_t\rangle = -t\langle F'(u), mv - u\rangle,$$

which implies that

$$\langle F'(v_t), mv - u\rangle \geq \langle F'(u), v - u\rangle. \tag{11}$$

Consider the auxiliary function

$$g(t) = e^{F(u+t(mv-u))},$$

from which, we have

$$g(1) = F(mv), \quad g(0) = F(u).$$

Then, from (11), we have

$$g'(t) = \langle F'(v_t, mv - u\rangle \geq \langle F'(u), mv - u\rangle. \tag{12}$$

Integrating (12) between 0 and 1, we have

$$g(1) - g(0) = \int_0^1 g'(t)dt \geq \langle F'(u), mv - u\rangle.$$

Thus it follows that

$$F(mv) - F(u) \geq \langle F'(u), mv - u\rangle,$$

which is the required (8).

3 Strongly m-Convex Functions

We now introduce the concept of higher order strongly m-convex functions and its variant forms.

Definition 9 A function F on the convex set K_m is said to be higher order strongly m-convex, if there exists a constant $\mu > 0$, such that

$$F(u + t(mv - u)) \leq (1-t)F(u)+tF(mv)-\mu\{t^p(1-t)+t(1-t)^p\}\|mv - u\|^p,$$
$$\forall u, mv \in K_m, t \in [0, 1], m \in (0, 1], p > 1. \tag{13}$$

A function F is said to higher order strongly m-concave, if and only if, $-F$ is higher order strongly m-convex function.

If $t = \frac{1}{2}$ and $\mu = 1$, then

$$F\left(\frac{u+mv}{2}\right) \leq \frac{F(u)+F(mv)}{2}-\mu\frac{1}{2^p}\|mv - u\|^p, \forall u, mv \in K_m, m \in (0, 1], p > 1. \tag{14}$$

The function F is said to be higher order strongly Jm-convex function.

For $m = 1$, Definition 9 reduces to:

Definition 10 A function F on the convex set K is said to be higher order strongly convex, if there exists a constant $\mu > 0$, such that

$$F(u+t(v-u)) \leq (1-t)F(u)+tF(v)-\mu\{t^p(1-t)+t(1-t)^p\}\|v-u\|^p,$$
$$\forall u, v \in K, t \in [0,1]. \tag{15}$$

A function F is said to higher order strongly concave, if and only if, $-F$ is higher order strongly convex.

If $t = \frac{1}{2}$ and $\mu = 1$, then

$$F\left(\frac{u+v}{2}\right) \leq \frac{F(u)+F(v)}{2} - \mu\frac{1}{2^p}\|v-u\|^p, \forall u, v \in K, p > 1. \tag{16}$$

The function F is said to be higher order strongly J-convex function.

Definition 11 A function F on the m-convex set K_m is said to be a higher order strongly affine m-convex, if there exists a constant $\mu > 0$, such that

$$F(u+t(mv-u))=(1-t)F(u)+tF(mv)-\mu\{t^p(1-t)+t(1-t)^p\}\|mv-u\|^p, \tag{17}$$
$$\forall u, mv \in K_m, t \in [0,1], P > 1.$$

Note that if a functions is both higher order strongly m-convex and higher order strongly m-concave, then it is higher order strongly affine m-convex function.

A function F is called higher order strongly quadratic equation, if there exists a constant $\mu > 0$, such that

$$F\left(\frac{u+mv}{2}\right)=\frac{F(u)+F(mv)}{2}-\mu\frac{1}{2^p}\|mv-u\|^p, \forall u, mv \in K_m, t \in [0,1]. \tag{18}$$

This function F is also called higher order strongly affine Jm-convex function.

We now discuss some special cases.

I. If $p = 2$, then the higher order strongly convex function becomes strongly convex functions, that is,

$$F(u+t(mv-u)) \leq (1-t)F(u)+tF(mv)-\mu t(1-t)\|mv-u\|^2, \forall u, v \in K_m, t \in [0,1].$$

For the properties of the strongly convex functions in variational inequalities and equilibrium problems, see Noor [15].

Definition 12 A function F on the m-convex set K_m is said to be higher order strongly quasi m-convex, if there exists a constant $\mu > 0$ such that

$$F(u+t(mv-u) \leq \max\{F(u), F(mv)\} - \mu\{t^p(1-t)+t(1-t)^p\}\|mv-u\|^p,$$
$$\forall u, v \in K_m, t \in [0,1], p > 1.$$

Definition 13 A function F on the m-convex set K_m is said to be higher order strongly log m-convex, if there exists a constant $\mu > 0$ such that

$$F(u + t(mv - u)) \leq (F(u))^{1-t}(F(mv))^t - \mu\{t^p(1-t) + t(1-t)^p\}\|mv - u\|^p,$$
$$\forall u, mv \in K_m, t \in [0, 1],$$

where $F(\cdot) > 0$.

From the above definitions, we have

$$\begin{aligned} F(u + t(mv - u)) &\leq (F(u))^{1-t}(F(mv))^t - \mu\{t^p(1-t) + t(1-t)^p\}\|mv - u\|^p \\ &\leq (1-t)F(u) + tF(mv) - \mu\{t^p(1-t) + t(1-t)^p\}\|mv - u\|^p \\ &\leq \max\{F(u), F(mv)\} - \mu\{t^p(1-t) + t(1-t)^p\}\|mv - u\|^p. \end{aligned}$$

This shows that every higher order strongly log m-convex function is a higher order strongly m-convex function and every higher order strongly m-convex function is a higher order quasi m-convex function. However, the converse is not true.

Definition 14 An operator $T : K_m \to H$ is said to be:

1. Higher order strongly m-monotone, if and only if, there exists a constant $\alpha > 0$ such that

$$\langle Tu - T(mv), u - mv\rangle \geq \alpha\|mv - u\|^p, \forall u, mv \in K_m.$$

2. Higher order strongly m-pseudomonotone, if and only if, there exists a constant $\nu > 0$ such that

$$\langle Tu, mv - u\rangle + \nu\|mv - u\|^p \geq 0$$
$$\Rightarrow$$
$$\langle T(mv), mv - u\rangle \geq 0, \forall u, mv \in K_m.$$

3. Higher order strongly relaxed m-pseudomonotone, if and only if, there exists a constant $\mu > 0$ such that

$$\langle Tu, mv - u\rangle \geq 0$$
$$\Rightarrow$$
$$-\langle T(mv), u - mv\rangle + \mu\|mv - u\|^p \geq 0, \forall u, v \in K_m.$$

Definition 15 A differentiable function F on the convex set K_m is said to be higher order strongly pseudo m-convex function, if and only if, if there exists a constant $\mu > 0$ such that

$$\langle F'(u), mv - u\rangle + \mu\|mv - u\|^p \geq 0 \Rightarrow F(mv) \geq F(u), \forall u, mv \in K_m.$$

4 Properties of Strongly m-Convex Functions

In this section, we consider some basic properties of higher order strongly m-convex functions.

Theorem 10 *Let F be a differentiable function on the m-convex set K_m. Then the function F is higher order strongly m-convex function, if and only if,*

$$F(mv) - F(u) \geq \langle F'(u), mv - u\rangle + \mu\|mv - u\|^p, \forall mv, u \in K_m. \tag{19}$$

Proof Let F be a higher order strongly m-convex function on the convex set K_m. Then

$$F(u+t(mv-u)) \leq (1-t)F(u)+tF(mv)-\mu\{t^p(1-t)+t(1-t)^p\}\|mv-u\|^p,$$
$$\forall u, mv \in K_m,$$

which can be written as

$$F(mv)-F(u) \geq \{\frac{F(u+t(mv-u))-F(u)}{t}\}+\{t^{p-1}(1-t)+(1-t)^p\}\|mv-u\|^p.$$

Taking the limit in the above inequality as $t \to 0,$, we have

$$F(mv) - F(u) \geq \langle F'(u), mv - u)\rangle + \mu\|mv - u\|^p, \forall u, mv \in K_m.$$

which is (19), the required result.

Conversely, let (19) hold. Then, $\forall u, mv \in K, t \in [0, 1]$, $v_t = u+t(mv-u) \in K$, we have

$$\begin{aligned} F(mv) - F(v_t) &\geq \langle F'(v_t), mv - v_t)\rangle + \mu\|mv - v_t\|^p \\ &= (1-t)F'(v_t), mv-u\rangle+\mu(1-t)^p\|mv-u\|^p, \forall u, mv \in K_m \end{aligned} \tag{20}$$

In a similar way, we have

$$\begin{aligned} F(u) - F(v_t) &\geq \langle F'(v_t), u - v_t)\rangle + \mu\|u - v_t\|^p \\ &= -tF'(v_t), mv - u\rangle + \mu t^p\|mv - u\|^p. \end{aligned} \tag{21}$$

Multiplying (20) by t and (21) by $(1 - t)$ and adding the resultant, we have

$$F(u+t(mv-u)) \leq (1-t)F(u)+tF(mv)-\mu\{t^p(1-t)+t(1-t)^p\}\|mv-u\|^p,$$
$$\forall u, mv \in K_m,$$

showing that F is a higher order strongly m-convex function.

Theorem 11 *Let F be a differentiable higher order strongly m-convex function on the m-convex set K_m. Then $F'(.)$ is a higher order strongly m-monotone operator.*

Proof Let F be a higher order strongly m-convex function on the m-convex set K_m. Then, from Theorem 10. we have

$$F(mv) - F(u) \geq \langle F'(u), mv - u\rangle + \mu\|mv - u\|^p, \quad \forall u, mv \in K_m. \tag{22}$$

Changing the role of u and mv in (22), we have

$$F(u) - F(mv) \geq \langle F'(mv), u - mv)\rangle + \mu\|mv - u\|^p, \quad \forall u, mv \in K_m. \tag{23}$$

Adding (22) and (23), we have

$$\langle F'(u) - F'(mv), u - mv\rangle \geq 2\mu\|mv - u\|^p, \quad \forall u, mv \in K_m. \tag{24}$$

which shows that $F'(.)$ is a higher order strongly monotone operator.

We remark that the converse of Theorem 11 is not true. However, we have the following result.

Theorem 12 *If the differential operator $F'(.)$ of a differentiable higher order strongly m-convex function F is higher order strongly m-monotone operator, then*

$$F(mv) - F(u) \geq \langle F'(u), mv - u\rangle + 2\mu\frac{1}{p}\|mv - u\|^p, \forall u, mv \in K_m. \tag{25}$$

Proof Let $F'(.)$ be a higher order strongly m-monotone operator. Then, from (24), we have

$$\langle F'(v), u - v\rangle \geq \langle F'(u), u - v)\rangle + 2\mu\|v - u\|^p. \quad \forall u, v \in K. \tag{26}$$

Since K is an convex set, $\forall u, mv \in K_m, t \in [0, 1]$, $v_t = u + t(mv - u) \in K_m$. Taking $v = v_t$ in (26, we have

$$\begin{aligned}\langle F'(v_t), u - v_t\rangle &\leq \langle F'(u), u - v_t\rangle - 2\mu\|mv - u\|^p \\ &= -t\langle F'(u), v - u\rangle - 2\mu t^p\|mv - u\|^p,\end{aligned}$$

which implies that

$$\langle F'(v_t), mv - u\rangle \geq \langle F'(u), mv - u\rangle + 2\mu t^{p-1}\|mv - u\|^P. \tag{27}$$

Consider the auxiliary function

$$g(t) = F(u + t(mv - u), \forall u, mv \in K_m,$$

from which, we have

$$g(1) = F(mv), \quad g(0) = F(u).$$

Then, from (27), we have

$$g'(t) = \langle F'(v_t, mv - u\rangle \geq \langle F'(u), mv - u\rangle + 2\mu t^{p-1}\|mv - u\|^p. \tag{28}$$

Integrating (28) between 0 and 1, we have

$$\begin{aligned} g(1) - g(0\ ) &= \int_0^1 g'(t)dt \\ &\geq \langle F'(u), mv - u\rangle + 2\mu\frac{1}{p}\|mv - u\|^p. \end{aligned}$$

Thus it follows that

$$F(mv) - F(u) \geq \langle F'(u), mv - u\rangle + 2\mu\frac{1}{p}\|v - u\|^p, \forall u, mv \in K_m,$$

which is the required (25).

We note that, if $p = 2$, then Theorem 12 can be viewed as the converse of Theorem 11.

We now give a necessary condition for higher order strongly pseudo m-convex function.

Theorem 13 *Let $F'(.)$ be a higher order strongly relaxed pseudo m-monotone operator. Then F is a higher order strongly pseudo m-connvex function.*

Proof Let F' be a higher order strongly relaxed pseudo m-monotone operator. Then, $\forall u, mv \in K_m$,

$$\langle F'(u), mv - u\rangle \geq 0.$$

implies that

$$\langle F'(mv), mv - u\rangle \geq \mu\|mv - u\|^p, \forall u, mv \in K_m. \tag{29}$$

Since K_m is an m-convex set, $\forall u, mv \in K_m, \quad t \in [0, 1]$, $v_t = u + t(mv - u) \in K$.

Taking $v = v_t$ in (29), we have

$$\langle F'(v_t), mv - u\rangle \geq \mu t^{p-1}\|mv. - u\|^p. \tag{30}$$

Consider the auxiliary function

$$g(t) = F(u + t(mv - u)) = F(v_t), \quad \forall u, mv \in K_m, t \in [0, 1],$$

which is differentiable, since F is differentiable function. Then, using (30), we have

$$g'(t) = \langle F'(v_t), mv - u)\rangle \geq \mu t^{p-1}\|mv - u\|^p.$$

Integrating the above relation between 0 to 1, we have

$$g(1) - g(0) = \int_0^1 g'(t)dt \geq \frac{\mu}{p}\|mv - u\|^p,$$

that is,

$$F(mv) - F(u) \geq \frac{\mu}{p}\|mv - u\|^p), \forall u, mv \in K_m,$$

showing that F is a higher order strongly pseudo m-convex function.

Definition 16 A function F is said to be sharply higher order strongly pseudo m-convex, if there exists a constant $\mu > 0$ such that

$$\langle F'(u), mv - u\rangle \geq 0$$

$$\Rightarrow$$

$$F(mv) \geq F(mv+t(u-mv))+\mu\{t^p(1-t)+t(1-t)^p\}\|mv - u\|^p, \forall u, mv \in K_m.$$

Theorem 14 *Let F be a sharply higher order strongly pseudo m-convex function on K_m with a constant $\mu > 0$. Then*

$$\langle F'(mv), mv - u\rangle \geq \mu\|mv - u\|^p, \forall u, mv \in K_m.$$

Proof Let F be a sharply higher order strongly pesudo m-convex function on K_m. Then

$$F(mv) \geq F(mv + t(u - mv)) + \mu\{t^p(1 - t) + t(1 - t)^p\}\|mv - u\|^p,$$
$$\forall u, mv \in K_m, t \in [0, 1],$$

from which, we have

$$\{\frac{F(mv + t(u - mv)) - F(mv)}{t}\} + \mu\{t^{p-1}(1 - t) + (1 - t)^p\}\|mv - u\|^p \geq 0.$$

Taking limit in the above inequality, as $t \to 0$, we have

$$\langle F'(mv), mv - u\rangle \geq \mu\|mv - u\|^p, \forall u, mv \in K_m,$$

the required result.

Definition 17 A function F is said to be a pseudo $m$convex function, if there exists a strictly positive bifunction $b(.,.)$, such that

$$F(mv) < F(u)$$

$$\Rightarrow$$

$$F(u + t(mv, u)) < F(u) + t(t-1)b(mv, u), \forall u, mv \in K_m, t \in [0, 1].$$

Theorem 15 *If the function F is higher order strongly m-convex function such that $F(mv) < F(u)$, then the function F is higher order strongly pseudo m-convex.*

Proof Since $F(mv) < F(u)$ and F is higher order strongly m-convex function, then $\forall u, mv \in K_m, \quad t \in [0, 1]$, we have

$$\begin{aligned} F(u + t(mv - u)) &\leq F(u) + t(F(mv) - F(u)) - \mu\{t^p(1-t) + t(1-t)^p\}\|mv - u\|^p \\ &< F(u) + t(1-t)(F(mv) - F(u)) - \mu\{t^p(1-t) + t(1-t)^p\}\|mv - u\|^p \\ &= F(u) + t(t-1)(F(u) - F(mv)) - \mu\{t^p(1-t) + t(1-t)^p\}\|mv - u\|^p \\ &< F(u) + t(t-1)b(u, mv) - \mu\{t^p(1-t)+t(1-t)^p\}\|mv - u\|^p, \forall u, mv \in K_m, \end{aligned}$$

where $b(u, mv) = F(u) - F(mv) > 0$. Thus, it show that the function$F$ is a higher order strongly m-convex function.

We now discuss the optimality for the differentiable generalized strongly convex functions, which is the main motivation of our next result.

Theorem 16 *Let F be a differentiable higher order strongly m-convex function with modulus $\mu > 0$. If $u \in K_m$ is the minimum of the function F, then*

$$F(mv) - F(u) \geq \mu\|mv - u\|^p, \quad \forall u, mv \in K_m. \tag{31}$$

Proof *Let $u \in K_m$ be a minimum of the function F. Then*

$$F(u) \leq F(mv), \forall mv \in K_m. \tag{32}$$

Since K is a m-convex set, so, $\forall u, mv \in K_m, \quad t \in [0, 1]$,

$$v_t = (1-t)u + tmv \in K_m.$$

Taking $v = v_t$ in (32), we have

$$0 \leq \lim_{t \to 0}\{\frac{F(u + t(mv - u)) - F(u)}{t}\} = \langle F'(u), mv - u\rangle. \tag{33}$$

Since F is differentiable higher order strongly m-convex function, so

$$F(u+t(mv-u)) \leq F(u)+t(F(mv)-F(u))-\mu\{t^p(1-t)+t(1-t)^p\}\|mv-u\|^p,$$
$$\forall u, mv \in K_m,$$

from which, using (33), we have

$$\begin{aligned} F(mv)-F(u) &\geq \lim_{t\to 0}\{\frac{F(u+t(mv-u))-F(u)}{t}\} \\ &\quad +\mu\{t^{p-1}(1-t)+(1-t)^p\}\|mv-u\|^p \\ &= \langle F'(u), mv-u\rangle + \mu\|mv-u\|^p, \end{aligned}$$

the required result (31).

Remark 2 We would like to mention that, if

$$\langle F'(u), mv-u\rangle + \mu\|mv-u\|^p \geq 0, \quad \forall u, mv \in K_m,$$

then $u \in K_m$ is the minimum of the function F.

Theorem 17 *Let f be a higher order strongly affine m-convex function. Then F is a higher order strongly m-convex function, if and only if, $g = F - f$ is a m-convex function.*

Proof Let f be a higher order strongly affine m-convex function, Then

$$f((1-t)u+tmv)=(1-t)f(u)+tf(mv)-\mu\{t^p(1-t)+t(1-t)^p\}\|mv-u\|^p,$$
$$\forall u, mv \in K_m. \tag{34}$$

From the higher order strongly convexity of F, we have

$$F((1-t)u+tmv) \leq (1-t)F(u)+tF(mv)-\mu\{t^p(1-t)+t(1-t)^p\}\|mv-u\|^p,$$
$$\forall u, mv \in K_m. \tag{35}$$

From (34) and (35), we have

$$\begin{aligned} F((1-t)u+tmv)-f((1-t)f(u)+tf(mv) &\leq (1-t)(F(u)-f(u)) \\ &\quad +t(F(mv)-f(mv)), \end{aligned} \tag{36}$$

from which it follows that

$$\begin{aligned} g((1-t)u+tmv) &= F((1-t)u+tmv)-f((1-t)u+tmv) \\ &\leq (1-t)F(u)+tF(mv)-(1-t)f(u(-tf(mv) \\ &= (1-t)(F(u)-f(u))+t(F(mv)-f(mv)), \end{aligned}$$

which show that $g = F - f$ is a m-convex function.

The inverse implication is obvious.

5 Applications

In this section, we show that the characterizations of uniformly Banach spaces involving the notion of higher order strong m-convexity.

Setting $F(u) = \|u\|^p$ in Definition 9, we have

$$\|u+t(mv-u)\|^p \leq (1-t)\|u\|^p+t\|mv\|^p-\mu\{t^p(1-t)+t(1-t)^p\}\|mv-u\|^p, \quad (37)$$
$$\forall u, mv \in K_m, t \in [0,1].$$

Taking $t = \frac{1}{2}$ in (37), we have

$$\|\frac{u+mv}{2}\|^p + \mu\frac{1}{2^p}\|mv-u\|^p \leq \frac{1}{2}\|u\|^p + \frac{1}{2}\|mv\|^p, \forall u, mv \in K_m, \quad (38)$$

which implies that

$$\|u+mv\|^p + \mu\|mv-u\|^p \leq 2^{p-1}\{\|u\|^p + \|mv\|^p\}, \forall u, mv \in K_m, \quad (39)$$

which is known as the lower parallelogram for the l^p-spaces. In a similar way, one can obtain the upper parallelogram law as

$$\|u+mv\|^p + \mu\|mv-u\|^p \geq 2^{p-1}\{\|u\|^p + \|mv\|^p\}, \forall u, mv \in K_m, \quad (40)$$

From Definition 11, we have

$$\|u+mv\|^p + \mu\|mv-u\|^p = 2^{p-1}\{\|u\|^p + \|mv\|^p\}, \forall u, mv \in K_m, \quad (41)$$

which is known as the parallelogram for the l^p-spaces.

Note that, for $m = 1$, we obtain the parallelogram laws for uniformly Banach spaces. To be more precise, Xi [21] obtained the characterizations of p-uniform convexity and q-uniform smoothness of a Banach space via the functionals $\|.\|^p$ and $\|.\|^q$, respectively. Bynum [5] and Chen et al. [6–8] have studied the properties and applications of the parallelogram laws for the Banach spaces. It is interesting to note that these parallelogram laws follow from the concepts of the higher order strongly m-convex functions, which is surprising and novel applications of the higher order strongly convex functions. For the applications of the parallelogram laws in Banach spaces in prediction theory and applied sciences, see [5–8, 20] and the references therein.

Conclusion

In this paper, we have introduced and studied a new class of convex functions, which is called higher order strongly m-convex function. It is shown that several new classes of strongly convex functions can be obtained as special cases of these higher order strongly m-convex functions. We have studied the basic properties of these functions. We have derived new parallelogram laws as applications of the higher order strongly m-convex functions. Some known results can be obtained for suitable and appropriate choices of m, which have been used to characterize the p-uniform convexity and q-uniform smoothness of a Banach spaces. The interested readers may explore the applications and other properties of the higher order strongly convex functions in various fields of pure and applied sciences. This is an interesting direction of future research.

Acknowledgments The authors would like to thank the Rector, COMSATS University Islamabad, Islamabad, Pakistan, for providing excellent research and academic environments. Authors are grateful to Prof. Dr. Themistocles M. Rassias for his kind invitation and support.

References

1. M. Adamek, On a problem connected with strongly convex functions. Math. Inequ. Appl. **19**(4), 1287–1293 (2016)
2. H. Angulo, J. Gimenez, A.M. Moeos, K. Nikodem, On strongly h-convex functions. Ann. Funct. Anal. **2**(2), 85–91 (2011)
3. M.U. Awan, M.A. Noor, T.-S. Du, K.I. Noor, New refinemnts of fractional Hermite–Hadamard inequality. RACSAM, **113**, 21–29 (2019)
4. A. Azcar, J. Gimnez, K. Nikodem. J.L. Snchez, On strongly midconvex functions. Opuscula Math. **31**(1), 15–26 (2011)
5. W.L. Bynum, Weak parallelogram laws for Banach spaces. Can. Math. Bull. **19**, 269–275 (1976)
6. R. Cheng, C.B. Harris, Duality of the weak parallelogram laws on Banach spaces. J. Math. Anal. Appl. **404**, 64–70 (2013)
7. R. Cheng, W.T. Ross, Weak parallelogram laws on Banach spaces and applications to prediction. Period. Math. Hung. **71**, 45–58 (2015)
8. R. Cheng, J. Mashreghi, W.T. Ross, Optimal weak parallelogram constants for L_p space. Math. Inequal. Appl. **21**(4), 1047–1058 (2018)
9. S. Karamardian, The nonlinear complementarity problems with applications, Part 2. J. Optim. Theory Appl. **4**(3), 167–181 (1969)
10. T. Lara, N. Merentes, R. Quintero, E. Rosales, Strongly m-convex functions. Math. Aeterna, **5**(3), 521–535 (2015)
11. G.H. Lin, M. Fukushima, Some exact penalty results for nonlinear programs and mathematical programs with equilibrium constraints. J. Optim. Theory Appl. **118**(1), 67–80 (2003)
12. S.M. Mishra, N. Sharma, On strongly generalized convex functions of higher order. Math. Inequal. Appl. **22**(1), 111–121 (2019)
13. C.P. Niculescu, L.E. Persson, *Convex Functions and their Applications* (Springer, New York, 2018)

14. K. Nikodem, Z.S. Pales, Characterizations of inner product spaces by strongly convex functions. Banach J. Math. Anal., **1**, 83–87 (2011)
15. M.A. Noor, Fundamentals of equilibrium problems. Math. Inequal. Appl. **9**(3), 529–566 (2006)
16. M.A. Noor, K.I. Noor, On generalized strongly convex functions involving bifunction. Appl. Math. Inf. Sci. **13**(3), 411–416 (2019)
17. M.A. Noor, K.I. Noor, Properties of exponentially m-convex functions, in *Nonlinear Analysis and Global Optimization* eds. by P. Pardalos, T.M. Rassias (Springer, Berlin, 2019)
18. Z. Pavic, A. Ardic, The most important inequalities of m-convex functions. Turkish J. Math. **41**, 625–635 (2017)
19. J. Pecric, F. Proschan, Y. I. Tong, *Convex Functions, Partial Ordering and Statistical Applications* (Academic Press, New York, 1992)
20. B.T. Polyak, Existence theorems and convergence of minimizing sequences in extremum problems with restrictions. Soviet Math. Dokl. **7**, 2–75 (1966)
21. G. Qu, N. Li. On the exponentially stability of primal-dual gradeint dynamics, IEEE Control Syst. Letters **3**(1), 43–48 (2019)
22. G. Toader, Some generalizations of the convexity. Proc. Colloq. Approx. Optim. Cluj-Naploca (Romania), 329–338 (1984)
23. S. Toader, The order of a star-convex function. Bull. Appl. Comp. Math. (Budapest) **85-B** (BAM-1473), 347–350 (1998)
24. D.L. Zu, P. Marcotte, Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. SIAM J. Optim. **6**(3), 714–726 (1996)

Characterizations of Higher Order Strongly Generalized Convex Functions

Muhammad Aslam Noor, Khalida Inayat Noor, and Michael Th. Rassias

Abstract In this paper, we define and consider some new concepts of the higher order strongly generalized convex functions with respect to two arbitrary functions. Some properties of the higher order strongly generalized convex functions are investigated under suitable conditions. It is shown that the operator parallelogram laws for the characterization of uniformly Banach spaces can be obtained as a novel applications of higher order strongly affine functions. It is shown that the optimality conditions of the higher order strongly generalized convex functions are characterized by a new class of variational inequalities. Some special cases also discussed. Results obtained in this paper can be viewed as significant refinement and improvement of previously known results.

1 Introduction

Convexity theory is a branch of mathematical sciences with a wide range of applications in industry, physical, social, regional and engineering sciences. The general theory of the convexity started soon after the introduction of differential and integral calculus by Newton and Leibnitz, although some individual optimization problems had been investigated before that. To be more specific, Bernoulli's brothers (1697) were the first, who considered the variational problems in mathematical terms. It

In: Nonlinear Analysis, Differential Equations and Applications (Edit: Themistocles M. Rassias)

M. A. Noor (✉) · K. I. Noor
COMSATS University Islamabad, Islamabad, Pakistan

M. T. Rassias
Institute of Mathematics, University of Zurich, Zurich, Switzerland
Moscow Institute of Physics and Technology, Dolgoprudny, Russia

Institute for Advanced Study, Program in Interdisciplinary Studies, Princeton, NJ, USA
e-mail: michail.rassias@math.uzh.ch

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_15

is worth mentioning that the first phase of the development of calculus of variations was characterized by a combination of philosophical concepts, mathematical methods and physical problems. Euler (eighteenth century) created a new branch of mathematics known as the calculus of variations. Motivated by geometrical considerations, Euler deduced his first principle which is now referred to as Eulers differential equation for the determination of maximizing and minimizing arcs. By variational principles, we mean maximum and minimum problems arising in game theory, mechanics, geometrical optics, general relativity theory, economics, transportation, differential geometry and related areas. The Hamiltonian-Jacobi theory represents a general framework for the mathematical description of the propagation of actions in nature and optimal modeling of control processes in daily life. It is known that the gauge field theories are a continuation of Einsteins concept of describing physical effects mathematically in terms of differential geometry. These theories play a fundamental role in the modern theory of elementary particles and are right tool of building up a unified theory of elementary particles, which includes all kind of known interactions. For example, the Weinberg-Salam theory unifies weak and electromagnetic interactions. It is also known that the variational formulation of field theories allows for a degree of unification absent their versions in terms of differential equations. Convexity plays an important part in the existence and stability of soliton, which occur in almost every branch of physics.

Variational inequalities represent the optimality conditions for the differentiable convex functions on the convex sets in a normed space. which were introduced and considered in early 1960s by Stampacchia [41], Variational inequalities combine both theoretical and algorithmic advances with new and novel domain of applications. Analysis of these problems requires a blend of techniques from convex analysis, functional analysis and numerical analysis. In recent years, considerable interest has been shown in developing various extensions and generalizations of variational inequalities, both for their own sake and their applications.

Mohsen et al. [19] and Noor and Noor [30–34, 37] introduced the concept of higher order strongly convex functions and studied their properties. These results can be viewed as a significant refinement of the results of Adamek [1], Alabdali et al. [2] and Lin and Fukushima [18]. Higher order strongly convex functions include the strongly convex functions, which were introduced and studied by Polyak [39]. Karmardian[15] used the strongly convex functions to discuss the unique existence of a solution of the nonlinear complementarity problems. Awan et al. [4, 5] have derived Hermite–Hadamard type inequalities for various classes of strongly convex functions, which provide upper and lower estimate for the integrand. For the applications of strongly convex functions in optimization, variational inequalities and other branches of pure and applied sciences, see [1–8, 8–18, 20, 22–32, 34, 43–45] and the references therein.

It is known that the properties of the convex functions may not hold, in general, when the convex set is non-convex. Recently, the concept of convexity has been generalized in several directions. A significant generalization of the convex set is the introduction of the (h, g)-convex set and (h, g)-convex [14, 28, 29] function involving two arbitrary functions h and g, (say). These (h, g) convex functions are

called generalized convex functions. It has been shown that that the generalized convex functions enjoy some nice properties which convex function have. We would like to emphasize that the (h, g)-convex set and generalized convex functions may not be convex sets and convex functions. For $h = g$, generalized functions reduces the g-convex functions, which were introduced and studied by Youness [43]. It is known [24] that the minimum of a differentiable g-convex function on the g-convex set can be characterized by a class of variational inequalities, which is known as general (Noor) variational inequalities introduced and studied by Noor [22, 24, 25] in 1988. For suitable and appropriate choices of the arbitrary functions h and g, one can obtain a wide classes of known and new classes of convex functions and related general variational inequalities. This clearly shows that the (h, g)-convex sets and the generalized convex functions are quite flexible and unifying ones. For the formulation, applications, numerical methods,sensitivity analysis and other aspects of general variational inequalities, see [11–14, 19–29, 34, 35, 37] and the references es therein.

Noor and Noor [32] have considered and studied the strongly generalized functions and studied their properties. Motivated by the work of Noor et al. [30–34], Alabdali et al. [2] and Lin and Fukushima [18], we introduce and investigate a new class of higher order strongly generalized convex functions with respect to two arbitrary functions, which is the main motivation of this paper. Several new concepts of monotonicity are introduced. We establish the relationship between these classes and derive some new results under some mild conditions. It is shown that (operator) parallelograms laws, which characterize the uniform Banach spaces can be obtained from these definitions. We have shown that the minimum of the a higher order strongly generalized convex functions on the generalized convex set can be characterized by higher order strongly generalized variational inequalities. Some new special cases are discussed, which can be viewed itself an novel and interesting applications of the higher order strongly generalized convex functions.

For the sake of readers convenience, we include all the relevant all the details. In Section 2, we recall the baric concepts and results. Generalized convex functions and their properties are discussed in Section 3. In Section 4, higher order strongly generalized convex are introduced. It is shown that several important special cases are discussed, which can be viewed significant refinement of the previous known results. Properties of the differential higher order strongly generalized convex are proved. It is shown that the parallelogram laws can be obtained from the generalized affine functions, which itself a novel application. These concepts are discussed in Section 5. Connection with generalized variational inequalities is investigated with the optimality conditions of the differentiable generalized convex functions, which is the main motivation of Section “Conclusion”. We would like to emphasize that the results obtained and discussed in this paper may motivate and bring a large number of novel, innovate and potential applications, extensions and interesting topics in these areas. We have given only a brief introduction of higher order strongly generalized convex functions and applications. The interested reader is advised to explore this field further and discover novel and fascinating applications of the generalized convex functions in other areas of sciences.

2 Formulations and Basic Facts

Let K be a nonempty closed set in a real Hilbert space H. We denote by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ be the inner product and norm, respectively.

Definition 1 ([10, 20]) A set K in H is said to be a convex set, if

$$u + t(v - u) \in K, \qquad \forall u, v \in K, t \in [0, 1].$$

Definition 2 A function F is said to be convex function, if

$$F((1 - t)u + tv) \leq (1 - t)F(u) + tF(v); \forall u, v \in K, t \in [0, 1]. \tag{1}$$

It is well known that $u \in K$ of a differential convex functions F is equivalent to finding $u \in K$ such that

$$\langle F'(u(, v - u\rangle \geq 0, \forall v \in K, \tag{2}$$

which is called the variational inequality, introduced and studied by Stampacchia [41]. Variational inequalities can be regarded as a novel and significant extension of variational principles, the origin of which can be traced back to Euler, Lagrange, Newton, and Bernoulli brothers.

We would like to mention that the underlying the set may not be a convex set in many important applications. To overcome this drawback, the set can be made convex set with respect to an arbitrary function, which is called general convex set.

We would like to mention that the underlying the set may not be a convex set in many important applications. To overcome this drawback, the set can be made convex set with respect to two arbitrary functions, which is called a generalized or (h, g)-convex set [14, 27, 28].

Definition 3 ([14, 28]) The set $K \subseteq H$is said to be a (h, g)-convex set, if there are two functions h and g such as

$$(1 - t)h(u) + tg(v) \in K; \quad \forall u, v \in K, t \in [0, 1]. \tag{3}$$

We now discuss some special cases of the (h, g)-convex set $K \subseteq D$.

(I). If $g(u) = I(u) = u = h(u)$, the identity operator, then (h, g)-convex set reduces to the classical convex set. Clearly every convex set is a (h, g)-convex set, but the converse is not true.

(II). If $h(u) = I(u) = u$, then the (h, g)-convex set becomes the g-convex set, that is,

Definition 4 The set K is said to be g-convex set, if

$$(1 - t)u + tg(v) \in K \subseteq D, \quad \forall u, v \in K \subseteq D, t \in [0, 1],$$

which was introduced and studied by Noor [27]. Cristescu et al. [9] discussed various applications of the general convex sets related to the necessity of adjusting investment or development projects out of environmental or social reasons. For example, the easiest manner of constructing this kind of convex sets comes from the problem of modernizing the railway transport system. Shape properties of the general convex sets with respect to a projection are investigated.

(III). If $g(u) = I(u) = u$, then the (h, g)-convex set becomes the h-convex set, that is,

Definition 5 The set K is said to be h-convex set, if

$$(1-t)h(u) + tv \in K \subseteq D, \qquad \forall u, v \in K \subseteq D, t \in [0,1],$$

which is mainly due to Noor [28].

For the sake of simplicity, we always assume that function $F : D \to R$ and $K \cup h(K) \cup g(K) \subseteq D$. If K is (g, h)-convex set, then this condition becomes $K \subseteq D$ and (h, g)-convex is called generalized convex functions unless otherwise specified.

Definition 6 A function F is said to be a generalized convex function on the (h, g)-convex set $K \subseteq D$, if there exist two arbitrary non-negative functions h, g such that

$$F((1-t)h(u)+tg(v)) \leq (1-t)F(h(u))+tF(g(v)), \ \forall u, v \in K \subseteq D, t \in [0,1]. \quad (4)$$

The generalized convex functions were introduced by Noor [28]. Noor [28] proved that the minimum $u \in K \subseteq D$ of a differentiable generalized convex functions F can be characterized by the class of variational inequalities of the type:

$$\langle F'(h(u)), g(v) - h(u) \rangle \geq 0, \quad \forall v \in K \subseteq D, \quad (5)$$

which is known as the extended general variational inequalities. For the applications of the general variational inequalities in various branches of pure and applied sciences, see [22–25, 27–29, 37] and the references therein.

We now define the generalized convex functions on the interval $K = I_{hg} = [h(a), g(b)]$.

Definition 7 Let $I_{hg} = [h(a), g(b)]$. Then F is a generalized convex function, if and only if,

$$\begin{vmatrix} 1 & 1 & 1 \\ h(a) & x & g(b) \\ F(h(a) & F(x) & F(g(b)) \end{vmatrix} \geq 0; \quad h(a) \leq x \leq g(b).$$

One can easily show that the following are equivalent:

1. F is a generalized convex function.
2. $F(x) \leq F(h(a)) + \frac{F(g(b)-F(h(a)}{g(b)-h(a)}(x - h(a))$.
3. $\frac{F(x)-F(h(a))}{x-h(a)} \leq \frac{F(g(b)-F(h(a))}{g(b)-h(a)}$.
4. $(g(b) - x)F(h(a)) + (h(a) - g(b))F(x) + (x - h(a))F(g(b)) \geq 0$.
5. $\frac{F(a)}{(g(b)-h(a))(h(a)-x)} + \frac{F(x)}{(x-g(b))(a-x)} + \frac{F(g(b))}{(g(b)-h(a))(x-g(b))} \leq 0$,

where $x = (1 - t)h(a) + tg(b) \in [h(a), g(b)]$.

Definition 8 The function F on the (h, g)-convex set K is said to be a generalized quasi-convex, if

$$F(h(u) + t(g(v) - h(u))) \leq \max\{F(h(u)), F(g(v))\}, \quad \forall u, v \in K, t \in [0, 1].$$

Definition 9 The function F on the (h, g)-convex set K is said to be a generalized log-convex, if

$$F(h(u) + t(g(v) - h(u))) \leq (F(h(u))^{1-t}(F(g(v)))^t, \quad \forall u, v \in K, t \in [0, 1],$$

where $F(\cdot) > 0$.

From the above definitions, we have

$$\begin{aligned} F(h(u) + t(g(v) - h(u)) &\leq (F(h(u))^{1-t}(F(g(v)))^t \\ &\leq (1 - t)F(h(u)) + tF(g(v))) \\ &\leq \max\{F(h(u)), F(g(v))\}, \quad \forall u, v \in K, t \in [0, 1]. \end{aligned}$$

This shows that every generalized log-convex function is a generalized convex function and every generalized convex function is a generalized quasi-convex function. However, the converse is not true.

3 Generalized Convex Functions

In section, we now consider some basic properties of generalized convex functions.

Theorem 1 *Let F be a strictly generalized convex function. Then any local minimum of F is a global minimum.*

Proof Let the strictly generalized convex function F have a local minimum at $u \in K$. Assume the contrary, that is, $F(g(v)) < F(h(u))$ for some $g(v) \in K$. Since F is strictly generalized convex function, so

$$F(h(u) + t(g(v) - h(u))) < tF(g(v)) + (1 - t)F(h(u)), \quad \text{for} \quad 0 < t < 1.$$

Thus

$$F(h(u) + t(g(v) - h(u))) - F(h(u)) < -t[F(g(v)) - F(h(u))] < 0,$$

from which it follows that

$$F(h(u) + t(g(v) - h(u)) < F(h(u)),$$

for arbitrary small $t > 0$, contradicting the local minimum.

Theorem 2 *If the function F on the (h, g)-convex set K is generalized convex, then the level set*

$$L_\alpha = \{u \in K : F(h(u)) \leq \alpha, \quad \alpha \in R\}$$

is a (h, g)-convex set.

Proof Let $h(u), g(v) \in L_\alpha$. Then $F(h(u)) \leq \alpha$ and $F(g(v)) \leq \alpha$. Now, $\forall t \in (0, 1)$, $g(w) = u + t(g(v) - h(u)) \in K$, since K is a (h, g)-convex set. Thus, by the generalized convexity of F, we have

$$F(h(u)) + t(g(v) - h(u))) \leq (1 - t)F(h(u) + tF(g(v)) \leq (1 - t)\alpha + t\alpha = \alpha,$$

from which it follows that $h(u)+t(g(v)-h(u)) \in L_\alpha$. Hence L_α is a (h, g)-convex convex set.

Theorem 3 *The function F is generalized convex function, if and only if,*

$$epi(F) = \{h(u), \alpha) : h(u) \in K : F(h(u)) \leq \alpha, \alpha \in R\}$$

is a (h, g)-convex set.

Proof Assume that F is generalized convex function. Let

$$(h(u), \alpha), \quad (g(v), \beta) \in epi(F).$$

Then it follows that $F(h(u() \leq \alpha$ and $F(g(v)) \leq \beta$. Hence, we have

$$F(h(u) + t(g(v) - h(u))) \leq (1 - t)F(h(u)) + tF(g(v)) \leq (1 - t)\alpha + t\beta,$$

which implies that

$$((1 - t)h(u) + tg(v)), (1 - t)\alpha + t\beta) \in epi(F).$$

Thus $epi(F)$ is a (h, g)-convex set. Conversely, let $epi(F)$ be a (h, g)-convex set. Let $h(u), g(v) \in K$. Then $(h(u), F(h(u) \in epi(F)$ and $(g(v), F(g(v))) \in epi(F)$. Since F is a (h, g)-convex set, we must have

$$(h(u) + t(g(v) - h(u)), (1 - t)F(h(u)) + tF(g(v)) \in epi(F),$$

which implies that

$$F((1-t)h(u)+tg(v)) \leq (1-t)F(h(u)+tF(g(v)).$$

This shows that F is a generalized convex function.

Theorem 4 *The function F is a generalized quasi-convex, if and only if, the level set*

$$L_\alpha = \{h(u) \in K, \alpha \in R : F(h(u)) \leq \alpha\}$$

is a (h, g)-convex set.

Proof Let $h(u), g(v) \in L_\alpha$. Then $h(u), g(v) \in K$ and $\max(F(h(u)), F(g(v))) \leq \alpha$. Now for $t \in (0, 1)$, $g(w) = h(u) + t(g(v) - h(u)) \in K$. We have to prove that $h(u) + t(g(v) - h(u)) \in L_\alpha$. By the generalized convexity of F, we have

$$F(h(u) + t(g(v) - h(u))) \leq \max(F(h(u)), F(g(v))) \leq \alpha,$$

which implies that $h(u) + t(g(v) - h(u)) \in L_\alpha$, showing that the level set L_α is indeed a (h, g)-convex set.

Conversely, assume that L_α is a (h, g)-convex set. Then, $\forall \quad h(u), g(v) \in L_\alpha, t \in [0, 1]$, $h(u) + t(g(v) - h(u)) \in L_\alpha$. Let $h(u), g(v) \in L_\alpha$ for

$$\alpha = max(F(h(u)), F(g(v))) \quad \text{and} \quad F(g(v)) \leq F(h(u)).$$

Then from the definition of the level set L_α, it follows that

$$F(h(u) + t(g(v) - h(u)) \leq \max(F(h(u)), F(g(v))) \leq \alpha.$$

Thus F is an generalized quasi-convex function. This completes the proof.

Theorem 5 *Let F be a generalized convex function. Let $\mu = \inf_{h(u)\in K} F(u)$. Then the set*

$$E = \{h(u) \in K : F(h(u)) = \mu\}$$

is a (h, g)-convex set of K. If F is strictly generalized convex function, then E is a singleton.

Proof Let $h(u), g(v) \in E$. For $0 < t < 1$, let $g(w) = h(u) + t(g(v) - h(u))$. Since F is a generalized convex function, then

$$\begin{aligned} F(g(w)) &= F(h(u) + t(g(v) - h(u)) \\ &\leq (1-t)F(h(u)) + tF(g(v)) = t\mu + (1-t)\mu = \mu, \end{aligned}$$

which implies $g(w) \in E$ and hence E is a (h, g)-convex set. For the second part, assume to the contrary that $F(h(u)) = F(g(v)) = \mu$. Since K is a (h, g)-convex set, then for $0 < t < 1, h(u) + t(g(v) - h(u)) \in K$. Further, since F is a strictly generalized convex function, so

$$F(h(u) + t(g(v) - h(u))) < (1 - t)F(h(u)) + tF(g(v)) = (1 - t)\mu + t\mu = \mu.$$

This contradicts the fact that $\mu = \inf_{h(u)\in K} F(u)$ and hence the result follows.

Theorem 6 *If the function F is a generalized convex such that*

$$F(g(v)) < F(h(u)), \forall h(u), g(v) \in K,$$

then F is a strictly generalized quasi-convex function.

Proof By the generalized convexity of the function F, we have

$$\begin{aligned} F(h(u)+t(g(v)-h(u))) &\leq (1-t)F(h(u))+tF(g(v)), \forall h(u), g(v) \in K, t \in [0, 1] \\ &< F(h(u)), \end{aligned}$$

since $F(g(v)) < F(h(u))$, which shows that the function F is a strictly generalized quasi-convex.

4 Higher Order Strongly Generalized Convex Functions

In this section, some new classes of higher order strongly generalized convex functions and higher order strongly affine (h, g) functions on the (h, g)-convex set $K \subseteq D$.

Definition 10 A function F on the (h, g)-convex set $K \subseteq D$ is said to be higher order strongly generalized convex with respect to two functions h and g, if there exists a constant $\mu > 0$, such that

$$\begin{aligned} &F(h(u) + t(g(v) - h(u))) \\ &\leq (1 - t)F(h(u)) + tF(g(v)) - \mu\{t^p(1 - t) + t(1 - t)^p\}\|g(v) - h(u)\|^p, \\ &\forall u, v \in K \subseteq D, t \in [0, 1], p > 1. \end{aligned} \tag{6}$$

A function F is said to be higher order strongly generalized convex with respect to two functions h and g, if and only if, $-F$ is a higher order strongly generalized convex function with respect to two functions h and g.

If $t = \frac{1}{2}$ in Definition 10, then one gets the generalized Jensen-type property called higher order strongly generalized convex with respect to two functions h and g.

We now discuss some special cases.

(IV). If $p = 2$. then the Definition 10 reduces to:

Definition 11 A function F on the (h, g)-convex set $K \subseteq D$ is said to be strongly generalized convex with respect to two functions h and g, if there exists a constant $\mu > 0$, such that

$$\begin{aligned} F(h(u) + t(g(v) - h(u))) \leq (1-t)F(h(u)) + tF(g(v)) \\ -\mu\{t(1-t)\}\|g(v) - h(u)\|^2, \\ \forall u, v \in K \subseteq D, t \in [0,1], p > 1. \end{aligned} \tag{7}$$

For the characterizations and properties of the strongly generalized convex functions, see Noor and Noor [32].

(V). If $h(u) = I(u) = u$, then the higher order strongly generalized convex with respect to two functions h and g, becomes strongly g-convex functions, that is,

$$\begin{aligned} F(u + t(g(v) - u)) \leq (1-t)F(u) + tF(g(v)) - \mu\{t^p(1-t) \\ +t(1-t)^p\}\|g(v) - u\|^p, \\ \forall u, v \in K \subseteq D, t \in [0,1]. \end{aligned}$$

For the properties of the higher order strongly generalized convex functions in variational inequalities and equilibrium problems, see Noor [27].

(VI). If $g(u) = I(u) = u$, then the higher order strongly generalized convex with respect to two functions h and g, becomes higher order strongly h-convex functions, that is,

$$\begin{aligned} F(h(u) + t(v - h(u)) \leq (1-t)F(h(u)) + tF(v) - \mu\{t^p(1-t) \\ +t(1-t)^p\}\|v - h(u)\|^p, \\ \forall u, v \in K \subseteq D, t \in [0,1]. \end{aligned}$$

Definition 12 A function F on the (h, g)-convex set $K \subseteq D$ is said to be a higher order strongly generalized quasi-convex with respect to two functions h and g, if there exists a constant $\mu > 0$ such that

$$\begin{aligned} F(h(u) + t(g(v) - h(u)) \leq \max\{F(h(u)), F(g(v))\} - \mu\{t^p(1-t) \\ +t(1-t)^p\}\|g(v) - h(u)\|^p, \\ \forall u, v \in K \subseteq D, t \in [0,1]. \end{aligned}$$

Definition 13 A function F on the (h, g)-convex set $K \subseteq D$ is said to be higher order strongly generalized log-convex with respect to two functions h and g, if there exists a constant $\mu > 0$ such that

$$F(h(u) + t(g(v) - h(u)) \leq (F(h(u)))^{1-t}(F(g(v)))^t - \mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p,$$
$$\forall u, v \in K \subseteq D, t \in [0, 1],$$

where $F(\cdot) > 0$.

From the above definitions, we have

$$\begin{aligned} F(h(u) + t(g(v) - h(u)) &\leq (F(h(u)))^{1-t}(F(g(v)))^t - \mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p \\ &\leq (1-t)F(h(u)) + tF(g(v)) - \mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p \\ &\leq \max\{F(h(u)), F(g(v))\} - \mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p. \end{aligned}$$

This shows that every higher order strongly generalized log-convex function is a higher order strongly generalized convex function and every higher order strongly generalized convex function is a higher order strongly generalized quasi-convex function. However, the converse is not true.

Definition 14 A function F on the (h, g)-convex set $K \subseteq D$ is said to be a higher order strongly generalized affine with respect to two functions h and g, if there exists a constant $\mu > 0$, such that

$$F(h(u) + t(g(v) - h(u)) = (1-t)F(h(u)) + t(F(g(v)) - \mu\{t^p(1-t)$$
$$+t(1-t)^p\}\|g(v) - h(u)\|^p,$$
$$\forall u, v \in K \subseteq D, t \in [0, 1].$$

We would like to remark that, if $t = 1/2$ in Definition 14, then one gets the higher order generalized Jensen type property called higher order strongly generalized J-affine function.

For appropriate and suitable choice of the arbitrary functions h, g, one can obtain several new and known classes of and their variant forms as special cases of higher order strongly generalized convex functions with respect to two functions h and g. This shows that the class of higher order strongly generalized convex functions with respect to two functions h and g, is quite broad and unifying one.

We now introduce some new concepts and definitions.

Definition 15 Let $K \subseteq D$ be a (h, g)-convex set. An operator $T : K \to H$ is said to be:

1. Higher order strongly monotone, if and only if, there exists a constant $\alpha > 0$ such that

$$\langle Tu - Tv, h(u) - g(v)\rangle \geq \alpha\|g(v) - h(u)\|^p, \forall u, v \in K \subseteq D.$$

2. Higher order strongly pseudomonotone, if and only if, there exists a constant $\nu > 0$ such that

$$\langle Tu, g(v) - h(u)\rangle + \nu\|g(v) - h(u)\|^p \geq 0$$
$$\Rightarrow$$
$$\langle Tv, g(v) - h(u)\rangle \geq 0, \forall u, v \in K \subseteq D.$$

3. Higher order strongly relaxed pseudomonotone, if and only if, there exists a constant $\mu > 0$ such that

$$\langle Tu, g(v) - h(u)\rangle \geq 0$$
$$\Rightarrow$$
$$-\langle Tv, h(u) - g(v)\rangle + \mu\|g(v) - h(u)\|^p \geq 0, \forall u, v \in K \subseteq D.$$

4. Generalized monotone with respect to two functions h and g, if

$$\langle T(h(u)) - T(g(v)), h(u) - g(v)\rangle \geq 0, \forall u, v \in K \subseteq D.$$

Definition 16 A differentiable function F on the (h, g)-convex set $K \subseteq D$ is said to be higher order strongly generalized pseudoconvex function, if and only if, if there exists a constant $\mu > 0$, such that

$$\langle F'(h(u)), g(v)-h(u)\rangle+\mu\|g(v)-h(u)\|^p \geq 0 \Rightarrow F(g(v)) \geq F(h(u)),$$
$$\forall u, v \in K \subseteq D.$$

For suitable and appropriate choices of the arbitrary functions h, g and the constant p, one can obtain various new and old concepts as special of the above definitions. Thus, it is obvious that these concepts are unifying ones and flexible.

We now consider some basic properties of higher order strongly generalized convex functions.

Theorem 7 *Let F be a differentiable function on the (h, g)-convex set $K \subseteq D$. Then the function F is higher order strongly generalized convex function, if and only if,*

$$F(g(v)) - F(h(u)) \geq \langle F'(h(u)), g(v) - h(u)\rangle + \mu\|g(v) - h(u)\|^p,$$
$$\forall u, v \in K \subseteq D, t \in [0, 1]. \tag{8}$$

Proof Let F be a higher order strongly generalized convex function on the (h, g)-convex set $K \subseteq D$. Then

$$F(h(u)+t(g(v)-h(u))) \leq (1-t)F(h(u))+tF(g(v)) \\ -\mu\{t^p(1-t)+t(1-t)^p\}\|g(v)-h(u)\|^p, \\ \forall u, v \in K, t \in [0.1].$$

which can be written as

$$F(g(v))-F(h(u)) \geq \{\frac{F(h(u)+t(g(v)-h(u))-F(h(u))}{t}\} \\ +\{t^{p-1}(1-t)+(1-t)^p\}\|g(v)-h(u)\|^p.$$

Taking the limit in the above inequality as $t \to 0$, we have

$$F(g(v))-F(h(u)) \geq \langle F'(h(u)), g(v)-h(u))\rangle + \mu\|g(v)-h(u)\|^p, \\ \forall u, v \in K \subseteq D.$$

which is (8), the required result.

Conversely, let (8) hold. Then, $\forall u, v \in K \subseteq D, t \in [0, 1]$, $g(v_t) = h(u) + t(g(v) - h(u)) \in K \subseteq D$, we have

$$F(g(v))-F(g(v_t)) \geq \langle F'(g(v_t)), g(v)-g(v_t))\rangle + \mu\|g(v)-g(v_t)\|^p \\ = (1-t)F'(g(v_t)), g(v)-h(u)\rangle+\mu(1-t)^p\|g(v)-h(u)\|^p \quad (9) \\ \forall v, u \in K \subseteq D.$$

In a similar way, we have

$$F(u)-F(v_t) \geq \langle F'(v_t), u-v_t)\rangle + \mu\|g(u)-g(v_t)\|^p \\ = -tF'(v_t), v-u\rangle + \mu t^p\|g(v)-g(u)\|^p. \quad (10)$$

Multiplying (9) by t and (10) by $(1-t)$ and adding the resultant, we have

$$F(h(u)+t(g(v)-h(u))) \leq (1-t)F(h(u))+tF(g(v))-\mu\{t^p(1-t) \\ +t(1-t)^p\}\|g(v)-h(u)\|^p, \\ \forall u, v \in K \subseteq D,$$

showing that F is a higher order strongly generalized convex function.

Theorem 8 *Let F be a differentiable higher order strongly generalized convex function on the (h, g)-convex set $K \subseteq D$. Then $F'(.)$ is a higher order strongly monotone operator.*

Proof Let F be a higher order strongly generalized convex function on the (h, g)-convex set $K \subseteq D$. Then, from Theorem 7, we have

$$F(g(v))-F(h(u)) \geq \langle F'(h(u)), g(v)-h(u)\rangle+\mu\|g(v)-h(u)\|^p, \ \forall u, v \in K \subseteq D. \tag{11}$$

Changing the role of $h(u)$ and $g(v)$ in (11), we have

$$F(h(u))-F(g(v)) \geq \langle F'(g(v)), h(u)-g(v))\rangle+\mu\|g(v)-h(u)\|^p, \ \forall u, v \in K \subseteq D. \tag{12}$$

Adding (11) and (12), we have

$$\langle F'(h(u))-F'(g(v)), h(u)-g(v)\rangle \geq 2\mu\|g(v)-h(u)\|^p, \quad \forall u, v \in K \subseteq D. \tag{13}$$

which shows that $F'(.)$ is a higher order strongly monotone operator.

We remark that the converse of Theorem 8 is not true. However, we have the following result.

Theorem 9 *If the differential operator $F'(.)$ of a differentiable higher order strongly generalized convex function F is higher order strongly monotone operator, then*

$$F(g(v)) - F(h(u)) \geq \langle F'(h(u)), g(v) - h(u)\rangle + 2\mu\frac{1}{p}\|g(v) - h(u)\|^p, \quad \forall u, v \in K \subseteq D. \tag{14}$$

Proof Let $F'(.)$ be a higher order strongly monotone operator. Then, from (13), we have

$$\langle F'(g(v)), h(u) - g(v)\rangle \geq \langle F'(h(u)), h(u) - g(v))\rangle + 2\mu\|g(v) - h(u)\|^p. \quad \forall u, v \in K \subseteq D. \tag{15}$$

Since K is the (h, g)-convex set, $\forall u, v \in K \subseteq D$, $t \in [0, 1]$,
$g(v_t) = h(u) + t(g(v) - h(u)) \in K \subseteq D$. Taking $g(v) = g(v_t)$ in (15, we have

$$\begin{aligned}\langle F'(g(v_t)), h(u) - g(v_t)\rangle &\leq \langle F'(h(u)), h(u) - g(v_t)\rangle - 2\mu\|g(v) - h(u)\|^p \\ &= -t\langle F'(h(u)), g(v) - h(u)\rangle - 2\mu t^p\|g(v) - h(u)\|^p,\end{aligned}$$

which implies that

$$\langle F'(g(v_t)), g(v)-h(u)\rangle \geq \langle F'(h(u)), g(v)-h(u)\rangle+2\mu t^{p-1}\|g(v)-h(u)\|^P. \tag{16}$$

Consider the auxiliary function

$$\xi(t) = F(h(u) + t(g(v) - h(u)), \forall u, v \in K \subseteq D,$$

from which, we have

$$\xi(1) = F(g(v)), \quad \xi(0) = F(h(u)).$$

Then, from (16), we have

$$\begin{aligned}\xi'(t)=&\langle F'(g(v_t), g(v)-h(u)\rangle \geq \langle F'(h(u)), g(v)-h(u)\rangle \\ &+2\mu t^{p-1}\|g(v)-h(u)\|^p. \end{aligned} \tag{17}$$

Integrating (17) between 0 and 1, we have

$$\xi(1) - \xi(0) = \int_0^1 \xi'(t)dt \geq \langle F'(h(u)), g(v) - h(u)\rangle + 2\mu\frac{1}{p}\|g(v) - h(u)\|^p.$$

Thus it follows that

$$\begin{aligned}F(g(v)) - F(h(u)) \geq \langle F'(h(u)), g(v) - h(u)\rangle + 2\mu\tfrac{1}{p}\|g(v) - h(u)\|^p,\\ \forall u, v \in K \subseteq D,\end{aligned}$$

which is the required (14).

We note that, if $p = 2$, then Theorem 9 can be viewed as the converse of Theorem 8.

We now give a necessary condition for higher order strongly generalized pseudo-convex function.

Theorem 10 *Let $F'(.)$ be a higher order strongly relaxed pseudomonotone operator. Then F is a higher order strongly generalized pseudo-connvex function.*

Proof Let F' be a higher order strongly relaxed pseudomonotone operator. Then,

$$\langle F'(h(u)), g(v) - h(u)\rangle \geq 0, \forall u, v \in K \subseteq D,$$

implies that

$$\langle F'(g(v)), g(v) - h(u)\rangle \geq \mu\|g(v) - h(u)\|^p, \forall u, v \in K \subseteq D. \tag{18}$$

Since K is an convex set, $\forall u, v \in K \subseteq D, \quad t \in [0, 1]$, $g(v_t) = h(u) + t(g(v) - h(u)) \in K \subseteq D$.

Taking $g(v) = g(v_t)$ in (18), we have

$$\langle F'(g(v_t), g(v) - h(u)\rangle \geq \mu t^{p-1}\|g(v) - h(u)u\|^p. \tag{19}$$

Consider the auxiliary function

$$\xi(t) = F(h(u) + t(g(v) - h(u))) = F(g(v_t)), \quad \forall u, v \in K \subseteq D, t \in [0, 1],$$

which is differentiable, since F is differentiable function. Then, using (19), we have

$$\xi'(t) = \langle F'(g(v_t)), g(v) - h(u))\rangle \geq \mu t^{p-1}\|g(v(-h(u)\|^p.$$

Integrating the above relation between 0 to 1, we have

$$\xi(1) - \xi(0) = \int_0^1 \xi'(t)dt \geq \frac{\mu}{p}\|g(v) - h(u)\|^p,$$

that is,

$$F(g(v)) - F(h(u)) \geq \frac{\mu}{p}\|g(v) - h(u)\|^p), \forall u, v \in K \subseteq D,$$

showing that F is a higher order strongly generalized pseudo-convex function.

Definition 17 A function F is said to be sharply higher order strongly generalized pseudo convex, if there exists a constant $\mu > 0$ such that

$$\langle F'(h(u)), g(v) - h(u)\rangle \geq 0$$

$$\Rightarrow$$

$$F(g(v)) \geq F(g(v)+t(h(u)-g(v)))+\mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p,$$

$$\forall u, v \in K \subseteq D.$$

Theorem 11 *Let F be a sharply higher order strongly generalized pseudo convex function on the (h, g)-convex set $K \subseteq D$ with a constant $\mu > 0$. Then*

$$\langle F'(g(v)), g(v) - h(u)\rangle \geq \mu\|g(v) - h(u)\|^p, \forall u, v \in K \subseteq D.$$

Proof Let F be a sharply higher order strongly generalized pesudo-convex function. Then

$$F(g(v)) \geq F(g(v) + t(h(u) - g(v))) + \mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p,$$

$$\forall u, v \in K \subseteq D, t \in [0, 1],$$

from which, we have

$$\{\frac{F(g(v)+t(h(u)-g(v))-F(g(v))}{t}\}+\mu\{t^{p-1}(1-t)+(1-t)^p\}\|g(v)-h(u)\|^p\geq 0.$$

Taking limit in the above inequality, as $t\to 0$, we have

$$\langle F'(g(v)),g(v)-g(u)\rangle\geq\mu\|g(v)-h(u)\|^p,\forall u,v\in K\subseteq D,$$

the required result.

Theorem 12 *Let f be a higher order strongly affine generalized convex function. Then F is a higher order strongly generalized convex function, if and only if, $H=F-f$ is a generalized convex function.*

Proof Let f be a higher order strongly affine generalized convex function, Then

$$\begin{aligned}f((1-t)h(u)+tg(v))=(1-t)f(h(u))+tf(g(v))-\mu\{t^p(1-t)\\+t(1-t)^p\}\|g(v)-h(u)\|^p,\\\forall u,v\in K\subseteq D.\end{aligned}\tag{20}$$

From the higher order strongly generalized convexity of F, we have

$$\begin{aligned}F((1-t)h(u)+tg(v))\leq(1-t)F(h(u))+tF(g(v))-\mu\{t^p(1-t)\\+t(1-t)^p\}\|g(v)-h(u)\|^p,\\\forall u,v\in K\subseteq D.\end{aligned}\tag{21}$$

From (20) and (21), we have

$$\begin{aligned}F((1-t)h(u)+tg(v))-f((1-t)f(h(u))+tf(g(v))\leq(1-t)(F(h(u))-f(h(u)))\\+t(F(g(v))-f(g(v))),\end{aligned}\tag{22}$$

from which it follows that

$$\begin{aligned}H((1-t)h(u)+tg(v))&=F((1-t)h(u)+tg(v))-f((1-t)h(u)+tg(v))\\&\leq(1-t)F(h(u))+tF(g(v))-(1-t)f(h(u)(-tf(g(v))\\&=(1-t)(F(h(u))-f(h(u)))+t(F(g(v))-f(g(v))),\end{aligned}$$

which show that $H=F-f$ is a convex function.

The inverse implication is obvious.

Definition 18 A function F is said to be a pseudoconvex function with respect to a strictly positive bifunction $B(.,.)$, if

$$F(g(v)) < F(h(u))$$

$$\Rightarrow$$

$$F(h(u) + (1-t)(g(v), h(u))) < F(h(u)) + t(t-1)B(g(v), h(u)),$$

$$\forall u, v \in K \subseteq D, t \in [0,1].$$

Theorem 13 *If the function F is higher order strongly generalized convex function such that $F(g(v)) < F(h(u))$, then the function F is higher order strongly generalized pseudo convex function*

Proof Since $F(g(v)) < F(h(u))$ and F is higher order strongly generalized convex function, then
$\forall u, v \in K \subseteq D, \quad t \in [0,1]$, we have

$$\begin{aligned}
&F(h(u) + t(g(v) - h(u))) \leq F(h(u)) + t(F(g(v)) - F(h(u)))\\
&-\mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - hu)\|^p\\
&< F(h(u)) + t(1-t)(F(g(v)) - F(h(u))) - \mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p\\
&= F(h(u)) + t(t-1)(F(h(u)) - F(g(v))) - \mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p\\
&< F(h(u)) + t(t-1)B(h(u), g(v)) - \mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p,\\
&\forall u, v \in K \subseteq D,
\end{aligned}$$

where $B(h(u), g(v)) = F(h(u)) - F(g(v)) > 0$. Hence the function F is higher order strongly generalized convex function, the required result.

From Definition 10, we have

$$\begin{aligned}
F((1-t)h(u) + tg(v)) + F(th(u) + (1-t)g(v))\\
\leq F(h(u)) + F(g(v)) - 2\mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p,\\
\forall u, v \in K \subseteq D, t \in [0,1], p > 1,
\end{aligned}$$

which is called the higher order strongly Wright generalized convex functions. One can investigate the properties and applications of higher order strongly Wright generalized convex functions.

5 Applications

We now show that uniformly Banach spaces can be characterized by the parallelogram laws, which can be obtained from the higher order strongly generalized affine functions.

Setting $F(u) = \|u\|^p$ in Definition 14, we have

$$\begin{aligned} &\|h(u) + t(g(v) - h(u))\|^p \\ &= (1-t)\|h(u)\|^p + t\|g(v)\|^p - \mu\{t^p(1-t) + t(1-t)^p\}\|g(v) - h(u)\|^p, \\ &\forall u, v \in K \subseteq D, t \in [0,1]. \end{aligned} \tag{23}$$

Taking $t = \frac{1}{2}$ in (23), we have

$$\|\frac{h(u)+g(v)}{2}\|^p + \mu\frac{1}{2^p}\|g(v)-h(u)\|^p = \frac{1}{2}\|h(u)\|^p + \frac{1}{2}\|g(v)\|^p, \forall u, v \in K \subseteq D, \tag{24}$$

which implies that

$$\|h(u)+g(v)\|^p + \mu\|g(v)-h(u)\|^p = 2^{p-1}\{\|h(u)\|^p + \|g(v)\|^p\}, \forall u, v \in K \subseteq D, \tag{25}$$

which is the generalized parallelogram law for the l^p-spaces.

For $p = 2$, the generalized parallelogram law (25) reduces to:

$$\|h(u) + g(v)\|^2 + \mu\|g(v) - h(u)\|^2 = 2\{\|h(u)\|^2 + \|g(v)\|^2\}, \forall u, v \in K \subseteq D, \tag{26}$$

which is a new parallelogram law involving two arbitrary functions characterizing the inner product spaces and can be viewed as a novel application of the strongly generalized affine functions.

For $g(u) = h(u) = I$, we obtain the well known parallelogram law, that is,

$$\|u + v\|^p + \mu\|v - u\|^p = 2^{p-1}\{\|u\|^p + \|v\|^p\}, \forall u, v \in K \subseteq D, \tag{27}$$

which was derived by Xu [42] via the functionals $\|.\|^p$ and $\|.\|^q$, respectively. These parallelogram laws characterize the p-uniform convexity and q-uniform smoothness of a Banach space. Bynum [6] and Chen et al. [7, 8] have studied the properties and applications of the parallelogram laws for the Banach spaces in the prediction theory. In brief, for suitable and appropriate choice of the arbitrary functions h and g, one can obtained a wide class of parallelogram laws, which can be used to characterize the inner product spaces and uniformly Banach spaces. It is an interesting problem to consider the applications in optimization and prediction theory.

Let $B(H)$ be the space of all bounded linear operators on a separable complex Hilbert space H. The absolute value of an operator $A \in B(H)$ is defined by $\|A\|^p = \langle A, A\rangle$.

By taking $F(u) = \|A\|^p$ in (27), we have

$$\|A + B\|^p + \mu\|A - B\|^p = 2^{p-1}\{\|A\|^p + \|B\|^p\}, \quad \forall A, B \in B(H) \subseteq D,$$

which is known as the Clarkson Inequalities for Operators and be viewed as Operator Parallelogram Laws characterizing the uniformly Banach spaces. For more details, see Hirzallah and Kittaneh [12].

5.1 Generalized Variational Inequalities

In this section, we show that the optimality conditions of the differentiable higher order strongly generalized convex functions can be characterized be a class of variational inequalities. This is the main motivation of our next result.

Theorem 14 *Let F be a differentiable higher order strongly generalized convex function with modulus $\mu > 0$. If $u \in K \subseteq D$ is the minimum of the function F, then*

$$F(g(v)) - F(h(u)) \geq \mu \|g(v) - h(u)\|^p, \quad \forall v \in K \subseteq D, \quad p \geq 1. \tag{28}$$

Proof *Let $u \in K \subseteq D$ be a minimum of the function F. Then*

$$F(h(u)) \leq F(g(v)), \forall v \in K \subseteq D. \tag{29}$$

Since $K \subseteq D$ is a (h, g)-convex set, so, $\forall u.v \in K \subseteq D, \quad t \in [0, 1]$,

$$g(v_t) = (1 - t)h(u) + tg(v) \in K \subseteq D.$$

Taking $g(v) = g(v_t)$ in (29), we have

$$0 \leq \lim_{t \to 0} \{ \frac{F(h(u)+t(g(v)-h(u)))-F(h(u))}{t} \} = \langle F'(h(u)), g(v) - h(u) \rangle. \tag{30}$$

Since F is differentiable higher order strongly convex generalized function, so

$$\begin{aligned} F(h(u) + t(g(v) - h(u))) \leq\ & F(h(u)) + t(F(g(v)) - F(h(u))) \\ & -\mu\{t^p(1 - t) + t(1 - t)^p\}\|g(v) - h(u)\|^p, \forall u.v \in K \subseteq D, \end{aligned}$$

from which, using (30), we have

$$\begin{aligned} F(g(v)) - F(h(u)) &\geq \lim_{t \to 0} \{ \frac{F(h(u) + t(g(v) - h(u))) - F(h(u))}{t} \} + \mu \|g(v) - h(u)\|^p \\ &= \langle F'(h(u)), g(v) - h(u) \rangle + \mu \|g(v) - h(u)\|^p, \\ &\geq \mu \|g(v) - h(u)\|^p, \end{aligned}$$

the required result (28).

Using the above technique and ideas of Theorem 7, one can obtain the following result.

Theorem 15 *Let $K \subseteq D$ be a (h, g)-convex set. If $u \in K \subseteq D$ satisfies the inequality*

$$\langle F'(h(u)), g(v) - h(u)\rangle + \mu\|g(v) - h(u)\|^p \geq 0, \forall u, v \in K \subseteq D, p \geq 1, \quad (31)$$

then $u \in K \subseteq D$ is the minimum of the higher order strongly generalized convex function F.

Remark 1 The inequality (31) is called the higher order strongly extended strongly general variational inequality, which include strongly extended general variational inequalities, higher order strongly general variational inequalities and general variational inequalities as special cases. It is an interesting problem to study the both quantitative and qualitative properties of these variational inequalities.

Theorem 16 *If the operator T is a generalized monotone with respect the arbitrary functions h, g and $u \in Ksubseteq D$ is the solution of the inequality (31), then $u \in K \subseteq D$ satisfies the inequality*

$$\langle F'(g(v)), g(v)-h(u)\rangle+\mu\|g(v)-h(u)\|^p \geq 0, \forall u, v \in K \subseteq D, \quad p \geq 1. \quad (32)$$

Proof Let $u \in K \subseteq D$ satisfies the inequality (31). Then

$$\begin{aligned} 0 &\leq \langle F'(h(u)), g(v) - h(u)\rangle + \mu\|g(v) - h(u)\|^p \\ &= \langle F'(h(u)) - F'(g(v)) + F'(g(v)), g(v) - h(u)\rangle + \mu\|g(v) - h(u)\|^p \\ &\leq F'(g(v)), g(v) - h(u)\rangle + \mu\|g(v) - h(u)\|^p, \end{aligned}$$

where we have used the generalized monotonicity of the operator T.

We note that the converse of Theorem 16 does not hold due the presence of the term $\|g(v) - h(u)\|^p$. However, for $p = 1$, one can prove the converse of Theorem 16 using the concept of hemicontinuity.

Remark 2 The inequality (32) is called the Minty higher order strongly extended strongly general variational inequality. We remark that the projection method and its variant forms can not be used to study the higher order strongly general variational inequalities (31) due to its inherent structure. To overcome this drawback, one can consider the auxiliary principle technique for solving higher order strongly extended strongly general variational inequality, which is mainly due to Glowinski et al. [11] and Lions and Stampacchia [17] as developed by Noor [25, 26]and Noor et al. [37]. We have only given a glimpse of the higher order strongly extended general variational inequalities. We are going to consider these problems in future. These problems may applications in various branches of pure and applied sciences and need further efforts.

Conclusion

In this paper, we have introduced and studied a new class of convex functions, which is called higher order strongly generalized convex function. It is shown that several new classes of strongly convex functions can be obtained as special cases of these higher order strongly generalized convex functions. We have studied the basic properties of these functions. We have shown that one can derive the parallelogram laws in Banach spaces, which have applications in prediction theory and stochastic analysis. It is shown that the minimum of the differentiable higher order strongly generalized convex functions can be characterized by a new class of variational inequalities. Several important special cases are also discussed, which can be obtained from our results. The interested readers may explore the applications and other properties of the higher order strongly convex functions in various fields of pure and applied sciences. This is an interesting direction of future research.

Acknowledgments We wish to express our deepest gratitude to our teachers, colleagues, collaborators and friends, who have direct or indirect contributions in the process of this paper. The authors would like to thank the Rector, COMSATS University Islamabad, Islamabad Pakistan, for providing excellent research and academic environments.

References

1. M. Adamek, On a problem connected with strongly convex functions. Math. Inequal. Appl. **19**(4), 1287–1293 (2016)
2. O. Alabdali, A. Guessab, G. Schmeisser, Characterization of uniform convexity for differentiable functions. Appl. Anal. Discrete Math. **13**, 721–732 (2019)
3. H. Angulo, J. Gimenez, A.M. Moeos, K. Nikodem, On strongly h-convex functions. Ann. Funct. Anal. **2**(2), 85–91 (2011)
4. M.U. Awan, M.A. Noor, M.V. Mihai, K.I. Noor, N. Akhtar, On approximately harmonic h-convex functions depending on a given function. Filomat **33**(12), 3783–3793 (2019)
5. M.U. Awan, M.A. Noor, T.-S. Du, K.I. Noor, New refinemnts of fractional Hermite-Hadamard inequality. RACSAM **113**, 21–29 (2019)
6. W.L. Bynum, Weak parallelogram laws for Banach spaces. Can. Math. Bull. **19**, 269–275 (1976)
7. R. Cheng, C.B. Harris, Duality of the weak parallelogram laws on Banach spaces. J. Math. Anal. Appl. **404**, 64–70 (2013)
8. R. Cheng, W.T. Ross, Weak parallelogram laws on Banach spaces and applications to prediction. Period. Math. Hung. **71**, 45–58 (2015)
9. G. Cristescu, M. Găianu, Shape properties of Noors convex sets, in *Proceedings of the Twelfth Symposium of Mathematics and its Applications, Timisoara* (2009), pp. 1–13
10. G. Cristescu, L. Lupsa, *Non-Connected Convexities and Applications* (Kluwer Academic Publisher, Dordrechet, 2002)
11. F. Glowinski, J.L. Lions, R. Tremileres, *Numerical Analysis of Variational Inequalities* (North Holland, Amsterdam, 1981)
12. O. Hirzallah, F. Kittaneh, Non-commutative Clarkson inequalities for n-tuples of operators. Integr. Equ. Oper. Theory **60**, 369–379 (2008)

13. S. Jabeen, B.B. Mohsen, M.A. Noor, K.I. Noor, Inertial projection methods for solving general quasi-variational inequalities. AIMS Math. **6**(2), 1075–1086 (2021)
14. J.-B. Jian, On (E, F) generalized convexity. Internat. J. Math. Sci. **2**, 121–132 (2003)
15. S. Karamardian, The nonlinear complementarity problems with applications, Part 2. J. Optim. Theory Appl. **4**(3), 167–181 (1969)
16. D. Kinderleher, G. Stampacchia, *An Introduction to Variational Inequalities and their Applications* (Academic Press, London, 1980)
17. J.L. Lions, G. Stampacchia, Variational inequalities. Commun. Pure Appl. Math. **20**, 492–512 (1967)
18. G.H. Lin, M. Fukushima, Some exact penalty results for nonlinear programs and mathematical programs with equilibrium constraints. J. Optim. Theory Appl. **118**(1), 67–80 (2003)
19. B.B. Mohsen, M.A. Noor, K.I. Noor, M. Postolache, Strongly convex functions of higher order involving bifunction. Mathematics **7**(11), 1028 (2019). https://doi.org/10.3390/math7111028
20. C.P. Niculescu, L.E. Persson, *Convex Functions and their Applications* (Springer, New York, 2018)
21. K. Nikodem, Z.S. Pales, Characterizations of inner product spaces by strongly convex functions. Banach J. Math. Anal. **1**, 83–87 (2011)
22. M. A. Noor, General variational inequalities. Appl. Math. Letts. **1**, 119–121(1988)
23. M.A. Noor, Quasi variational inequalities. Appl. Math. Letts. **1**(4), 367–370 (1988)
24. M.A. Noor, New approximations schemes for general variational inequalities. J. Math. Anal. Appl. **251**, 217–230 (2000)
25. M.A. Noor, Some developments in general variational inequalities. Appl. Math. Comput. **152**, 199–277 (2004)
26. M.A. Noor, Fundamentals of equilibrium problems. Math. Inequal. Appl. **9**(3), 529–566 (2006)
27. M.A. Noor, Differentiable nonconvex functions and general variational inequalities. Appl. Math. Comput. **199**(2), 623–630 (2008)
28. M.A. Noor, Extended general variational inequalities. Appl. Math. Lett. **22**, 182–186 (2009)
29. M.A. Noor, Some aspects of extended general variational inequalities. Abstract Appl. Anal. **2012**, 16 (2012). ID 303569
30. M.A. Noor, K.I. Noor, On generalized strongly convex functions involving bifunction. Appl. Math. Inform. Sci. **13**(3), 411–416 (2019)
31. M.A. Noor, K.I. Noor, Higher order strongly general convex functions and variational inequalities. AIMS Math. **5**(4), 3646–3663 (2020)
32. M.A. Noor, K.I. Noor, Characterizations of strongly generalized convex functions. TJMM **11**(2), 117–127 (2020)
33. M.A. Noor, K.I. Noor, Higher order general convex functions and general variational inequalities. Canada J. Appl. Math. **3**(1), 1–17 (2021)
34. M.A. Noor, K.I. Noor, New classes of preinvex functions and variational-like inequalities. Filomat **35** (2021)
35. M.A. Noor, K.I. Noor, T.M. Rassias, Some aspects of variational inequalities. J. Comput. Appl. Math. **47**(3), 285–312 (1993)
36. M.A. Noor, K.I. Noor, A. Bnouhachem, Some new iterative methods for solving variationaal inequalities. Canad. J. Appl. Math. **2**(2), 1–7 (2020)
37. M.A. Noor, K.I. Noor, M.T. Rassias, New trends in general variational inequalities. Acta Appl. Math. **170**(1), 981–1064 (2020)
38. J. Pecric, F. Proschan, Y.I. Tong, *Convex Functions, Partial Ordering and Statistical Applications* (Academic Press, New York, 1992)
39. B.T. Polyak, Existence theorems and convergence of minimizing sequences in extremum problems with restrictions. Soviet Math. Dokl. **7**, 2–75 (1966)
40. G. Qu, N. Li, On the exponentially stability of primal-dual gradeint dynamics. IEEE Control Syst. Letters, **3**(1), 43–48 (2019)
41. G. Stampacchia, Formes bilineaires coercitives sur les ensembles convexes. C. R. Acad. Sci. Paris **258**, 4423–4426 (1964)

42. H.-K. Xu, Inequalities in Banach spaces with applications. Nonlinear Anal. Theory Meth. Appl. **16**(12), 1127–1138 (1991)
43. E.A. Youness, E-convex sets, E-convex functions and E-convex programming. J. Optim. Theory Appl. **102**, 439–450 (1999)
44. Y. Zhao, D. Sun, Alternative theorems for nonlinear projection equations and applications to generalized complementarity problems. Nonl. Anal. **46**, 853–868 (2001)
45. D.L. Zu, P. Marcotte, Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. SIAM J. Optim. **6**(3), 714–726 (1996)

A Note on Generalized Nash Games Played on Networks

Mauro Passacantando and Fabio Raciti

Abstract We investigate a generalized Nash equilibrium problem where players are modeled as nodes of a network and the utility function of each player depends on his/her own action as well as on the actions of his/her neighbors in the network. In the case of a quadratic reference model with shared constraints we are able to derive the variational solution of the game as a series expansion which involves the powers of the adjacency matrix, thus extending a previous result. Our analysis is illustrated by means of some numerical examples.

1 Introduction

Economic and social sciences are probably the areas that have benefited the most from the mathematical development of Game Theory (GT), although in the last decades, specific game-theoretical models have also been applied to various problems from engineering, transportation and communication networks, biology, and other fields (see, e.g., [1, 19, 21]. The pervasive role of physical and virtual social interactions in the actions taken by individuals or groups, described through a network model, has led to consider Network Games as a powerful tool to describe the process of decision making. Indeed, in many social or economic environments our actions are influenced by the actions of our friends, acquaintances or colleagues. In network game models, each individual (player) is identified with the node of a graph and the players that can interact directly are connected through links of the graph. The specificity of these games is the central role played by the graph structure in the description of the patterns of interactions, and in the final social or economic

M. Passacantando
Department of Computer Science, University of Pisa, Pisa, Italy
e-mail: mauro.passacantando@unipi.it

F. Raciti (✉)
Department of Mathematics and Computer Science, University of Catania, Catania, Italy
e-mail: fabio.raciti@unict.it

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_16

outcome. The mathematical consequence of this description is that some interesting results depend on quantities such as the spectral radius of the adjacency matrix, its minimum eigenvalue, and its powers. With the huge variety of possible networks and interactions, it is difficult to make progress in the analysis without positing some specific structure of the problems under consideration. In these regards it is interesting to investigate the two classes of *games with strategic complements and substitutes*. Roughly speaking, in the first case, the incentive for a player to take an action increases when the number of his/her social contacts who take the action increases while in the second case this monotonic relation in reversed. The linear-quadratic model, although its simplicity, deserves particular attention and has been investigated in detail because it can be solved exactly and the solution formula can be nicely interpreted from a graph-theoretical point of view. Among the numerous references we suggest that the reader who wishes to become familiar with the topic of network games see the beautiful survey by Jackson and Zenou [9], along with the seminal paper by Ballester et al. [3]. Very recently, a different approach, based on variational inequality theory, has been put forward to tackle this kind of problems and, in this respect, the interested reader can refer to the interesting and very detailed paper by Parise and Ozdaglar [20].

In this paper we investigate, within the simple frame of the linear-quadratic model, a generalized Nash equilibrium problem (GNEP) with shared constraints. This class of games was proposed a long time ago by Rosen [22], but the last decade has witnessed a renewed interest in the subject, due to its wide range of applications and to its connection with the theory of variational inequalities [5, 7, 16, 17]. By adding a shared constrained to the original quadratic problem we thus obtain a GNEP, and loose uniqueness of the solution. Among the solutions of the new problem, we select the so called *variational solution* and provide a closed form expression for it. Furthermore, the new formula can be written by a series expansion of the adjacency matrix, thus extending one of the results in [3] and allowing for a graph-theoretic (as well as socio-economic) interpretation. Namely, this expansion shows that although players only interact with their neighbors, the solution also depends, to a certain extent, on indirect contacts (i.e., neighbors of neighbors, neighbors of neighbors of neighbors, and so on).

The paper is organized as follows. In the following Section 2 we summarize the basic background material on network games and focus on the exactly solvable linear-quadratic model. Section 3 is devoted to a brief description of generalized Nash equilibrium problems with shared constrained, and to the solution of our specific problem, while Section 4 is dedicated to illustrate our result by means of two worked-out examples. The paper ends with a concluding section where we outline some promising avenues of research.

2 Network Games

2.1 Elements of Graph Theory and Game Classes

We begin this section by recalling a few concepts and definitions of graph theory that will be used in the sequel. We warn the reader that the terminology is not uniform in the related literature. Formally, a graph g is a pair of sets (V, E), where V is the set of nodes (or vertexes) and E is the set of arcs (or edges), formed by pairs of nodes (v, w). Arcs which have the same end nodes are called parallel, while arcs of the form (v, v) are called loops. We consider here *simple* graphs, that is graphs with no parallel arcs or loops. In our setting, the players will be represented by the n nodes in the graph. Moreover, we consider here indirect graphs: arcs (v, w) and (w, v) are the same. Two nodes v and w are adjacent if they are connected by an arc, i.e., if (v, w) is an arc. The information about the adjacency of nodes can be stored in the adjacency matrix G whose elements g_{ij} are equal to 1 if (v_i, v_j) is an arc, 0 otherwise. G is thus a symmetric and zero diagonal matrix. Given a node v, the nodes connected to v with an arc are called the *neighbors* of v and are grouped in the set $N_v(g)$. The number of elements of $N_v(g)$ is the *degree* of v. A *walk* in the graph g is a finite sequence of the form

$$v_{i_0}, e_{j_1}, v_{i_1}, e_{j_2}, \ldots, e_{j_k}, v_{j_k},$$

which consists of alternating nodes and arcs of the graph, such that $v_{i_{t-1}}$ and v_{i_t} are end nodes of e_{j_t}. The *length* of a walk is the number of its arcs. Let us remark that it is allowed to visit a node or go through an arc more than once. A *path* is a walk with all different nodes (except possibly the initial and terminal ones if the walk is closed). The indirect connections between any two nodes in the graph are described by means of the powers of the adjacency matrix G. Indeed, it can be proved that the element $g_{ij}^{[k]}$ of G^k gives the number of walks of length k between v_i and v_j.

We now proceed to specify the type of game that we will consider. For simplicity, the set of players will be denoted by $\{1, 2, \ldots, n\}$ instead of $\{v_1, v_2, \ldots, v_n\}$. We denote with $A_i \subset \mathbb{R}$ the action space of player i, while $A = A_1 \times \cdots \times A_n$ and the notation $a = (a_i, a_{-i})$ will be used when we want to distinguish the action of player i from the action of all the other players. Each player i is endowed with a payoff function $u_i : A \to \mathbb{R}$ that he/she wishes to maximize. The notation $u_i(a, g)$ is often utilized when one wants to emphasize the influence of the graph structure (e.g., when studying perturbation with respect to the removal of an arc). The solution concept that we consider here is the Nash equilibrium of the game, that is, we seek an element $a^* \in A$ such that for each $i \in \{1, \ldots, n\}$:

$$u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*), \qquad \forall\, a_i \in A_i. \tag{1}$$

A peculiarity of network games is that the vector a_{-i} is only made up of components a_j such that $j \in N_i(g)$, that is, j is a neighbor of i.

We mentioned in the introduction that it is convenient to consider two specific classes of games which allow a deeper investigation of the patterns of interactions among players. For any given player i it is interesting to distinguish how variations of the actions of player's i neighbors affect his/her marginal utility. In the case where the utility functions are twice continuously differentiable the following definitions clarify this point.

Definition 1 We say that the network game has the property of strategic substitutes if for each player i the following condition holds:

$$\frac{\partial^2 u_i(a_i, a_{-i})}{\partial a_j \partial a_i} < 0, \qquad \forall (i, j) : g_{ij} = 1, \ \forall a \in A.$$

Definition 2 We say that the network game has the property of strategic complements if for each player i the following condition holds:

$$\frac{\partial^2 u_i(a_i, a_{-i})}{\partial a_j \partial a_i} > 0, \qquad \forall (i, j) : g_{ij} = 1, \forall a \in A.$$

For the subsequent development it is important to recall that if the u_i are continuously differentiable functions on A, the Nash equilibrium problem is equivalent to the variational inequality $VI(F, A)$: find $a^* \in A$ such that

$$[F(a^*)]^\top (a - a^*) \geq 0, \qquad \forall a \in A, \tag{2}$$

where

$$[F(a)]^\top := -\left(\frac{\partial u_1}{\partial a_1}(a), \ldots, \frac{\partial u_n}{\partial a_n}(a)\right) \tag{3}$$

is also called the pseudo-gradient of the game, according to the terminology introduced by Rosen. For an account of variational inequalities the interested reader can refer to [6, 13, 18]. We recall here some useful monotonicity properties.

Definition 3 $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone on A iff:

$$[F(x) - F(y)]^\top (x - y) \geq 0, \qquad \forall x, y \in A.$$

If the equality holds only when $x = y$, F is said to be strictly monotone.

A stronger type of monotonicity is given by the following

Definition 4 Let $\beta > 0$. $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be β-strongly monotone on A iff:

$$[F(x) - F(y)]^\top (x - y) \geq \beta \|x - y\|^2, \qquad \forall x, y \in A.$$

For linear operators on $\mathbb{R}^n$ the two concepts of strict and strong monotonicity coincide and are equivalent to the positive definiteness of the corresponding matrix.

Conditions that ensure the unique solvability of a variational inequality problem are given by the following theorem (see, e.g. [6, 13, 18]).

Theorem 1 *If $K \subset \mathbb{R}^n$ is a compact convex set and $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous on K, then the variational inequality problem $VI(F, K)$ admits at least one solution. In the case that K is unbounded, existence of a solution may be established under the following coercivity condition:*

$$\lim_{\|x\| \to +\infty} \frac{[F(x) - F(x_0)]^\top (x - x_0)}{\|x - x_0\|} = +\infty,$$

for $x \in K$ and some $x_0 \in K$.

Furtheremore, if F is strictly monotone on K the solution is unique.

In the following subsection, we describe in detail the linear-quadratic reference model on which we will build our generalized Nash equilibrium problem.

2.2 The Linear-Quadratic Model

Let $A_i = \mathbb{R}_+$ for any $i \in \{1, \ldots, n\}$, hence $A = \mathbb{R}^n_+$. The payoff of player i is given by:

$$u_i(a, g) = -\frac{1}{2} a_i^2 + \alpha a_i + \phi \sum_{j=1}^{n} g_{ij} a_i a_j, \qquad \alpha, \phi > 0. \tag{4}$$

In this simplified model α and ϕ take on the same value for all players, which then only differ according to their position in the network. The last term describes the interaction between neighbors and since $\phi > 0$ this interaction falls in the class of strategic complements. The pseudo-gradient's components of this game are easily computed as:

$$F_i(a) = a_i - \alpha - \phi \sum_{j=1}^{n} g_{ij} a_j, \qquad i \in \{1, \ldots, n\},$$

which can be written in compact form as:

$$F(a) = (I - \phi G)a - \alpha \mathbf{1},$$

where $\mathbf{1} = (1, \ldots, 1)^\top \in \mathbb{R}^n$. We will seek Nash equilibrium points by solving the variational inequality:

$$[F(a^*)]^\top (a - a^*) \geq 0, \qquad \forall\, a \in \mathbb{R}^n_+. \tag{5}$$

Since the constraint set is unbounded, to ensure solvability we require that F be strongly monotone, which (implying coercivity, for linear operators) also guarantees the uniqueness of the solution.

Lemma 1 (see e.g. [9]) *The matrix $I - \phi G$ is positive definite iff $\phi\rho(G) < 1$, where $\rho(G)$ is the spectral radius of G.*

Proof The symmetric matrix $I - \phi G$ is positive definite if and only if $\lambda_{min}(I - \phi G) > 0$. On the other hand, $\lambda_{min}(I - \phi G) = 1 - \phi\lambda_{max}(G)$. Since G is a symmetric non-negative matrix, the Perron-Frobenius Theorem guarantees that $\lambda_{max}(G) = \rho(G)$, hence $I - \phi G$ is positive definite if and only if $\phi\rho(G) < 1$. □

To be self consistent, in the next lemma we recall the following result about series of matrices.

Lemma 2 (see e.g. [2]) *Let T be a square matrix and consider the series:*

$$I + T + T^2 + \cdots + T^k + \ldots$$

The series converges provided that $\lim_k T^k = 0$, which is equivalent to $\rho(T) < 1$. In such case the matrix $I - T$ is non singular and we have:

$$(I - T)^{-1} = I + T + T^2 + \cdots + T^k + \ldots$$

Theorem 2 (see e.g. [9]) *If $\phi\rho(G) < 1$, then the unique Nash equilibrium is*

$$a^* = \alpha(I - \phi G)^{-1}\mathbf{1} = \alpha \sum_{p=0}^{\infty} \phi^p G^p \mathbf{1}. \tag{6}$$

Proof Since $\phi\rho(G) < 1$, Lemma 1 guarantees that F is strongly monotone. Hence, Theorem 1 applies and we get a unique solution of (5). On the other hand, Lemma 2 implies that the matrix $I - \phi G$ is non singular, thus the linear system $F(a) = 0$, which reads

$$(I - \phi G)a = \alpha\mathbf{1},$$

has a unique solution a^* given by (6). Moreover, looking at the expansion we get, by construction, that any component of a^* is strictly positive. Therefore, a^* is the unique solution of (5), thus it is the unique Nash equilibrium. □

Remark 1 The expansion in (6) suggests an interesting interpretation. Indeed, it can be shown that the (i, j) entry, $g_{ij}^{[p]}$, of the matrix G^p gives the number of walks of length p between nodes i and j. Based on this observation, a measure of centrality on the network was proposed by Katz and Bonacich (see e.g. [4]). Specifically, for

any weight $w \in \mathbb{R}^n_+$, the weighted vector of Katz-Bonacich is given by:

$$b_w(G, \phi) = M(G, \phi) = (I - \phi G)^{-1} w = \sum_{p=0}^{\infty} \phi^p G^p w.$$

In the case where $w = \mathbf{1}$, the (non weighted) centrality measure of Katz-Bonacich of node i is given by:

$$b_{1,i}(G, \phi) = \sum_{j=1}^{n} M_{ij}(G, \phi)$$

and counts the total number of walks in the graph, which start at node i, exponentially damped by ϕ.

Remark 2 The game under consideration also falls in the class of potential games according to the definition introduced by Monderer and Shapley [15]. Indeed, a potential function is given by:

$$P(a, G, \phi) = \sum_{i=1}^{n} u_i(a, G) - \frac{\phi}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} a_i a_j.$$

Monderer and Shapley have proved that, in general, the solutions of the problem

$$\max_{a \in A} \; P(a, G, \phi)$$

form a subset of the solution set of the Nash game. Because under the condition $\phi \rho(G) < 1$ both problems have a unique solution, it follows that the two problems share the same solution.

3 Generalized Nash Equilibrium Problems on Networks

3.1 An Overview of GNEPs and the Variational Inequality Approach to their Solution

In GNEPs each player's startegy set may depend on the strategies of the other players. We consider here the simplified framework where $A_i \subseteq \mathbb{R}_+$ and we are given a function $g : \mathbb{R}^n \to \mathbb{R}^m$ which describes the shared constraints. The strategy set of player i is then written as

$$K_i(a_{-i}) = \{a_i \in \mathbb{R}_+ : \; g(a) = g(a_i, a_{-i}) \leq 0\}.$$

Thus, players share a common constraint g and have an additional individual nonnegativity constraint. With these ingredients, the GNEP is the problem of finding $a^* \in \mathbb{R}^n$ such that, for any $i \in \{1, \dots, n\}$, $a_i^* \in K_i(a_{-i}^*)$ and

$$u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*), \qquad \forall\, a_i \in K_i(a_{-i}^*). \tag{7}$$

We will work under the common (although not minimal) assumptions that for each fixed $a_{-i} \in A_{-i}$, the functions $u_i(\cdot, a_{-i})$ are concave and continuously differentiable, and the components of g are convex and continuously differentiable. As a consequence, a necessary and sufficient condition for $a_i^* \in K_i(a_{-i}^*)$ to satisfy (7) is

$$-\frac{\partial u_i(a_i^*, a_{-i}^*)}{\partial a_i}(a_i - a_i^*) \geq 0, \qquad \forall\, a_i \in K_i(a_{-i}^*). \tag{8}$$

Thus, if we define $F(a)$ as in (3), and

$$K(a) = K_1(a_{-1}) \times \cdots \times K_n(a_{-n}),$$

it follows that a^* is a GNE if and only if $a^* \in K(a^*)$ and

$$[F(a^*)]^\top (a - a^*) \geq 0, \qquad \forall\, a \in K(a^*). \tag{9}$$

The problem above, where also the feasible set depends on the solution, is called a quasi-variational inequality and its solution is as difficult as the original GNEP.

Assume now that a^* is a solution of GNEP. Hence, for each i, a_i^* solves the maximization problem

$$\max_{a_i} \{u_i(a_i, a_{-i}^*) :\ g(a_i, a_{-i}^*) \leq 0,\ a_i \geq 0\}.$$

Under some standard constraint qualification we can then write the KKT conditions for each maximization problem. We then introduce the Lagrange multiplier $\lambda^i \in \mathbb{R}^m$ associated with the costraint $g(a_i, a_{-i}^*) \leq 0$ and the multiplier $\mu_i \in \mathbb{R}$ associated with the nonnegativity constraint $a_i \geq 0$. The Lagrangian function for each player i reads as:

$$L_i(a_i, a_{-i}^*, \lambda^i, \mu_i) = u_i(a_i, a_{-i}^*) - [g(a_i, a_{-i}^*)]^\top \lambda^i + \mu_i a_i$$

and the KKT conditions for all players are given by:

$$\nabla_{a_i} L_i(a_i^*, a_{-i}^*, \lambda^{i*}, \mu_i^*) = 0, \qquad i = 1, \dots, n, \tag{10}$$

$$\lambda_j^{i*} g_j(a^*) = 0,\ \lambda_j^{i*} \geq 0,\ g_j(a^*) \leq 0, \qquad i = 1, \dots, n,\ j = 1, \dots, m \tag{11}$$

$$\mu_i^* a_i^* = 0,\ \mu_i^* \geq 0,\ a_i^* \geq 0, \qquad i = 1, \dots, n. \tag{12}$$

Conversely, under the assumptions made, if a^*, λ, μ^*, where $\lambda^* = (\lambda^{1*}, \ldots, \lambda^{n*})$ and $\mu^* = (\mu_1^*, \ldots, \mu_n^*)$, satisfy the KKT system (10)–(12), then a^* is a GNE.

Definition 5 Let a^* be a GNE which together with the Lagrange multipliers $\lambda^* = (\lambda^{1*}, \ldots, \lambda^{n*})$ and $\mu^* = (\mu_1^*, \ldots, \mu_n^*)$ satisfies the KKT system of all players. We call a^* a *normalized equilibrium* if there exists a vector $r \in \mathbb{R}^n_{++}$ and a vector $\bar{\lambda} \in \mathbb{R}^m_+$ such that

$$\lambda^{i*} = \frac{\bar{\lambda}}{r_i}, \qquad \forall\, i = 1, \ldots, n,$$

which means that, for a normalized equilibrium, the multipliers of the constraints shared by all players are proportional to a common multiplier. In the special case $r_i = 1$ for any i, i.e., the multipliers coincide for each player, a^* is called *variational equilibrium* (VE). Rosen [22] proved that if the feasible set, which in our case is:

$$K = \{a \in \mathbb{R}^n_+ :\ g(a) \leq 0\}$$

is compact and convex, then there exists a normalized equilibrium for each $r \in \mathbb{R}^n_{++}$.

Now, let us define, for each $r \in \mathbb{R}^n_{++}$, the vector function $F^r : \mathbb{R}^n \to \mathbb{R}^n$ as follows:

$$[F^r(a)]^\top := -\left(r_1 \frac{\partial u_1}{\partial a_1}(a), \ldots, r_n \frac{\partial u_n}{\partial a_n}(a)\right).$$

The variational inequality approach for finding the normalized equilibria of the GNEP is expressed by the following theorem which can be viewed as a special case of Proposition 3.2 in [17] or of Theorem 6.1 in [14].

Theorem 3

1. *Suppose that a^* is a solution of $VI(F^r, K)$, where $r \in \mathbb{R}^n_{++}$, a constraint qualification holds at a^* and $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^m \times \mathbb{R}^n$ are the multipliers associated to a^*. Then, a^* is a normalized equilibrium such that the multipliers (λ^{i*}, μ_i^*) of each player i satisfy the following conditions:*

$$\lambda^{i*} = \frac{\bar{\lambda}}{r_i}, \quad \mu_i^* = \frac{\bar{\mu}_i}{r_i}, \qquad \forall\, i = 1, \ldots, n.$$

2. *If a^* is a normalized equilibrium such that the multipliers (λ^{i*}, μ_i^*) of each player i satisfy the following conditions:*

$$\lambda^{i*} = \frac{\bar{\lambda}}{r_i}, \qquad \forall\, i = 1, \ldots, n,$$

for some vector $\bar{\lambda} \in \mathbb{R}^m_+$ *and* $r \in \mathbb{R}^n_{++}$*, then* a^* *is a solution of* $VI(F^r, K)$ *and* $(\bar{\lambda}, r_1\mu_1^*, \ldots, r_n\mu_n^*)$ *are the corresponding multipliers.*

3.2 A Linear-Quadratic Network GNEP

In this section we investigate an extension to a GNEP of the linear-quadratic network game described in Section 2.2. Specifically, we assume the same network structure given by the adjacency matrix G and the same payoff functions defined as in (4), while the strategy set of player i is given by the usual individual constraint $a_i \geq 0$ and an additional constraint, shared by all the players, on the total quantity of activities of all players, that is

$$K_i(a_{-i}) = \left\{ a_i \in \mathbb{R}_+ : \sum_{j=1}^{n} a_j \leq C \right\}, \qquad i = 1, \ldots, n,$$

where $C > 0$ is a given parameter. Depending on the specific application, the additional constraint can have the meaning of a collective budget upper bound or of a limited availability of a certain commodity.

We know from Theorem 2 that if $\phi\rho(G) < 1$, then the linear-quadratic network game (without the new shared constraint) has a unique Nash equilibrium a^* given by (6). However, if a^* does not satisfy the shared constraint, i.e., it does not belong to the set

$$K = \left\{ a \in \mathbb{R}^n_+ : \sum_{i=1}^{n} a_i \leq C \right\},$$

it cannot be a GNE for the new game. On the other hand, under the assumption $\phi\rho(G) < 1$, Theorem 1 guarantees that the pseudo-gradient F of the game, defined as

$$F(a) = (I - \phi G)a - \alpha \mathbf{1},$$

is strongly monotone, hence there exists a unique solution of $VI(F, K)$, i.e., there exists a unique variational equilibrium of the linear-quadratic network GNEP. The following result gives an explicit formula for such variational equilibrium and an expansion similar to (6).

Theorem 4 *If* $\phi\rho(G) < 1$*, then the unique variational equilibrium* $\bar{a}$ *of the linear-quadratic network GNEP, that is the unique solution of* $VI(F, K)$*, is given by the following formula:*

$$\bar{a} = \begin{cases} a^* = \alpha \sum_{p=0}^{\infty} \phi^p G^p \mathbf{1} & \text{if } \sum_{i=1}^{n} a_i^* \leq C, \\ \dfrac{Ca^*}{\sum_{i=1}^{n} a_i^*} = \dfrac{C \sum_{p=0}^{\infty} \phi^p G^p \mathbf{1}}{\sum_{p=0}^{\infty} \phi^p \mathbf{1}^\top G^p \mathbf{1}} & \text{if } \sum_{i=1}^{n} a_i^* > C, \end{cases} \tag{13}$$

where $a^* = \alpha(I - \phi G)^{-1}\mathbf{1}$ *is the Nash equilibrium of the linear-quadratic network game.*

Proof Theorem 1 guarantees that the matrix $I - \phi G$ is positive definite and the map F is strongly monotone. Therefore, $VI(F, K)$ has a unique solution. If $a^* \in K$, then a^* solves $VI(F, K)$ since $F(a^*) = 0$. Otherwise, if $a^* \notin K$, then $\bar{a} = Ca^*/\sum_{i=1}^{n} a_i^* \in K$ since $\bar{a}_i > 0$ for any $i = 1, \ldots, n$ and $\sum_{i=1}^{n} \bar{a}_i = C$. Moreover $\bar{a}$ is a solution of the KKT system related to $VI(F, K)$ with multipliers

$$\bar{\lambda} = \alpha \left(1 - \frac{C}{\sum_{i=1}^{n} a_i^*} \right) > 0,$$

associated to the shared constraint, and $\bar{\mu}_i = 0$ associated to $a_i \geq 0$ for any $i = 1, \ldots, n$. In fact, we have

$$F(\bar{a}) + \bar{\lambda}\mathbf{1} - \bar{\mu} = \frac{C\alpha(I - \phi G)(I - \phi G)^{-1}\mathbf{1}}{\sum_{i=1}^{n} a_i^*} - \alpha\mathbf{1} + \alpha \left(1 - \frac{C}{\sum_{i=1}^{n} a_i^*} \right) \mathbf{1} = 0$$

$$\bar{\lambda} \geq 0, \quad \sum_{i=1}^{n} \bar{a}_i \leq C, \quad \bar{\lambda} \left(C - \sum_{i=1}^{n} \bar{a}_i \right) = 0$$

$$\bar{\mu}_i \geq 0, \quad \bar{a}_i \geq 0, \quad \bar{\mu}_i \bar{a}_i = 0, \quad \forall\, i = 1, \ldots, n.$$

Therefore, $\bar{a}$ solves $VI(F, K)$. Finally, the ratio between the expansions in (13) follows from the one for a^* given in (6):

$$\bar{a} = \frac{Ca^*}{\mathbf{1}^\top a^*} = \frac{C\alpha \sum_{p=0}^{\infty} \phi^p G^p \mathbf{1}}{\alpha \sum_{p=0}^{\infty} \phi^p \mathbf{1}^\top G^p \mathbf{1}} = \frac{C \sum_{p=0}^{\infty} \phi^p G^p \mathbf{1}}{\sum_{p=0}^{\infty} \phi^p \mathbf{1}^\top G^p \mathbf{1}}.$$

This concludes the proof. □

Notice that if $a^* \in K$, then the formula giving $\bar{a}$ contains α but does not C, while if $a^* \notin K$, it contains C but not α.

4 Numerical Experiments

In this section, we show some preliminary numerical experiments for the linear-quadratic network GNEP described in Section 3.2 by means of two small-size test problems.

Example 1 We consider the network shown in Figure 1 (see also [3]) with 11 nodes (players). The spectral radius of the adjacency matrix G is $\rho(G) \simeq 4.4040$. We set parameter $\alpha = 1$ and chose five different values for ϕ:

$$\phi = 0.3/\rho(G),\ \ \phi = 0.5/\rho(G),\ \ \phi = 0.7/\rho(G),\ \ \phi = 0.9/\rho(G),\ \ \phi = 0.95/\rho(G),$$

to guarantee the assumption of Theorem 4 holds. Moreover, we set $C = 20$ in the shared constraint so that the variational equilibrium $\bar{a}$ of the GNEP is different from the Nash equilibrium a^* of the classical network game. It follows from expansions in (13) that a^* and $\bar{a}$ can be approximated by the sequences $\{a_k^*\}$ and $\{\bar{a}_k\}$, respectively:

$$a_k^* = \alpha \sum_{p=0}^{k} \phi^p G^p \mathbf{1}, \qquad \bar{a}_k = \frac{C \sum_{p=0}^{k} \phi^p G^p \mathbf{1}}{\sum_{p=0}^{k} \phi^p \mathbf{1}^\top G^p \mathbf{1}}.$$

Table 1 shows, for both sequences, the number of sums needed to get an approximation error less than 10^{-t} for any $t = 1, \ldots, 10$. Specifically, for any value of ϕ, the numbers

$$\min\left\{k :\ \|a_k^* - a^*\|_\infty < 10^{-t}\right\} \qquad \text{and} \qquad \min\left\{k :\ \|\bar{a}_k - \bar{a}\|_\infty < 10^{-t}\right\}$$

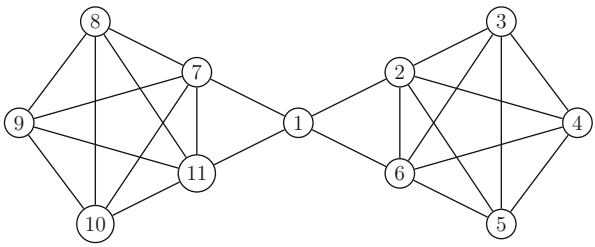


Fig. 1 Network topology of Example 1

Table 1 Speed of convergence of the sequences $\{a_k^*\}$ and $\{\bar{a}_k\}$ to a^* and $\bar{a}$, respectively

	$\phi = \frac{0.3}{\rho(G)}$		$\phi = \frac{0.5}{\rho(G)}$		$\phi = \frac{0.7}{\rho(G)}$		$\phi = \frac{0.9}{\rho(G)}$		$\phi = \frac{0.95}{\rho(G)}$	
Error	NE	VE	NE	VE	NE	VE	NE	VE	NE	VE
10^{-1}	2	1	4	1	10	1	44	1	105	1
10^{-2}	4	2	7	3	16	5	66	9	150	11
10^{-3}	6	4	11	6	23	11	88	27	194	43
10^{-4}	8	5	14	9	29	17	110	48	239	86
10^{-5}	9	7	17	13	35	23	132	70	284	131
10^{-6}	11	9	21	16	42	30	153	92	329	176
10^{-7}	13	11	24	19	48	36	175	114	374	221
10^{-8}	15	13	27	23	55	43	197	136	419	265
10^{-9}	17	15	31	26	61	49	219	157	464	310
10^{-10}	19	17	34	29	68	56	241	179	509	355

For any value of ϕ, the first column (NE) reports $\min\left\{k : \|a_k^* - a^*\|_\infty < 10^{-t}\right\}$, while the second column (VE) reports $\min\left\{k : \|\bar{a}_k - \bar{a}\|_\infty < 10^{-t}\right\}$, for any $t = 1, \ldots, 10$

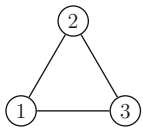


Fig. 2 Network topology of Example 2

are reported in the first (NE) and second (VE) column, respectively, for any $t = 1, \ldots, 10$.

The results in Table 1 show that the convergence of $\{\bar{a}_k\}$ to the variational equilibrium $\bar{a}$ seems to be faster than the convergence of $\{a_k^*\}$ to the Nash equilibrium a^*. Moreover, the more the value of ϕ is close to $1/\rho(G)$, the more this gap is evident.

Example 2 We now consider the network shown in Figure 2 with 3 nodes (players). The spectral radius of the adjacency matrix G is $\rho(G) = 2$. We set parameter $\alpha = 1$, $\phi = 0.25$ and $C = 3$, so that the Nash equilibrium is $a^* = (2, 2, 2)^\top$ and the variational equilibrium is $\bar{a} = (1, 1, 1)^\top$. We exploit Theorem 3 to approximate the set of normalized equilibria. Consider the simplex

$$W = \left\{r \in \mathbb{R}^3_{++} : \; r_1 + r_2 + r_3 = 1\right\}$$

of weights of the parametrized $VI(F^r, K)$ and its discretization given by the finite set of vectors

$$\left(\frac{q_1}{D}, \frac{q_2}{D}, \frac{q_3}{D}\right)$$

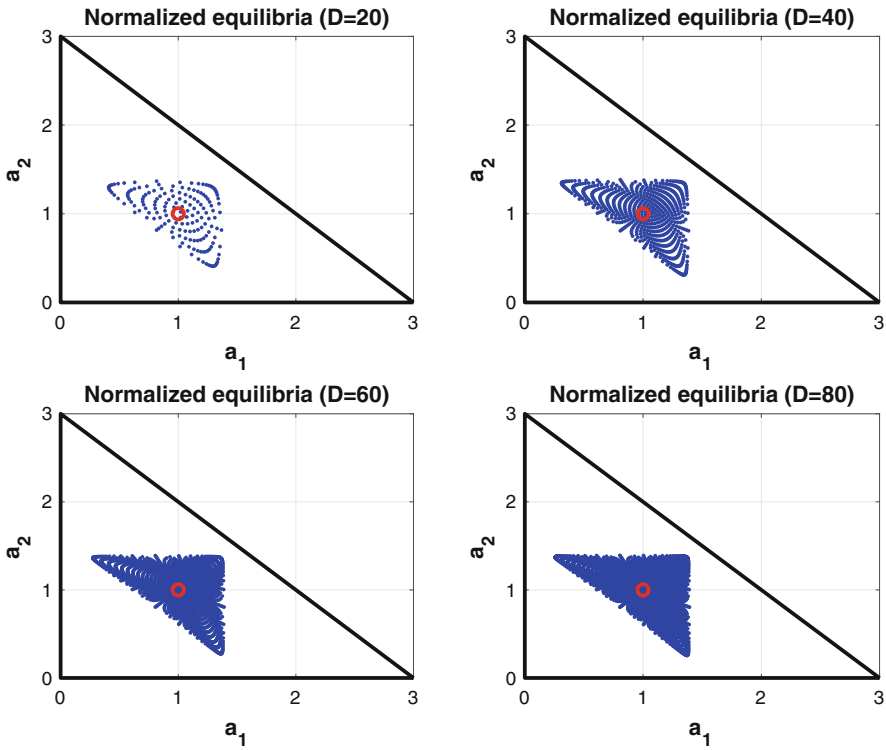


Fig. 3 Normalized equilibria (blue points) and variational equilibrium (red circle) of the linear-quadratic network GNEP

such that q_1, q_2, q_3 and D are positive integers and $q_1 + q_2 + q_3 = D$. Figure 3 shows the set of normalized equilibria, projected on the plane (a_1, a_2), for different values of D. Notice that all the found normalized equilibria belong to the plane $a_1 + a_2 + a_3 = 3$.

The results in Figure 3 suggest that the set of normalized equilibria is equal to

$$\left\{a \in \mathbb{R}^3_+ : \; a_1 + a_2 + a_3 = 3, \quad a_i \leq 1.3924 \quad i = 1, 2, 3\right\}$$
$$= \text{conv}\left\{\begin{pmatrix} 1.3924 \\ 1.3924 \\ 0.2152 \end{pmatrix}, \begin{pmatrix} 1.3924 \\ 0.2152 \\ 1.3924 \end{pmatrix}, \begin{pmatrix} 0.2152 \\ 1.3924 \\ 1.3924 \end{pmatrix}\right\}.$$

Notice that, due to the symmetry of the considered network, the variational equilibrium $\bar{a}$ is equal to the barycenter of the set of normalized equilibria.

5 Conclusions and Further Research Perspectives

In this note, we dealt with a network GNEP and derived a closed formula for its solution which involves the powers of the adjacency matrix, thus extending a previous result. To the best of our knowledge, this kind of formulas have been derived only in a few special cases and, because of their very interesting interpretation, it would be desirable to obtain similar results for more general problem classes. Another promising direction of research is the inclusion of random data in the model (see e.g. [10, 11]), which could be done by using tools from infinite-dimensional duality theory (see e.g. [8, 12, 14]).

Acknowledgments The authors are members of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA—National Group for Mathematical Analysis, Probability and their Applications) of the Istituto Nazionale di Alta Matematica (INdAM—National Institute of Higher Mathematics).

References

1. T. Alpcan, T. Başar, A game-theoretic framework for congestion control in general topology networks, in *Proceedings of the IEEE 41st Conference on Decision and Control Las Vegas, December 10–13* (2002)
2. K. Atkinson, W. Han, *Theoretical Numerical Analysis* (Springer, Berlin, 2007)
3. C. Ballester, A. Calvo-Armengol, Y. Zenou, Who's who in networks. Wanted: the key player. Econometrica **74**, 1403–1417 (2006)
4. P. Bonacich, Power and centrality: a family of measures. Am. J. Sociol. **92**, 1170–1182 (1987)
5. F. Facchinei, C. Kanzov, Generalized Nash equilibrium problems. Ann. Oper. Res. **175**, 177–211 (2010)
6. F. Facchinei, J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems* (Springer, Berlin, 2003)
7. F. Facchinei, A. Fischer, V. Piccialli, On generalized Nash games and variational inequalities. Oper. Res. Lett. **35**, 159–164 (2007)
8. F. Faraci, F. Raciti, On generalized Nash equilibrium problems in infinite dimension: the Lagrange multipliers approach. Opitimization **64**, 321–338 (2015)
9. M.O. Jackson, Y. Zenou, Games on networks, in *Handbook of Game Theory with Economic Applications* (Elsevier, Amsterdam, 2015), pp. 95–163
10. B. Jadamba, F. Raciti, On the modelling of some environmental games with uncertain data. J. Optim. Theory Appl. **167**, 959–968 (2015)
11. B. Jadamba, A.A. Khan, F. Raciti, Regularization of stochastic variational inequalities and a comparison of an L_p and a sample-path approach. Nonlinear Anal. Theory Methods Appl. **94**, 65–83 (2014)
12. A.A. Khan, M. Sama, A new conical regularization for some optimization and optimal control problems: convergence analysis and finite element discretization. Numer. Func. Anal. Opt. **34**(8), 861–895 (2013)
13. D. Kinderleher, G. Stampacchia, *An Introduction to Variational Inequalities and their Applications* (Academic Press, New York, 1980)
14. G. Mastroeni, M. Pappalardo, F. Raciti, Generalized Nash equilibrium problems and variational inequalities in Lebesgue spaces. Minimax Theory Appl. **5**, 47–64 (2020)
15. D. Monderer, L.S. Shapley, Potential games. Games Econ. Behav. **14**, 124–143 (1996)

16. K. Nabetani, *Variational Inequality Approaches to Generalized Nash Equilibrium Problems*, Master Thesis. Department of Applied Mathematics and Physics Graduate School of Informatics, Kyoto University (2008)
17. K. Nabetani, P. Tseng, M. Fukushima, Parametrized variational inequality approaches to generalized Nash equilibrium problems with shared constraints. Comput. Optim. Appl. **48**, 423–452 (2011)
18. A. Nagurney, *Network Economics a Variational Inequality Approach* (Springer, Berlin, 1999)
19. A. Orda, R. Rom, N. Shimkin, Competitive routing in multiuser communication networks. IEEE/ACM Trans. Netw. **1**, 510–521 (1993)
20. F. Parise, A. Ozdaglar, A variational inequality framework for network games: Existence, uniqueness, convergence and sensitivity analysis. Games Econ. Behav. **114**, 47–82 (2019)
21. M. Passacantando, F. Raciti, Optimal road maintenance investment in traffic networks with random demands. Optim. Lett. (2019). https://doi.org/10.1007/s11590-019-01493-y
22. J.B. Rosen, Existence and uniqueness of equilibrium points for concave n-person games. Econometrica **33**, 520–534 (1965)

Piecewise Polynomial Inversion of the Radon Transform in Three Space Dimensions via Plane Integration and Applications in Positron Emission Tomography

Nicholas E. Protonotarios, George A. Kastis, Nikolaos Dikaios, and Athanassios S. Fokas

Abstract The inversion of the celebrated Radon transform in three dimensions involves two-dimensional plane integration. This inversion provides the mathematical foundation of the important field of medical imaging, known as three-dimensional positron emission tomography (3D PET). In this chapter, we present an analytical expression for the inversion of the three-dimensional Radon transform, as well as a novel numerical implementation of this formula, based on piecewise polynomials of the third degree.

N. E. Protonotarios (✉)
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK

Research Center of Mathematics, Academy of Athens, Athens, Greece
e-mail: np558@cam.ac.uk

G. A. Kastis · N. Dikaios
Research Center of Mathematics, Academy of Athens, Athens, Greece
e-mail: gkastis@academyofathens.gr; ndikaios@academyofathens.gr

A. S. Fokas
Research Center of Mathematics, Academy of Athens, Athens, Greece

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK

Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, USA
e-mail: t.fokas@damtp.cam.ac.uk

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*, Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_17

1 Introduction

The celebrated Radon transform of a two-dimensional function is defined as the set of all its line integrals [1, 2]. The transition to three space dimensions yields a certain generalization of line integration, namely *plane integration*. Indeed, the Radon transform of a three-dimensional function is defined as the set of all its plane integrals.[1] The inversion of the three-dimensional Radon transform provides the mathematical foundation of the important field of medical imaging, known as 3D positron emission tomography (3D PET). The 3D Radon transform gives rise to an associated *inverse problem*, namely to "reconstruct" a function from its plane integrals. The main task in 3D PET imaging is the numerical implementation of the inversion of the 3D Radon transform.

In 3D PET, contrary to the conventional 2D PET, there is a certain generalization of the notion of image reconstruction: in the 2D case, the integration occurs in planes instead of lines. The difficulties arising in the 2D cases and their generalizations [3] are overcome in the 3D case. The inversion of the 3D Radon transform seems more straightforward than the one of the conventional Radon transform [4]. There are several numerical implementation methods in the literature, including: (i) the introduction of the concept of three-dimensional image reconstruction from "complete" projections [5]; (ii) the formulation of the 3D Radon transform for discrete 3D images (volumes), based on the summation over planes with small absolute slopes [6]; and (iii) the reconstruction of conductivities in the context of electric impedance tomography (EIT) [7]. The differences between 2D and 3D Radon transform inversion are emphasized in [8], and [9], where an analytic filter-backprojection method is introduced based on the spatially invariant detector point spread function. The authors of [10] proposed a spline-based inversion of the Radon transform in two and three dimensions; also the PET image reconstruction algorithms proposed in [11] show that analytic algorithms in 3D are linear and therefore allow easier control of the spatial resolution and noise correlations than in the case of the 2D reconstructions.

In this chapter, we present a novel formula for the inversion of the 3D Radon transform, as well as a novel numerical implementation of this formula, based on piecewise polynomial interpolation. We expect that our novel numerical implementation will enhance three-dimensional medical image reconstruction, especially in the case of 3D PET.

[1] Plane integrals are special cases of surface integrals, where the surface of integration is a plane.

2 The Radon Transform in Two Space Dimensions

In order to elucidate the properties of the three-dimensional Radon transform, it is essential to review the corresponding properties of the two-dimensional Radon transform.

A line L on the plane can be specified by the signed distance from the origin ρ, with $-\infty < \rho < \infty$, and the angle with the x_1-axis θ, with $0 \leq \theta < 2\pi$, as in Figure 1. We denote the corresponding unit vectors perpendicular and parallel to L by $\mathbf{n}$ and $\mathbf{p}$, respectively. These unit vectors are given by

$$\mathbf{n} = (-\sin\theta, \cos\theta)^{\mathrm{T}} \quad \text{and} \quad \mathbf{p} = (\cos\theta, \sin\theta)^{\mathrm{T}}, \tag{1}$$

with

$$\mathbf{n} \cdot \mathbf{p} = 0. \tag{2}$$

Every point $\mathbf{x} = (x_1, x_2)^{\mathrm{T}}$ lying on the line L in Cartesian coordinates can be expressed in terms of the so-called *local coordinates* (ρ, τ) via

$$\mathbf{x} = \rho\,\mathbf{n} + \tau\mathbf{p},$$

where τ denotes the arc length. Therefore, we parameterize each point $\mathbf{x}$ on the line L in the following manner:

$$\mathbf{x} := \mathbf{x}(\rho, \tau; \theta) = \begin{bmatrix} x_1(\rho, \tau; \theta) \\ x_2(\rho, \tau; \theta) \end{bmatrix} = \begin{bmatrix} \tau\cos\theta - \rho\sin\theta \\ \tau\sin\theta + \rho\cos\theta \end{bmatrix} \tag{3}$$

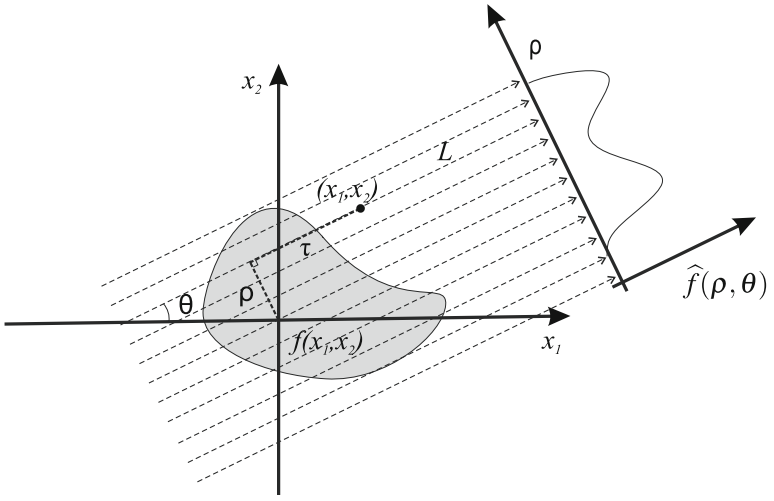


Fig. 1 A two-dimensional function $f(x_1, x_2)$ expressed in Cartesian coordinates, and its projections $\widehat{f}(\rho, \theta)$, expressed in local coordinates

Through Equation (3), we can express the local coordinates (ρ, τ) in terms of Cartesian coordinates (x_1, x_2) and the associated angle θ:

$$\begin{bmatrix} \rho \\ \tau \end{bmatrix} := \begin{bmatrix} \rho(x_1, x_2; \theta) \\ \tau(x_1, x_2; \theta) \end{bmatrix} = \begin{bmatrix} x_2 \cos\theta - x_1 \sin\theta \\ x_2 \sin\theta + x_1 \cos\theta \end{bmatrix} \tag{4}$$

We define the line integral over all lines L, defined in Equation (1), of a two-dimensional Schwartz function $f : \mathbb{R}^2 \to \mathbb{R}$, $f \in S(\mathbb{R}^2)$, as its *two-dimensional Radon transform*, $\mathscr{R}_2 f$. In the context of 2D PET, the 2D Radon transform of the function f is usually stored in the form of the so-called *sinogram*, denoted by $\widehat{f}(\rho, \theta)$

$$\mathscr{R}_2 f = \widehat{f}(\rho, \theta) = \int_L f \, \mathrm{d}s, \tag{5}$$

where ds denotes an arc length differential, and $S(\mathbb{R}^2)$ denotes the space of Schwartz functions in $\mathbb{R}^2$,

$$S(\mathbb{R}^2) = \left\{ f \in C^\infty(\mathbb{R}^2) : ||f||_{\alpha,\beta} < \infty \right\} \subset C^\infty(\mathbb{R}^2), \tag{6}$$

and

$$||f||_{\alpha,\beta} = \sup_{x \in \mathbb{R}^2} |x^\alpha D^\beta f(x)|, \quad \forall \text{ multi} - \text{index } \alpha, \beta, \quad |x^\alpha D^\beta f(x)| \to 0, \quad \text{as } |x| \to \infty. \tag{7}$$

Equation (5) may be rewritten via a parameterization $\mathbf{x} := \mathbf{x}(\tau)$ of the line L, with $\mathbf{x} : \mathbb{R}^2 \to L$, as follows:

$$\widehat{f}(\rho, \theta) = \int_{-\infty}^{\infty} f\left(\mathbf{x}(\tau)\right) \left|\left|\mathbf{x}'(\tau)\right|\right|_2 \mathrm{d}\tau, \tag{8}$$

where $||\cdot||_2$ denotes the L^2-norm in $\mathbb{R}^2$. The parameterization provided by (3) will be proven to be very convenient and easy-to-manipulate, especially for the description of parallel lines. In this case, it is worth noting that

$$\left|\left|\mathbf{x}'(\tau)\right|\right|_2 = \sqrt{\left(\frac{\mathrm{d}x_1}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}x_2}{\mathrm{d}\tau}\right)^2} = \cos^2\theta + \sin^2\theta = 1. \tag{9}$$

Hence, Equation (3) is a natural parameterization of the set of parallel lines L. Therefore, the 2D Radon transform defined in Equation (5) may be expressed as follows:

$$\mathscr{R}_2 f = \widehat{f}(\rho, \theta) = \int_{-\infty}^{\infty} f(\tau \cos\theta - \rho \sin\theta, \tau \sin\theta + \rho \cos\theta) \mathrm{d}\tau, \tag{10}$$

with $0 \leq \theta < 2\pi$ and $-\infty < \rho < \infty$. If we use a Dirac delta function, or a line impulse, then the Radon transform, denoted by $\mathscr{R}_{2\mathrm{D}}$ defined in (10) may be rewritten in the form:

$$\mathscr{R}_2 f = \widehat{f}(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2)\delta(\rho + x_1 \sin\theta - x_2 \cos\theta)\mathrm{d}x_1\mathrm{d}x_2, \tag{11}$$

taking into account Equation (4).

The 2D Radon transform (10) gives rise to one of the most significant inverse problems in emission tomography. This specific inverse problem implies the "reconstruction" of the function $f(x_1, x_2)$, from its two-dimensional Radon transform, i.e. the function $\widehat{f}(\rho, \theta)$.

3 The Radon Transform in Three Space Dimensions

In the two-dimensional case, the Radon transform is considered on sets of parallel lines. This consideration implies the involvement of line integrals. However, in the three-dimensional case, the Radon transform is restricted on two-dimensional planes. In this direction, the transition to three space dimensions yields a certain generalization of line integration, namely *plane integration*.[2]

Therefore, we define the surface integral over all planes P of a three-dimensional Schwartz function $f : \mathbb{R}^3 \to \mathbb{R}$, $f \in S(\mathbb{R}^3)$, as its *three-dimensional Radon transform*, $\mathscr{R}_3 f$. In the context of 3D PET, the 3D Radon transform of the function f is usually stored in the form of the so-called *3D sinogram*, denoted by $\widehat{f}(\rho, \theta, \phi)$:

$$\mathscr{R}_3 f = \widehat{f}(\rho, \theta, \phi) = \iint_P f\mathrm{d}s, \tag{12}$$

where $\mathrm{d}s$ denotes an area differential and $S(\mathbb{R}^3)$ denotes the space of Schwartz functions in $\mathbb{R}^3$:

$$S(\mathbb{R}^3) = \left\{ f \in C^\infty(\mathbb{R}^3) : ||f||_{\alpha,\beta} < \infty \right\} \subset C^\infty(\mathbb{R}^3), \tag{13}$$

and

$$||f||_{\alpha,\beta} = \sup_{x \in \mathbb{R}^3} |x^\alpha D^\beta f(x)|, \quad \forall \,\text{multi} - \text{index}\, \alpha, \beta, \quad |x^\alpha D^\beta f(x)| \to 0, \quad \text{as } |x| \to \infty. \tag{14}$$

Equation (12) may be rewritten via a parameterization $\mathbf{x} := \mathbf{x}(u, v)$ of the plane P, with $\mathbf{x} : \mathbb{R}^3 \to P$, as follows:

[2] Plane integrals are special cases of surface integrals, where the surface of integration is a plane.

$$\widehat{f}(\rho,\theta,\phi)=\int_{-\infty}^{\infty} f\left(\mathbf{x}(u,v)\right)\left|\frac{\partial\mathbf{x}}{\partial u}\times\frac{\partial\mathbf{x}}{\partial v}\right|\mathrm{d}u\mathrm{d}v, \tag{15}$$

where, for the area differential we employed

$$\mathrm{d}s=\left|\frac{\partial\mathbf{x}}{\partial u}\times\frac{\partial\mathbf{x}}{\partial v}\right|\mathrm{d}u\mathrm{d}v. \tag{16}$$

In the three-dimensional setting, for convenience, we characterize each two-dimensional plane P by a vector and a scalar, namely:

(i) the unit normal vector $\mathbf{n}$

$$\mathbf{n}=\begin{bmatrix}n_1\\ n_2\\ n_3\end{bmatrix},\quad \text{with}\quad \sqrt{n_1^2+n_2^2+n_3^2}=1,\quad \text{and} \tag{17}$$

(ii) the signed distance from the origin ρ.

The normal from the origin to the plane intersects the plane at the point $\rho\mathbf{n}$. Thus, if $\mathbf{x}=(x_1,x_2,x_3)^{\mathrm{T}}$ is a point on the plane under investigation, then

$$(\rho\mathbf{n}-\mathbf{x})\cdot\mathbf{n}=0 \tag{18}$$

The above implies

$$\begin{bmatrix}\rho n_1-x_1\\ \rho n_2-x_2\\ \rho n_3-x_3\end{bmatrix}\cdot\begin{bmatrix}n_1\\ n_2\\ n_3\end{bmatrix}=0,$$

or

$$\rho(n_1^2+n_2^2+n_3^2)-(n_1x_1+n_2x_2+n_3x_3)=0.$$

Hence, the equation of the plane is

$$\rho-\mathbf{n}\cdot\mathbf{x}=0,\quad \forall\,\mathbf{x}\in P. \tag{19}$$

We suppose that the plane of integration P, specified by its signed distance from the origin ρ and its unit normal vector $\mathbf{n}$, intersects the x_1x_2-plane in an angle θ, and the x_2x_3-plane in an angle ϕ (spherical angles). In this connection, the unit normal vector $\mathbf{n}$ is uniquely specified by the two spherical angles, i.e. $\mathbf{n}:=\mathbf{n}(\theta,\phi)$. Thus $f(\rho,\mathbf{n})$ involves three variables, namely $f(\rho,\mathbf{n})=f(\rho,\theta,\phi)$. In this direction, we characterize $\mathbf{n}$ in terms of spherical angles, as follows:

$$\mathbf{n}(\theta,\phi)=\begin{bmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{bmatrix},\quad 0\le\theta\le\pi\quad 0\le\phi\le 2\pi. \tag{20}$$

If a point $\mathbf{x} = (x_1, x_2, x_3)^{\mathrm{T}}$ lies on the plane of integration P, i.e. $\mathbf{x} \in P$, then Equation (19) implies

$$\rho - x_1 \sin\theta \cos\phi - x_2 \sin\theta \sin\phi - x_3 \cos\theta = 0. \tag{21}$$

Equation (21) provides a convenient parameterization of the plane under investigation via x_1 and x_2, treated hereafter as independent variables. In this setting, x_3 will be considered as a dependent variable. Taking into account the parameterization induced by Equation (21), we rewrite the equation of a point $\mathbf{x}$ lying on the plane of integration P as $\mathbf{x} = \mathbf{x}(x_1, x_2)$. More specifically, if $\mathbf{x} \in P$, then:

$$\mathbf{x} := \mathbf{x}(x_1, x_2; \rho, \theta, \phi) = (x_1, x_2, \csc\theta\,(\rho - x_1 \sin\theta \cos\phi - x_2 \sin\theta \sin\phi))^{\mathrm{T}}, \tag{22}$$

where $\csc\theta$ denotes the cosecant of the angle θ, i.e.,

$$\csc\theta = \frac{1}{\cos\theta}. \tag{23}$$

Hence, the area differential $\mathrm{d}s$ is given by

$$\mathrm{d}s = \left| \frac{\partial \mathbf{x}(x_1, x_2)}{\partial x_1} \times \frac{\partial \mathbf{x}(x_1, x_2)}{\partial x_2} \right| \mathrm{d}x_1 \mathrm{d}x_2, \tag{24}$$

where

$$\frac{\partial \mathbf{x}}{\partial x_1} \times \frac{\partial \mathbf{x}}{\partial x_2} = \begin{vmatrix} \hat{\mathbf{x}}_1 & \hat{\mathbf{x}}_2 & \hat{\mathbf{x}}_3 \\ \frac{\partial x_1}{\partial x_1} & \frac{\partial x_2}{\partial x_1} & \frac{\partial x_3}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2} & \frac{\partial x_2}{\partial x_2} & \frac{\partial x_3}{\partial x_2} \end{vmatrix} = \begin{vmatrix} \hat{\mathbf{x}}_1 & \hat{\mathbf{x}}_2 & \hat{\mathbf{x}}_3 \\ 1 & 0 & -\tan\theta\cos\phi \\ 0 & 1 & \tan\theta\cos\phi \end{vmatrix} = \begin{bmatrix} \tan\theta\cos\phi \\ \tan\theta\cos\phi \\ 1 \end{bmatrix}, \tag{25}$$

and $\hat{\mathbf{x}}_1$, $\hat{\mathbf{x}}_2$, and $\hat{\mathbf{x}}_3$ are the corresponding unit vectors in the x_1, x_2, and x_3 directions, respectively. Taking into account Equation (25), the magnitude of the above "Jacobian" vector is,

$$\left| \frac{\partial \mathbf{x}}{\partial x_1} \times \frac{\partial \mathbf{x}}{\partial x_2} \right| = \sqrt{\tan^2\theta + 1} = \csc\theta. \tag{26}$$

Hence Equation (24) yields

$$\mathrm{d}s = \csc\theta \mathrm{d}x_1 \mathrm{d}x_2. \tag{27}$$

Thus, Equation (15) becomes

$$\widehat{f}(\rho, \theta, \phi) = \int_{-\infty}^{\infty} \mathrm{d}x_1 \int_{-\infty}^{\infty} \csc\theta f\,(x_1, x_2, \csc\theta(\rho - x_1 \sin\theta \cos\phi - x_2 \sin\theta \sin\phi))\,\mathrm{d}x_2. \tag{28}$$

An alternative way to express the three-dimensional Radon transform of a function $f : \mathbb{R}^3 \to \mathbb{R}$ involves a Dirac delta, or "plane impulse", namely

$$\mathscr{R}_3 f = \widehat{f}(\rho, \theta, \phi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2, x_3) \times \delta(\rho - x_1 \sin\theta\cos\phi - x_2 \sin\theta \sin\phi - x_3 \cos\theta)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3. \tag{29}$$

The alternative definition given by Equation (29) will be proven very useful (see Theorem 1) for the inversion and the numerical implementation of the three-dimensional Radon transform, as discussed in Sections 4 and 5.

4 The Inversion of the Radon Transform in Three Space Dimensions via Plane Integration

For the analytical inversion of the Radon transform in three space dimensions defined in Equation (28) we shall employ plane integration. In this direction, we will make use of the so-called *central slice theorem* (CST). This specific theorem, applied in the three-dimensional case, provides a fundamental tool for the Fourier-based inversion of the 3D Radon transform.

Theorem 1 (Central Slice Theorem in 3D) *The three-dimensional Fourier transform $\mathscr{F}_3$ of a function $f(x_1, x_2, x_3)$, usually denoted by $\widetilde{f} = \mathscr{F}_3 f$, equals the one-dimensional Fourier transform with respect to the signed distance from the origin $\mathscr{F}_1^{(\rho)}$ of the three-dimensional Radon transform $\mathscr{R}_3$ of the same function $\widehat{f} = \mathscr{R}_3 f$, i.e.*

$$\mathscr{F}_3 f = \mathscr{F}_1^{(\rho)} \mathscr{R}_3 f, \quad \text{or} \quad \widetilde{f} = \mathscr{F}_1^{(\rho)} \widehat{f}, \tag{30}$$

where

$$\widetilde{f}(k_1, k_2, k_3) := (\mathscr{F}_3 f)(k_1, k_2, k_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2, x_3) e^{-2\pi \mathrm{i}(k_1 x_1 + k_2 x_2 + k_3 x_3)} \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3, \tag{31}$$

and

$$\mathscr{F}_1^{(\rho)} \widehat{f} = \int_{-\infty}^{\infty} \widehat{f}(\rho, \theta, \phi) e^{-2\pi \mathrm{i} k \rho} \mathrm{d}\rho. \tag{32}$$

Proof For the proof of the central slice theorem in three dimensions, it is convenient to employ the alternative definition of the three-dimensional Radon transform as

provided via a delta function in Equation (29). In this case, we expand Equation (32) as follows:

$$
\begin{aligned}
\mathscr{F}_1^{(\rho)}\widehat{f} &= \int_{-\infty}^{\infty} \widehat{f}(\rho,\theta,\phi)e^{-2\pi \mathrm{i}k\rho}\mathrm{d}\rho \qquad (33)\\
&= \int_{-\infty}^{\infty}\Bigg(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2,x_3)\\
&\quad \times \delta(\rho - x_1\sin\theta\cos\phi - x_2\sin\theta\sin\phi - x_3\cos\theta)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\Bigg)e^{-2\pi \mathrm{i}k\rho}\mathrm{d}\rho\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2,x_3)\Bigg(\int_{-\infty}^{\infty} e^{-2\pi \mathrm{i}k\rho}\\
&\quad \times \delta(\rho - x_1\sin\theta\cos\phi - x_2\sin\theta\sin\phi - x_3\cos\theta)\mathrm{d}\rho\Bigg)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2,x_3)e^{-2\pi \mathrm{i}k(x_1\sin\theta\cos\phi + x_2\sin\theta\sin\phi + x_3\cos\theta)}\\
&\quad \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2,x_3)e^{-2\pi \mathrm{i}[(k\sin\theta\cos\phi)x_1 + (k\sin\theta\sin\phi)x_2 + (k\cos\theta)x_3]}\\
&\quad \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2,x_3)e^{-2\pi \mathrm{i}(k_1x_1 + k_2x_2 + k_3x_3)}\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\\
&= \widetilde{f},
\end{aligned}
$$

where we introduced the new k-variables in the Fourier space vector $\mathbf{k} := (k_1, k_2, k_3)^{\mathrm{T}}$ of spatial frequencies as follows:

$$k_1 = k\sin\theta\cos\phi, \quad k_2 = k\sin\theta\sin\phi, \quad k_3 = k\cos\theta, \qquad (34)$$

as the new k-variables in the Fourier space. □

Hence, the Fourier transform with respect to ρ of the "data" equals the three-dimensional Fourier transform of the function under investigation evaluated at the new set of variables. The inversion of Equation (30) yields

$$f = \mathscr{F}_3^{-1}\widetilde{f}. \qquad (35)$$

For the inversion of the 3D Radon transform we will utilize the following corollary.

Corollary 1 (One-Dimensional Fourier Transform of the Second Derivative) *For any twice differentiable function g with respect to ρ, the one-dimensional Fourier transform of the second derivative of g, $\mathscr{F}_1^{(\rho)} g''$, is related with the one-dimensional Fourier transform of g, $\mathscr{F}_1^{(\rho)} g$, by the following expression:*

$$\left(\mathscr{F}_1^{(\rho)} g''\right)(\xi) = -4\pi^2 \xi^2 G(\xi), \tag{36}$$

where $\mathscr{F}_1^{(\rho)}$ denotes the one-dimensional Fourier transform with respect to ρ defined in Equation (32), g'' denotes the second derivative of g with respect to ρ, i.e.,

$$g'' = \frac{\partial^2 g}{\partial \rho^2}, \tag{37}$$

and G denotes the one-dimensional Fourier transform of g,

$$G = \mathscr{F}_1^{(\rho)} g. \tag{38}$$

Proof Inverting Equation (32) and employing g instead of $\widetilde{f}$ yields

$$g = \left\{\mathscr{F}_1^{(\rho)}\right\}^{-1} G = \int_{-\infty}^{\infty} G(\xi) e^{2\pi \mathrm{i} \rho \xi} \mathrm{d}\xi, \tag{39}$$

where G is defined in Equation (38). As in [12], we take the second derivative of both sides of Equation (39):

$$\begin{aligned} g'' := \frac{\partial^2 g}{\partial \rho^2} &= \frac{\partial^2}{\partial \rho^2} \left(\int_{-\infty}^{\infty} G(\xi) e^{2\pi \mathrm{i} \rho \xi} \mathrm{d}\xi \right) \\ &= \int_{-\infty}^{\infty} G(\xi) \left[\frac{\partial^2}{\partial \rho^2} \left(e^{2\pi \mathrm{i} \rho \xi} \right) \right] \mathrm{d}\xi \\ &= \int_{-\infty}^{\infty} G(\xi) \left[(2\pi \mathrm{i} \xi)^2 e^{2\pi \mathrm{i} \rho \xi} \right] \mathrm{d}\xi \\ &= \int_{-\infty}^{\infty} \left[-4\pi^2 \xi^2 G(\xi) \right] e^{2\pi \mathrm{i} \rho \xi} \mathrm{d}\xi. \end{aligned} \tag{40}$$

From the above it is clear that the functions g'' and $-4\pi^2 \xi^2 G(\xi)$ form a Fourier transform pair. Hence, Equation (40), combined with Equation (39), imply Equation (36). □

Theorem 2 (Inversion of the Three-Dimensional Fourier Transform) *The inverse of the three dimensional Radon transform $\widetilde{f} = \mathscr{R}_3 f$, defined in Equations (28) and (29), of a Schwartz function $f \in \mathscr{S}(\mathbb{R}^3)$ is given by*

$$f(x_1, x_2, x_3) = -\frac{1}{4\pi^2} \int_0^{\pi} \sin\theta \mathrm{d}\theta \int_0^{2\pi} \widetilde{f}''(\rho^*, \theta, \phi)\mathrm{d}\phi, \tag{41}$$

where, as in Equation (37)*, prime denotes differentiation with respect to* ρ*, i.e.*

$$\widetilde{f}''(\rho^*, \theta, \phi) = \frac{\partial^2}{\partial \rho^2} \widehat{f}(\rho, \theta, \phi)\Big|_{\rho=\rho^*}, \tag{42}$$

and ρ^* *is given by*

$$\rho^* = x_1 \sin\theta \cos\phi + x_2 \sin\theta \sin\phi + x_3 \cos\theta. \tag{43}$$

Proof Equation (35) implies

$$\left(\mathscr{F}_3^{-1} g\right)(x_1, x_2, x_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2, x_3) e^{2\pi \mathrm{i}(k_1 x_1 + k_2 x_2 + k_3 x_3)} \mathrm{d}k_1 \mathrm{d}k_2 \mathrm{d}k_3. \tag{44}$$

In this direction, the inversion of the three-dimensional Fourier transform will reveal the unknown function f in the sense that:

$$f(x_1, x_2, x_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \widetilde{f}(k_1, k_2, k_3) e^{2\pi \mathrm{i}(k_1 x_1 + k_2 x_2 + k_3 x_3)} \mathrm{d}k_1 \mathrm{d}k_2 \mathrm{d}k_3. \tag{45}$$

We proceed by making a change of variables from (k_1, k_2, k_3) to (k, θ, ϕ) defined by Equation (34). The corresponding Jacobian is given by

$$J(k, \theta, \phi) = \begin{vmatrix} \frac{\partial k_1}{\partial k} & \frac{\partial k_1}{\partial \theta} & \frac{\partial k_1}{\partial \phi} \\ \frac{\partial k_2}{\partial k} & \frac{\partial k_2}{\partial \theta} & \frac{\partial k_2}{\partial \phi} \\ \frac{\partial k_3}{\partial k} & \frac{\partial k_3}{\partial \theta} & \frac{\partial k_3}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \sin\theta\cos\phi & k\cos\theta\cos\phi & -k\sin\theta\sin\phi \\ \sin\theta\sin\phi & k\cos\theta\sin\phi & k\sin\theta\cos\phi \\ \cos\theta & -k\sin\theta & 0 \end{vmatrix} = k^2 \sin\theta. \tag{46}$$

Thus we modify Equation (45) in the following manner:

$$\begin{aligned} f(x_1, x_2, x_3) = \int_0^{\pi} \int_0^{2\pi} \int_{-\infty}^{\infty} & \widetilde{f}(k\sin\theta\cos\phi, k\sin\theta\sin\phi, k\cos\theta) \\ & \times e^{2\pi \mathrm{i} k(\sin\theta\cos\phi x_1 + \sin\theta\sin\phi x_2 + \cos\theta x_3)} J(k, \theta, \phi) \mathrm{d}k \mathrm{d}\theta \mathrm{d}\phi. \end{aligned} \tag{47}$$

However, a point $\mathbf{x} = (x_1, x_2, x_3)^{\mathrm{T}}$ lying on the plane of integration P ($\mathbf{x} \in P$), according to Equation (21), satisfies

$$\rho^* - x_1 \sin\theta\cos\phi - x_2 \sin\theta\sin\phi - x_3 \cos\theta = 0, \tag{48}$$

where ρ^* is the signed distance of the plane P from the origin, see Equation (43). Hence, taking into account Equations (46), (48), and (47) may be rewritten as

$$f(x_1, x_2, x_3) = \int_0^{\pi} \sin\theta \mathrm{d}\theta \int_0^{2\pi} \left[\int_{-\infty}^{\infty} \widetilde{f}(k\sin\theta\cos\phi, k\sin\theta\sin\phi, k\cos\theta) \times e^{2\pi \mathrm{i} k\rho^*} k^2 \mathrm{d}k \right] \mathrm{d}\phi. \tag{49}$$

We combine Equations (30) with (32) and (34) to obtain

$$\widetilde{f}(k\sin\theta\cos\phi, k\sin\theta\sin\phi, k\cos\theta) = \int_{-\infty}^{\infty} \widehat{f}(\rho, \theta, \phi) e^{-2\pi \mathrm{i} k\rho} \mathrm{d}\rho. \tag{50}$$

In Equation (49), we replace $\widetilde{f}$ by the right-hand side of Equation (50)

$$f(x_1, x_2, x_3) = \int_0^{\pi} \sin\theta \mathrm{d}\theta \int_0^{2\pi} \left[\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \widehat{f}(\rho, \theta, \phi) e^{-2\pi \mathrm{i} k\rho} \mathrm{d}\rho \right) e^{2\pi \mathrm{i} k\rho^*} k^2 \mathrm{d}k \right] \mathrm{d}\phi. \tag{51}$$

We denote the one-dimensional Fourier transform of the three-dimensional Radon transform of f by $\widehat{F}$,

$$\widehat{F}(k) := \int_{-\infty}^{\infty} \widehat{f}(\rho, \theta, \phi) e^{-2\pi \mathrm{i} k\rho} \mathrm{d}\rho, \tag{52}$$

and insert Equation (52) into Equation (51):

$$f(x_1, x_2, x_3) = \int_0^{\pi} \sin\theta \mathrm{d}\theta \int_0^{2\pi} \left[\int_{-\infty}^{\infty} k^2 \widehat{F}(k) e^{2\pi \mathrm{i} k\rho^*} \mathrm{d}k \right] \mathrm{d}\phi. \tag{53}$$

The final step involves the rewriting of Equation (53) in the following manner:

$$f(x_1, x_2, x_3) = -\frac{1}{4\pi^2} \int_0^{\pi} \sin\theta \mathrm{d}\theta \int_0^{2\pi} \left[\int_{-\infty}^{\infty} \left(-4\pi^2 k^2 \widehat{F}(k) \right) e^{2\pi \mathrm{i} k\rho^*} \mathrm{d}k \right] \mathrm{d}\phi. \tag{54}$$

We employ Corollary 1, and replace the integral inside the brackets on the left-hand side of Equation (54), by the left-hand side of the first line of Equation (40) to obtain

$$f(x_1, x_2, x_3) = -\frac{1}{(4\pi)^2} \int_0^{\pi} \sin\theta \mathrm{d}\theta \int_0^{2\pi} \left. \frac{\partial^2 \widetilde{f}(\rho, \theta, \phi)}{\partial \rho^2} \right|_{\rho=\rho^*} \mathrm{d}\phi, \tag{55}$$

which, via Equation (42), is Equation (41). □

5 Numerical Implementation of the Inversion of the Radon Transform in Three Space Dimensions via Piecewise Cubic Polynomials

For the numerical implementation of the inversion of the Radon transform in three space dimensions we will employ piecewise continuous cubic polynomials, namely cubic splines. It is important to note that all integrals involving the second derivative with respect to ρ of the 3D Radon transform will be evaluated at $\rho = \rho^*$, namely at

$$\rho^* = x_1 \sin\theta \cos\phi + x_2 \sin\theta \sin\phi + x_3 \cos\theta. \tag{56}$$

As shown in the previous section, the 3D inverse Radon transform can be expressed as

$$f(x_1, x_2, x_3) = -\frac{1}{4\pi^2} \int_0^{\pi} \sin\theta \mathrm{d}\theta \int_0^{2\pi} \widehat{f}''(\rho, \theta, \phi)\Big|_{\rho=\rho^*} \mathrm{d}\phi \tag{57}$$

where ρ^*, $\widehat{f}$ and $\widehat{f}''$ are defined in Equations (56), (28), and (42), respectively.

We assume that the three-dimensional Radon transform, $\widehat{f}$, is given for every θ and every ϕ at the n knots $\{\rho_i\}_1^n$. We denote the value of $\widehat{f}$ at ρ_i by $\widehat{f}_i$, namely

$$\widehat{f}_i = \widehat{f}(\rho_i, \theta, \phi), \quad \theta \in [0, \pi], \quad \phi \in [0, 2\pi], \quad i = 1, \ldots, n-1. \tag{58}$$

We also assume that both $\widehat{f}(\rho, \theta, \phi)$ and $\widehat{f}'(\rho, \theta, \phi)$, where

$$\widehat{f}'(\rho, \theta, \phi) = \frac{\partial \widehat{f}(\rho, \theta, \phi)}{\partial \rho}, \tag{59}$$

vanish at the endpoints $\rho_1 = -1$ and $\rho_n = 1$, i.e.

$$\widehat{f}(\rho_1, \theta, \phi) = \widehat{f}(\rho_n, \theta, \phi) = 0, \quad \theta \in [0, \pi], \quad \phi \in [0, 2\pi], \tag{60}$$

and

$$\frac{\partial}{\partial \rho} \widehat{f}(\rho_1, \theta, \phi) = \frac{\partial}{\partial \rho} \widehat{f}(\rho_n, \theta, \phi) = 0, \quad \theta \in [0, \pi], \quad \phi \in [0, 2\pi]. \tag{61}$$

In each interval $[\rho_i, \rho_{i+1}]$, $i = 1, \ldots, n-1$, we approximate $\widehat{f}(\rho, \theta, \phi)$ by the third-degree spline $S_i^{(3)}$, namely

$$\widehat{f}(\rho, \theta, \phi) \sim S_i^{(3)}(\rho, \theta, \phi), \quad \rho \in [\rho_i, \rho_{i+1}] \quad \theta \in [0, \pi], \quad \phi \in [0, 2\pi]. \tag{62}$$

The cubic spline $S_i^{(3)}$ interpolates $\widetilde{f}$ at the points $\{\rho_i\}_1^n$:

$$S_i^{(3)}(\rho_i, \theta, \phi) = f_i, \quad i = 1, \ldots, n. \tag{63}$$

Therefore, for $\rho \in [\rho_i, \rho_{i+1}]$

$$S_i^{(3)}(\rho, \theta, \phi) = a_i(\theta, \phi) + b_i(\theta, \phi)\rho + c_i(\theta, \phi)\rho^2 + d_i(\theta, \phi)\rho^3. \tag{64}$$

Then, following Equation (62),

$$\frac{\partial}{\partial \rho}\widehat{f}(\rho, \theta, \phi) \sim \frac{\partial}{\partial \rho} S_i^{(3)}(\rho, \theta, \phi) =: S_i^{(2)}(\rho, \theta, \phi), \tag{65}$$

where

$$S_i^{(2)}(\rho, \theta, \phi) = b_i(\theta, \phi) + 2c_i(\theta, \phi)\rho + 3d_i(\theta, \phi)\rho^2. \tag{66}$$

Similarly,

$$\frac{\partial^2}{\partial \rho^2}\widehat{f}(\rho, \theta, \phi) \sim \frac{\partial^2}{\partial \rho^2} S_i^{(3)}(\rho, \theta, \phi) = \frac{\partial}{\partial \rho} S_i^{(2)}(\rho, \theta, \phi) =: S_i^{(1)}(\rho, \theta, \phi), \tag{67}$$

where

$$S_i^{(1)}(\rho, \theta, \phi) = 2c_i(\theta, \phi) + 6d_i(\theta, \phi)\rho \tag{68}$$

Hence, Equation (57) becomes

$$f(x, y, z) = -\frac{1}{4\pi^2}\int_0^{\pi}\int_0^{\pi} [2c_i(\theta, \phi) + 6d_i(\theta, \phi)\rho] \sin\theta \mathrm{d}\phi \mathrm{d}\theta \tag{69}$$

The constants $c_i(\theta, \phi)$ and $d_i(\theta, \phi)$ involved in the above inversion integral, are given by the following expressions, see [13]:

$$c_i(\theta) = \frac{1}{2\Delta_i}(\rho_{i+1}\widehat{f}_i'' - \rho_i \widehat{f}_{i+1}''), \tag{70a}$$

$$d_i(\theta) = \frac{\widehat{f}_{i+1}'' - \widehat{f}_i''}{6\Delta_i}, \tag{70b}$$

where

$$\Delta_i = \rho_{i+1} - \rho i, \tag{70c}$$

and

$$\widehat{f}_i'' := \frac{\partial^2}{\partial \rho^2} \widehat{f}(\rho_i, \theta, \phi). \tag{70d}$$

It is worth noting that the inversion formula (57) involves the known constants $\{\widehat{f}\}_1^n$ and the unknown constants $\{\widehat{f}''\}_1^n$. For the computation of $\{\widehat{f}''\}_1^n$, we employ the continuity of the first derivative of the cubic spline, i.e.

$$S_i^{(2)}(\rho_{i+1}, \theta, \phi) = S_{i+1}^{(2)}(\rho_i, \theta, \phi), \quad i = 1, 2, \ldots, n-2, \quad \theta \in [0, \pi], \quad \phi \in [0, 2\pi], \tag{71a}$$

and

$$S_1^{(2)}(\rho_1, \theta, \phi) = S_{n-1}^{(2)}(\rho_n, \theta, \phi) = 0, \quad \theta \in [0, \pi], \quad \phi \in [0, 2\pi]. \tag{71b}$$

The above consists of a system of n unknowns and of n equations, namely $n-2$ equations arising from Equation (71a) for $i = 1, 2, \ldots, n-2$), and 2 equations arising from Equation (71b). The continuity of the cubic spline itself, i.e.

$$S_i^{(3)}(\rho_{i+1}, \theta, \phi) = S_{i+1}^{(3)}(\rho_i, \theta, \phi) = 0, \quad i = 1, 2, \ldots, n-2, \quad \theta \in [0, \pi], \quad \phi \in [0, 2\pi] \tag{72a}$$

and

$$S_1^{(2)}(\rho_1, \theta, \phi) = S_{n-1}^{(2)}(\rho_n, \theta, \phi) = 0, \quad \theta \in [0, \pi], \quad \phi \in [0, 2\pi]. \tag{72b}$$

The continuity of the cubic spline, $S_i^{(3)}(\rho, \theta, \phi)$, as expressed in Equations (72), implies that the knots $\{\rho_i\}_1^n$ are *removable* logarithmic singularities.

Acknowledgments This work was partially supported by the research programme "Inverse Problems and Medical Imaging" (200/947) of the Research Committee of the Academy of Athens. A.S. Fokas has been supported by EPSRC, UK in the form of a senior fellowship.

References

1. J. Radon, Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. Akad. Wiss. **69**, 262–277 (1917)
2. P. Kuchment, The Radon transform and medical imaging, in *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (2014)
3. N.E. Protonotarios, G.A. Kastis, A.S. Fokas, *A New Approach for the Inversion of the Attenuated Radon Transform* (Springer, Cham, 2019), pp. 433–457. https://doi.org/10.1007/978-3-030-31339-5_16
4. M.-Y. Chiu, H.H. Barrett, R.G. Simpson, Three-dimensional reconstruction from planar projections. JOSA **70**(7), 755–762 (1980). https://doi.org/10.1364/JOSA.70.000755
5. M. Defrise, D. Townsend, R. Clack, Three-dimensional image reconstruction from complete projections. Phys. Med. Biol. **34**(5), 573 (1989). https://doi.org/10.1088/0031-9155/34/5/002

6. A. Averbuch, Y. Shkolnisky, 3d fourier based discrete radon transform. Appl. Comput. Harmon. Anal. **15**(1), 33–69 (2003). https://doi.org/10.1016/S1063-5203(03)00030-7
7. J. Bikowski, K. Knudsen, J.L. Mueller, Direct numerical reconstruction of conductivities in three dimensions using scattering transforms. Inverse Probl. **27**(1), 015002 (2010). https://doi.org/10.1088/0266-5611/27/1/015002
8. F.H. Fahey, Data acquisition in PET imaging. J. Nucl. Med. Technol. **30**(2), 39–49 (2002)
9. P. Kinahan, J. Rogers, Analytic 3D image reconstruction using all detected events. IEEE Trans. Nucl. Sci. **36**, 964–968 (1989). https://doi.org/10.1109/23.34585
10. P. La Rivière, X. Pan, Spline-based inverse radon transform in two and three dimensions. IEEE Trans. Nucl. Sci. **45**(4), 2224–2231 (1998). https://doi.org/10.1109/23.708352
11. M. Defrise, P.E. Kinahan, C.J. Michel, Image reconstruction algorithms in PET, in *Positron Emission Tomography* (Springer, Berlin, 2005), pp. 63–91. https://doi.org/10.1007/1-84628-007-9_4
12. J. Gaskill, J.W. Sons, *Linear Systems, Fourier Transforms, and Optics*. *Wiley Series in Pure and Applied Optics* (Wiley, New York, 1978)
13. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd edn. (Cambridge University, New York, 2007)

Factorization and Solution of Linear and Nonlinear Second Order Differential Equations with Variable Coefficients and Mixed Conditions

E. Providas

Abstract This chapter deals with the factorization and solution of initial and boundary value problems for a class of linear and nonlinear second order differential equations with variable coefficients subject to mixed conditions. The technique for nonlinear differential equations is based on their decomposition into linear components of the same or lower order and the factorization of the associated second order linear differential operators. The implementation and efficiency of the procedure is shown by solving several examples.

1 Introduction

One of the most important categories of ordinary differential equations is the second order differential equations with variable coefficients. Many problems from engineering and science are within this large class of differential equations. These equations, in addition to their natural significance, have also been used as a vehicle for the study of other higher order differential equations. Both exact and numerical methods have been developed for the their solution [2]. Most of the explicit techniques rely on the knowledge of fundamental solutions. The factorization method does not require any fundamental solution of the given second order differential equation, but its applicability is limited to certain problems. For a review of the factorization of differential operators the interested reader can look at the selected articles [1, 3–9, 14–16].

Following the work in [10–13] and [17], this paper is concerned with the exact solution of a class of linear and nonlinear differential equations of second order with variable coefficients subject to nonlocal boundary conditions by direct factorization of the differential equation as well as the boundary conditions. Specifically, in Section 2, we recall some basic results and consider linear first order problems with

E. Providas (✉)
Department of Environmental Sciences, University of Thessaly, Larissa, Greece
e-mail: providas@uth.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_18

a mixed boundary condition. In Section 3, we present the operator factorization method for solving, under certain conditions, the linear second order differential equation

$$u''(x) + p(x)u(x) + q(x)u(x) = f(x), \quad x \in (a, b), \tag{1}$$

where the coefficients $p(x)$, $q(x) \in C[a, b]$ and the forcing function $f(x) \in C[a, b]$, subject to general boundary conditions

$$\begin{aligned} \mu_{11}u(a) + \mu_{12}u(b) &= \beta_1, \\ \mu_{21}u'(a) + \mu_{22}u'(b) + \mu_{23}u(a) + \mu_{24}u(b) &= \beta_2, \end{aligned} \tag{2}$$

where μ_{ij}, $\beta_i \in \mathbb{R}$, $i = 1, 2$, $j = 1, 2, 3, 4$. In Section 4, we deal with the construction of explicit solutions to two kinds of nonlinear differential equations of second order, which can be decomposed initially into linear second order differential equations. First, we consider the equation of the form

$$u''(x)u'(x) + \left[q(x)u'(x) + g(x)u''(x)\right]u(x) + q(x)g(x)u^2(x) = 0, \tag{3}$$

for $x \in (a, b)$ and $q(x)$, $g(x) \in C[a, b]$, along with the general boundary conditions (2). Also, we consider the nonlinear differential equation of the type

$$F\left(\frac{u''(x)}{u(x)}, x\right) = F\left(w(x), x\right) = w^2(x) + a(x)w(x) + b(x) = 0, \quad x \in (a, b), \tag{4}$$

where the nonlinear function F is a second degree polynomial of $w(x) = u''(x)/u(x)$ and the coefficients $a(x)$, $b(x) \in C[a, b]$, subject to general boundary conditions (2). Finally, some conclusions are quoted in Section 5.

2 Preliminaries

We first recall some basic results. A linear operator $P : C[a, b] \to C[a, b]$ is said to be *correct* if P is injective, $R(P) = C[a, b]$ and its inverse P^{-1} is bounded on $C[a, b]$. Let $A : C[a, b] \to C[a, b]$ be the linear first order operator

$$Ay(x) = y'(x) + a(x)y(x), \quad D(A) = C^1[a, b], \tag{5}$$

where $a(x) \in C[a, b]$, and $\widehat{A}$ be its restriction on

$$D(\widehat{A}) = \{y(x) \in D(A) : \ y(x_0) = y_0\}, \tag{6}$$

where $x_0 \in [a,b]$ and y_0 is an arbitrary real initial value. Then the following fundamental theorem holds.

Theorem 1 *The linear operator $\widehat{A}$ in (5) and (6) is correct and the unique solution of the initial value problem*

$$\widehat{A}y(x) = f(x), \quad \forall f(x) \in C[a,b], \tag{7}$$

is given by

$$y(x) = \widehat{A}^{-1}f(x) = e^{-\int_{x_0}^{x} a(t)dt}\left(y_0 + \int_{x_0}^{x} f(t)e^{\int_{t_0}^{t} a(\tau)d\tau}dt\right). \tag{8}$$

Accordingly, let $B : C[a,b] \to C[a,b]$ be the linear second order operator

$$By(x) = y''(x) + b_1(x)y'(x) + b_2(x)y(x), \quad D(B) = C^2[a,b], \tag{9}$$

where $b_1(x),\ b_2(x) \in C[a,b]$, and $\widehat{B}$ be its restriction on

$$D(\widehat{B}) = \left\{y(x) \in D(B) :\ y(x_0) = y_0,\ y'(x_0) = y_0'\right\}, \tag{10}$$

where $x_0 \in [a,b]$ and $y_0,\ y_0'$ is a couple of given real numbers. Then we have the next fundamental theorem.

Theorem 2 *The linear operator $\widehat{B}$ in (9) and (10) is correct and the initial value problem*

$$\widehat{B}y(x) = f(x), \quad \forall f(x) \in C[a,b], \tag{11}$$

has exactly one solution $y(x) = \widehat{B}^{-1}f(x)$.

We now consider a problem for a first order differential equation and a nonlocal boundary condition, which we will encounter below. For this, we prove the next theorem.

Theorem 3 *Let the general linear first order problem with a nonlocal boundary condition*

$$\begin{aligned} &\widehat{Q}y(x) = y'(x) + q(x)y(x) = f(x), \\ &D(\widehat{Q}) = \left\{y(x) \in C^1[a,b] :\ \mu_1 y(a) + \mu_2 y(b) = \beta\right\}, \end{aligned} \tag{12}$$

where the operator $\widehat{Q} : C[a,b] \to C[a,b]$, the given functions $q(x),\ f(x) \in C[a,b]$, and the constants $\mu_1, \mu_2, \beta \in \mathbb{R}$. If

$$\mu_1 + \mu_2 e^{-\int_a^b q(t)dt} \neq 0, \tag{13}$$

then the operator $\widehat{Q}$ is correct and the unique solution of problem (12) is given by

$$y(x) = \widehat{Q}^{-1} f(x) = e^{-\int_a^x q(t)dt} \left(C + \int_a^x f(t) e^{\int_a^t q(\tau)d\tau} dt \right), \tag{14}$$

where

$$C = \left(\mu_1 + \mu_2 e^{-\int_a^b q(t)dt} \right)^{-1} \left(\beta - \mu_2 e^{-\int_a^b q(t)dt} \int_a^b f(t) e^{\int_a^t q(\tau)d\tau} dt \right).$$

Proof It is known that the general solution of the first order differential equation in (12) is

$$y(x) = e^{-\int_{x_0}^x q(t)dt} \left(C + \int_{x_0}^x f(t) e^{\int_{t_0}^t q(\tau)d\tau} dt \right), \tag{15}$$

where $x_0 \in [a, b]$. For $x_0 = a$, we have

$$y(a) = C, \quad y(b) = e^{-\int_a^b q(t)dt} \left(C + \int_a^b f(t) e^{\int_a^t q(\tau)d\tau} dt \right).$$

Substituting these values into the boundary condition in (12), we obtain

$$\left(\mu_1 + \mu_2 e^{-\int_a^b q(t)dt} \right) C = \beta - \mu_2 e^{-\int_a^b q(t)dt} \int_a^b f(t) e^{\int_a^t q(\tau)d\tau} dt.$$

If relation (13) holds, then

$$C = \left(\mu_1 + \mu_2 e^{-\int_a^b q(t)dt} \right)^{-1} \left(\beta - \mu_2 e^{-\int_a^b q(t)dt} \int_a^b f(t) e^{\int_a^t q(\tau)d\tau} dt \right). \tag{16}$$

From (15) and (16) it is implied (14). □

3 Factorization Method for Linear Differential Equations

Let the linear differential operators of first order $L_1 : C[a, b] \to C[a, b]$ and $L_2 : C[a, b] \to C[a, b]$ be defined by

$$L_1 u(x) = [D + r(x)] u(x), \quad D(L_1) = C^1[a, b], \tag{17}$$

$$L_2 u(x) = [D + s(x)] u(x), \quad D(L_2) = C^1[a, b], \tag{18}$$

respectively, where $D = \frac{d}{dx}$, and the coefficients $r(x) \in C[a,b]$ and $s(x) \in C^1[a,b]$. Consider the composition,

$$\begin{aligned} L_1 L_2 u(x) &= L_1 \left(L_2 u(x)\right) \\ &= \left[D + r(x)\right] \left(\left[D + s(x)\right] u(x)\right) \\ &= \left[D^2 + \left(r(x) + s(x)\right) D + \left(s'(x) + r(x)s(x)\right)\right] u(x). \end{aligned} \tag{19}$$

This gives rise to the following proposition.

Proposition 1 *Let the linear differential operator of second order $L : C[a,b] \to C[a,b]$ be defined by*

$$Lu(x) = \left[D^2 + p(x)D + q(x)\right] u(x), \quad D(L) = C^2[a,b], \tag{20}$$

where the coefficients $p(x), q(x) \in C[a,b]$. If there exist two functions $r(x) \in C[a,b]$ and $s(x) \in C^1[a,b]$ satisfying the relations

$$r(x) + s(x) = p(x), \tag{21}$$

$$s'(x) + r(x)s(x) = q(x), \tag{22}$$

then the operator L can be factorized into a product of the two linear differential operators of first order L_1, L_2 in (17) and (18), respectively, such that

$$Lu(x) = L_1 L_2 u(x). \tag{23}$$

Remark 1 By solving equation (21) with respect to $r(x)$ and then substituting into (22), we get

$$r(x) = p(x) - s(x), \tag{24}$$

$$s'(x) + p(x)s(x) - s^2(x) = q(x), \tag{25}$$

where (25) is the nonlinear Riccati equation.

Consider the linear second order initial value problem

$$Lu(x) = f(x), \quad u(x_0) = \beta_1, \quad u'(x_0) = \beta_2, \tag{26}$$

where $f(x) \in C[a,b]$ is a forcing function, x_0 is a point in $[a,b]$, $\beta_i \in \mathbb{R}$, $i = 1, 2$, and $u(x) \in C^2[a,b]$ is the unknown function describing the response of the system modeled by (26). If (21) and (22) hold true, then this problem can be factorized and solved in closed form as it is shown in the next theorem.

Theorem 4 *Let L be the linear differential operator of second order in (20) and $\widehat{L}$ be its restriction on*

$$D(\widehat{L}) = \{u(x) \in D(L) : u(x_0) = \beta_1,\ u'(x_0) = \beta_2\}, \tag{27}$$

where $x_0 \in [a, b]$ and $\beta_1, \beta_2 \in \mathbb{R}$. If the prerequisites (21) and (22) are fulfilled then:

(i) The operator $\widehat{L}$ can be factorized as

$$\widehat{L}u(x) = \widehat{L}_1 \widehat{L}_2 u(x), \tag{28}$$

where $\widehat{L}_1, \widehat{L}_2$ are correct restrictions of the linear first order differential operators L_1, L_2, defined in (17) and (18), on

$$D(\widehat{L}_1) = \{z(x) \in D(L_1) : z(x_0) = \beta_2 + s(x_0)\beta_1\}, \tag{29}$$

$$D(\widehat{L}_2) = \{u(x) \in D(L_2) : u(x_0) = \beta_1\}, \tag{30}$$

respectively.

(ii) The operator $\widehat{L}$ is correct and the unique solution of the initial value problem

$$\widehat{L}u(x) = f(x), \quad \forall f(x) \in C[a, b], \tag{31}$$

is given in closed form by

$$\begin{aligned} u(x) &= \widehat{L}^{-1}u(x) = \widehat{L}_2^{-1}\widehat{L}_1^{-1}f(x) = \widehat{L}_2^{-1}z(x) \\ &= e^{-\int_{x_0}^{x} s(t)dt}\left(\beta_1 + \int_{x_0}^{x} z(t)e^{\int_{t_0}^{t} s(\tau)d\tau}dt\right), \end{aligned} \tag{32}$$

where

$$z(x) = \widehat{L}_1^{-1}f(x) = e^{-\int_{x_0}^{x} r(t)dt}\left(\beta_2 + s(x_0)\beta_1 + \int_{x_0}^{x} f(t)e^{\int_{t_0}^{t} r(\tau)d\tau}dt\right). \tag{33}$$

Proof

(i) From the definition of $\widehat{L}$ and Proposition 1, we have

$$Lu(x) = L_1L_2u(x) = f(x), \quad u(x_0) = \beta_1, \quad u'(x_0) = \beta_2. \tag{34}$$

By setting

$$L_2u(x) = u'(x) + s(x)u(x) = z(x), \tag{35}$$

we get

$$L_1 z(x) = z'(x) + r(x)z(x) = f(x). \tag{36}$$

From (35) it is implied that

$$u'(x_0) = z(x_0) - s(x_0)u(x_0),$$

which when is substituted into the second condition in (34) yields

$$z(x_0) = \beta_2 + s(x_0)\beta_1.$$

Whence, we have the two linear first order initial value problems

$$L_1 z(x) = f(x), \quad z(x_0) = \beta_2 + s(x_0)\beta_1, \tag{37}$$

$$L_2 u(x) = z(x), \quad u(x_0) = \beta_1. \tag{38}$$

That is $\widehat{L}u(x) = \widehat{L}_1\widehat{L}_2 u(x)$. It remains to show that $D(\widehat{L}) = D(\widehat{L}_1\widehat{L}_2)$. By using (29) and (30), we obtain

$$\begin{aligned} D(\widehat{L}_1\widehat{L}_2) &= \left\{u(x) \in D(\widehat{L}_2) : \widehat{L}_2 u(x) \in D(\widehat{L}_1)\right\} \\ &= \left\{u(x) \in D(L_2) : u(x_0) = \beta_1,\ u'(x) + s(x)u(x) \in D(\widehat{L}_1)\right\} \\ &= \left\{u(x) \in C^2[a,b] : u(x_0) = \beta_1,\ u'(x_0) + s(x_0)u(x_0) = \beta_2 + s(x_0)\beta_1\right\} \\ &= \left\{u(x) \in C^2[a,b] : u(x_0) = \beta_1,\ u'(x_0) = \beta_2\right\}. \end{aligned} \tag{39}$$

(ii) The linear first order initial value problem (37) possesses exactly one solution $z(x)$, which can be found by using the standard means, such as the method of integrating factors [2], and is given in (33). Having obtained $z(x)$, we can solve the linear first order initial value problem (38) in like manner to obtain the solution $u(x)$ in (32), which is the solution of the linear second order initial value problem (31). The operator $\widehat{L} = \widehat{L}_1\widehat{L}_2$ is correct because $\widehat{L}_1$ and $\widehat{L}_2$ are correct.

□

The factorization method also applies to some types of boundary value problems, although it is more complicated. Let the linear second order differential equation,

$$Lu(x) = u''(x) + p(x)u(x) + q(x)u(x) = f(x), \quad x \in (a,b), \tag{40}$$

where the coefficients $p(x),\ q(x) \in C[a,b]$ and $f(x) \in C[a,b]$, and assume that the operator $L : C[a,b] \to C[a,b]$ is factorable, i.e. there exist $r(x) \in C[a,b]$ and

$s(x) \in C^1[a,b]$ such that $r(x) + s(x) = p(x)$ and $s'(x) + r(x)s(x) = q(x)$. Let also the two boundary conditions

$$\begin{aligned} \mu_{11}u(a) + \mu_{12}u(b) &= \beta_1, \\ \mu_{21}[u'(a) + s(a)u(a)] + \mu_{22}[u'(b) + s(b)u(b)] &= \beta_2, \end{aligned} \tag{41}$$

where $\mu_{ij},\ \beta_i \in \mathbb{R}, i = 1, 2,\ j = 1, 2$. Notice that (41) are the boundary conditions as in (2) when

$$\mu_{23} = s(a)\mu_{21}, \quad \mu_{24} = s(b)\mu_{22}. \tag{42}$$

We claim that the boundary value problem for the differential equation (40) and the boundary conditions (41) can be factorized and solved explicitly. We prove the following theorem.

Theorem 5 *Let L be the linear second order differential operator in (40) and assume that there exist two functions $r(x) \in C[a,b]$ and $s(x) \in C^1[a,b]$ which satisfy (21) and (22). Let $\bar{L}$ be a restriction of L on*

$$\begin{aligned} D(\bar{L}) = \{u(x) : u(x) \in D(L),\ \mu_{11}u(a) + \mu_{12}u(b) = \beta_1, \\ \mu_{21}[u'(a) + s(a)u(a)] + \mu_{22}[u'(b) + s(b)u(b)] = \beta_2\}, \end{aligned} \tag{43}$$

where $\mu_{ij},\ \beta_i \in \mathbb{R}, i = 1, 2,\ j = 1, 2$. Then:

(i) The operator $\bar{L}$ can be factorized as follows

$$\bar{L}u(x) = \bar{L}_1\bar{L}_2u(x), \tag{44}$$

where $\bar{L}_1, \bar{L}_2$ are restrictions of the two first order linear differential operators L_1, L_2, defined in (17) and (18), on

$$D(\bar{L}_1) = \left\{z(x) \in C^1[a,b] :\ \mu_{21}z(a) + \mu_{22}z(b) = \beta_2\right\}, \tag{45}$$

$$D(\bar{L}_2) = \left\{u(x) \in C^1[a,b] :\ \mu_{11}u(a) + \mu_{12}u(b) = \beta_1\right\}, \tag{46}$$

respectively.

(ii) If

$$\mu_{21} + \mu_{22}e^{-\int_a^b r(t)dt} \neq 0, \quad \mu_{11} + \mu_{12}e^{-\int_a^b s(t)dt} \neq 0, \tag{47}$$

then the operator $\bar{L}$ is correct and the unique solution of the boundary value problem

$$\bar{L}u(x) = f(x), \quad \forall f(x) \in C[a,b], \tag{48}$$

is given by

$$u(x) = \bar{L}^{-1} f(x) = \bar{L}_2^{-1} \bar{L}_1^{-1} f(x) = \bar{L}_2^{-1} z(x)$$
$$= e^{-\int_a^x s(t)dt} \left(C_2 + \int_a^x z(t) e^{\int_a^t s(\tau)d\tau} dt \right), \tag{49}$$

where

$$z(x) = \bar{L}_1^{-1} f(x) = e^{-\int_a^x r(t)dt} \left(C_1 + \int_a^x f(t) e^{\int_a^t r(\tau)d\tau} dt \right), \tag{50}$$

and

$$C_1 = \left(\mu_{21} + \mu_{22} e^{-\int_a^b r(t)dt} \right)^{-1} \left(\beta_2 - \mu_{22} e^{-\int_a^b r(t)dt} \int_a^b f(t) e^{\int_a^t r(\tau)d\tau} dt \right),$$

$$C_2 = \left(\mu_{11} + \mu_{12} e^{-\int_a^b s(t)dt} \right)^{-1} \left(\beta_1 - \mu_{12} e^{-\int_a^b s(t)dt} \int_a^b z(t) e^{\int_a^t s(\tau)d\tau} dt \right).$$

Proof

(i) From the definition of $\bar{L}$ and Proposition 1, we have

$$Lu(x) = L_1 L_2 u(x) = f(x), \tag{51}$$

and

$$\mu_{11} u(a) + \mu_{12} u(b) = \beta_1,$$
$$\mu_{21}[u'(a) + s(a)u(a)] + \mu_{22}[u'(b) + s(b)u(b)] = \beta_2. \tag{52}$$

Let

$$L_2 u(x) = u'(x) + s(x)u(x) = z(x). \tag{53}$$

It follows that

$$u'(a) + s(a)u(a) = z(a), \quad u'(b) + s(b)z(b) = z(b),$$

and upon substitution into the second boundary condition in (52), we get

$$\mu_{21} z(a) + \mu_{22} z(b) = \beta_2.$$

Thus, we have

$$L_1 z(x) = z'(x) + r(x)z(x) = f(x), \quad \mu_{21} z(a) + \mu_{22} z(b) = \beta_2, \tag{54}$$

$$L_2u(x) = u'(x) + s(x)u(x) = z(x), \quad \mu_{11}u(a) + \mu_{12}u(b) = \beta_1. \quad (55)$$

That is $\bar{L}u(x) = \bar{L}_1\bar{L}_2u(x)$. It remains to show that $D(\bar{L}) = D(\bar{L}_1\bar{L}_2)$. By using the definition of $D(\bar{L}_1\bar{L}_2)$ we obtain

$$\begin{aligned}
D(\bar{L}_1\bar{L}_2) &= \left\{u(x) \in D(\bar{L}_2) : \bar{L}_2u(x) \in D(\bar{L}_1)\right\} \\
&= \Big\{u(x) \in C^1[a,b] : \ \mu_{11}u(a) + \mu_{12}u(b) = \beta_1, \\
&\qquad u'(x) + s(x)u(x) \in D(\bar{L}_1)\Big\} \\
&= \Big\{u(x) \in C^1[a,b] : \ \mu_{11}u(a) + \mu_{12}u(b) = \beta_1, \\
&\qquad z(x) = u'(x) + s(x)u(x) \in C^1[a,b], \\
&\qquad \mu_{21}[u'(a) + s(a)u(a)] + \mu_{22}[u'(b) + s(b)u(b)] = \beta_2\Big\} \\
&= \Big\{u(x) \in C^2[a,b] : \ \mu_{11}u(a) + \mu_{12}u(b) = \beta_1, \\
&\qquad \mu_{21}[u'(a) + s(a)u(a)] + \mu_{22}[u'(b) + s(b)u(b)] = \beta_2\Big\} \\
&= D(\bar{L}). \qquad (56)
\end{aligned}$$

(ii) Application of Theorem 3 to solve boundary value problem (54) yields (50). Substituting this unique solution $z(x) = \bar{L}_1^{-1}f(x)$ into (55) and applying Theorem 3 once more, we obtain (49), which is the solution to boundary value problem (48). The correctness of $\bar{L} = \bar{L}_1\bar{L}_2$ follows from the correctness of $\bar{L}_1$ and $\bar{L}_2$.

□

To elucidate the implementation of the above procedure, we solve the following example problem.

Example 1 Let the boundary value problem

$$\begin{gathered}
u''(x) - \tfrac{x+2}{x+1}u'(x) + \tfrac{1}{x+1}u(x) = 3(x+1), \quad 0 < x < 1, \\
u(0) - 5u(1) = 0, \\
3u'(0) - 4u'(1) - 3u(0) + 4u(1) = 2. \qquad (57)
\end{gathered}$$

We take

$$p(x) = -\frac{x+2}{x+1}, \quad \text{and} \quad q(x) = \frac{1}{x+1},$$

which are continuous on [0, 1]. Notice that equations (21) and (22) are satisfied by

$$r(x) = -\frac{1}{x+1}, \quad \text{and} \quad s(x) = -1,$$

which are continuous on $[0, 1]$ and $s'(x) = 0$. Lastly, the second of the boundary conditions (57) can be put in the form

$$3[u'(0) + (-1)u(0)] - 4[u'(1) + (-1)u(1)] = 2.$$

Thus (57) is carried to

$$\begin{aligned} \bar{L}u(x) &= u''(x) - \frac{x+2}{x+1}u'(x) + \frac{1}{x+1}u(x) = f(x), \\ D(\bar{L}) &= \Big\{u(x) : u(x) \in C^2[0,1],\ u(0) - 5u(1) = 0, \\ &\qquad 3[u'(0) + s(0)u(0)] - 4[u'(1) + s(1)u(1)] = 2\Big\}, \end{aligned} \tag{58}$$

where $f(x) = 3(x+1)$. By Theorem 5, the operator $\bar{L}$ can be factorized as $\bar{L}u(x) = \bar{L}_1\bar{L}_2u(x)$, where

$$\bar{L}_1 z(x) = z'(x) - \frac{1}{x+1}z(x), \quad D(\bar{L}_1) = \Big\{z(x) \in C^1[0,1] :\ 3z(0) - 4z(1) = 2\Big\},$$

$$\bar{L}_2 u(x) = u'(x) - u(x), \quad D(\bar{L}_2) = \Big\{u(x) \in C^1[0,1] :\ u(0) - 5u(1) = 0\Big\}.$$

Furthermore,

$$\mu_{21} + \mu_{22}e^{-\int_0^1 r(t)dt} = -5 \neq 0, \quad \mu_{11} + \mu_{12}e^{-\int_0^1 s(t)dt} = 1 - 5e \neq 0, \tag{59}$$

and therefore (58) has only one solution. To construct the solution, we first solve the problem $\bar{L}_1u(x) = f(x)$ by means of (50), which yields

$$z(x) = (x+1)(3x - \frac{26}{5}). \tag{60}$$

Then by utilizing (60) and solving $\bar{L}_2u(x) = z(x)$ by (49), we get

$$u(x) = \frac{142e^x}{5(5e-1)} - \frac{15x^2 + 19x - 7}{5}. \tag{61}$$

This is the unique solution of the given boundary value problem (57).

4 Factorization Method for Nonlinear Differential Equations

In this section, we deal with the solution of a class of nonlinear boundary value problems for second order differential equations. Let the nonlinear differential equation of the form

$$u''(x)u'(x) + \left[q(x)u'(x) + g(x)u''(x)\right]u(x) + q(x)g(x)(u(x))^2 = 0, \tag{62}$$

for $x \in (a, b)$, and where $q(x)$, $g(x) \in C[a, b]$, together with the boundary conditions

$$\begin{aligned} \mu_{11}u(a) + \mu_{12}u(b) &= \beta_1, \\ \mu_{21}u'(a) + \mu_{22}u'(b) + \mu_{23}u(a) + \mu_{24}u(b) &= \beta_2, \end{aligned} \tag{63}$$

where μ_{ij}, $\beta_i \in \mathbb{R}$, $i = 1, 2$, $j = 1, 2, 3, 4$.

The nonlinear equation (62) can be decomposed as the product

$$\left[u''(x) + q(x)u(x)\right]\left[u'(x) + g(x)u(x)\right] = 0,$$

and hence, either

$$u''(x) + q(x)u(x) = 0, \tag{64}$$

or

$$u'(x) + g(x)u(x) = 0. \tag{65}$$

As a consequence, the solutions of the nonlinear boundary value problem (62) and (63) may be obtained by solving the linear second order problem (64) and (63) and the linear first order problem (65) and (63).

For the solution of the linear second order problem (64) and (63), we may employ Theorem 5 provided that prerequisites (21) and (22) are satisfied, i.e. there exist $r(x) \in C[a, b]$ and $s(x) \in C^1[a, b]$ such that

$$r(x) = -s(x), \quad s'(x) - (s(x))^2 = q(x), \tag{66}$$

and if

$$\mu_{23} = s(a)\mu_{21}, \quad \mu_{24} = s(b)\mu_{22}. \tag{67}$$

In this case problem (64), (63) can be put in the form

$$\bar{L}u(x) = u''(x) + q(x)u(x) = 0,$$

$$D(\bar{L}) = \Big\{ u(x) \in C^2[a,b] : \mu_{11}u(a) + \mu_{12}u(b) = \beta_1, \\ \mu_{21}[u'(a) + s(a)u(a)] + \mu_{22}[u'(b) + s(b)u(b)] = \beta_2 \Big\}. \quad (68)$$

Problem (68) can be now solved by means of Theorem 5.

The linear first order problem (65) and (63) is subjected to more conditions than the order of the differential equation and it is most likely to possess no solution. Nevertheless, we can proceed as follows. By utilizing (65) evaluate $u'(a)$ and $u'(b)$ and substitute into the second of the boundary conditions in (63). Taking into account (67), we get

$$\mu_{21}[s(a) - g(a)]u(a) + \mu_{22}[s(b) - g(b)]u(b) = \beta_2. \quad (69)$$

Thus, problem (65) and (63) may be formulated as

$$Tu(x) = u'(x) + g(x)u(x) = 0, \\ D(T) = \{u(x) \in C^1[0,1] : \mu_{11}u(a) + \mu_{12}u(b) = \beta_1, \\ \mu_{21}[s(a) - g(a)]u(a) + \mu_{22}[s(b) - g(b)]u(b) = \beta_2\}. \quad (70)$$

By employing Theorem 3, we find the unique solution of the problem

$$T_0u(x) = u'(x) + g(x)u(x) = 0, \\ D(T_0) = \{u(x) \in C^1[0,1] : \mu_{11}u(a) + \mu_{12}u(b) = \beta_1\}. \quad (71)$$

If the solution $u(x)$ of this problem satisfies the second boundary condition in (70), then $u(x)$ is a solution of (70); otherwise (70) has no solution.

Example 2 Let us find the solutions of the nonlinear second order boundary value problem

$$u''(x)u'(x) - \left[\frac{2}{(x+1)^2}u'(x) + \frac{1}{x+3}u''(x)\right]u(x) + \frac{2}{(x+3)(x+1)^2}(u(x))^2 = 0, \\ u(0) + 5u(1) = 0, \\ -u'(0) + 6u'(1) - u(0) + 3u(1) = 4, \quad (72)$$

where $x \in [0,1]$ and $u(x) \in C^2[0,1]$.

The nonlinear second order differential equation (72) is of the type (62) with

$$q(x) = -\frac{2}{(x+1)^2}, \quad g(x) = -\frac{1}{x+3},$$

and it can be decomposed as

$$\left[u''(x) - \frac{2}{(x+1)^2}u(x)\right]\left[u'(x) - \frac{1}{x+3}u(x)\right] = 0.$$

Thus, we get the following two linear problems

$$\begin{aligned} \bar{L}u(x) &= u''(x) - \frac{2}{(x+1)^2}u(x) = 0, \\ D(\bar{L}) &= \{u(x) \in C^2[0,1] : u(0) + 5u(1) = 0, \\ &\qquad -u'(0) + 6u'(1) - u(0) + 3u(1) = 4\}, \end{aligned} \tag{73}$$

and

$$\begin{aligned} Tu(x) &= u'(x) - \frac{1}{x+3}u(x) = 0, \\ D(T) &= \{u(x) \in C^2[0,1] : u(0) + 5u(1) = 0, \\ &\qquad -u'(0) + 6u'(1) - u(0) + 3u(1) = 4\}. \end{aligned} \tag{74}$$

In solving the boundary value problem (73), notice that the functions

$$r(x) = -\frac{1}{x+1}, \quad s(x) = \frac{1}{x+1},$$

obey (66), $r(x) \in C[0,1]$, $s(x) \in C^1[0,1]$ and $s(0) = 1$, $s(1) = \frac{1}{2}$, and that the preconditions (67) are met. Hence, problem (73) may be written in the form (68), namely

$$\begin{aligned} \bar{L}u(x) &= u''(x) - \frac{2}{(x+1)^2}u(x) = 0, \\ D(\bar{L}) &= \{u(x) \in C^2[0,1] : u(0) + 5u(1) = 0, \\ &\qquad -[u'(0) + s(0)u(0)] + 6[u'(1) + s(1)u(1)] = 4\}. \end{aligned} \tag{75}$$

By Theorem 5, the boundary value problem (75) is factorized into the following two first order problems

$$\begin{aligned} \bar{L}_1 z(x) &= z'(x) - \frac{1}{x+1}z(x) = 0, \\ D(\bar{L}_1) &= \left\{z(x) \in C^1[0,1] : \ -z(0) + 6z(1) = 4\right\}, \end{aligned} \tag{76}$$

and

$$\bar{L}_2 u(x) = u'(x) + \frac{1}{x+1} u(x) = z(x),$$
$$D(\bar{L}_2) = \left\{ u(x) \in C^1[0,1] : \; u(0) + 5u(1) = 0 \right\}. \tag{77}$$

The first of the uniqueness requirements (47) is fulfilled, viz.

$$\mu_{21} + \mu_{22} e^{-\int_0^1 r(t)dt} = -1 + 6\left(e^{\int_0^1 \frac{1}{t+1} dt}\right) = 11 \neq 0, \tag{78}$$

and therefore the operator $\bar{L}_1$ is correct and the unique solution of (76) is derived through (50), which is

$$z(x) = \frac{4}{11}(x+1). \tag{79}$$

By substituting (79) into (77) and verifying that the second of the uniqueness conditions (47) is also satisfied, viz.

$$\mu_{11} + \mu_{12} e^{-\int_0^1 s(t)dt} = 1 + 5\left(e^{-\int_0^1 \frac{1}{t+1} dt}\right) = \frac{7}{2} \neq 0, \tag{80}$$

it follows that the operator $\bar{L}_2$ is correct and the unique solution of (77), obtained via (49), is

$$u(x) = \frac{4(x^3 + 3x^2 + 3x - 5)}{33(x+1)}. \tag{81}$$

The function $u(x)$ in (81) is a solution to nonlinear second order boundary value problem (72).

We now examine the existence of a solution of the linear first order problem (74). By applying Theorem 3, we find that the problem

$$T_1 u(x) = u'(x) - \frac{1}{x+3} u(x) = 0,$$
$$D(T_1) = \{ u(x) \in C^2[0,1] : \; u(0) + 5u(1) = 0 \} \tag{82}$$

has no solution except the trivial $u(x) = 0$, which however does not satisfy the second of the boundary conditions in (74).

Summing up, the nonlinear second order boundary value problem (72) admits only the solution (81).

The technique presented above and explained in Example 2 can be extended to solve and other types of nonlinear boundary value problems. For example, consider the nonlinear differential equation of second order of the form

$$(u''(x))^2 + a(x)u''(x)u(x) + b(x)(u(x))^2 = 0, \quad x \in (a, b), \tag{83}$$

subject to two general boundary constraints

$$\begin{aligned} \mu_{11}u(a) + \mu_{12}u(b) &= \beta_1, \\ \mu_{21}u'(a) + \mu_{22}u'(b) + \mu_{23}u(a) + \mu_{24}u(b) &= \beta_2, \end{aligned} \tag{84}$$

where $a(x),\ b(x) \in C[a, b]$ and $\mu_{ij},\ \beta_i \in \mathbb{R},\ i = 1, 2,\ j = 1, 2, 3, 4$.

The differential equation (83) can be put in the form

$$F\left(\frac{u''(x)}{u(x)}, x\right) = F\left(w(x), x\right) = (w(x))^2 + a(x)w(x) + b(x) = 0,$$

where $w(x) = u''(x)/u(x)$ and the nonlinear function F is a second degree polynomial of $w(x)$. Hence, it can be decomposed as

$$\left[w(x) + q^-(x)\right]\left[w(x) + q^+(x)\right] = 0,$$

where $q^-(x),\ q^+(x) \in C[a, b]$ and $a(x) = q^-(x) + q^+(x),\ b(x) = q^-(x)q^+(x)$. By substituting back $w(x) = u''(x)/u(x)$, we get

$$\left[u''(x) + q^-(x)u(x)\right]\left[u''(x) + q^+(x)u(x)\right] = 0, \tag{85}$$

from where follows that, either

$$u''(x) + q^-(x)u(x) = 0, \quad x \in (a, b), \tag{86}$$

or

$$u''(x) + q^+(x)u(x) = 0, \quad x \in (a, b). \tag{87}$$

Thus, the solution of the nonlinear boundary value problem (83) and (84) is reduced to the solution of the two linear second order boundary value problems (86), (84), and (87), (84). Whenever the conditions (21), (22) and (42) are met, Theorem 5 may be applied to acquire the solutions in closed form.

5 Conclusions

A practical technique has been presented for factorizing and solving linear initial and boundary value problems for second order differential equations with nonlocal boundary conditions. Two types of nonlinear boundary value problems for second order differential equations have also been considered where the factorization

method was used to construct their solutions in closed form. The main advantage of the factorization method is that no fundamental or particular solutions are required. Its main disadvantage is that it cannot be applied to all boundary value problems except to those where certain conditions are satisfied. The efficiency of the method encourages the pursuit of further research for the extension of the method to problems with fully mixed boundary conditions and multipoint conditions.

References

1. L.M. Berkovich, Factorization and transformations of linear and nonlinear ordinary differential equations. Nucl. Instrum. Methods Phys. Res. A **502**, 646–648 (2003). https://doi.org/10.1016/S0168-9002(03)00531-X
2. W.E. Boyce, R.C. DiPrima, *Elementary Differential Equations and Boundary Value Problems* (Wiley, London, 2012)
3. D.I. Caruntu, Factorization of self-adjoint ordinary differential equations. Appl. Math. Comput. **219**, 7622–7631 (2013). https://doi.org/10.1016/j.amc.2013.01.049
4. J. Clegg, A new factorisation of a general second order differential equation. Int. J. Math. Edu. Sci. Technol. **37**, 51–64 (2006). https://doi.org/10.1080/00207390500186339
5. E. García, L. Littlejohn, J.L. López, E.P. Sinusía, Factorization of second-order linear differential equations and Liouville–Neumann expansions. Math. Comput. Modell. **57**, 1514–1530 (2013). https://doi.org/10.1016/j.mcm.2012.12.012
6. M.N. Hounkonnou, A. Ronveaux, Factorization of generalized Lamé and Heun's differential equations. Commun. Math. Anal. **11**, 121–136 (2011). https://projecteuclid.org/euclid.cma/1293054278
7. L. Infeld, T.E. Hull, The factorization method. Rev. Mod. Phys. **23**, 21–68 (1951)
8. K. Janglajew, E. Schmeidel, The factorization of the linear differential operator. Tatra Mt. Math. Publ. **63**, 139–151 (2015). https://doi.org/10.1515/tmmp-2015-0026
9. J.J. Kovacic, An algorithm for solving second order linear homogeneous differential equations. J. Symb. Comput. **2**, 3–43 (1986). https://doi.org/10.1016/S0747-7171(86)80010-4
10. I.N. Parasidis, P. Hahamis, Factorization method for solving multipoint problems for second order difference equations with polynomial coefficients, in *Discrete Mathematics and Applications*. Springer Optimization and Its Applications, vol. 165, eds. by A.M. Raigorodskii, M.T. Rassias (Springer, Cham, 2020). http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-030-55857-4_17
11. I.N. Parasidis, E. Providas, Factorization method for the second order linear nonlocal difference equations, in *International Conference Polynomial Computer Algebra '2018*, ed. by N.N. Vassiliev (Euler International Mathematical Institute, St. Petersburg, 2018), pp. 85–89
12. I.N. Parasidis, E. Providas, P.C. Tsekrekos, Factorization of linear operators and some eigenvalue problems of special operators. Vestn. Bashkir. Univ. **17**, 830–839 (2012)
13. E. Providas, Operator factorization and solution of second-order nonlinear difference equations with variable coefficients and multipoint constraints, in *Nonlinear Analysis and Global Optimization*. Springer Optimization and Its Applications, vol. 167, eds. by T.M. Rassias, P.M. Pardalos, pp. 427–443 (Springer, Cham, 2021). https://doi.org/10.1007/978-3-030-61732-5_20
14. W. Robin, Operator factorization and the solution of second-order linear ordinary differential equations. Int. J. Math. Educ. Sci. Technol. **38**, 189–211 (2007). https://doi.org/10.1080/00207390601002815
15. F. Schwarz, Decomposition of ordinary differential equations. Bull. Math. Sci. **7**, 575–613 (2017). https://doi.org/10.1007/s13373-017-0110-0

16. S. Tsarev, Symbolic manipulation of integro differential expressions and factorization of linear ordinary differential operators over transcendental extensions of a differential field, in *ISSAC '97: Proceedings of the 1997 International Symposium on Symbolic and Algebraic Computation*, vol. 1997 (1997), pp. 310–315. https://doi.org/10.1145/258726.258824
17. N.N. Vassiliev, I.N. Parasidis, E. Providas, Exact solution method for Fredholm integro-differential equations with multipoint and integral boundary conditions. Part 2. Decomposition-extension method for squared operators. Inf. Control Syst. **2**, 2–9 (2019). https://doi.org/10.31799/1684-8853-2019-2-2-9

A General Framework for Studying Certain Generalized Topologically Open Sets in Relator Spaces

Themistocles M. Rassias and Árpád Száz

Abstract A family $\mathscr{R}$ of binary relations on a set X is called a relator on X, and the ordered pair $X(\mathscr{R}) = (X, \mathscr{R})$ is called a relator space. Sometimes relators on X to Y are also considered.

By using an obvious definition of the generated open sets, each generalized topology $\mathscr{T}$ on X can be easily derived from the family $\mathscr{R}_{\mathscr{T}}$ of all Pervin's preorder relations $R_V = V^2 \cup V^c \times X$ with $V \in \mathscr{T}$, where $V^2 = V \times V$ and $V^c = X \setminus V$.

For a subset A of the relator space $X(\mathscr{R})$, we define

$$A^\circ = \operatorname{int}_{\mathscr{R}}(A) = \left\{ x \in X : \quad \exists\, R \in \mathscr{R} : \quad R(x) \subseteq A \right\}$$

and $A^- = \operatorname{cl}_{\mathscr{R}}(A) = \operatorname{int}_{\mathscr{R}}(A^c)^c$. And, for instance, we write $A \in \mathscr{T}_{\mathscr{R}}$ if $A \subseteq A^\circ$.

Moreover, following some basic definitions in topological spaces, for a subset A of the relator space $X(\mathscr{R})$ we write

(1) $A \in \mathscr{T}^r_{\mathscr{R}}$ if $A = A^{-\circ}$;

(2) $A \in \mathscr{T}^p_{\mathscr{R}}$ if $A \subseteq A^{-\circ}$;

(3) $A \in \mathscr{T}^s_{\mathscr{R}}$ if $A \subseteq A^{\circ-}$;

(4) $A \in \mathscr{T}^\alpha_{\mathscr{R}}$ if $A \subseteq A^{\circ-\circ}$;

(5) $A \in \mathscr{T}^\beta_{\mathscr{R}}$ if $A \subseteq A^{-\circ-}$;

(6) $A \in \mathscr{T}^a_{\mathscr{R}}$ if $A \subseteq A^{-\circ} \cap A^{\circ-}$;

(7) $A \in \mathscr{T}^b_{\mathscr{R}}$ if $A \subseteq A^{-\circ} \cup A^{\circ-}$.

The members of the above families will be called the topologically regular open, preopen, semi-open, α-open, β-open, a-open and b-open subsets of the relator space $X(\mathscr{R})$, respectively.

Th. M. Rassias
Department of Mathematics Zografou Campus, National Technical University of Athens, Athens, Greece
e-mail: trassias@math.ntua.gr

Á. Száz (✉)
Department of Mathematics, University of Debrecen, Debrecen, Hungary
e-mail: szaz@science.unideb.hu

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_19

In our former papers, having in mind the original definitions of N. Levine [49] and H. H. Corson and E. Michael [11], we have also investigated four further, closely related, families of generalized topologically open sets in $X(\mathscr{R})$.

Now, we shall offer a general framework for studying these families. Moreover, motivated by a definition of Á. Császár [15] and his predecessors, we shall also consider a further important class of generalized topologically open sets.

For the latter purpose, for a subset A of the relator space $X(\mathscr{R})$, we shall write

(8) $A \in \mathscr{A}_{\mathscr{R}}$ if $A^{-\circ} \subseteq A^{\circ-}$.

Thus, according to Császár's terminology, the members of the family $\mathscr{A}_{\mathscr{R}}$ should be called the topologically quasi-open subsets of the relator space $X(\mathscr{R})$. However, in the earlier literature, these sets have been studied under different names.

While, for the former purpose, for any two subsets A and B of the relator space $X(\mathscr{R})$ we shall write

(9) $A \in \mathrm{Ln}_{\mathscr{R}}(B)$ and $B \in \mathrm{Un}_{\mathscr{R}}(A)$ if $A \subseteq B \subseteq A^{-}$.

Moreover, for a family $\mathscr{A}$ of subsets of $X(\mathscr{R})$ we shall define

(10) $\mathscr{A}^{\ell} = \mathrm{cl}_{\mathrm{Ln}_{\mathscr{R}}}(\mathscr{A}) = \mathrm{Ln}_{\mathscr{R}}^{-1}[\mathscr{A}]$ and $\mathscr{A}^{u} = \mathrm{cl}_{\mathrm{Un}_{\mathscr{R}}}(\mathscr{A}) = \mathrm{Un}_{\mathscr{R}}^{-1}[\mathscr{A}]$.

Thus, $\mathscr{A}^{\ell}$ and $\mathscr{A}^{u}$ may be called the lower and upper nearness closures of $\mathscr{A}$, respectively. Namely, if $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then we may naturally say that A is near to B from below and B is near to A from above.

The most important particular cases are when $\mathscr{A}$ is a minimal structure or a generalized topology on X. Or even more specially, $\mathscr{A}$ is one of the families $\mathscr{T}_{\mathscr{R}}$, $\mathscr{T}_{\mathscr{R}}^{\ell}$ or $\mathscr{T}_{\mathscr{R}}^{u}$.

1 Motivations

If $\mathscr{T}$ is a family of subsets of a set X such that $\mathscr{T}$ is closed under finite intersections and arbitrary unions, then the family $\mathscr{T}$ is called a *topology* on X, and the ordered pair $X(\mathscr{T}) = (X, \mathscr{T})$ is called a *topological space*.

The members of $\mathscr{T}$ are called the *open subsets* of X. While, the members of $\mathscr{F} = \{A^c : A \in \mathscr{T}\}$, where $A^c = X \setminus A$, are called the *closed subsets* of X. Moreover, the members of $\mathscr{T} \cap \mathscr{F}$ are called the *clopen subsets* of X.

Since, $\emptyset = \bigcup \emptyset$ and $X = \bigcap \emptyset$, we necessary have $\{\emptyset, X\} \subseteq \mathscr{T} \cap \mathscr{F}$. Therefore, if in particular $\mathscr{T} = \{\emptyset, X\}$, then $\mathscr{T}$ is called *minimal* [69] instead of indiscrete. While, if $\mathscr{T} \cap \mathscr{F} = \{\emptyset, X\}$, then $\mathscr{T}$ is called *connected* [97, p. 31].

For a subset A of $X(\mathscr{T})$, the sets $A^{\circ} = \mathrm{int}(A) = \bigcup \mathscr{T} \cap \mathscr{P}(A)$,

$$A^{-} = \mathrm{cl}(A) = \mathrm{int}(A^c)^c \qquad \text{and} \qquad A^{\dagger} = \mathrm{res}(A) = \mathrm{cl}(A) \setminus A$$

are called the *interior, closure and residue of* A, respectively.

Thus, – is a *Kuratowski closure operation* on X. That is, $\emptyset^- = \emptyset$, and – is *extensive, idempotent and additive* in the sense that, for any A, $B \subseteq X$, we have $A \subseteq A^-$, $A^{--} = A^-$ and $(A \cup B)^- = A^- \cup B^-$.

In particular, the members of the families

$$\mathscr{D} = \{ A \subseteq X : \ A^- = X \} \qquad \text{and} \qquad \mathscr{N} = \{ A \subseteq X : \ A^{-\circ} = \emptyset \}$$

are called the *dense and rare (or nowhere dense) subsets* of $X(\mathscr{T})$, respectively.

In 1922, a subset A of a closure space $X(-)$ was called *regular open* by Kuratowski [44] if $A = A^{-\circ}$. While, in 1937, a subset A of a topological space $X(\mathscr{T})$ was called regular open by Stone [72] if $A = B^\circ$ for some $B \in \mathscr{F}$.

The importance of regular open subsets of $X(\mathscr{T})$ lies mainly in the fact that their family forms a complete *Boolean agebra* [32, p. 66] with respect to the operations defined by $A' = A^{-c}$, $A \wedge B = A \cap B$ and $A \vee B = (A \cup B)''$.

In 1982, a subset A of $X(\mathscr{T})$ was called *preopen* by Mashhour et al. [56] if $A \subseteq A^{-\circ}$. However, by Dontchev [22], preopen sets, under different names, were much earlier studied by several mathematicians.

For instance, in 1964, Corson and Michael [11] called a subset A of $X(\mathscr{T})$ *locally dense* if it is a dense subset of some $V \in \mathscr{T}$ in the sense that $A \subseteq V \subseteq A^-$. Moreover, they noted that this property is equivalent to the inclusion $A \subseteq A^{-\circ}$.

This equivalence was later also stated by Jun at al. [38]. Moreover, Ganster [28] proved that A is preopen if and only if there exist $V \in \mathscr{T}$ and $B \in \mathscr{D}$ such that $A = V \cap B$. (See also Dontchev [22].)

In 1963, a subset A of $X(\mathscr{T})$ was called *semi-open* by Levine [49] if there exists $V \in \mathscr{T}$ such that $V \subseteq A \subseteq V^-$. First of all, he showed that the set A is semi-open if and only if $A \subseteq A^{\circ -}$.

Moreover, he also proved that if A is a semi-open subset of $X(\mathscr{T})$, then there exist $V \in \mathscr{T}$ and $B \in \mathscr{N}$ such that $A = V \cup B$ and $V \cap B = \emptyset$. In addition, he also noted that the converse statement is false.

Levine's statement closely resembles to a famous stability theorem of Hyers [36] which says that an ε-approximately additive function of one Banach space to another is the sum of an additive function and an ε-small function.

Analogously to the paper of Hyers, Levine's paper has also attracted the interest of a surprisingly great number of mathematicians. For instance, by the Google Scholar, it has been cited by 2985 works.

Moreover, the above statement of Levine was improved by Dlaska et al. [21] who observed that a subset A of $X(\mathscr{T})$ is semi-open if and only if there exist $V \in \mathscr{T}$ and $B \subseteq V^\dagger$ such that $A = V \cup B$.

The latter observation was later reformulated, in a more convenient form, by Duszyński and Noiri [23] who noted that a subset A of $X(\mathscr{T})$ is semi-open if and only if there exists $B \subseteq A^{\circ\dagger}$ such that $A = A^\circ \cup B$.

In particular, in 1965 and 1971, Njåstad [62] and Isomichi [37], being not aware of the paper of Levine, studied semi-open sets under the names *β-sets* and *subcondensed sets*, respectively.

Moreover, Njåstad called a subset A of $X(\mathscr{T})$ an *α-set* if $A \subseteq A^{\circ-\circ}$. And, he proved that the set A is an α-set if and only if there exist $V \in \mathscr{T}$ and $B \in \mathscr{N}$ such that $A = V \setminus B$.

In 1983, the subset A was called *β-open* by Abd El-Monsef et al. [1] if $A \subseteq A^{-\circ-}$. Moreover, in 1986 Andrijević [3] used the term *semi-preopen* instead of β-open without knowing of [1].

Actually, Andrijević called a subset A of $X(\mathscr{T})$ semi-preopen if there exists a preopen subset V of $X(\mathscr{T})$ such that $V \subseteq A \subseteq V^{-}$. And, he showed that this is equivalent to the inclusion $A \subseteq A^{-\circ-}$.

Moreover, in 1996, a subset A of $X(\mathscr{T})$ was called *b-open* by Andrijević [4] if $A \subseteq A^{\circ-} \cup A^{-\circ}$. He proved that A is b-open if and only if there exist a preopen subset B and a semi-open subset C of $X(\mathscr{T})$ such that $A = B \cup C$.

In 1961, a subset A of a topological space $X(\mathscr{T})$ was said to have *property Q* by Levine [48] if $A^{\circ-} = A^{-\circ}$. He proved that A has property Q if and only if there exist $V \in \mathscr{T} \cap \mathscr{F}$ and $B \in \mathscr{N}$ such that $A = V \Delta B$. (See also [7, 10].)

While, in 1991, a subset A of $X(\mathscr{T})$ was called a *δ-set* by Chattopadhyay and Bandyopadhyay [8] if $A^{-\circ} \subseteq A^{\circ-}$. Moreover, in 2001, δ-open sets, under the name *quasi-open sets*, were more systematically studied by Császár [15, 16].

In 1992, Ganster et al. [29] already proved that A is a δ-set if and only if $A = V \cup N$ for some $V \in \mathscr{T}$ and $B \in \mathscr{N}$. Thus, δ-sets coincide with the *simply open sets* of Biswas [5] and Neubrunnová [61]. (See also [43, 59, 60].)

Actually, such sets were also first studied by Kuratowski [45, p. 69] in a more general framework. By his definition, a subset A of $X(\mathscr{T})$ has to be called *open modulo nowhere dense sets* if there exists $V \in \mathscr{T}$ such that $A \Delta V \in \mathscr{N}$.

2 Preliminaries

In our former papers [67, 68], we have shown that the above definitions and several theorems on generalized open sets can be naturally extended not only to generalized topological and closure spaces, but also to relator spaces.

In the sequel, following a terminology introduced by the second author [73], a family $\mathscr{R}$ of binary relations on a set X will be called a *relator* on X, and the ordered pair $X(\mathscr{R}) = (X, \mathscr{R})$ will be called a *relator space*.

Thus, relator spaces are generalizations of not only *ordered sets* [19] and *uniform spaces* [27], but also *topological, closure and proximity spaces* [57]. However, to include *context spaces* [30] relators on X to Y are also needed [81, 82].

For instance, by [85], each *generalized topology* $\mathscr{T}$ on X can be easily derived from the family $\mathscr{R}_{\mathscr{T}}$ of all *Pervin's preorder relations* $R_V = V^2 \cup V^c \times X$ with $V \in \mathscr{T}$. Thus, generalized topologies need not be studied separately.

For a subset A of the relator space $X(\mathscr{R})$, we define

$$A^{\circ} = \operatorname{int}_{\mathscr{R}}(A) = \left\{ x \in X : \quad \exists\ R \in \mathscr{R} : \quad R(x) \subseteq A \right\}$$

and $A^- = \mathrm{cl}_{\mathscr{R}}(A) = \mathrm{int}_{\mathscr{R}}(A^c)^c$. And, for instance, $A \in \mathscr{T}_{\mathscr{R}}$ if $A \subseteq A^\circ$.

Now, according to the former definitions on open-like subsets of topologically spaces mentioned in the Motivations, we may also naturally write

(1) $A \in \mathscr{T}^r_{\mathscr{R}}$ if $A = A^{-\circ}$;
(2) $A \in \mathscr{T}^p_{\mathscr{R}}$ if $A \subseteq A^{-\circ}$; (3) $A \in \mathscr{T}^s_{\mathscr{R}}$ if $A \subseteq A^{\circ-}$;
(4) $A \in \mathscr{T}^\alpha_{\mathscr{R}}$ if $A \subseteq A^{\circ-\circ}$; (5) $A \in \mathscr{T}^\beta_{\mathscr{R}}$ if $A \subseteq A^{-\circ-}$;
(6) $A \in \mathscr{T}^a_{\mathscr{R}}$ if $A \subseteq A^{-\circ} \cap A^{\circ-}$; (7) $A \in \mathscr{T}^b_{\mathscr{R}}$ if $A \subseteq A^{-\circ} \cup A^{\circ-}$;
(8) $A \in \mathscr{T}^q_{\mathscr{R}}$ if there exists $V \in \mathscr{T}_{\mathscr{R}}$ such that $V \subseteq A \subseteq V^-$;
(9) $A \in \mathscr{T}^{ps}_{\mathscr{R}}$ if there exists $V \in \mathscr{T}_{\mathscr{R}}$ such that $A \subseteq V \subseteq A^-$;
(10) $A \in \mathscr{T}^\gamma_{\mathscr{R}}$ if there exists $V \in \mathscr{T}^s_{\mathscr{R}}$ such that $A \subseteq V \subseteq A^-$;
(11) $A \in \mathscr{T}^\delta_{\mathscr{R}}$ if there exists $V \in \mathscr{T}^p_{\mathscr{R}}$ such that $V \subseteq A \subseteq V^-$.

Moreover, the members of the above families may be called the *topologically regular open, preopen, semi-open*, α-open, β-open, *a-open, b-open, quasi-open, pseudo-open*, γ*-open and* δ*-open subsets* of the relator space $X(\mathscr{R})$, respectively.

Here, the use of the extra term "topologically" can only be motivated by the fact, for any two subsets A and B of the relator space $X(\mathscr{R})$, we may also naturally write $B \in \mathrm{Int}_{\mathscr{R}}(A)$ if $R[B] \subseteq A$ for some $R \in \mathscr{R}$.

Thus, by using the plausible notations $\mathrm{Cl}_{\mathscr{R}}(A) = \mathrm{Int}_{\mathscr{R}}(A^c)^c$, and $A \in \tau_{\mathscr{R}}$ if $A \in \mathrm{Int}_{\mathscr{R}}(A)$, we may also naturally introduce some reasonable notions of certain *generalized proximally open sets*.

By using the *topological closure (refinement)*

$$\mathscr{R}^\wedge = \left\{ S \subseteq X^2 : \quad \forall\, x \in X : \quad x \in S(x)^\circ \right\}$$

of the relator $\mathscr{R}$, it can be shown that $\mathrm{Int}_{\mathscr{R}^\wedge}(A) = \mathscr{P}\big(\mathrm{int}_{\mathscr{R}}(A)\big)$ and $\tau_{\mathscr{R}^\wedge} = \mathscr{T}_{\mathscr{R}}$.

Therefore, the topological properties of $\mathscr{R}$ can, in principle, be immediately derived from some proximal ones. For instance, $\mathscr{R}$ may be called *topologically compact* if $\mathscr{R}^\wedge$ is *properly compact* in the sense that for each $S \in \mathscr{R}^\wedge$ there exists a finite subset A of X such that $X = S[A]$. (See [78] and [80].)

Unfortunately, up till now we have not been able to find the right proximal versions of definitions (1)–(11). However, we can now offer a general framework for introducing and studying the families given by (8)–(11).

Moreover, following the ideas of Chattopadhyay and Bandyopadhyay [8], Császár [15] and others mentioned in the Motivations, we shall now also consider a further important class of generalized topologically open sets.

For the latter purpose, for a subset A of the relator space $X(\mathscr{R})$, we shall write

(12) $A \in \mathscr{A}_{\mathscr{R}}$ if $A^{-\circ} \subseteq A^{\circ-}$.

Thus, the members of the family $\mathscr{A}_{\mathscr{R}}$ may be called the *topologically simply open subsets* subsets of the relator space $X(\mathscr{R})$.

While, for the former purpose, for any two subsets A and B of the relator space $X(\mathscr{R})$ we shall write

(13) $A \in \mathrm{Ln}_{\mathscr{R}}(B)$ and $B \in \mathrm{Un}_{\mathscr{R}}(A)$ if $A \subseteq B \subseteq A^{-}$.

Moreover, for a family $\mathscr{A}$ of subsets of $X(\mathscr{R})$ we shall define

(14) $\mathscr{A}^{\ell} = \mathrm{cl}_{\mathrm{Ln}_{\mathscr{R}}}(\mathscr{A}) = \mathrm{Ln}_{\mathscr{R}}^{-1}[\mathscr{A}]$ and $\mathscr{A}^{u} = \mathrm{cl}_{\mathrm{Un}_{\mathscr{R}}}(\mathscr{A}) = \mathrm{Un}_{\mathscr{R}}^{-1}[\mathscr{A}]$.

Thus, $\mathscr{A}^{\ell}$ and $\mathscr{A}^{u}$ may be called the *lower and upper nearness closures* of $\mathscr{A}$, respectively. Namely, if $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then we may naturally say that A is near to B from below and B is near to A from above.

The most important particular cases are when $\mathscr{A}$ is a minimal structure or a generalized topology on X. Or even more specially, $\mathscr{A}$ is one of the families $\mathscr{T}_{\mathscr{R}}$, $\mathscr{T}_{\mathscr{R}}{}^{\ell}$ or $\mathscr{T}_{\mathscr{R}}{}^{u}$.

In a subsequent paper, we shall also try to work out a general framework for introducing and studying the families given by definitions (2)–(7). However, for this, in addition to ordinary and hyper relators we shall also need super relators.

The necessary prerequisites on relations and relators, which are certainly unfamiliar to the reader, will be briefly laid out in the subsequent preparatory sections which will also contain several new observations.

These sections may also be useful for all those readers who are not very much interested in the various generalizations of open sets having been studied recently by a surprisingly great number of topologists.

3 A Few Basic Facts on Relations

A subset F of a product set $X \times Y$ is called a *relation on X to Y*. In particular, a relation on X to itself is called a *relation on X*. And, $\Delta_X = \{(x, x) : x \in X\}$ is called the *identity relation of X*.

If F is a relation on X to Y, then by the above definitions we can also state that F is a relation on $X \cup Y$. However, for several purposes, the latter view of the relation F would be quite unnatural.

If F is a relation on X to Y, then for any $x \in X$ and $A \subseteq X$ the sets $F(x) = \{y \in Y : (x, y) \in F\}$ and $F[A] = \bigcup\{F(x) : x \in A\}$ are called the *images or neighbourhoods of x and A under F*, respectively.

If $(x, y) \in F$, then instead of $y \in F(x)$, we may also write $x\, F\, y$. However, instead of $F[A]$, we cannot write $F(A)$. Namely, it may occur that, in addition to $A \subseteq X$, we also have $A \in X$.

Now, the sets $D_F = \{x \in X : F(x) \neq \emptyset\}$ and $R_F = F[X]$ may be called the *domain* and *range* of F, respectively. If in particular $D_F = X$, then we may say that F is a *relation of X to Y*, or that F is a *non-partial relation on X to Y*.

In particular, a relation f on X to Y is called a *function* if for each $x \in D_f$ there exists $y \in Y$ such that $f(x) = \{y\}$. In this case, by identifying singletons with their elements, we may simply write $f(x) = y$ instead of $f(x) = \{y\}$.

Moreover, a function $\star$ of X to itself is called a *unary operation on* X. While, a function $*$ of X^2 to X is called a *binary operation on* X. And, for any $x, y \in X$, we usually write $x^\star$ and $x * y$ instead of $\star(x)$ and $*\big((x, y)\big)$, respectively.

If F is a relation on X to Y, then a function f of D_F to Y is called a *selection function* of F if $f(x) \in F(x)$ for all $x \in D_F$. By using the Axiom of Choice, it can be shown that every relation is the union of its selection functions.

For a relation F on X to Y, we may naturally define two *set-valued functions* φ_F of X to $\mathscr{P}(Y)$ and Φ_F of $\mathscr{P}(X)$ to $\mathscr{P}(Y)$ such that $\varphi_F(x) = F(x)$ for all $x \in X$ and $\Phi_F(A) = F[A]$ for all $A \subseteq X$.

Functions of X to $\mathscr{P}(Y)$ can be naturally identified with relations on X to Y. While, functions of $\mathscr{P}(X)$ to $\mathscr{P}(Y)$ are more general objects than relations on X to Y. In [88, 93, 94], they were briefly called *corelations* on X to Y.

However, a relation on $\mathscr{P}(X)$ to Y should be rather called a *super relation* on X to Y, and a relation on $\mathscr{P}(X)$ to $\mathscr{P}(Y)$ should be rather called a *hyper relation* on X to Y. Thus, $\mathrm{cl}_{\mathscr{R}}$ is a super relation and $\mathrm{Cl}_{\mathscr{R}}$ is a hyper relation on X.

If F is a relation on X to Y, then one can easily see that $F = \bigcup_{x \in X} \{x\} \times F(x)$. Therefore, the images $F(x)$, where $x \in X$, uniquely determine F. Thus, a relation F on X to Y can also be naturally defined by specifying $F(x)$ for all $x \in X$.

For instance, the *complement* F^c and the *inverse* F^{-1} can be defined such that $F^c(x) = F(x)^c = Y \setminus F(x)$ for all $x \in X$ and $F^{-1}(y) = \{x \in X : y \in F(x)\}$ for all $y \in Y$. Thus, it can be easily seen that $F^c = X \times Y \setminus F$.

Moreover, if in addition G is a relation on Y to Z, then the *composition* $G \circ F$ can be defined such that $(G \circ F)(x) = G[F(x)]$ for all $x \in X$. Thus, it can be easily seen that $(G \circ F)[A] = G\big[F[A]\big] = \bigcup_{y \in F[A]} G(y)$ for all $A \subseteq X$.

While, if G is a relation on Z to W, then the *box product* $F \boxtimes G$ can be defined such that $(F \boxtimes G)(x, z) = F(x) \times G(z)$ for all $x \in X$ and $z \in Z$. Thus, it can be shown that $(F \boxtimes G)[A] = G \circ A \circ F^{-1}$ for all $A \subseteq X \times Z$ [87].

Hence, by taking $A = \{(x, z)\}$, and $A = \Delta_Y$ if $Y = Z$, one can at once see that the box and composition products are actually equivalent tools. However, the box product can be immediately defined for any family of relations.

Now, a relation R on X may be briefly defined to be *reflexive* on X if $\Delta_X \subseteq R$, and *transitive* if $R \circ R \subseteq R$. Moreover, R may be briefly defined to be *symmetric* if $R^{-1} \subseteq R$, and *antisymmetric* if $R \cap R^{-1} \subseteq \Delta_X$.

Thus, a reflexive and transitive (symmetric) relation may be called a *preorder (tolerance) relation*. And, a symmetric (antisymmetric) preorder relation may be called an *equivalence (partial order) relation*.

For any relation R on X, we may also naturally define $R^0 = \Delta_X$ and $R^n = R \circ R^{n-1}$ if $n \in \mathbb{N}$. Moreover, we may also naturally define $R^\infty = \bigcup_{n=0}^{\infty} R^n$. Thus, R^∞ is the smallest preorder relation on X containing R [33].

For $A \subseteq X$, the Pervin relation $R_A = A^2 \cup A^c \times X$ is an important preorder on X [66]. While, for a *pseudometric* d on X, the *Weil surrounding* $B_r = \{(x, y) \in X^2 : d(x, y) < r\}$, with $r > 0$, is an important tolerance on X [99].

Note that $S_A = R_A \cap R_A^{-1} = R_A \cap R_{A^c} = A^2 \cap \big(A^c\big)^2$ is already an equivalence relation on X. And, more generally if $\mathscr{A}$ is a *cover (partition)* of X, then $S_{\mathscr{A}} = \bigcup_{A \in \mathscr{A}} A^2$ is a tolerance (equivalence) relation on X.

As an important generalization of the Pervin relation R_A, for any $A \subseteq X$ and $B \subseteq Y$, we may also naturally consider the *Hunsaker-Lindgren relation* $R_{(A,B)} = A \times B \cap A^c \times Y$ [35]. Namely, thus we evidently have $R_A = R_{(A,A)}$.

The Pervin relations R_A and the Hunsaker-Lindgren relations $R_{(A,B)}$ were actually first used by Davis [20] and Császár [12, pp. 42 and 351] in some less explicit and convenient forms, respectively.

4 A Few Basic Facts on Relators

A family $\mathscr{R}$ of relations on one set X to another Y is called a *relator on X to Y*, and the ordered pair $(X, Y)(\mathscr{R}) = \big((X, Y), \mathscr{R}\big)$ is called a *relator space*. For the origins of this notion, see [73, 81], and the references in [73].

If in particular $\mathscr{R}$ is a relator on X to itself, then $\mathscr{R}$ is simply called a *relator on X*. Thus, by identifying singletons with their elements, we may naturally write $X(\mathscr{R})$ instead of $(X, X)(\mathscr{R})$. Namely, $(X, X) = \{\{X\}, \{X, X\}\} = \{\{X\}\}$.

Relator spaces of this simpler type are already substantial generalizations of the various *ordered sets* [19] and *uniform spaces* [27]. However, they are insufficient for some important purposes. (See, for instance, [30] and [81, 91].)

A relator $\mathscr{R}$ on X to Y, or the relator space $(X, Y)(\mathscr{R})$, is called *simple* if $\mathscr{R} = \{R\}$ for some relation R on X to Y. Simple relator spaces $(X, Y)(R)$ and $X(R)$ were called *formal contexts* and *gosets* in [30] and [90], respectively.

Moreover, a relator $\mathscr{R}$ on X, or the relator space $X(\mathscr{R})$, may, for instance, be naturally called *reflexive* if each member of $\mathscr{R}$ is reflexive on X. Thus, we may also naturally speak of *preorder, tolerance, and equivalence relators*.

For instance, for a family $\mathscr{A}$ of subsets of X, the family $\mathscr{R}_{\mathscr{A}} = \{R_A : A \in \mathscr{A}\}$, where $R_A = A^2 \cup A^c \times X$, is an important preorder relator on X. Such relators were first used by Pervin [66] and Levine [52].

While, for a family $\mathscr{D}$ of *pseudo-metrics* on X, the family $\mathscr{R}_{\mathscr{D}} = \{B_r^d : r > 0, d \in \mathscr{D}\}$, where $B_r^d = \{(x, y) : d(x, y) < r\}$, is an important tolerance relator on X. Such relators were first considered by Weil [99].

Moreover, if $\mathfrak{S}$ is a family of *partitions* of X, then the family $\mathscr{R}_{\mathfrak{S}} = \{S_{\mathscr{A}} : \mathscr{A} \in \mathfrak{S}\}$, where $S_{\mathscr{A}} = \bigcup_{A \in \mathscr{A}} A^2$, is an equivalence relator on X. Such practically important relators were first investigated by Levine [51].

If $\star$ is a unary operation for relations on X to Y, then for any relator $\mathscr{R}$ on X to Y we may naturally define $\mathscr{R}^\star = \big\{R^\star : R \in \mathscr{R}\big\}$. However, this plausible notation may cause some confusions whenever, for instance, $\star = c$.

In particular, for any relator $\mathscr{R}$ on X, we may naturally define $\mathscr{R}^\infty = \big\{R^\infty : R \in \mathscr{R}\big\}$. Moreover, we may also naturally define $\mathscr{R}^\partial = \big\{S \subseteq X^2 : S^\infty \in \mathscr{R}\big\}$. These operations were first introduced by Mala [53, 55] and Pataki [64, 65].

While, if $*$ is a binary operation for relations, then for any two relators $\mathscr{R}$ and $\mathscr{S}$ we may naturally define $\mathscr{R} * \mathscr{S} = \big\{R * S : R \in \mathscr{R}, S \in \mathscr{S}\big\}$. However, this plausible notation may again cause some confusions whenever, for instance, $* = \cap$.

Therefore, in general we rather write $\mathscr{R} \wedge \mathscr{S} = \{R \cap S : \ R \in \mathscr{R}, \ S \in \mathscr{S}\}$. Moreover, for instance, we also write $\mathscr{R} \triangle \mathscr{R}^{-1} = \{R \cap R^{-1} : \ R \in \mathscr{R}\}$. Note that thus $\mathscr{R} \triangle \mathscr{R}^{-1}$ is a symmetric relator such that $\mathscr{R} \triangle \mathscr{R}^{-1} \subseteq \mathscr{R} \wedge \mathscr{R}^{-1}$.

A function $\square$ of the family of all relators on X to Y is called a *direct (indirect) unary operation for relators* if, for every relator $\mathscr{R}$ on X to Y, the value $\mathscr{R}^{\square} = \square(\mathscr{R})$ is a relator on X to Y (on Y to X).

For instance, c and -1 are *involution operations* for relators. While, ∞ and ∂ are *projection operations* for relators. Moreover, the operation $\square = c, \ \infty$ or ∂ is *inversion compatible* in the sense that $\mathscr{R}^{\square -1} = \mathscr{R}^{-1 \square}$.

More generally, a function $\mathfrak{F}$ of the family of all relators on X to Y is called a *structure for relators* if, for every relator $\mathscr{R}$ on X to Y, the value $\mathfrak{F}_{\mathscr{R}} = \mathfrak{F}(\mathscr{R})$ is in a power set depending only on X and Y.

For instance, if $\mathrm{cl}_{\mathscr{R}}(B) = \bigcap \{R^{-1}[B] : \ R \in \mathscr{R}\}$ for every relator $\mathscr{R}$ on X to Y and $B \subseteq Y$, then the function $\mathfrak{F}$, defined by $\mathfrak{F}(\mathscr{R}) = \mathrm{cl}_{\mathscr{R}}$, is a structure for relators such that $\mathfrak{F}(\mathscr{R}) \subseteq \mathscr{P}(Y) \times X$, and thus $\mathfrak{F}(\mathscr{R}) \in \mathscr{P}\big(\mathscr{P}(Y) \times X\big)$.

A structure $\mathfrak{F}$ for relators is called *increasing* if $\mathscr{R} \subseteq \mathscr{S}$ implies $\mathfrak{F}_{\mathscr{R}} \subseteq \mathfrak{F}_{\mathscr{S}}$ for any two relators $\mathscr{R}$ and S on X to Y. And, $\mathfrak{F}$ is called *quasi-increasing* if $R \in \mathscr{R}$ implies $\mathfrak{F}_R \subseteq \mathfrak{F}_{\mathscr{R}}$ for any relator $\mathscr{R}$ on X to Y. Note that here $\mathfrak{F}_R = \mathfrak{F}_{\{R\}}$.

Moreover, the structure $\mathfrak{F}$ is called *union-preserving* if $\mathfrak{F}_{\bigcup_{i \in I} \mathscr{R}_i} = \bigcup_{i \in I} \mathfrak{F}_{\mathscr{R}_i}$ for any family $(\mathscr{R}_i)_{i \in I}$ of relators on X to Y. It can be shown that $\mathfrak{F}$ is union-preserving if and only if $\mathfrak{F}_{\mathscr{R}} = \bigcup_{R \in \mathscr{R}} \mathfrak{F}_R$ for every relator $\mathscr{R}$ on X to Y [88].

In particular, an increasing operation $\square$ for relators on X to Y is called a *projection or modification operation* for relators if it is idempotent in the sense that $\mathscr{R}^{\square \square} = \mathscr{R}^{\square}$ holds for any relator $\mathscr{R}$ on X to Y.

Moreover, a projection operation $\square$ for relators on X to Y is called a *closure or refinement operation* for relators if it is extensive in the sense that $\mathscr{R} \subseteq \mathscr{R}^{\square}$ holds for any relator $\mathscr{R}$ on X to Y.

By using Pataki connections [64, 95], several closure operations can be derived from union-preserving structures. However, more generally, one can find first the Galois adjoint $\mathfrak{G}$ of such a structure $\mathfrak{F}$, and then take $\square_{\mathfrak{F}} = \mathfrak{G} \circ \mathfrak{F}$ [84].

Now, for an operation $\square$ for relators, a relator $\mathscr{R}$ on X to Y may be naturally called $\square$-*fine* if $\mathscr{R}^{\square} = \mathscr{R}$. And, for some structure $\mathfrak{F}$ for relators, two relators $\mathscr{R}$ and $\mathscr{S}$ on X to Y may be naturally called $\mathfrak{F}$-*equivalent* if $\mathfrak{F}_{\mathscr{R}} = \mathfrak{F}_{\mathscr{S}}$.

Moreover, for a structure $\mathfrak{F}$ for relators, a relator $\mathscr{R}$ on X to Y may, for instance, be naturally called $\mathfrak{F}$-*simple* if $\mathfrak{F}_{\mathscr{R}} = \mathfrak{F}_R$ for some relation R on X to Y. Thus, in particular singleton relators have to be actually called *properly simple*.

5 Structures Derived from Relators

Definition 1 If $\mathscr{R}$ is a relator on X to Y, then for any $A \subseteq X$, $B \subseteq Y$ and $x \in X$, $y \in Y$ we define:

(1) $A \in \mathrm{Int}_{\mathscr{R}}(B)$ if $R[A] \subseteq B$ for some $R \in \mathscr{R}$;
(2) $A \in \mathrm{Cl}_{\mathscr{R}}(B)$ if $R[A] \cap B \neq \emptyset$ for all $R \in \mathscr{R}$;

(3) $x \in \mathrm{int}_{\mathscr{R}}(B)$ if $\{x\} \in \mathrm{Int}_{\mathscr{R}}(B)$; (4) $x \in \sigma_{\mathscr{R}}(y)$ if $x \in \mathrm{int}_{\mathscr{R}}(\{y\})$;
(5) $x \in \mathrm{cl}_{\mathscr{R}}(B)$ if $\{x\} \in \mathrm{Cl}_{\mathscr{R}}(B)$; (6) $x \in \rho_{\mathscr{R}}(y)$ if $x \in \mathrm{cl}_{\mathscr{R}}(\{y\})$;
(7) $B \in \mathscr{E}_{\mathscr{R}}$ if $\mathrm{int}_{\mathscr{R}}(B) \neq \emptyset$; (8) $B \in \mathscr{D}_{\mathscr{R}}$ if $\mathrm{cl}_{\mathscr{R}}(B) = X$.

Remark 1 The relations $\mathrm{Int}_{\mathscr{R}}$, $\mathrm{int}_{\mathscr{R}}$ and $\sigma_{\mathscr{R}}$ are called *the proximal, topological and infinitesimal interiors* generated by $\mathscr{R}$, respectively. While, the members of the families, $\mathscr{E}_{\mathscr{R}}$ and $\mathscr{D}_{\mathscr{R}}$ are called the *fat and dense subsets* of the relator space $(X, Y)(\mathscr{R})$, respectively.

The origins of the relations $\mathrm{Cl}_{\mathscr{R}}$ and $\mathrm{Int}_{\mathscr{R}}$ go back to Efremović's proximity δ [24] and Smirnov's strong inclusion $\Subset$ [71], respectively. While, the convenient notations $\mathrm{Cl}_{\mathscr{R}}$ and $\mathrm{Int}_{\mathscr{R}}$, and family $\mathscr{E}_{\mathscr{R}}$, together with its dual $\mathscr{D}_{\mathscr{R}}$, was first explicitly used by the second author [73, 75, 76, 83].

The following simple, but important theorem shows that the big interior and closure are equivalent tools in a relator space.

Theorem 1 *If $\mathscr{R}$ is a relator on X to Y, then for any $B \subseteq Y$ we have*

(1) $\mathrm{Cl}_{\mathscr{R}}(B) = \mathrm{Int}_{\mathscr{R}}(B^c)^c$; (2) $\mathrm{Int}_{\mathscr{R}}(B) = \mathrm{Cl}_{\mathscr{R}}(B^c)^c$.

Remark 2 By using the notation $\mathscr{C}_Y(B) = B^c$, assertion (1) can be expressed in the more concise form that $\mathrm{Cl}_{\mathscr{R}} = \left(\mathrm{Int}_{\mathscr{R}} \circ \mathscr{C}_Y\right)^c = \left(\mathrm{Int}_{\mathscr{R}}\right)^c \circ \mathscr{C}_Y$.

From Theorem 1, we can easily derive the following more familiar

Theorem 2 *If $\mathscr{R}$ is a relator on X to Y, then for any $B \subseteq Y$ we have*

(1) $\mathrm{cl}_{\mathscr{R}}(B) = \mathrm{int}_{\mathscr{R}}(B^c)^c$; *(2)* $\mathrm{int}_{\mathscr{R}}(B) = \mathrm{cl}_{\mathscr{R}}(B^c)^c$.

Remark 3 By using the convenient notations $B^- = \mathrm{cl}_{\mathscr{R}}(B)$ and $B^\circ = \mathrm{int}_{\mathscr{R}}(B)$, assertion (1) can be expressed in the more concise form that $- = c \circ c$, or equivalently $-c = c\circ$.

The small closure and interior are usually much weaker tools than the big ones. Namely, in general, we can only prove the following

Theorem 3 *If $\mathscr{R}$ is a relator on X to Y, then for any $A \subseteq X$ and $B \subseteq Y$*

(1) $A \in \mathrm{Int}_{\mathscr{R}}(B)$ *implies* $A \subseteq \mathrm{int}_{\mathscr{R}}(B)$;
(2) $A \cap \mathrm{cl}_{\mathscr{R}}(B) \neq \emptyset$ *implies* $A \in \mathrm{Cl}_{\mathscr{R}}(B)$.

Concerning closures and interiors, we can also prove the following two theorems which show that, despite their equivalences, closures are sometimes more convenient tools than interiors.

Theorem 4 *For any relator $\mathscr{R}$ on X to Y, we have*

(1) $\mathrm{Cl}_{\mathscr{R}^{-1}} = \mathrm{Cl}_{\mathscr{R}}^{-1}$; *(2)* $\mathrm{Int}_{\mathscr{R}^{-1}} = \mathscr{C}_Y \circ \mathrm{Int}_{\mathscr{R}}^{-1} \circ \mathscr{C}_X$.

Theorem 5 *If $\mathscr{R}$ is a relator on X to Y, then for any $B \subseteq Y$, we have (1)* $\mathrm{cl}_{\mathscr{R}}(B) = \bigcap_{R \in \mathscr{R}} R^{-1}[B]$; *(2)* $\mathrm{int}_{\mathscr{R}}(B) = \bigcup_{R \in \mathscr{R}} R^{-1}[B^c]^c$.

From the $B = \{y\}$ particular case of this theorem, we can easily derive

Corollary 1 *For any relator $\mathscr{R}$ on X to Y, we have*

$$\rho_{\mathscr{R}} = \bigcap \mathscr{R}^{-1} = \left(\bigcap \mathscr{R}\right)^{-1}.$$

Moreover, by using the $\mathscr{R} = \{R\}$ particular case of Theorem 5, we can prove

Theorem 6 *If R is a relation on X to Y, then for any $A \subseteq X$ and $B \subseteq Y$*

$$A \subseteq \operatorname{int}_R(B) \iff \operatorname{cl}_{R^{-1}}(A) \subseteq B.$$

Remark 4 This shows that the mappings $A \mapsto \operatorname{cl}_{R^{-1}}(A)$ and $B \mapsto \operatorname{int}_R(B)$ establish a Galois connection between the posets $\mathscr{P}(X)$ and $\mathscr{P}(Y)$.

The above important closure-interior Galois connection, used first in [92], is not independent from the well-known upper and lower bound one [86].

The following two closely related theorems show that the fat and dense sets are also equivalent tools in a relator space.

Theorem 7 *If $\mathscr{R}$ is a relator on X to Y, then for any $B \subseteq Y$ we have (1) $B \in \mathscr{D}_{\mathscr{R}} \iff B^c \notin \mathscr{E}_{\mathscr{R}}$; (2) $B \in \mathscr{E}_{\mathscr{R}} \iff B^c \notin \mathscr{D}_{\mathscr{R}}$.*

Theorem 8 *If $\mathscr{R}$ is a relator on X to Y, then for any $B \subseteq Y$ we have*

(1) $B \in \mathscr{D}_{\mathscr{R}}$ if and only if $B \cap E \neq \emptyset$ for all $E \in \mathscr{E}_{\mathscr{R}}$;
(2) $B \in \mathscr{E}_{\mathscr{R}}$ if and only if $B \cap D \neq \emptyset$ for all $D \in \mathscr{D}_{\mathscr{R}}$.

Remark 5 By the corresponding definitions, we have $R(x) \in \mathscr{E}_{\mathscr{R}}$ and thus also $R(x)^c \notin \mathscr{D}_{\mathscr{R}}$ for all $x \in X$ and $R \in \mathscr{R}$.

While, by using the notation $\mathscr{U}_{\mathscr{R}}(x) = \operatorname{int}^{-1}_{\mathscr{R}}(x) = \{B \subseteq Y : \ x \in \operatorname{int}_{\mathscr{R}}(B)\}$, we can note that $\mathscr{E}_{\mathscr{R}} = \bigcup_{x \in X} \mathscr{U}_{\mathscr{R}}(x)$.

By using Definition 1, we may easily introduce several further important definitions. For instance, we may also naturally have the following

Definition 2 If $\mathscr{R}$ is a relator on X to Y, then for any $B \subseteq Y$, we define (1) $\operatorname{bnd}_{\mathscr{R}}(B) = \operatorname{cl}_{\mathscr{R}}(B) \setminus \operatorname{int}_{\mathscr{R}}(B)$.

Moreover, if in particular $\mathscr{R}$ is a relator on X, then for any $A \subseteq X$ we also define (2) $\operatorname{res}_{\mathscr{R}}(A) = \operatorname{cl}_{\mathscr{R}}(A) \setminus A$; (3) $\operatorname{bor}_{\mathscr{R}}(A) = A \setminus \operatorname{int}_{\mathscr{R}}(A)$.

Remark 6 Somewhat differently, the *border, boundary and residue of a set* in neighbourhood and closure spaces were also introduced by Hausdorff and Kuratowski [44, pp. 4–5]. (See also Elez and Papaz [26] for a recent treatment.)

If in particular $\mathscr{R}$ is a reflexive relator on X, then by Definition 1, for any $A \subseteq X$, we have $A^\circ \subseteq A \subseteq A^-$ Therefore,

$$\operatorname{bnd}_{\mathscr{R}}(A) = \operatorname{res}_{\mathscr{R}}(A) \cup \operatorname{bor}_{\mathscr{R}}(A) = \operatorname{res}_{\mathscr{R}}(A) \cup \operatorname{res}_{\mathscr{R}}(A^c).$$

Namely, by using Definition 2 and Theorem 2, we can easily see that

$$\mathrm{res}_{\mathscr{R}}(A^c) = A^{c-} \setminus A^c = A^{c-} \cap A^{cc} = A^{\circ c} \cap A = A \setminus A^\circ = \mathrm{bor}_{\mathscr{R}}(A).$$

Note that if in particular $A \in \mathscr{T}_{\mathscr{R}}$ in the sense that $A \subseteq A^\circ$, then $\mathrm{bor}_{\mathscr{R}}(A) = \emptyset$. Therefore, in this particular case, by the above equality, we can simply state that $\mathrm{bnd}_{\mathscr{R}}(A) = \mathrm{res}_{\mathscr{R}}(A)$.

6 Further Structures Derived from Relators

By using Definition 1, we may also naturally introduce the following

Definition 3 If $\mathscr{R}$ is a relator on X, then for any $A \subseteq X$ we also define:
(1) $A \in \tau_{\mathscr{R}}$ if $A \in \mathrm{Int}_{\mathscr{R}}(A)$; (2) $A \in \tau\!\!\!-_{\mathscr{R}}$ if $A^c \notin \mathrm{Cl}_{\mathscr{R}}(A)$;
(3) $A \in \mathscr{T}_{\mathscr{R}}$ if $A \subseteq \mathrm{int}_{\mathscr{R}}(A)$; (4) $A \in \mathscr{F}_{\mathscr{R}}$ if $\mathrm{cl}_{\mathscr{R}}(A) \subseteq A$;
(5) $A \in \mathscr{N}_{\mathscr{R}}$ if $\mathrm{cl}_{\mathscr{R}}(A) \notin \mathscr{E}_{\mathscr{R}}$; (6) $A \in \mathscr{M}_{\mathscr{R}}$ if $\mathrm{int}_{\mathscr{R}}(A) \in \mathscr{D}_{\mathscr{R}}$.

Remark 7 The members of the families, $\tau_{\mathscr{R}}$ and $\mathscr{T}_{\mathscr{R}}$ and $\mathscr{N}_{\mathscr{R}}$ are called the *proximally open, topologically open and rare (or nowhere dense) subsets* of the relator space $X(\mathscr{R})$, respectively.

The family $\tau_{\mathscr{R}}$ was first introduced by the second author in [75, 76]. While, the practical notation $\tau\!\!\!-_{\mathscr{R}}$ was suggested by János Kurdics who first noticed that "connectedness" is a particular case of "well-chainedness". (See [46, 47, 65, 69].)

By using the corresponding results of Section 5, we can easily establish the following theorems.

Theorem 9 *If $\mathscr{R}$ is a relator on X, then for any $A \subseteq X$, we have (1) $A \in \tau_{\mathscr{R}} \iff A^c \in \tau_{\mathscr{R}}$; (2) $A \in \tau\!\!\!-_{\mathscr{R}} \iff A^c \in \tau\!\!\!-_{\mathscr{R}}$.*

Theorem 10 *For any relator $\mathscr{R}$ on X, we have (1) $\tau_{\mathscr{R}} = \tau_{\mathscr{R}^{-1}}$; (2) $\tau\!\!\!-_{\mathscr{R}} = \tau\!\!\!-_{\mathscr{R}^{-1}}$.*

Theorem 11 *If $\mathscr{R}$ is a relator on X, then for any $A \subseteq X$, we have (1) $A \in \mathscr{F}_{\mathscr{R}} \iff A^c \in \mathscr{T}_{\mathscr{R}}$; (2) $A \in \mathscr{T}_{\mathscr{R}} \iff A^c \in \mathscr{F}_{\mathscr{R}}$.*

Corollary 2 *If $\mathscr{R}$ is a relator on X and $A \subseteq X$ and $V \in \mathscr{T}_{\mathscr{R}}$ such that $A \cap V = \emptyset$, then $\mathrm{cl}_{\mathscr{R}}(A) \cap V = \emptyset$ also hold.*

Proof By Theorem 11, we have $V^c \in \mathscr{F}_{\mathscr{R}}$. Thus, by Definition 3, we also have $V^{c-} \subseteq V^c$. Hence, by using the increasingness of the operation $-$, we can see that $A \cap V = \emptyset \implies A \subseteq V^c \implies A^- \subseteq V^{c-} \implies A^- \subseteq V^c \implies A^- \cap V = \emptyset$.

Remark 8 Note that if $\mathscr{R}$ is a reflexive relator on X, then $A \subseteq A^-$ for any $A \subseteq X$. Therefore, $A^- \cap V = \emptyset$ trivially implies $A \cap V = \emptyset$ for any $A, V \subseteq X$.

Theorem 12 *For any relator $\mathscr{R}$ on X, we have (1) $\tau_{\mathscr{R}} \subseteq \mathscr{T}_{\mathscr{R}}$; (2) $\tau_{\mathscr{R}} \subseteq \mathscr{F}_{\mathscr{R}}$.*

Remark 9 In particular, for any relation R on X, we have
(1) $\tau_R = \mathscr{T}_R$; (2) $\tau_R = \mathscr{F}_R$.

Theorem 13 *For any relator $\mathscr{R}$ on X, we have (1) $\mathscr{T}_{\mathscr{R}} \setminus \{\emptyset\} \subseteq \mathscr{E}_{\mathscr{R}}$; (2) $\mathscr{D}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}} \subseteq \{X\}$.*

Remark 10 Hence, by using global complementations, we can easily infer that $\mathscr{F}_{\mathscr{R}} \subseteq \left(\mathscr{D}_{\mathscr{R}}\right)^c \cup \{X\}$ and $\mathscr{D}_{\mathscr{R}} \subseteq \left(\mathscr{F}_{\mathscr{R}}\right)^c \cup \{X\}$.

Theorem 14 *If $\mathscr{R}$ is a relator on X, then for any $A \subseteq X$ we have (1) $A \in \mathscr{E}_{\mathscr{R}}$ if $V \subseteq A$ for some $V \in \mathscr{T}_R \setminus \{\emptyset\}$;*
(2) $A \in \mathscr{D}_{\mathscr{R}}$ only if $A \setminus W \neq \emptyset$ for all $W \in \mathscr{F}_{\mathscr{R}} \setminus \{X\}$.

Remark 11 The fat sets are frequently more convenient tools than the topologically open ones. For instance, if $\leq$ is a relation on X, then $\mathscr{T}_{\leq}$ and $\mathscr{E}_{\leq}$ are the families of all ascending and residual subsets of the goset $X(\leq)$, respectively.

Moreover, if in particular $X = \mathbb{R}$ and $R(x) = \{x - 1\} \cup [x, +\infty[$ for all $x \in X$, then R is a reflexive relation on X such that $\mathscr{T}_R = \{\emptyset, X\}$, but $\mathscr{E}_R$ is quite a large family. Namely, the supersets of each $R(x)$ are also contained in $\mathscr{E}_R$.

However, the importance of fat and dense lies mainly in the following

Definition 4 If $\mathscr{R}$ is a relator on X to Y, and φ and ψ are functions of a relator space $\Gamma(\mathscr{U})$ to X and Y, respectively, then by using the notation

$$(\varphi, \psi)(\gamma) = \left(\varphi(\gamma), \psi(\gamma)\right)$$

for all $\gamma \in \Gamma$, we may also naturally define (1) $\varphi \in \mathrm{Lim}_{\mathscr{R}}(\psi)$ if $(\varphi, \psi)^{-1}[R] \in \mathscr{E}_{\mathscr{U}}$ for all $R \in \mathscr{R}$,
(2) $\varphi \in \mathrm{Adh}_{\mathscr{R}}(\psi)$ if $(\varphi, \psi)^{-1}[R] \in \mathscr{D}_{\mathscr{U}}$ for all $R \in \mathscr{R}$.

Moreover, for any $x \in X$, we may also naturally define:
(3) $x \in \lim_{\mathscr{R}}(\psi)$ if $x_\Gamma \in \mathrm{Lim}_{\mathscr{R}}(\psi)$, (4) $x \in \mathrm{adh}_{\mathscr{R}}(\psi)$ if $x_\Gamma \in \mathrm{Adh}_{\mathscr{R}}(\psi)$, where x_Γ is a function of Γ to X such that $x_\Gamma(\gamma) = x$ for all $\gamma \in \Gamma$.

Remark 12 Fortunately, the small limit and adherece relations are equivalent to the small closure and interior ones.

However, the big limit and adherence relations, suggested by Efremović and Švarc [25], are usually stronger tools than the big closure and interior ones.

In this respect, it seems convenient to only mention here the following

Theorem 15 *If $\mathscr{R}$ is a relator on X to Y, then for any $A \subseteq X$ and $B \subseteq Y$ the following assertions are equivalent:*

(1) $A \in \mathrm{Cl}_{\mathscr{R}}(B)$;
(2) there exist functions φ and ψ of the poset $\mathscr{R}(\supseteq)$ to A and B, respectively, such that $\varphi \in \mathrm{Lim}_{\mathscr{R}}(\psi)$;
(3) there exist functions φ and ψ of a relator space $\Gamma(\mathscr{U})$ to A and B, respectively, such that $\varphi \in \mathrm{Lim}_{\mathscr{R}}(\psi)$.

Proof For instance, if (1) holds, then for each $R \in \mathscr{R}$, we have $R[A] \cap B \neq \emptyset$. Therefore, there exist $\varphi(R) \in A$ and $\psi(R) \in B$ such that $\psi(R) \in R\big(\varphi(R)\big)$. Hence, we can already infer that $(\varphi, \psi)(R) = \big(\varphi(R), \psi(R)\big) \in R$, and thus also $R \in (\varphi, \psi)^{-1}[R]$.

Therefore, if $R \in \mathscr{R}$, then for any $S \in \mathscr{R}$, with $R \supseteq S$, we have

$$S \in (\varphi, \psi)^{-1}[S] \subseteq (\varphi, \psi)^{-1}[R].$$

This shows that $(\varphi, \psi)^{-1}[R]$ is a fat subset of $\mathscr{R}(\supseteq)$, and thus $\varphi \in \mathrm{Lim}_{\mathscr{R}}(\psi)$.

Remark 13 Finally, we note that if $\mathscr{R}$ is a relator on X to Y, then according to [82], for any $A \subseteq X$ and $B \subseteq Y$, we may also naturally write $A \in \mathrm{Lb}_{\mathscr{R}}(B)$ and $B \in \mathrm{Ub}_{\mathscr{R}}(A)$ if there exists $R \in \mathscr{R}$ such that $A \times B \subseteq R$.

However, the algebraic structures $\mathrm{Lb}_{\mathscr{R}}$ and $\mathrm{Ub}_{\mathscr{R}}$, and the structures derivable from them, are not independent of the former topological ones. Namely, it can be easily shown that $\mathrm{Lb}_{\mathscr{R}} = \mathrm{Int}_{\mathscr{R}^c} \circ \mathscr{C}_Y$ and $\mathrm{Int}_{\mathscr{R}} = \mathrm{Lb}_{\mathscr{R}^c} \circ \mathscr{C}_Y$.

7 Closure Operations for Relators

Similar operations for relators have formerly been studied by Kenyon [42], Nakano–Nakano [58], Száz [77, 79] and Pataki [64].

Definition 5 For any relator $\mathscr{R}$ on X to Y, the relators

$$\mathscr{R}^* = \big\{S \subseteq X \times Y : \quad \exists\ R \in \mathscr{R} : \quad R \subseteq S\big\};$$

$$\mathscr{R}^\# = \big\{S \subseteq X \times Y : \quad \forall\ A \subseteq X : \quad \exists\ R \in \mathscr{R} : \quad R[A] \subseteq S[A]\big\};$$

$$\mathscr{R}^\wedge = \big\{S \subseteq X \times Y : \quad \forall\ x \in X : \quad \exists\ R \in \mathscr{R} : \quad R(x) \subseteq S(x)\big\};$$

and

$$\mathscr{R}^\triangle = \big\{S \subseteq X \times Y : \quad \forall\, x \in X : \quad \exists\ u \in X : \quad \exists\ R \in \mathscr{R} : \quad R(u) \subseteq S(x)\big\}$$

are called the *uniform, proximal, topological and paratopological closures (or refinements)* of the relator $\mathscr{R}$, respectively.

Remark 14 Thus, we evidently have $\mathscr{R} \subseteq \mathscr{R}^* \subseteq \mathscr{R}^{\#} \subseteq \mathscr{R}^{\wedge} \subseteq \mathscr{R}^{\triangle}$. Moreover, if $\mathscr{R}$ is a relator on X, then it can be easily shown that $\mathscr{R}^{\infty} \subseteq \mathscr{R}^{*\infty} \subseteq \mathscr{R}^{\infty *} \subseteq \mathscr{R}^*$.

Remark 15 However, it is now more important to note that, because of Definition 1, we also have

$$\begin{aligned}
\mathscr{R}^{\#} &= \left\{ S \subseteq X \times Y : \quad \forall\, A \subseteq X : \quad A \in \mathrm{Int}_{\mathscr{R}}\left(S[A]\right) \right\}, \\
\mathscr{R}^{\wedge} &= \left\{ S \subseteq X \times Y : \quad \forall\ x \in X : \quad x \in \mathrm{int}_{\mathscr{R}}\left(S(x)\right) \right\}, \\
\mathscr{R}^{\triangle} &= \left\{ S \subseteq X \times Y : \quad \forall\, x \in X : \quad S(x) \in \mathscr{E}_{\mathscr{R}} \right\}.
\end{aligned}$$

Morover, by using Pataki connections [64, 69, 95], the following equivalences and their consequences can be proved in a unified way.

Theorem 16 *$\#$, $\wedge$ and $\triangle$ are closure operations for relators on X to Y such that, for any two relators $\mathscr{R}$ and $\mathscr{S}$ on X to Y, we have*

(1) $\mathscr{S} \subseteq \mathscr{R}^{\#} \iff \mathscr{S}^{\#} \subseteq \mathscr{R}^{\#} \iff \mathrm{Int}_{\mathscr{S}} \subseteq \mathrm{Int}_{\mathscr{R}} \iff \mathrm{Cl}_{\mathscr{R}} \subseteq \mathrm{Cl}_{\mathscr{S}}$,
(2) $\mathscr{S} \subseteq \mathscr{R}^{\wedge} \iff \mathscr{S}^{\wedge} \subseteq \mathscr{R}^{\wedge} \iff \mathrm{int}_{\mathscr{S}} \subseteq \mathrm{int}_{\mathscr{R}} \iff \mathrm{cl}_{\mathscr{R}} \subseteq \mathrm{cl}_{\mathscr{S}}$,
(3) $\mathscr{S} \subseteq \mathscr{R}^{\triangle} \iff \mathscr{S}^{\triangle} \subseteq \mathscr{R}^{\triangle} \iff \mathscr{E}_{\mathscr{S}} \subseteq \mathscr{E}_{\mathscr{R}} \iff \mathscr{D}_{\mathscr{R}} \subseteq \mathscr{D}_{\mathscr{S}}$.

Corollary 3 *For any relator $\mathscr{R}$ on X to Y,*

(1) $\mathscr{S} = \mathscr{R}^{\#}$ *is the largest relator on X to Y such that* $\mathrm{Int}_{\mathscr{S}} \subseteq \mathrm{Int}_{\mathscr{R}}$ $\left(\mathrm{Int}_{\mathscr{S}} = \mathrm{Int}_{\mathscr{R}}\right)$, *or equivalently* $\mathrm{Cl}_{\mathscr{R}} \subseteq \mathrm{Cl}_{\mathscr{S}}$ $\left(\mathrm{Cl}_{\mathscr{S}} = \mathrm{Cl}_{\mathscr{R}}\right)$;
(2) $\mathscr{S} = \mathscr{R}^{\wedge}$ *is the largest relator on X to Y such that* $\mathrm{int}_{\mathscr{S}} \subseteq \mathrm{int}_{\mathscr{R}}$ $\left(\mathrm{int}_{\mathscr{S}} = \mathrm{int}_{\mathscr{R}}\right)$, *or equivalently* $\mathrm{cl}_{\mathscr{R}} \subseteq \mathrm{cl}_{\mathscr{S}}$ $\left(\mathrm{cl}_{\mathscr{S}} = \mathrm{cl}_{\mathscr{R}}\right)$;
(3) $\mathscr{S} = \mathscr{R}^{\triangle}$ *is the largest relator on X to Y such that* $\mathscr{E}_{\mathscr{S}} \subseteq \mathscr{E}_{\mathscr{R}}$ $\left(\mathscr{E}_{\mathscr{S}} = \mathscr{E}_{\mathscr{R}}\right)$, *or equivalently* $\mathscr{D}_{\mathscr{R}} \subseteq \mathscr{D}_{\mathscr{S}}$ $\left(\mathscr{D}_{\mathscr{S}} = \mathscr{D}_{\mathscr{R}}\right)$.

Remark 16 To prove some similar statements for the operation $*$, the structures $\mathrm{Lim}_{\mathscr{R}}$ and $\mathrm{Adh}_{\mathscr{R}}$ have to be used [73].

Moreover, for instance, to investigate the structures $\mathrm{Lb}_{\mathscr{R}}$ and $\mathrm{Ub}_{\mathscr{R}}$ the compound operation $\circledast = c\#c$ is needed [89].

Concerning the above basic closure operations, we can also prove the following two theorems.

Theorem 17 *For any relator $\mathscr{R}$ on X to Y, we have*

(1) $\mathscr{R}^{\#} = \mathscr{R}^{\Diamond\#} = \mathscr{R}^{\#\Diamond}$ *with* $\Diamond = *$ *and* $\#$;
(2) $\mathscr{R}^{\wedge} = \mathscr{R}^{\Diamond\wedge} = \mathscr{R}^{\wedge\Diamond}$ *with* $\Diamond = *, \#$ *and* $\wedge$;
(3) $\mathscr{R}^{\triangle} = \mathscr{R}^{\Diamond\triangle} = \mathscr{R}^{\triangle\Diamond}$ *with* $\Diamond = *, \#, \wedge$ *and* $\triangle$.

Proof To prove (1), note that, by Remark 14 and the closure properties, we have $\mathscr{R}^{\#} \subseteq \mathscr{R}^{\#*} \subseteq \mathscr{R}^{\#\#} = \mathscr{R}^{\#}$ and $\mathscr{R}^{\#} \subseteq \mathscr{R}^{*\#} \subseteq \mathscr{R}^{\#\#} = \mathscr{R}^{\#}$.

Theorem 18 *For any relator $\mathscr{R}$ on X to Y, we have (1)* $\mathscr{R}^{*-1} = \mathscr{R}^{-1*}$; *(2)* $\mathscr{R}^{\#-1} = \mathscr{R}^{-1\#}$.

Proof To prove (2), note that by Theorems 4 and 16 we have

$$\mathrm{Cl}_{\mathscr{R}^{\#-1}} = \mathrm{Cl}^{-1}_{\mathscr{R}^{\#}} = \mathrm{Cl}^{-1}_{\mathscr{R}} = \mathrm{Cl}_{\mathscr{R}^{-1}}.$$

and thus in particular $\mathrm{Cl}_{\mathscr{R}^{-1}} \subseteq \mathrm{Cl}_{\mathscr{R}^{\#-1}}$. Hence, by using Theorem 16, we can infer that $\mathscr{R}^{\#-1} \subseteq \mathscr{R}^{-1\#}$.

Now, by writing $\mathscr{R}^{-1}$ in place of $\mathscr{R}$, we can see that $\mathscr{R}^{-1\#-1} \subseteq \mathscr{R}^{\#}$, and thus $\mathscr{R}^{-1\#} \subseteq \mathscr{R}^{\#-1}$. Therefore, (2) is also true.

Remark 17 For instance, the elementwise operations c and ∞ are also inversion compatible. Moreover, the operation ∂ is also inversion compatible.

However, unfortunately, the operations $\wedge$ and $\triangle$ are not inversion compatible. Therefore, in addition to Definition 5, we must also have the following

Definition 6 For any relator $\mathscr{R}$ on X to Y, we define

$$\mathscr{R}^{\vee} = \mathscr{R}^{\wedge-1} \qquad \text{and} \qquad \mathscr{R}^{\nabla} = \mathscr{R}^{\triangle-1}.$$

Remark 18 The latter operations have very curious properties. For instance, if $\mathscr{R}$ is nonvoid, then $\mathscr{R}^{\vee\wedge} = \{\rho_{\mathscr{R}}\}^{\wedge}$ [54].

Thus, in particular, the relator $\mathscr{R}^{\vee}$ is topologically simple in the sense that it is topologically equivalent to a singleton relator.

8 Some Further Theorems on the Operations $\wedge$ and $\triangle$

A preliminary form of the following theorem was already proved in [73].

Theorem 19 *If $\mathscr{R}$ is nonvoid relator on X to Y, then for any $B \subseteq Y$ we have: (1)* $\mathrm{Int}_{\mathscr{R}^{\wedge}}(B) = \mathscr{P}\big(\mathrm{int}_{\mathscr{R}}(B)\big)$; *(2)* $\mathrm{Cl}_{\mathscr{R}^{\wedge}}(B) = \mathscr{P}\big(\mathrm{cl}_{\mathscr{R}}(B)^{c}\big)^{c}$.

Proof If $A \in \mathrm{Int}_{\mathscr{R}^{\wedge}}(B)$, then by Theorems 3 and 16 we have

$$A \subseteq \mathrm{int}_{\mathscr{R}^{\wedge}}(B) = \mathrm{int}_{\mathscr{R}}(B),$$

and thus $A \in \mathscr{P}\big(\mathrm{int}_{\mathscr{R}}(B)\big)$. Therefore, $\mathrm{Int}_{\mathscr{R}^{\wedge}}(B) \subseteq \mathscr{P}\big(\mathrm{int}_{\mathscr{R}}(B)\big)$.

While, if $A \in \mathscr{P}\big(\mathrm{int}_{\mathscr{R}}(B)\big)$, then $A \subseteq \mathrm{int}_{\mathscr{R}}(B)$. Therefore, for each $x \in A$, there exists $R_x \in \mathscr{R}$ such that $R_x(x) \subseteq B$. Now, by defining

$$S(x) = R_x(x) \quad \text{forall} \quad x \in A \qquad \text{and} \qquad S(x) = Y \quad \text{forall} \quad x \in A^{c},$$

we can easily see that $S \in \mathscr{R}^{\wedge}$ such that $S[A] \subseteq B$. Therefore, we also have $A \in \mathrm{Int}_{\mathscr{R}^{\wedge}}(B)$. Consequently, $\mathscr{P}\big(\mathrm{int}_{\mathscr{R}}(B)\big) \subseteq \mathrm{Int}_{\mathscr{R}^{\wedge}}(B)$, and thus (1) also holds.

Now, by using Theorems 1 and 2, we can also easily see that

$$\mathrm{Cl}_{\mathscr{R}^{\wedge}}(B) = \mathrm{Int}_{\mathscr{R}^{\wedge}}(B^c)^c = \mathscr{P}\big(\mathrm{int}_{\mathscr{R}}(B^c)\big)^c = \mathscr{P}\big(\mathrm{cl}_{\mathscr{R}}(B)^c\big)^c.$$

Remark 19 Thus, for any $A \subseteq X$, we have

$$A \in \mathrm{Cl}_{\mathscr{R}^{\wedge}}(B) \iff A \cap \mathrm{cl}_{\mathscr{R}}(B) \neq \emptyset.$$

From Theorem 19, by using Definition 3, we can immediately derived

Corollary 4 *If $\mathscr{R}$ is a nonvoid relator on X, then (1) $\tau_{\mathscr{R}^{\wedge}} = \mathscr{T}_{\mathscr{R}}$; (2) $\tau_{\mathscr{R}^{\wedge}} = \mathscr{F}_{\mathscr{R}}$.*

Remark 20 Hence, since $\tau_{\mathscr{R}} = \bigcup_{R\in\mathscr{R}} \tau_R = \bigcup_{R\in\mathscr{R}} \mathscr{T}_R$, we can infer that

$$\mathscr{T}_{\mathscr{R}} = \bigcup_{R\in\mathscr{R}^{\wedge}} \mathscr{T}_R.$$

Unfortunately, in contrast to the structures Int, int, $\mathscr{E}$ and τ, the increasing structure $\mathscr{T}$ is already not union-preserving.

Example 1 If $\mathrm{card}(X) > 2$, $x_1 \in X$ and $x_2 \in X \setminus \{x_1\}$, and

$$R_i = \{x_i\}^2 \cup \big(X \setminus \{x_i\}\big)^2$$

for all $i = 1, 2$, then $\mathscr{R} = \{R_1, R_2\}$ is an equivalence relator on X such that

$$\{x_1, x_2\} \in \mathscr{T}_{\mathscr{R}} \setminus \big(\mathscr{T}_{R_1} \cup \mathscr{T}_{R_2}\big), \qquad \text{and thus} \qquad \mathscr{T}_{\mathscr{R}} \not\subseteq \mathscr{T}_{R_1} \cup \mathscr{T}_{R_1}.$$

From Corollary 4, by using Theorem 17, we can immediately derive

Corollary 5 *If $\mathscr{R}$ is a nonvoid relator on X, then (1) $\tau_{\mathscr{R}^{\triangle}} = \mathscr{T}_{\mathscr{R}^{\triangle}}$; (2) $\tau_{\mathscr{R}^{\triangle}} = \mathscr{F}_{\mathscr{R}^{\triangle}}$.*

Concerning the operation $\triangle$, we can also prove the following

Theorem 20 *If $\mathscr{R}$ is a nonvoid relator on X to Y, then for any $B \subseteq Y$ we have:*

(1) $\mathrm{Int}_{\mathscr{R}^{\triangle}}(B) = \{\emptyset\}$ *if* $B \notin \mathscr{E}_{\mathscr{R}}$ *and* $\mathrm{Int}_{\mathscr{R}^{\triangle}}(B) = \mathscr{P}(X)$ *if* $B \in \mathscr{E}_{\mathscr{R}}$;
(2) $\mathrm{Cl}_{\mathscr{R}^{\triangle}}(B) = \emptyset$ *if* $B \notin \mathscr{D}_{\mathscr{R}}$ *and* $\mathrm{Cl}_{\mathscr{R}^{\triangle}}(B) = \mathscr{P}(X) \setminus \{\emptyset\}$ *if* $B \in \mathscr{D}_{\mathscr{R}}$.

Proof If $A \in \mathrm{Int}_{\mathscr{R}^{\triangle}}(B)$, then there exists $S \in \mathscr{R}^{\triangle}$ such that $S[A] \subseteq B$. Therefore, if $A \neq \emptyset$, then there exists $x \in X$ such that $S(x) \subseteq B$. Hence, since $S(x) \in \mathscr{E}_{\mathscr{R}}$, it follows that $B \in \mathscr{E}_{\mathscr{R}}$. Therefore, the first part of (1) is true.

To prove the second part of (1), it is enough to note only that if $B \in \mathscr{E}_{\mathscr{R}}$, then $R = X \times B \in \mathscr{R}^{\triangle}$ such that $R[A] \subseteq B$, and thus $A \in \mathrm{Int}_{\mathscr{R}^{\triangle}}(B)$ for all $A \subseteq X$.

Assertion (2) can again be derived from (1) by using Theorem 1.

From this theorem, by Definition 1, it is clear that in particular we also have

Corollary 6 *If $\mathscr{R}$ is nonvoid relator on X to Y, then for any $B \subseteq Y$:*

(1) $\mathrm{cl}_{\mathscr{R}^\triangle}(B) = \emptyset$ *if* $B \notin \mathscr{D}_{\mathscr{R}}$ *and* $\mathrm{cl}_{\mathscr{R}^\triangle}(B) = X$ *if* $B \in \mathscr{D}_{\mathscr{R}}$;
(2) $\mathrm{int}_{\mathscr{R}^\triangle}(B) = \emptyset$ *if* $B \notin \mathscr{E}_{\mathscr{R}}$ *and* $\mathrm{int}_{\mathscr{R}^\triangle}(B) = X$ *if* $B \in \mathscr{E}_{\mathscr{R}}$.

Hence, by using Definitions 1 and 3, we can immediately derive

Corollary 7 *If $\mathscr{R}$ is a relator on X, then (1)* $\mathscr{T}_{\mathscr{R}^\triangle} = \mathscr{E}_{\mathscr{R}} \cup \{\emptyset\}$; *(2)* $\mathscr{F}_{\mathscr{R}^\triangle} = \big(\mathscr{P}(X) \setminus \mathscr{D}_{\mathscr{R}}\big) \cup \{X\}$.

Remark 21 Note that if in particular $\mathscr{R} = \emptyset$, then $\mathscr{E}_{\mathscr{R}} = \emptyset$. Moreover, $\mathscr{R}^\triangle = \emptyset$ if $X \neq \emptyset$, and $\mathscr{R}^\triangle = \{\emptyset\}$ if $X = \emptyset$. Therefore, $\mathscr{T}_{\mathscr{R}^\triangle} = \{\emptyset\}$, and thus (1) is still true.

Now, since $\emptyset \notin \mathscr{E}_{\mathscr{R}}$ if $\mathscr{R}$ is non-partial, we can also state

Corollary 8 *If $\mathscr{R}$ is a non-partial relator on X, then (1)* $\mathscr{E}_{\mathscr{R}} = \mathscr{T}_{\mathscr{R}^\triangle} \setminus \{\emptyset\}$, *(2)* $\mathscr{D}_{\mathscr{R}} = \big(\mathscr{P}(X) \setminus \mathscr{F}_{\mathscr{R}^\triangle}\big) \cup \{X\}$.

9 Projection Operations for Relators

By using the basic properties of the operation ∞, in addition to a particular case of Theorem 16, we can also prove the following

Theorem 21 *∞ is a closure operation for relations on X such that, for any two relations R and S on X, we have*

$$S \subseteq R^\infty \iff S^\infty \subseteq R^\infty \iff \tau_R \subseteq \tau_S \iff \tau_R \subseteq \tau_S;$$

Proof To prove that $\tau_R \subseteq \tau_S \iff S \subseteq R^\infty$, note that if $x \in X$, then because of the inclusion $R \subseteq R^\infty$ and the transitivity of R^∞ we have

$$R[\,R^\infty(x)\,] \subseteq R^\infty[\,R^\infty(x)\,] = \big(R^\infty \circ R^\infty\big)(x) \subseteq R^\infty(x).$$

Thus, by the definition of $\tau_{\mathscr{R}}$, we have $R^\infty(x) \in \tau_{\mathscr{R}}$. Now, if $\tau_R \subseteq \tau_S$ holds, then we can see that $R^\infty(x) \in \tau_S$, and thus $S[\,R^\infty(x)\,] \subseteq R^\infty(x)$. Hence, by using the reflexivity of R^∞, we can already infer that $S(x) \subseteq R^\infty(x)$. Therefore, $S \subseteq R^\infty$ also holds.

While, if $A \in \tau_{\mathscr{R}}$, then by the definition of $\tau_{\mathscr{R}}$ we have $R[\,A\,] \subseteq A$. Hence, by induction, we can see that $R^n[\,A\,] \subseteq A$ for all $n \in \mathbb{N}$. Now, since $R^0[\,A\,] = \Delta_X[\,A\,] = A$ also holds, we can already state that

$$R^{\infty}[A] = \Big(\bigcup_{n=0}^{\infty} R^n\Big)[A] = \bigcup_{n=0}^{\infty} R^n[A] \subseteq \bigcup_{n=0}^{\infty} A = A.$$

Therefore, if $S \subseteq R^{\infty}$ holds, then we have $S[A] \subseteq R^{\infty}[A] \subseteq A$, and thus $A \in \tau_S$ also holds.

Now, analogously to Corollary 3, we can also state

Corollary 9 *For any relation R on X, $S = R^{\infty}$ is the largest relation on X such that $\tau_R \subseteq \tau_S$ $\big(\tau_R = \tau_S\big)$, or equivalently $\tau_R \subseteq \tau_S$ $\big(\tau_R = \tau_S\big)$.*

Remark 22 Preliminary forms of the above theorem and its corollary were first proved by Mala [53].

Moreover, he also proved that $R^{\infty}(x) = \bigcap \{ A \in \tau_R : \ x \in A \}$ for all $x \in X$, and thus $R^{\infty} = \bigcap \{ R_A : \ A \in \tau_R \}$.

By using Theorem 21, as an analogue of Theorem 16, we can also prove

Theorem 22 *$\#\partial$ is a closure operation for relators on X such that, for any two relators $\mathscr{R}$ and $\mathscr{S}$ on X, we have*

$$\mathscr{S} \subseteq \mathscr{R}^{\#\partial} \iff \mathscr{S}^{\#\partial} \subseteq \mathscr{R}^{\#\partial} \iff \tau_{\mathscr{S}} \subseteq \tau_{\mathscr{R}} \iff \tau_{\mathscr{S}} \subseteq \tau_{\mathscr{R}};$$

Thus, analogously to Corollary 3, we can also state

Corollary 10 *For any relator $\mathscr{R}$ on X, $\mathscr{S} = \mathscr{R}^{\#\partial}$ is the largest relator on X such that $\tau_{\mathscr{S}} \subseteq \tau_{\mathscr{R}}$ $\big(\tau_{\mathscr{S}} = \tau_{\mathscr{R}}\big)$, or equivalently $\tau_{\mathscr{S}} \subseteq \tau_{\mathscr{R}}$ $\big(\tau_{\mathscr{S}} = \tau_{\mathscr{R}}\big)$.*

By using the Galois property of the operations ∞ and ∂, Theorem 22 can be reformulated in the following more convenient form.

Theorem 23 *$\#\infty$ is a projection operation for relators on X such that, for any two relators $\mathscr{R}$ and $\mathscr{S}$ on X, we have*

$$\mathscr{S}^{\infty} \subseteq \mathscr{R}^{\#} \iff \mathscr{S}^{\#\infty} \subseteq \mathscr{R}^{\#\infty} \iff \tau_{\mathscr{S}} \subseteq \tau_{\mathscr{R}} \iff \tau_{\mathscr{S}} \subseteq \tau_{\mathscr{R}}.$$

Remark 23 Moreover, it can be shown that the inclusions $\mathscr{S}^{\infty} \subseteq \mathscr{R}^{\#}$, $\mathscr{S}^{\#\infty} \subseteq \mathscr{R}^{\#}$ and $\mathscr{S}^{\infty\#} \subseteq \mathscr{R}^{\infty\#}$ are also equivalent.

Now, analogously to our former corollaries, we can also state

Corollary 11 *For any relator $\mathscr{R}$ on X, $\mathscr{S} = \mathscr{R}^{\#\infty}$ is the largest preorder relator on X such that $\tau_{\mathscr{S}} \subseteq \tau_{\mathscr{R}}$ $\big(\tau_{\mathscr{S}} = \tau_{\mathscr{R}}\big)$, or equivalently $\tau_{\mathscr{S}} \subseteq \tau_{\mathscr{R}}$ $\big(\tau_{\mathscr{S}} = \tau_{\mathscr{R}}\big)$.*

Remark 24 The advantage of the projection operation $\#\infty$ over the closure operation $\#\partial$ lies mainly in the fact that, in contrast to $\#\partial$, it is *stable* in the sense $\{X^2\}^{\#\infty} = \{X^2\}$.

Since the structure $\mathscr{T}$ is not union-preserving, by using some parts of the theory of Pataki connections [64, 69, 95], we can only prove the following

Theorem 24 *$\wedge\partial$ is a preclosure operation for relators such that, for any two relators $\mathscr{R}$ and $\mathscr{S}$ on X, we have*

$$\mathscr{T}_{\mathscr{S}} \subseteq \mathscr{T}_{\mathscr{R}} \iff \mathscr{F}_{\mathscr{S}} \subseteq \mathscr{F}_{\mathscr{R}} \implies \mathscr{S}^{\wedge} \subseteq \mathscr{R}^{\wedge\partial} \implies \mathscr{S}^{\wedge\partial} \subseteq \mathscr{R}^{\wedge\partial}.$$

Remark 25 If $\operatorname{card}(X) > 2$, then by using the equivalence relator $\mathscr{R} = \{X^2\}$ Mala [53, Example 5.3] proved that there does not exist a largest relator $\mathscr{S}$ on X such that $\mathscr{T}_{\mathscr{R}} = \mathscr{T}_{\mathscr{S}}$.

Moreover, Pataki [64, Example 7.2] proved that $\mathscr{T}_{\mathscr{R}^{\wedge\partial}} \not\subseteq \mathscr{T}_{\mathscr{R}}$ and $\wedge\partial$ is not idempotent. (Actually, it can be proved that $\mathscr{R}^{\wedge\partial\wedge} \not\subseteq \mathscr{R}^{\wedge\partial}$ also holds [84, Example 10.11].)

Fortunately, as an analogue of Theorem 23, we can also prove

Theorem 25 *$\wedge\infty$ is a projection operation for relators on X such that, for any two nonvoid relators $\mathscr{R}$ and $\mathscr{S}$ on X, we have*

$$\mathscr{S}^{\wedge\infty} \subseteq \mathscr{R}^{\wedge} \iff \mathscr{S}^{\wedge\infty} \subseteq \mathscr{R}^{\wedge\infty} \iff \mathscr{T}_{\mathscr{S}} \subseteq \mathscr{T}_{\mathscr{R}} \iff \mathscr{F}_{\mathscr{S}} \subseteq \mathscr{F}_{\mathscr{R}}.$$

Thus, in particular, we can also state

Corollary 12 *For any nonvoid relator $\mathscr{R}$ on X, $\mathscr{S} = \mathscr{R}^{\wedge\infty}$ is the largest preorder relator on X such that $\mathscr{T}_{\mathscr{S}} \subseteq \mathscr{T}_{\mathscr{R}}$ $\left(\mathscr{T}_{\mathscr{S}} = \mathscr{T}_{\mathscr{R}}\right)$, or equivalently $\mathscr{F}_{\mathscr{S}} \subseteq \mathscr{F}_{\mathscr{R}}$ $\left(\mathscr{F}_{\mathscr{S}} = \mathscr{F}_{\mathscr{R}}\right)$.*

Remark 26 In the light of the several disadvantages of the structure $\mathscr{T}$, it is rather curious that most of the works in general topology and abstract analysis have been based on open sets suggested by Tietze [98] and Alexandroff [2], and standardized by Bourbaki [6] and Kelley [41]. (See Thron [97, p. 18].)

Moreover, it also a very striking fact that, despite the results of Davis [20], Pervin [66], Hunsaker and Lindgren [35] and the second author [76, 85], generalized proximities and closures, minimal structures, generalized topologies and stacks (ascending systems) are still intensively investigated by a great number of mathematicians without using generalized uniformities.

10 Reflexive, Non-partial and Non-degenerated Relators

Definition 7 A relator $\mathscr{R}$ on X is called *reflexive* if each member R of $\mathscr{R}$ is a reflexive relation on X.

Remark 27 Thus, the following assertions are equivalent:

(1) $\mathscr{R}$ is reflexive;
(2) $x \in R(x)$ for all $x \in X$ and $R \in \mathscr{R}$;
(3) $A \subseteq R[A]$ for all $A \subseteq X$ and $R \in \mathscr{R}$.

The importance of reflexive relators is also apparent from the following two obvious theorems.

Theorem 26 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent: (1) $\rho_{\mathscr{R}}$ is reflexive; (2) $\mathscr{R}$ is reflexive;*
(3) $A \subseteq \mathrm{cl}_{\mathscr{R}}(A)$ $\left(\mathrm{int}_{\mathscr{R}}(A) \subseteq A\right)$ for all $A \subseteq X$.

Proof To prove the equivalence of (1) and (2), recall that $\rho_{\mathscr{R}} = \left(\bigcap \mathscr{R}\right)^{-1}$.

Remark 28 Thus, the relator $\mathscr{R}$ is reflexive if and only if $A^{\circ} \subseteq A$ $(A \subseteq A^{-})$ for all $A \subseteq X$.

Therefore, if $\mathscr{R}$ is a reflexive relator on X, then for any $A \subseteq X$ we have $A \in \mathscr{T}_{\mathscr{R}}$ $(A \in \mathscr{F}_{\mathscr{R}})$ if and only if $A^{\circ} = A$ $(A^{-} = A)$.

Theorem 27 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent:*

(1) $\mathscr{R}$ is reflexive;
(2) $A \in \mathrm{Int}_{\mathscr{R}}(B)$ implies $A \subseteq B$ for all $A, B \subseteq X$;
(3) $A \cap B \neq \emptyset$ implies $A \in \mathrm{Cl}_{\mathscr{R}}(B)$ for all $A, B \subseteq X$.

Remark 29 In addition to the above two theorems, it is also worth mentioning that if $\mathscr{R}$ is a reflexive relator on X, then

(1) $\mathrm{Int}_{\mathscr{R}}$ is a transitive relation on $\mathscr{P}(X)$;
(2) $B \in \mathrm{Cl}_{\mathscr{R}}(A)$ implies $\mathscr{P}(X) = \mathrm{Cl}_{\mathscr{R}}(A)^{c} \cup \mathrm{Cl}_{\mathscr{R}}^{-1}(B)$;
(3) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{bor}_{\mathscr{R}}(A)\right) = \emptyset$ and $\mathrm{int}_{\mathscr{R}}\left(\mathrm{res}_{\mathscr{R}}(A)\right)\big) = \emptyset$ for all $A \subseteq X$.

Thus, for instance, for any $A \subseteq X$ we have $\mathrm{res}_{\mathscr{R}}(A) \in \mathscr{T}_{\mathscr{R}}$ if and only if $A \in \mathscr{F}_{\mathscr{R}}$.

Analogously to Definition 7, we may also naturally have the following

Definition 8 A relator $\mathscr{R}$ on X to Y is called *non-partial* if each member R of $\mathscr{R}$ is a non-partial relation on X to Y.

Remark 30 Thus, the following assertions are equivalent:

(1) $\mathscr{R}$ is non-partial;
(2) $R^{-1}[Y] = X$ for all $R \in \mathscr{R}$;
(3) $R(x) \neq \emptyset$ for all $x \in X$ and $R \in \mathscr{R}$.

The importance of non-partial relators is apparent from the following

Theorem 28 *For a relator $\mathscr{R}$ on X to Y, the following assertions are equivalent: (1) $\mathscr{R}$ is non-partial;*

(2) $\emptyset \notin \mathscr{E}_{\mathscr{R}}$; (3) $\mathscr{D}_{\mathscr{R}} \neq \emptyset$; (4) $Y \in \mathscr{D}_{\mathscr{R}}$; (5) $\mathscr{E}_{\mathscr{R}} \neq \mathscr{P}(Y)$.

Sometimes, we also need the following localized form of Definition 8.

Definition 9 A relator $\mathscr{R}$ on X is called *locally non-partial* if for each $x \in X$ there exists $R \in \mathscr{R}$ such that for any $y \in R(x)$ and $S \in \mathscr{R}$ we have $S(y) \neq \emptyset$.

Remark 31 Thus, if either $X = \emptyset$ or $\mathscr{R}$ is nonvoid and non-partial, then $\mathscr{R}$ is locally non-partial.

Moreover, by using the corresponding definitions, we can also easily prove

Theorem 29 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent: (1) $\mathscr{R}$ is locally non-partial;* *(2)* $X = \operatorname{int}_{\mathscr{R}}\big(\operatorname{cl}_{\mathscr{R}}(X)\big)$.

Proof To prove the implication (1) $\Longrightarrow$ (2), note that if (1) holds, then for each $x \in X$ there exists $R \in \mathscr{R}$ such that for any $y \in R(x)$ and for any $S \in \mathscr{R}$ we have $S(y) \cap X = S(y) \neq \emptyset$, and thus $y \in \operatorname{cl}_{\mathscr{R}}(X)$.

Therefore, for each $x \in X$ there exists $R \in \mathscr{R}$ such that $R(x) \subseteq \operatorname{cl}_{\mathscr{R}}(X)$, and thus $x \in \operatorname{int}_{\mathscr{R}}\big(\operatorname{cl}_{\mathscr{R}}(X)\big)$. Hence, we can already see that $X \subseteq \operatorname{int}_{\mathscr{R}}\big(\operatorname{cl}_{\mathscr{R}}(X)\big)$, and thus (2) also holds. (Therefore, by a former notation, $X \in \mathscr{T}^{r}_{\mathscr{R}}$.)

In addition to Definition 8, it is also worth introducing the following

Definition 10 A relator $\mathscr{R}$ on X to Y is called *non-degerated* if both $X \neq \emptyset$ and $\mathscr{R} \neq \emptyset$.

Thus, analogously to Theorem 28, we can also easily establish the following

Theorem 30 *For a relator $\mathscr{R}$ on X to Y, the following assertions are equivalent: (1) $\mathscr{R}$ is non-degenerated;*

(2) $\emptyset \notin \mathscr{D}_{\mathscr{R}}$; *(3)* $\mathscr{E}_{\mathscr{R}} \neq \emptyset$; *(4)* $Y \in \mathscr{E}_{\mathscr{R}}$; *(5)* $\mathscr{D}_{\mathscr{R}} \neq \mathscr{P}(Y)$.

Remark 32 In addition to Theorems 28 and 30, it is also worth mentioning that if a relator $\mathscr{R}$ on X to Y is paratopologically simple in the sense that $\mathscr{E}_{\mathscr{R}} = \mathscr{E}_S$ for some relation S on X to Y, then the stack $\mathscr{E}_{\mathscr{R}}$ has a base $\mathscr{B}$ with $\operatorname{card}(\mathscr{B}) \leq \operatorname{card}(X)$. (See [63, Theorem 5.9] of Pataki.)

The existence of a non-paratopologically simple (actually finite equivalence) relator, proved first by Pataki [63, Example 5.11], shows that in our definitions of the relations $\operatorname{Lim}_{\mathscr{R}}$ and $\operatorname{Adh}_{\mathscr{R}}$ we cannot restrict ourselves to functions of gosets (generalized ordered sets) without some loss of generality.

11 Topological and Quasi-Topological Relators

The following improvement of [74, Definition 2.1] was first considered in [75].

Definition 11 A relator $\mathscr{R}$ on X is called:

(1) *quasi-topological* if $x \in \operatorname{int}_{\mathscr{R}}\big(\operatorname{int}_{\mathscr{R}}\big(R(x)\big)\big)$ for all $x \in X$ and $R \in \mathscr{R}$;
(2) *topological* if for any $x \in X$ and $R \in \mathscr{R}$ there exists $V \in \mathscr{T}_{\mathscr{R}}$ such that $x \in V \subseteq R(x)$.

The appropriateness of these definitions is already quite obvious from the following four theorems.

Theorem 31 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent:*

(1) $\mathscr{R}$ *is quasi-topological;*
(2) $\operatorname{int}_{\mathscr{R}}\big(R(x)\big) \in \mathscr{T}_{\mathscr{R}}$ *for all* $x \in X$ *and* $R \in \mathscr{R}$.
(3) $\operatorname{cl}_{\mathscr{R}}(A) \in \mathscr{F}_{\mathscr{R}}$ $\big(\operatorname{int}_{\mathscr{R}}(A) \in \mathscr{T}_{\mathscr{R}}\big)$ *for all* $A \subseteq X$.

Remark 33 Hence, by Definition 3, we can see that the relator $\mathscr{R}$ is quasi-topological if and only if $A^{\circ} \subseteq A^{\circ\circ}$ $(A^{--} \subseteq A^{-})$ for all $A \subseteq X$.

Theorem 32 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent:*

(1) $\mathscr{R}$ *is topological;* *(2)* $\mathscr{R}$ *is reflexive and quasi-topological.*

Remark 34 By Theorem 31, the relator $\mathscr{R}$ may be called *weakly (strongly) quasi-topological* if $\rho_{\mathscr{R}}(x) \in \mathscr{F}_{\mathscr{R}}$ $\big(R(x) \in \mathscr{T}_{\mathscr{R}}\big)$ for all $x \in X$ and $R \in \mathscr{R}$.

Moreover, by Theorem 32, the relator $\mathscr{R}$ may be called *weakly (strongly) toplogical* if it is reflexive and weakly (strongly) quasi-topological.

The following theorem shows that in a topological relator space $X(\mathscr{R})$, the relation $\operatorname{int}_{\mathscr{R}}$ and the family $\mathscr{T}_{\mathscr{R}}$ are also equivalent tools.

Theorem 33 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent:*

(1) $\mathscr{R}$ *is topological;*
(2) $\operatorname{int}_{\mathscr{R}}(A) = \bigcup \mathscr{T}_{\mathscr{R}} \cap \mathscr{P}(A)$ *for all* $A \subseteq X$;
(3) $\operatorname{cl}_{\mathscr{R}}(A) = \bigcap \mathscr{F}_{\mathscr{R}} \cap \mathscr{P}^{-1}(A)$ *for all* $A \subseteq X$.

Now, as an immediate consequence of Theorems 14 and 33, we can also state

Corollary 13 *If $\mathscr{R}$ is topological relator on X, then for any $A \subset X$, we have*

(1) $A \in \mathscr{E}_{\mathscr{R}}$ *if and only if there exists* $V \in \mathscr{T}_{\mathscr{R}} \setminus \{\emptyset\}$ *such that* $V \subseteq A$;
(2) $A \in \mathscr{D}_{\mathscr{R}}$ *if and only if for all* $W \in \mathscr{F}_{\mathscr{R}} \setminus \{X\}$ *we have* $A \setminus W \neq \emptyset$.

However, it is now more important to note that we can also prove the following

Theorem 34 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent:*

(1) $\mathscr{R}$ *is topological;*
(2) $\mathscr{R}$ *is topologically equivalent to* $\mathscr{R}^{\wedge\infty}$;
(3) $\mathscr{R}$ *is topologically equivalent to a preorder relator on* X.

Proof To prove the implication (1) $\Longrightarrow$ (3), note that if (1) holds, then by Definition 11, for any $x \in X$ and $R \in \mathscr{R}$, there exists $V \in \mathscr{T}_{\mathscr{R}}$ such that $x \in V \subseteq R(x)$. Thus, by using the Pervin preorder relator

$$\mathscr{S} = \mathscr{R}_{\mathscr{T}_{\mathscr{R}}} = \big\{ R_V : \ V \in \mathscr{T}_{\mathscr{R}} \big\}, \qquad \text{where} \qquad R_V = V^2 \cup V^c \times X,$$

we can show that $\operatorname{int}_{\mathscr{R}}(A) = \operatorname{int}_{\mathscr{S}}(A)$ for all $A \subseteq X$, and thus (3) also holds.

For this, we have to note that

$$R_V(x) = V \quad \text{if} \quad x \in V \qquad \text{and} \qquad R_V(x) = X \quad \text{if} \quad x \in V^c.$$

In addition to Theorem 31, it is also worth proving the following

Theorem 35 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent: (1) $\mathscr{R}$ is quasi-topological;*
(2) $\mathscr{R} \subseteq \left(\mathscr{R}^{\wedge} \circ \mathscr{R}\right)^{\wedge}$; (3) $\mathscr{R}^{\wedge} \subseteq \left(\mathscr{R}^{\wedge} \circ \mathscr{R}^{\wedge}\right)^{\wedge}$.

Remark 35 By [74], a relator $\mathscr{R}$ on X may be naturally called *topologically transitive* if, for each $x \in X$ and $R \in \mathscr{R}$ there exist $S, T \in \mathscr{R}$ such that $T[S(x)] \subseteq R(x)$.

This property can be easily reformulated in the more concise form that $\mathscr{R} \subseteq \left(\mathscr{R} \circ \mathscr{R}\right)^{\wedge}$. Thus, the equivalence (1) and (3) can be expressed by saying that $\mathscr{R}$ is quasi-topological if and only if $\mathscr{R}^{\wedge}$ is topologically transitive.

12 Proximal and Quasi-Proximal Relators

Analogously to Definition 11, we may also naturally have the following

Definition 12 A relator $\mathscr{R}$ on X is called

(1) *quasi-proximal* if $A \in \operatorname{Int}_{\mathscr{R}}\left[\tau_{\mathscr{R}} \cap \operatorname{Int}_{\mathscr{R}}\left(R[A]\right)\right]$ for all $A \subseteq X$ and $R \in \mathscr{R}$;
(2) *proximal* if for any $A \subseteq X$ and $R \in \mathscr{R}$ there exists $V \in \tau_{\mathscr{R}}$ such that $A \subseteq V \subseteq R[A]$.

Remark 36 Thus, the relator $\mathscr{R}$ is quasi-proximal if and only if, for any $A \subseteq X$ and $R \in \mathscr{R}$, there exists $V \in \tau_{\mathscr{R}}$ such that $A \in \operatorname{Int}_{\mathscr{R}}(V)$ and $V \in \operatorname{Int}_{\mathscr{R}}\left(R[A]\right)$.

Now, by using the corresponding definitions, we can also easily prove the following analogues of Theorems 32 and 33.

Theorem 36 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent: (1) $\mathscr{R}$ is proximal; (2) $\mathscr{R}$ is reflexive and quasi-proximal.*

Proof To prove the implication (1) $\Longrightarrow$ (2), note that if (1) holds, then for any $A \subseteq X$ and $R \in \mathscr{R}$, there exists $V \in \tau_{\mathscr{R}}$ such that $A \subseteq V \subseteq R[A]$. Hence, by taking $A = \{x\}$ for $x \in X$, we can see that $\mathscr{R}$ is reflexive.

Moreover, since $V \in \tau_{\mathscr{R}}$, we can also note that $V \in \operatorname{Int}_{\mathscr{R}}(V)$. Hence, by using that $A \subseteq V$ and $V \subseteq R[A]$, we can already infer that $A \in \operatorname{Int}_{\mathscr{R}}(V)$ and $V \in \operatorname{Int}_{\mathscr{R}}\left(R[A]\right)$. Therefore, by Remark 36, $\mathscr{R}$ is quasi-proximal.

Remark 37 Note that if $\mathscr{R}$ is only a *weakly proximal* relator on X in the sense that, for any $x \in X$ and $R \in \mathscr{R}$, there exists $V \in \tau_{\mathscr{R}}$ such that $x \in V \subseteq R(x)$, then because of $\tau_{\mathscr{R}} \subseteq \mathscr{T}_{\mathscr{R}}$ we can already state that $\mathscr{R}$ is topological.

Theorem 37 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent:*

(1) $\mathscr{R}$ *is proximal;*
(2) $\mathrm{Int}_{\mathscr{R}}(A) = \mathscr{P}\left[\tau_{\mathscr{R}} \cap \mathscr{P}(A)\right]$ *for all* $A \subseteq X$*;*
(3) $\mathrm{Cl}_{\mathscr{R}}(A) = \bigcap \left\{ \mathscr{P}(W^c)^c : \ W \in \tau_{\mathscr{R}} \cap \mathscr{P}^{-1}(A) \right\}$ *for all* $A \subseteq X$.

Proof Note that if $A \subseteq X$ and $B \in \mathscr{P}\left[\tau_{\mathscr{R}} \cap \mathscr{P}(A)\right]$, then there exists $V \in \tau_{\mathscr{R}}$ such that $B \in \mathscr{P}(V)$ and $V \in \mathscr{P}(A)$, and thus $B \subseteq V \subseteq A$. Hence, by using that $V \in \mathrm{Int}_{\mathscr{R}}(V)$ we can already infer that $B \in \mathrm{Int}_{\mathscr{R}}(A)$. Thus, the inclusion $\mathscr{P}\left[\tau_{\mathscr{R}} \cap \mathscr{P}(A)\right] \subseteq \mathrm{Int}_{\mathscr{R}}(A)$ is always true.

Therefore, to obtain (1), it is enough to assume only the converse inclusion. For this, note that if $A \subseteq X$ and $R \in \mathscr{R}$, then because of $R[A] \subseteq R[A]$, we always have $A \in \mathrm{Int}_{\mathscr{R}}\left(R[A]\right)$. Therefore, if $\mathrm{Int}_{\mathscr{R}}\left(R[A]\right) \subseteq \mathscr{P}\left[\tau_{\mathscr{R}} \cap \mathscr{P}\left(R[A]\right)\right]$, then we also have $A \in \mathscr{P}\left[\tau_{\mathscr{R}} \cap \mathscr{P}\left(R[A]\right)\right]$. Thus, there exists $V \in \tau_{\mathscr{R}}$ such that $A \in \mathscr{P}(V)$ and $V \in \mathscr{P}\left(R[A]\right)$, and thus $A \subseteq V \subseteq R[A]$.

Remark 38 Note that $\mathscr{P}(A) = \mathrm{Int}_{\Delta_X}(A)$ for all $A \subseteq X$. Therefore, instead of (2) we may write that $\mathrm{Int}_{\mathscr{R}}(A) = \mathrm{Int}_{\Delta_X}\left[\tau_{\mathscr{R}} \cap \mathrm{Int}_{\Delta_X}(A)\right]$ for all $A \subseteq X$.

However, it is now more important to note that we also have the following

Theorem 38 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent:*

(1) $\mathscr{R}$ *is proximal;*
(2) $\mathscr{R}$ *is proximally equivalent to* $\mathscr{R}^{\infty}$ *or* $\mathscr{R}^{\#\infty}$*;*
(3) $\mathscr{R}$ *is proximally equivalent to a preorder relator on* X.

In principle, each theorem on topological and quasi-topological relators can be immediately derived from a corresponding theorem on proximal and quasi-proximal relators by using the following two theorems.

Theorem 39 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent: (1) $\mathscr{R}$ is quasi-topological; (2) $\mathscr{R}^{\wedge}$ is quasi-proximal.*

Proof To prove the implication (1) $\Longrightarrow$ (2), assume that (1) holds, and moreover $A \subseteq X$ and $S \in \mathscr{R}^{\wedge}$. Define $V = \mathrm{int}_{\mathscr{R}}\left(S[A]\right)$. Then, if $\mathscr{R} \neq \emptyset$, by Theorem 31 and Corollary 4, we have $V \in \mathscr{T}_{\mathscr{R}} = \tau_{\mathscr{R}^{\wedge}}$. Moreover, since $V \subseteq \mathrm{int}_{\mathscr{R}}\left(S[A]\right)$, by Theorem 19 we also have $V \in \mathrm{Int}_{\mathscr{R}^{\wedge}}\left(S[A]\right)$. Therefore, $V \in \tau_{\mathscr{R}^{\wedge}} \cap \mathrm{Int}_{\mathscr{R}^{\wedge}}\left(S[A]\right)$.

Furthermore, since $S[A] \subseteq S[A]$, we can also note that $A \in \mathrm{Int}_{\mathscr{R}^{\wedge}}\left(S[A]\right)$. Hence, by Theorem 19, we can infer that $A \subseteq \mathrm{int}_{\mathscr{R}}\left(S[A]\right) = V$. Moreover, since $V \in \tau_{\mathscr{R}^{\wedge}}$, we can also note that $V \in \mathrm{Int}_{\mathscr{R}^{\wedge}}(V)$. Hence, since $A \subseteq V$, we can infer that $A \in \mathrm{Int}_{\mathscr{R}^{\wedge}}(V)$. Therefore, since $V \in \tau_{\mathscr{R}^{\wedge}} \cap \mathrm{Int}_{\mathscr{R}^{\wedge}}\left(S[A]\right)$, we also have $A \in \mathrm{Int}_{\mathscr{R}^{\wedge}}\left[\tau_{\mathscr{R}^{\wedge}} \cap \mathrm{Int}_{\mathscr{R}^{\wedge}}\left(S[A]\right)\right]$.

This shows that (1) implies (2) whenever $\mathscr{R} \neq \emptyset$. However, if $\mathscr{R} = \emptyset$, then it can be easily seen that $\mathscr{R}$ is topological and $\mathscr{R}^{\wedge}$ is proximal.

Remark 39 If assertion (2) holds, then $\mathscr{R}^{\wedge}$ is *semi-proximal* in the sense that $A \in \mathrm{Int}_{\mathscr{R}^{\wedge}}\left[\mathrm{Int}_{\mathscr{R}^{\wedge}}\left(S[A]\right)\right]$ for all $A \subseteq X$ and $S \in \mathscr{R}^{\wedge}$.

Moreover, if in particular $\{x\} \in \mathrm{Int}_{\mathscr{R}^{\wedge}}\left[\mathrm{Int}_{\mathscr{R}^{\wedge}}\left(R(x)\right)\right]$ for all $x \in X$ and $R \in \mathscr{R}$, then we can already prove that assertion (1) also holds.

From Theorem 39, by using Theorems 32 and 36, we can immediately derive

Theorem 40 *For a relator $\mathscr{R}$ on X, the following assertions are equivalent: (1) $\mathscr{R}$ is topological, (2) $\mathscr{R}^{\wedge}$ is proximal.*

Remark 40 By the corresponding definitions, it is clear that the relator $\mathscr{R}^{\wedge}$ is reflexive if and only if $\mathscr{R}$ is reflexive.

However, if $\mathscr{R} \not\subseteq \{X^2\}$, then there exists $R \in \mathscr{R}$ such that $R \neq X^2$. Therefore, there exist $x, y \in X$ such that $x \notin R(y)$. Thus, $S = \{x\} \times R(y) \cup \{x\}^c \times X$ is a non-reflexive relation on X such that $S \in \mathscr{R}^{\vartriangle}$. Therefore, $\mathscr{R}^{\vartriangle}$ cannot be reflexive.

Note that if in particular either $\mathscr{R} = \emptyset$ or $\mathscr{R} = \{X^2\}$, then $\mathscr{R}^{\vartriangle}$ is also reflexive.

13 A Few Basic Facts on Filtred Relators

Intersection properties of relators were also first investigated in [74, 75].

Definition 13 A relator $\mathscr{R}$ on X to Y is called

(1) *properly filtered* if for any $R, S \in \mathscr{R}$ we have $R \cap S \in \mathscr{R}$;
(2) *uniformly filtered* if for any $R, S \in \mathscr{R}$ there exists $T \in \mathscr{R}$ such that $T \subseteq R \cap S$;
(3) *proximally filtered* if for any $A \subseteq X$ and $R, S \in \mathscr{R}$ there exists $T \in \mathscr{R}$ such that $T[A] \subseteq R[A] \cap S[A]$;
(4) *topologically filtered* if for any $x \in X$ and $R, S \in \mathscr{R}$ there exists $T \in \mathscr{R}$ such that $T(x) \subseteq R(x) \cap S(x)$.

Remark 41 By using the binary operation $\wedge$ and the basic closure operations on relators, the above properties can be reformulated in some more concise forms.

For instance, we can see that $\mathscr{R}$ is topologically filtered if and only if any one of the properties $\mathscr{R} \wedge \mathscr{R} \subseteq \mathscr{R}^{\wedge}$, $(\mathscr{R} \wedge \mathscr{R})^{\wedge} = \mathscr{R}^{\wedge}$ and $\mathscr{R}^{\wedge} \wedge \mathscr{R}^{\wedge} = \mathscr{R}^{\wedge}$ holds.

However, in general, we only have $(R \cap S)[A] \subseteq R[A] \cap S[A]$. Therefore, the corresponding proximal filteredness properties are, unfortunately, not equivalent.

Despite this, we can easily prove the following theorem which shows the appropriateness of the above proximal filteredness property.

Theorem 41 *For a relator $\mathscr{R}$ on X to Y, the following assertions are equivalent:*

(1) $\mathscr{R}$ is proximally filtered;
(2) $\mathrm{Cl}_{\mathscr{R}}(A \cup B) = \mathrm{Cl}_{\mathscr{R}}(A) \cup \mathrm{Cl}_{\mathscr{R}}(B)$ for all $A, B \subseteq Y$;

(3) $\operatorname{Int}_{\mathscr{R}}(A \cap B) = \operatorname{Int}_{\mathscr{R}}(A) \cap \operatorname{Int}_{\mathscr{R}}(B)$ *for all* $A, B \subseteq Y$.

Proof To prove the implication (3) $\Longrightarrow$ (1), note that if $A \subseteq X$ and $R, S \in \mathscr{R}$, then by the definition of $\operatorname{Int}_{\mathscr{R}}$ we trivially have $A \in \operatorname{Int}_{\mathscr{R}}\big(R[A]\big)$ and $A \in \operatorname{Int}_{\mathscr{R}}\big(S[A]\big)$. Therefore, if (3) holds, then we also have $A \in \operatorname{Int}_{\mathscr{R}}\big(R[A] \cap S[A]\big)$. Thus, by the definition of $\operatorname{Int}_{\mathscr{R}}$, there exists $T \in \mathscr{R}$ such that $T[A] \subseteq R[A] \cap S[A]$.

Now, as an immediate consequence of this theorem, we can also state

Corollary 14 *If $\mathscr{R}$ is a proximally filtered relator on X, then the families $\tau_{\mathscr{R}}$ and $\tau_{\mathscr{R}}$ are closed under binary unions and intersections, respectively.*

From Theorem 41, we can also easily derive the following

Theorem 42 *For a relator $\mathscr{R}$ on X to Y, the following assertions are equivalent:*

(1) *$\mathscr{R}$ is topologically filtered;*
(2) $\operatorname{cl}_{\mathscr{R}}(A \cup B) = \operatorname{cl}_{\mathscr{R}}(A) \cup \operatorname{cl}_{\mathscr{R}}(B)$ *for all* $A, B \subseteq Y$;
(3) $\operatorname{int}_{\mathscr{R}}(A \cap B) = \operatorname{int}_{\mathscr{R}}(A) \cap \operatorname{int}_{\mathscr{R}}(B)$ *for all* $A, B \subseteq Y$.

Thus, in particular, we can also state the following

Corollary 15 *If $\mathscr{R}$ is a topologically filtered relator on X, then the families $\mathscr{F}_{\mathscr{R}}$ and $\mathscr{T}_{\mathscr{R}}$ are closed under binary unions and intersections, respectively.*

The following example shows that, for a non-topological relator $\mathscr{R}$, the converse of the above corollary need not be true.

Example 2 If $X = \{1, 2, 3\}$ and R_i is relation on X, for each $i = 1, 2$, such that

$$R_i(1) = \{1,\ i+1\} \qquad \text{and} \qquad R_i(2) = R_i(3) = \{2,\ 3\},$$

then $\mathscr{R} = \{R_1, R_2\}$ is a reflexive relator on X such that $\mathscr{T}_{\mathscr{R}}$ is closed under arbitrary intersections, but $\mathscr{R}$ is still not topologically filtered.

By the corresponding definitions, it is clear that $\mathscr{T}_{\mathscr{R}} = \{\emptyset, \{2, 3\}, X\}$. Moreover, we can note that $R_i(1) \not\subseteq R_1(1) \cap R_2(1)$ for each $i = 1, 2$. Thus, by Definition 13, the relator $\mathscr{R}$ is not topologically filtered.

In addition to Theorem 42, we can also prove the following generalization of [48, Lemma 7] of Levine.

Theorem 43 *If $\mathscr{R}$ is a topologically filtered relator on X, $A, B \subseteq X$ and there exists $V \in \mathscr{T}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}}$ such that $A \subseteq V$ and $B \subseteq V^c$, then*

$$\operatorname{int}_{\mathscr{R}}(A \cup B) = \operatorname{int}_{\mathscr{R}}(A) \cup \operatorname{int}_{\mathscr{R}}(B).$$

Proof Because of the increasingness of $\circ$, we evidently have $A^\circ \cup B^\circ \subseteq (A \cup B)^\circ$. Therefore, we need actually prove the converse inclusion.

For this, note that if $x \in (A \cup B)^\circ$, then by the definition of the operation $\circ$, there exists $R \in \mathscr{R}$ such that $R(x) \subseteq A \cup B$.

Now, if $x \in V$, then by the definition of $\mathscr{T}_{\mathscr{R}}$, we can see that there exists $S \in \mathscr{R}$ such that $S(x) \subseteq V$. Moreover, since $\mathscr{R}$ is topologically filtered, there exists $T \in \mathscr{R}$ such that $T(x) \subseteq R(x) \cap S(x)$. Hence, we can already infer that

$$T(x) \subseteq R(x) \cap S(x) \subseteq (A \cup B) \cap V = (A \cap V) \cup (B \cap V) = A \cup \emptyset = A.$$

Therefore, if $x \in V$, then $x \in A^\circ$.

A quite similar argument shows that if $x \in V^c$, then $x \in B^\circ$. Therefore, in both cases, we have $x \in A^\circ \cup B^\circ$. Thus, the inclusion $(A \cup B)^\circ \subseteq A^\circ \cup B^\circ$ is also true.

Remark 42 More difficult conditions for the dual equality $(A \cap B)^- = A^- \cap B^-$ to hold were given by Gottschalk [34] and Jung and Nam [39, 40].

Concerning the latter problem, we shall only mention here the following

Theorem 44 *If $\mathscr{R}$ is a nonvoid, reflexive relator on X such that*

$$\mathrm{cl}_{\mathscr{R}}(A \cap B) = \mathrm{cl}_{\mathscr{R}}(A) \cap \mathrm{cl}_{\mathscr{R}}(B)$$

for all $A, B \subseteq X$, then $\mathscr{R}^{\wedge\infty} = \mathscr{P}(X^2)^\infty$.

Proof For this, it is enough to prove only that $\mathscr{T}_{\mathscr{R}} = \mathscr{P}(X)$. Namely, in this case we have $\mathscr{T}_{\mathscr{R}} = \mathscr{T}_{\Delta_X}$. Hence, by using Theorem 25 and the corresponding definitions, we can already infer that $\mathscr{R}^{\wedge\infty} = \{\Delta_X\}^{\wedge\infty} = \{\Delta_X\}^{*\infty} = \mathscr{P}(X^2)^\infty$.

To prove the equality $\mathscr{T}_{\mathscr{R}} = \mathscr{P}(X)$, note that if this not true, then there exists $A \subseteq X$ such that $A \notin \mathscr{T}_{\mathscr{R}}$, and thus $B = A^c \notin \mathscr{F}_{\mathscr{R}}$. Therefore, $B^- \not\subseteq B$, and thus there exists $x \in B^-$ such that $x \notin B$. Hence, by using the assumptions of the theorem, we can arrive at the contradiction that $x \in \{x\}^- \cap B^- = \big(\{x\} \cap B\big)^- = \emptyset^- = \emptyset$.

14 A Few Basic Facts on Quasi-Filtered Relators

Since $R \subseteq R^\infty$ for every relation R on X, in addition to Definition 13, we may also naturally introduce the following

Definition 14 A relator $\mathscr{R}$ on X is called

(1) *quasi-uniformly filtered* if for any $R, S \in \mathscr{R}$ there exists $T \in \mathscr{R}$ such that $T \subseteq R^\infty \cap S^\infty$;
(2) *quasi-proximally filtered* if for any $A \subseteq X$ and $R, S \in \mathscr{R}$ there exists $T \in \mathscr{R}$ such that $T[A] \subseteq R^\infty[A] \cap S^\infty[A]$;

(3) *quasi-topologically filtered* if for any $x \in X$ and $R, S \in \mathscr{R}^{\wedge}$ there exists $T \in \mathscr{R}$ such that $T(x) \subseteq R^{\infty}(x) \cap S^{\infty}(x)$.

Remark 43 Analogously to Remark 41, the above quasi-filteredness properties can also be reformulated in some more concise forms.

For instance, we can see that $\mathscr{R}$ is quasi-topologically filtered if and only if $\mathscr{R}^{\wedge\infty} \wedge \mathscr{R}^{\wedge\infty} \subseteq \mathscr{R}^{\wedge}$, $(\mathscr{R}^{\wedge\infty} \wedge \mathscr{R}^{\wedge\infty})^{\wedge\infty} = \mathscr{R}^{\wedge\infty}$ or $\mathscr{R}^{\wedge\infty} \wedge \mathscr{R}^{\wedge\infty} = \mathscr{R}^{\wedge\infty}$.

However, it is now more important to note that, by using some former results, we can also prove the following two theorems which show the appropriateness of the above quasi-proximal and quasi-topological filteredness properties.

Theorem 45 *For any relator $\mathscr{R}$ on X, the following assertions are equivalent:*

(1) $\mathscr{R}$ is a quasi-proximally filtered;
(2) $\tau_{\mathscr{R}}$ is closed under binary unions;
(3) $\tau_{\mathscr{R}}$ is closed under binary intersections.

Theorem 46 *For any relator $\mathscr{R}$ on X, the following assertions are equivalent:*

(1) $\mathscr{R}$ is a quasi-topologically filtered;
(2) $\mathscr{F}_{\mathscr{R}}$ is closed under binary unions;
(3) $\mathscr{T}_{\mathscr{R}}$ is closed under binary intersections.

Remark 44 In this respect it is also worth mentioning that if $\mathscr{R}$ is a relator on X to Y, then the family $\mathscr{E}_{\mathscr{R}}$ is closed under binary intersections if and only if $\mathscr{R}$ is *quasi-directed* in the sense that for any $x, y \in X$ and $R, S \in \mathscr{R}$ we have $R(x) \cap S(y) \in \mathscr{E}_{\mathscr{R}}$.

From the above two theorems, by using Corollaries 14 and 15, we can derive

Corollary 16 *If $\mathscr{R}$ is a proximally (topologically) filtered relator on X, then $\mathscr{R}$ is also quasi-proximally (quasi-topologically) filtered.*

Now, by using Theorem 45, we can also easily prove the following

Theorem 47 *If $\mathscr{R}$ is a quasi-proximally filtered, proximal relator on X, then $\mathscr{R}$ is proximally filtered.*

Proof Suppose that $A \subseteq X$ and $R, S \in \mathscr{R}$. Then, by Definition 12, there exist $U, V \in \tau_{\mathscr{R}}$ such that $A \subseteq U \subseteq R[A]$ and $A \subseteq V \subseteq S[A]$.

Moreover, by Theorem 45, we can state that $U \cap V \in \tau_{\mathscr{R}}$. Therefore, by the definition of $\tau_{\mathscr{R}}$, there exists $T \in \mathscr{R}$ such that $T[U \cap V] \subseteq U \cap V$. Hence, we can already see that $T[A] \subseteq T[U \cap V] \subseteq U \cap V \subseteq R[A] \cap S[A]$.

Moreover, by using Theorem 46, we can quite similarly prove the following

Theorem 48 *If $\mathscr{R}$ is a quasi-topologically filtered, topological relator on X, then $\mathscr{R}$ is topologically filtered.*

Remark 45 Our former Example 2 shows that even a quasi-proximally filtered, reflexive relator need not be topologically filtered.

Namely, if X and $\mathscr{R}$ are as in Example 2, then by the corresponding definitions it is clear that $\tau_{\mathscr{R}} = \{\emptyset, \{2,3\}, X\}$, and thus by Theorem 45 the relator $\mathscr{R}$ is quasi-proximally filtered.

15 Some Further Theorems on Topologically Filtered Relators

In our former papers [67, 68], by using the arguments of Kuratowski [45, pp. 39, 45], we have also proved the following two basic theorems.

Theorem 49 *If $\mathscr{R}$ is a topologically filtered relator on X to Y, then for any $A, B \subseteq Y$ we have*

$$\mathrm{cl}_{\mathscr{R}}(A) \setminus \mathrm{cl}_{\mathscr{R}}(B) = \mathrm{cl}_{\mathscr{R}}(A \setminus B) \setminus \mathrm{cl}_{\mathscr{R}}(B).$$

Proof By using Theorem 42, we can see that

$$A^- \cup B^- = (A \cup B)^- = \big((A \setminus B) \cup B\big)^- = (A \setminus B)^- \cup B^-.$$

Hence, because of the identity $(U \cup V) \setminus V = U \setminus V$, the required equality follows.

Thus, in particular we can also state the following

Corollary 17 *If $\mathscr{R}$ is a topologically filtered relator on X to Y, then for any $A, B \subseteq Y$ we have* $\mathrm{cl}_{\mathscr{R}}(A) \setminus \mathrm{cl}_{\mathscr{R}}(B) \subseteq \mathrm{cl}_{\mathscr{R}}(A \setminus B)$.

This corollary already allows us to easily prove the following

Theorem 50 *If $\mathscr{R}$ is a topologically filtered relator on X, then for any $A \subseteq X$ and $U \in \mathscr{T}_{\mathscr{R}}$ we have*

$$\mathrm{cl}_{\mathscr{R}}(A) \cap U = \mathrm{cl}_{\mathscr{R}}(A \cap U) \cap U.$$

Proof By Definition 3 and Theorem 2, we have $U \subseteq U^\circ = U^{c-c}$. Hence, by using Corollary 17, we can infer that

$$A^- \cap U \subseteq A^- \cap U^{c-c} = A^- \setminus U^{c-} \subseteq (A \setminus U^c)^- = (A \cap U)^-.$$

Therefore, $A^- \cap U = A^- \cap (U \cap U) = (A^- \cap U) \cap U \subseteq (A \cap U)^- \cap U$.

Moreover, by using the increasingness of $-$, we can see that $(A \cap U)^- \subseteq A^-$, and thus $(A \cap U)^- \cap U \subseteq A^- \cap U$ is always true. Therefore, we actually have $A^- \cap U = (A \cap U)^- \cap U$.

This theorem can also be easily derived from its subsequent corollary which can also be easily proved directly, without using Corollary 17. (See [68].)

Corollary 18 *If $\mathscr{R}$ is a topologically filtered relator on X, then for any $A \subseteq X$ and $U \in \mathscr{T}_{\mathscr{R}}$ we have* $\mathrm{cl}_{\mathscr{R}}(A) \cap U \subseteq \mathrm{cl}_{\mathscr{R}}(A \cap U)$.

Remark 46 The importance of the closure space counterpart of Corollary 18 was also recognized Császár [13–18] and Sivagami [70] who assumed it as an axiom for an increasing set-to-set function γ.

Now, as a dual form of Theorem 50, it is also worth proving the following

Theorem 51 *If $\mathscr{R}$ is a topologically filtered relator on X, then for any $A \subseteq X$ and $V \in \mathscr{F}_{\mathscr{R}}$ we have*

$$\mathrm{int}_{\mathscr{R}}(A) \cup V = \mathrm{int}_{\mathscr{R}}(A \cup V) \cup V.$$

Proof By using Theorems 2 and 50, we can see that

$$A^{\circ} \cup V = A^{c-c} \cup V^{cc} = (A^{c-} \cap V^{c})^{c} = \left((A^{c} \cap V^{c})^{-} \cap V^{c}\right)^{c}$$
$$= \left((A \cup V)^{c-} \cap V^{c}\right)^{c} = (A \cup V)^{c-c} \cup V = (A \cup V)^{\circ} \cup V.$$

Thus, in particular we can also state the following

Corollary 19 *If $\mathscr{R}$ is a topologically filtered relator on X, then for any $A \subseteq X$ and $V \in \mathscr{F}_{\mathscr{R}}$ we have* $\mathrm{int}_{\mathscr{R}}(A \cup V) \subseteq \mathrm{int}_{\mathscr{R}}(A) \cup V$.

The importance of this corollary is apparent from the following

Theorem 52 *If $\mathscr{R}$ is a topologically filtered, quasi-topological relator on X, then for any $A, B \in \mathscr{N}_{\mathscr{R}}$ we have $A \cup B \in \mathscr{N}_{\mathscr{R}}$.*

Proof By using Theorem 42 and Corollary 19, we can see that

$$(A \cup B)^{-\circ} = (A^{-} \cup B^{-})^{\circ} \subseteq A^{-\circ} \cup B^{-} = \emptyset \cup B^{-} = B^{-}.$$

Hence, by using Theorem 31 and the definition of $\mathscr{T}_{\mathscr{R}}$, we can infer that

$$(A \cup B)^{-\circ} \subseteq (A \cup B)^{-\circ\circ} \subseteq B^{-\circ} = \emptyset.$$

Therefore, $(A \cup B)^{-\circ} = \emptyset$, and thus $A \cup B \in \mathscr{N}_{\mathscr{R}}$.

Now, by using this theorem, we can also easily establish the following

Corollary 20 *If $\mathscr{R}$ is a nonvoid, non-partial, topologically filtered, quasi-topological relator on X, then $\mathscr{N}_{\mathscr{R}}$ is an ideal on X.*

Proof By the definition of $\mathscr{N}_{\mathscr{R}}$ and the increasingness of $-$ and $\circ$, it is clear that $\mathscr{N}_{\mathscr{R}}$ is always descending. Moreover, since $\mathscr{R}$ is nonvoid and non-partial, we can also see that $\emptyset^{-\circ} = \emptyset^{\circ} = \emptyset$. Therefore, $\emptyset \in \mathscr{N}_{\mathscr{R}}$, and thus $\mathscr{N}_{\mathscr{R}} \neq \emptyset$. Furthermore, from Theorem 52, we know that $\mathscr{N}_{\mathscr{R}}$ is closed under pairwise unions.

Remark 47 Note that if $\mathscr{R}$ is a locally non-partial relator on X, then by Theorem 29 we have $X^{-\circ} = X$. Therefore, if $X \neq \emptyset$, then we can also state that $X \notin \mathscr{N}_{\mathscr{R}}$, and thus $\mathscr{N}_{\mathscr{R}} \neq \mathscr{P}(X)$.

While, if $\mathscr{R}$ is a quasi-topological relator on X and $A \in \mathscr{N}_{\mathscr{R}}$, then by using Theorem 31 and the increasingness of $\circ$ we can also see that $A^{--\circ} \subseteq A^{-\circ} = \emptyset$. Therefore, $A^{--\circ} = \emptyset$, and thus $A^{-} \in \mathscr{N}_{\mathscr{R}}$.

16 Some More Particular Theorems on Topologically Filtered Relators

By using Corollary 18, we can also easily prove the following

Theorem 53 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then for any $A \subseteq X$ and $U \in \mathscr{T}_{\mathscr{R}}$ we have*

$$\mathrm{cl}_{\mathscr{R}}(A \cap U) = \mathrm{cl}_{\mathscr{R}}\big(\mathrm{cl}_{\mathscr{R}}(A) \cap U\big).$$

Proof By Corollary 18 we have $A^{-} \cap B \subseteq (A \cap B)^{-}$. Hence, by using Theorems 32 and 31, we can infer that

$$(A^{-} \cap B)^{-} \subseteq (A \cap B)^{--} \subseteq (A \cap B)^{-}.$$

On the other hand, by Theorem 26, we have $A \subseteq A^{-}$, and thus also $A \cap B \subseteq A^{-} \cap B$. Hence, we can infer that $(A \cap B)^{-} \subseteq \big(A^{-} \cap B\big)^{-}$. Therefore, the corresponding equality is also true.

From this theorem, we can immediate derive

Corollary 21 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then for any $A \in \mathscr{D}_{\mathscr{R}}$ and $U \in \mathscr{T}_{\mathscr{R}}$ we have* $\mathrm{cl}_{\mathscr{R}}(U) = \mathrm{cl}_{\mathscr{R}}(A \cap U)$.

Proof By Definition 1 and Theorem 53, we evidently have

$$U^{-} = (X \cap U)^{-} = (A^{-} \cap U)^{-} = (A \cap U)^{-}.$$

Now, by modifying an argument of Levine [50], we can also prove

Theorem 54 *If $\mathscr{R}$ is a nonvoid, topological relator on X and $A \subseteq X$ such that* $\mathrm{cl}_{\mathscr{R}}(U) = \mathrm{cl}_{\mathscr{R}}(A \cap U)$ *for all $U \in \mathscr{T}_{\mathscr{R}}$, then $A \in \mathscr{D}_{\mathscr{R}}$.*

Proof Assume on the contrary that $A \notin \mathscr{D}_{\mathscr{R}}$. Then, by Definition 1, there exists $x \in X$ such that $x \notin A^{-}$. Thus, by Definition 1, there exists $R \in \mathscr{R}$ such that $A \cap R(x) = \emptyset$. Moreover, by Definition 11, there exists $U \in \mathscr{T}_{\mathscr{R}}$ such that $x \in U \subseteq R(x)$. Thus, in particular we also have $A \cap U = \emptyset$.

Hence, by using the assumptions of the theorem, we can already infer that $U^{-} = (A \cap U)^{-} = \emptyset^{-} = \emptyset$. Note that the latter equality already requires that $\mathscr{R} \neq \emptyset$.

On the other hand, from the inclusion $x \in U$, by using Theorems 32 and 26 and the increasingness of $-$, we can infer that $x \in \{x\}^{-} \subseteq U^{-}$, and thus $U^{-} \neq \emptyset$. This contradiction proves that $A \in \mathscr{D}_{\mathscr{R}}$.

Remark 48 If $\mathscr{R}$ is a nonvoid, reflexive relator on X and $A \subseteq X$ such that $\mathrm{cl}_{\mathscr{R}}\big(R(x)\big) = \mathrm{cl}_{\mathscr{R}}\big(A \cap R(x)\big)$ for all $x \in X$ and $R \in \mathscr{R}$, then we can even more easily prove that $A \in \mathscr{D}_{\mathscr{R}}$.

In addition to Theorem 54, we can also prove the following

Theorem 55 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then for any $U \in \mathscr{T}_{\mathscr{R}}$ we have*

$$\mathrm{res}_{\mathscr{R}}(U) \in \mathscr{F}_{\mathscr{R}} \setminus \mathscr{E}_{\mathscr{R}}.$$

Proof By Theorem 11, we have $U^{c} \in \mathscr{F}_{\mathscr{R}}$. Moreover, by Theorems 32 and 31, we have $U^{-} \in \mathscr{F}_{\mathscr{R}}$. Hence, by using the notation $U^{\dagger} = \mathrm{res}_{\mathscr{R}}(U)$, we can infer that

$$U^{\dagger} = U^{-} \setminus U = U^{-} \cap U^{c} \in \mathscr{F}_{\mathscr{R}}.$$

Moreover, by using Theorems 42, 2, 32, and 26, we can also see that

$$U^{\dagger\circ} = \big(U^{-} \setminus U\big)^{\circ} = U^{-\circ} \cap U^{c\circ} = U^{-\circ} \cap U^{-c} \subseteq U^{-} \cap U^{-c} = \emptyset,$$

and thus $U^{\dagger\circ} = \emptyset$. Therefore, $U^{\dagger} \notin \mathscr{E}_{\mathscr{R}}$, and thus $U^{\dagger} \in \mathscr{F}_{\mathscr{R}} \setminus \mathscr{E}_{\mathscr{R}}$.

Now, as an immediate consequence of this theorem, we can also state

Corollary 22 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then $\mathrm{res}_{\mathscr{R}}(U) \in \mathscr{N}_{\mathscr{R}}$ for all $U \in \mathscr{T}_{\mathscr{R}}$.*

Remark 49 Note that if $\mathscr{R}$ is a topological relator on X and $U \in \mathscr{T}_{\mathscr{R}}$, then by Definition 3 and Theorems 32 and 26 we have $U = U^{\circ}$. Therefore, under the notation $U^{\ddagger} = \mathrm{bnd}_{\mathscr{R}}(U)$, we have $U^{\dagger} = U^{-} \setminus U = U^{-} \setminus U^{\circ} = U^{\ddagger}$.

Moreover, it is also worth noticing that in Theorem 55 and Corollary 22, it is enough to assume only that $\mathscr{R}$ is a quasi-topologically filtered, topological relator on X. Namely, in this case, $\mathscr{R}$ is already topologically filtered by Theorem 48.

17 Some Generalized Topologically Open Sets

Notation 1 In the sequel, we shall always assume that X is a set and $\mathscr{R}$ is a relator on X.

Moreover, to shorten the subsequent proofs, we shall again use the notations

$$A^- = \mathrm{cl}_{\mathscr{R}}(A), \qquad A^\circ = \mathrm{int}_{\mathscr{R}}(A) \qquad \text{and} \qquad A^\dagger = \mathrm{res}_{\mathscr{R}}(A).$$

In our first joint paper [67], motivated by the corresponding definitions on generalized open subsets of topological spaces mentioned in the Motivations, we have introduced the following

Definition 15 For a subset A of the relator space $X(\mathscr{R})$, we write

(1) $A \in \mathscr{T}^s_{\mathscr{R}}$ if $A \subseteq \mathrm{cl}_{\mathscr{R}}\big(\mathrm{int}_{\mathscr{R}}(A)\big)$;
(2) $A \in \mathscr{T}^p_{\mathscr{R}}$ if $A \subseteq \mathrm{int}_{\mathscr{R}}\big(\mathrm{cl}_{\mathscr{R}}(A)\big)$;
(3) $A \in \mathscr{T}^\alpha_{\mathscr{R}}$ if $A \subseteq \mathrm{int}_{\mathscr{R}}\big(\mathrm{cl}_{\mathscr{R}}\big(\mathrm{int}_{\mathscr{R}}(A)\big)\big)$;
(4) $A \in \mathscr{T}^\beta_{\mathscr{R}}$ if $A \subseteq \mathrm{cl}_{\mathscr{R}}\big(\mathrm{int}_{\mathscr{R}}\big(\mathrm{cl}_{\mathscr{R}}(A)\big)\big)$;
(5) $A \in \mathscr{T}^a_{\mathscr{R}}$ if $A \subseteq \mathrm{cl}_{\mathscr{R}}\big(\mathrm{int}_{\mathscr{R}}(A)\big) \cap \mathrm{int}_{\mathscr{R}}\big(\mathrm{cl}_{\mathscr{R}}(A)\big)$;
(6) $A \in \mathscr{T}^b_{\mathscr{R}}$ if $A \subseteq \mathrm{cl}_{\mathscr{R}}\big(\mathrm{int}_{\mathscr{R}}(A)\big) \cup \mathrm{int}_{\mathscr{R}}\big(\mathrm{cl}_{\mathscr{R}}(A)\big)$.

And, the members of the above families are called the *topologically semi-open, preopen*, α*-open*, β*-open*, *a-open and b-open subsets* of the relator space $X(\mathscr{R})$, respectively.

Remark 50 Note that, for instance, $\circ -$ is always an increasing operation on $\mathscr{P}(X)$. Moreover, if $\mathscr{R}$ is nonvoid and non-partial, then $\emptyset^{\circ -} = \emptyset$ and $X^{\circ -} = X$.

Therefore, by [76, Theorem 9.4], there exists a nonvoid and non-partial relator $\mathscr{S}$ on X such that $A^{\circ -} = \mathrm{int}_{\mathscr{S}}(A)$ for all $A \subseteq X$, and thus $\mathscr{T}^s_{\mathscr{R}} = \mathscr{T}_{\mathscr{S}}$.

However, this fact can be used to establish only that if $\mathscr{R}$ is a nonvoid and non-partial relator on X, then $\mathscr{T}^s_{\mathscr{R}}$ is a generalized topology on X [85].

In [67], by using the enormous literature on generalized open sets in topological spaces and their generalizations, we have, for instance, proved the following theorems.

Theorem 56 *We have (1)* $\mathscr{T}^a_{\mathscr{R}} = \mathscr{T}^s_{\mathscr{R}} \cap \mathscr{T}^p_{\mathscr{R}}$; *(2)* $\mathscr{T}^s_{\mathscr{R}} \cup \mathscr{T}^p_{\mathscr{R}} \subseteq \mathscr{T}^b_{\mathscr{R}}$.

Theorem 57 *If $\mathscr{R}$ is a reflexive relator on X, then (1)* $\mathscr{T}^\alpha_{\mathscr{R}} \subseteq \mathscr{T}^a_{\mathscr{R}}$; *(2)* $\mathscr{T}^b_{\mathscr{R}} \subseteq \mathscr{T}^\beta_{\mathscr{R}}$.
(3) $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}^\kappa_{\mathscr{R}}$ *for all* $\kappa = s, p, \alpha, \beta, a$ *and* b.

Remark 51 Note that, by the former inclusions, it is enough to prove the inclusion $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}^\kappa_{\mathscr{R}}$ only for $\kappa = \alpha$.

Theorem 58 *If $\mathscr{R}$ is a topological relator on X, then* $\mathscr{T}^\alpha_{\mathscr{R}} = \mathscr{T}^a_{\mathscr{R}}$.

Theorem 59 *If $\mathscr{R}$ is a topological relator on X, then for any $A \subseteq X$, the following assertions are equivalent:*
(1) $A \in \mathscr{T}^{s}_{\mathscr{R}}$; *(2)* $\mathrm{cl}_{\mathscr{R}}(A) \subseteq \mathrm{cl}_{\mathscr{R}}\big(\mathrm{int}_{\mathscr{R}}(A)\big)$; *(3)* $\mathrm{cl}_{\mathscr{R}}(A) = \mathrm{cl}_{\mathscr{R}}\big(\mathrm{int}_{\mathscr{R}}(A)\big)$.

Theorem 60 *If $\mathscr{R}$ is a topological relator on X, then for any $A \subseteq X$ the following assertions are equivalent:*

(1) $A \in \mathscr{T}^{s}_{\mathscr{R}}$;
(2) there exists $V \in \mathscr{T}_{\mathscr{R}}$ *such that* $V \subseteq A$ *and* $\mathrm{cl}_{\mathscr{R}}(A) = \mathrm{cl}_{\mathscr{R}}(V)$;
(3) there exist $V \in \mathscr{T}_{\mathscr{R}}$ *and* $B \subseteq X$ *such that* $A = V \cup B$ *and* $B \subseteq \mathrm{res}_{\mathscr{R}}(V)$.

Corollary 23 *If $\mathscr{R}$ is a topologically filtered, topological relator on X and $A \in \mathscr{T}^{s}_{\mathscr{R}}$, then there exist $V \in \mathscr{T}_{\mathscr{R}}$ and $B \in \mathscr{N}_{\mathscr{R}}$ such that $A = V \cup B$ and $V \cap B = \emptyset$.*

Theorem 61 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then for any $A \subseteq X$ the following assertions are equivalent:*

(1) $A \in \mathscr{T}^{p}_{\mathscr{R}}$;
(2) there exist $V \in \mathscr{T}_{\mathscr{R}}$ *and* $B \in \mathscr{D}_{\mathscr{R}}$ *such that* $A = V \cap B$;
(3) there exists $V \in \mathscr{T}_{\mathscr{R}}$ *such that* $A \subseteq V$ *and* $\mathrm{cl}_{\mathscr{R}}(A) = \mathrm{cl}_{\mathscr{R}}(V)$.

Theorem 62 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then for any $A \subseteq X$ the following assertions are equivalent:*

(1) $A \in \mathscr{T}^{\alpha}_{\mathscr{R}}$;
(2) there exist $V \in \mathscr{T}_{\mathscr{R}}$ *and* $B \in \mathscr{N}_{\mathscr{R}}$ *such that* $A = V \setminus B$;
(3) there exist $V \in \mathscr{T}_{\mathscr{R}}$ *and* $B \subseteq \mathrm{res}_{\mathscr{R}}\big(\mathrm{int}_{\mathscr{R}}(A)\big)$ *such that* $A = V \setminus B$.

Theorem 63 *If $\mathscr{R}$ is a topological relator on X, then for any $A \subseteq X$ the following assertions are equivalent: (1)* $A \in \mathscr{T}^{\beta}_{\mathscr{R}}$; *(2)* $\mathrm{cl}_{\mathscr{R}}(A) \in \mathscr{T}^{s}_{\mathscr{R}}$;

(3) there exists $V \in \mathscr{T}_{\mathscr{R}}$ *such that* $\mathrm{cl}_{\mathscr{R}}(A) = \mathrm{cl}_{\mathscr{R}}(V)$.

Remark 52 Hence, by using Theorems 59 and 60, we can derive some further characterizations of the family $\mathscr{T}^{\beta}_{\mathscr{R}}$.

On the other hand, from Theorem 56, by using Theorems 59, 60, and 61, we can derive several characterizations of the family $\mathscr{T}^{a}_{\mathscr{R}}$.

Theorem 64 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then for any $A \subseteq X$ the following assertions are equivalent:*

(1) $A \in \mathscr{T}^{b}_{\mathscr{R}}$;
(2) there exist $B \in \mathscr{T}^{s}_{\mathscr{R}}$ *and* $C \in \mathscr{T}^{p}_{\mathscr{R}}$ *such that* $A = B \cup C$.

18 Some Further Theorems on Generalized Topologically Open Sets

In our second joint paper [68], by using the enormous literature on generalized open sets in topological spaces and their generalizations, we have, for instance, proved the following theorems.

Theorem 65 *The families $\mathscr{T}_{\mathscr{R}}^{\kappa}$, with $\kappa = s, p, \alpha, \beta, a$ and b, are closed under arbitrary unions.*

Remark 53 Thus, in particular we have $\emptyset \in \mathscr{T}_{\mathscr{R}}^{\kappa}$ for all $\kappa = s, p, \alpha, \beta, a$ and b.

Theorem 66 *The following assertions are equivalent: (1) $X \in \mathscr{T}_{\mathscr{R}}^{s}$; (2) $X \in \mathscr{T}_{\mathscr{R}}^{\beta}$; (3) $\mathscr{R}$ is non-partial.*

Theorem 67 *The following assertions are equivalent: (1) $X \in \mathscr{T}_{\mathscr{R}}^{p}$; (2) $X \in \mathscr{T}_{\mathscr{R}}^{\alpha}$; (3) $\mathscr{R}$ is locally non-partial.*

Corollary 24 *The following assertions are equivalent: (1) $X \in \mathscr{T}_{\mathscr{R}}^{a}$; (2) $\mathscr{R}$ is non-partial and locally non-partial.*

Remark 54 If $\mathscr{R}$ is either non-partial or locally non-partial, then by Theorems 56, 66, and 67, we can also state that $X \in \mathscr{T}_{\mathscr{R}}^{b}$.

Theorem 68 *If $\mathscr{R}$ is a nonvoid, non-partial relator on X, then $\mathscr{T}_{\mathscr{R}}^{\kappa}$ is a generalized topology on X for all $\kappa = s, p, \alpha, \beta, a$ and b.*

Theorem 69 *If $\mathscr{R}$ is a topologically filtered relator on X and $U \in \mathscr{T}_{\mathscr{R}}$, then $U \cap A \in \mathscr{T}_{\mathscr{R}}^{\kappa}$ for all $A \in \mathscr{T}_{\mathscr{R}}^{\kappa}$ with $\kappa = s, p, \alpha, \beta, a$ and b.*

Theorem 70 *If $\mathscr{R}$ is a topologically filtered, topological relator on X and $A \in \mathscr{T}_{\mathscr{R}}^{\alpha}$, then $A \cap B \in \mathscr{T}_{\mathscr{R}}^{\kappa}$ for all $A \in \mathscr{T}_{\mathscr{R}}^{\kappa}$ with $\kappa = s, p, \alpha, \beta, a$ and b.*

Corollary 25 *If $\mathscr{R}$ is a nonvoid, topologically filtered, topological relator on X, then $\mathscr{T}_{\mathscr{R}}^{\alpha}$ is a topology on X.*

In [68], by using an argument of Njåstad [62, p. 962], we have also proved

Theorem 71 *If $\mathscr{R}$ is a nonvoid, topologically filtered, topological relator on X and $A \subseteq X$ such that $A \cap B \in \mathscr{T}_{\mathscr{R}}^{s}$ for all $B \in \mathscr{T}_{\mathscr{R}}^{s}$, then $A \in \mathscr{T}_{\mathscr{R}}^{\alpha}$.*

Moreover, to introduce the corresponding generalized topologically closed sets, we have also used the following

Definition 16 For any $\kappa = s, p, \alpha, \beta, a$ and b, we define

$$\mathscr{F}_{\mathscr{R}}^{\kappa} = \left\{ A \subseteq X : \quad A^{c} \in \mathscr{T}_{\mathscr{R}}^{\kappa} \right\}.$$

Thus, by using Theorem 2, we could easily prove the following

Theorem 72 *For any $A \subseteq X$, we have*

(1) $A \in \mathscr{F}^{s}_{\mathscr{R}}$ *if and only if* $\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right) \subseteq A$;
(2) $A \in \mathscr{F}^{p}_{\mathscr{R}}$ *if and only if* $\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}(A)\right) \subseteq A$;
(3) $A \in \mathscr{F}^{\alpha}_{\mathscr{R}}$ *if and only if* $\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right)\right) \subseteq A$;
(4) $A \in \mathscr{F}^{\beta}_{\mathscr{R}}$ *if and only if* $\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}(A)\right)\right) \subseteq A$;
(5) $A \in \mathscr{F}^{a}_{\mathscr{R}}$ *if and only if* $\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right) \cup \operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}(A)\right) \subseteq A$;
(6) $A \in \mathscr{F}^{b}_{\mathscr{R}}$ *if and only if* $\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right) \cap \operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}(A)\right) \subseteq A$.

In [67] and [68], following Kuratowski [44], we have also introduced

Definition 17 A subset A of the relator space $X(\mathscr{R})$ is called *topologically regular open* if

$$A = \operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right).$$

And, the family of all such subsets of $X(\mathscr{R})$ is denoted by $\mathscr{T}^{r}_{\mathscr{R}}$.

Thus, by using several papers on regular open sets in topological spaces, we have, for instance, proved the following theorems.

Theorem 73 *If $\mathscr{R}$ is a quasi-topological relator on X, then $\mathscr{T}^{r}_{\mathscr{R}} \subseteq \mathscr{T}_{\mathscr{R}}$.*

Theorem 74 *If $\mathscr{R}$ is a reflexive relator on X, then $\mathscr{T}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}} \subseteq \mathscr{T}^{r}_{\mathscr{R}}$.*

Theorem 75 *We always have $\mathscr{T}^{r}_{\mathscr{R}} = \mathscr{T}^{p}_{\mathscr{R}} \cap \mathscr{F}^{s}_{\mathscr{R}}$.*

Theorem 76 *If $\mathscr{R}$ is a topological relator on X, then (1) $\mathscr{T}^{r}_{\mathscr{R}} = \mathscr{T}_{\mathscr{R}} \cap \mathscr{F}^{s}_{\mathscr{R}}$;*

(2) $\mathscr{T}^{r}_{\mathscr{R}} = \mathscr{T}^{\alpha}_{\mathscr{R}} \cap \mathscr{F}^{\beta}_{\mathscr{R}}$.

Theorem 77 *If $\mathscr{R}$ is a topological relator on X, then (1) $\operatorname{cl}_{\mathscr{R}}(A) \in \mathscr{T}^{r}_{\mathscr{R}}$ for all $A \in \mathscr{T}^{s}_{\mathscr{R}}$;*
(2) $\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right) \in \mathscr{T}^{r}_{\mathscr{R}}$ for all $A \subseteq X$.

Theorem 78 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then $\mathscr{T}^{r}_{\mathscr{R}}$ is closed under pairwise intersections.*

The following example shows that the counterparts of Theorems 65 and 69 fail to hold for the family $\mathscr{T}^{r}_{\mathscr{R}}$.

Example 3 If $X = \mathbb{R}$ and

$$R_n = \left\{(x, y) \in X^2 : \quad d(x, y) < n^{-1}\right\}$$

for all $n \in \mathbb{N}$, then $\mathscr{R} = \{ R_n : n \in \mathbb{N} \}$ is a properly filtered, strongly topological, tolerance relator on X such that, for the sets

$$A =]0, 1[, \qquad B =]1, 2[, \qquad C =]0, 2[\qquad \text{and} \qquad U = \{1\}^c ,$$

we have $A \cup B = C \cap U \notin \mathscr{T}^r_{\mathscr{R}}$ despite that $A, B, C \in \mathscr{T}^r_{\mathscr{R}}$ and $U \in \mathscr{T}_{\mathscr{R}}$.

19 A Further Family of Generalized Topologically Open Sets

The following definitions and some of the forthcoming theorems have been mainly suggested to us by Levine [48], Chattopadhyay and Bandyopadhyay [8] and Császár [15, 16].

In [48], a subset A of a topological space is said to have *property* Q if $A^{-\circ} = A^{\circ-}$. While, in [8] and [15], the set A is called a *δ-set* and a *quasi-open set*, respectively, if $A^{-\circ} \subseteq A^{\circ-}$.

Definition 18 If Φ and Ψ are relation on $\mathscr{P}(X)$ to X, then for any $A \subseteq X$ we shall write

$$A \in \mathscr{A}(\Phi, \Psi) \qquad \text{if} \qquad \Psi\big(\Phi(A)\big) \subseteq \Phi\big(\Psi(A)\big).$$

Remark 55 Thus, for any $A \subseteq X$, we have

$$A \in \mathscr{A}(\Phi, \Psi) \cap \mathscr{A}(\Psi, \Phi) \qquad \text{if and only if} \qquad \Psi\big(\Phi(A)\big) = \Phi\big(\Psi(A)\big).$$

That is, the associated set-valued functions commute at the set A.

In the sequel, we shall also restrict ourselves to the particular case when $\Phi = \mathrm{cl}_{\mathscr{R}}$ and $\Psi = \mathrm{int}_{\mathscr{R}}$. And, to shorten the subsequent statements and proofs, we shall use the following

Definition 19 In particular, we define

$$\mathscr{A}_{\mathscr{R}} = \mathscr{A}_{\mathscr{R}}(-, \circ) = \mathscr{A}\left(\mathrm{cl}_{\mathscr{R}}, \mathrm{int}_{\mathscr{R}} \right).$$

Remark 56 Thus, for any $A \subseteq X$, we have

$$A \in \mathscr{A}_{\mathscr{R}} \qquad \text{if and only if} \qquad A^{-\circ} \subseteq A^{\circ-}.$$

By using the latter property, we can easily prove the following theorems which could, to some extent, be generalized to the more general case in Definition 18.

Theorem 79 *We have* $\mathscr{N}_{\mathscr{R}} \subseteq \mathscr{A}_{\mathscr{R}}$.

Proof If $A \in \mathscr{N}_{\mathscr{R}}$, then by Definition 3, we have $A^{-\circ} = \emptyset$. Thus, $A^{-\circ} \subseteq A^{\circ-}$ trivially holds. Hence, by Remark 56, we can see that $A \in \mathscr{A}_{\mathscr{R}}$. Therefore, the required inclusion is true.

Remark 57 If $\mathscr{R}$ is a nonvoid, reflexive relator on X, then we can also state that $\mathscr{N}_{\mathscr{R}} \subseteq \mathscr{A}_{\mathscr{R}}(-, \circ) \cap \mathscr{A}_{\mathscr{R}}(\circ, -)$.

Namely, if $\mathscr{R}$ is reflexive, then by Theorem 26 we have $A \subseteq A^{-}$, and thus also $A^{\circ} \subseteq A^{-\circ}$. Therefore, if $A \in \mathscr{N}_{\mathscr{R}}$, i.e., $A^{-\circ} = \emptyset$, then we also have $A^{\circ} = \emptyset$. Moreover, if $\mathscr{R}$ is nonvoid, then we also have $\emptyset^{-} = \emptyset$. Therefore, in this particular case, we actually have $A^{\circ-} = \emptyset^{-} = \emptyset = A^{-\circ}$, and thus also $A \in \mathscr{A}_{\mathscr{R}}(\circ, -)$.

Theorem 80 *If $\mathscr{R}$ is a nonvoid relator on X, then*

(1) $\mathscr{A}_{\mathscr{R}} \setminus \mathscr{N}_{\mathscr{R}} \subseteq \mathscr{E}_{\mathscr{R}}$; *(2)* $A \in \mathscr{A}_{\mathscr{R}} \cap \mathscr{D}_{\mathscr{R}}$ *implies* $\operatorname{int}_{\mathscr{R}}(A) \in \mathscr{D}_{\mathscr{R}}$.

Proof If $A \in \mathscr{A}_{\mathscr{R}} \setminus \mathscr{N}_{\mathscr{R}}$, then $A \in \mathscr{A}_{\mathscr{R}}$ and $A \notin \mathscr{N}_{\mathscr{R}}$. Therefore, by Remark 56 and Definition 3, we have $A^{-\circ} \subseteq A^{\circ-}$ and $A^{-\circ} \neq \emptyset$, and thus $A^{\circ-} \neq \emptyset$. Hence, since $\mathscr{R} \neq \emptyset$, and thus $\emptyset^{-} = \emptyset$, we can infer that $A^{\circ} \neq \emptyset$. Therefore, $A \in \mathscr{E}_{\mathscr{R}}$, and thus assertion (1) is true.

While, if $A \in \mathscr{A}_{\mathscr{R}} \cap \mathscr{D}_{\mathscr{R}}$, then we have $A \in \mathscr{A}_{\mathscr{R}}$ and $A \in \mathscr{D}_{\mathscr{R}}$. Therefore, by Remark 56 and Definition 1, we have $A^{-\circ} \subseteq A^{\circ-}$ and $A^{-} = X$. Hence, since $\mathscr{R} \neq \emptyset$, and thus $X^{\circ} = X$, we can infer that $X = X^{\circ} = A^{-\circ} \subseteq A^{\circ-}$, and thus $A^{\circ-} = X$. Therefore, $A^{\circ} \in \mathscr{D}_{\mathscr{R}}$, and thus assertion (2) is also true.

Remark 58 Assertions (1) and (2) can be reformulated in the forms that:

(1) $\mathscr{A}_{\mathscr{R}} \setminus \mathscr{E}_{\mathscr{R}} \subseteq \mathscr{N}_{\mathscr{R}}$; (2) $\left(\mathscr{A}_{\mathscr{R}} \cap \mathscr{D}_{\mathscr{R}}\right)^{\circ} \subseteq \mathscr{D}_{\mathscr{R}}$.

Theorem 81 *If $\mathscr{R}$ is a reflexive relator on X, then* $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{A}_{\mathscr{R}}$.

Proof If $A \in \mathscr{T}_{\mathscr{R}}$, then by Definition 3 and Theorem 26 we have $A \subseteq A^{\circ}$ and $A^{\circ} \subseteq A$, and thus also $A = A^{\circ}$.

Now, by Theorem 26, we can also see that $A^{-\circ} \subseteq A^{-} = A^{\circ-}$, and thus by Remark 56 we also have $A \in \mathscr{A}_{\mathscr{R}}$.

Theorem 82 *If $\mathscr{R}$ is a topological relator on X, then* $\mathscr{T}^{s}_{\mathscr{R}} \subseteq \mathscr{A}_{\mathscr{R}}$.

Proof If $A \in \mathscr{T}^{s}_{\mathscr{R}}$, then by Definition 15, we have $A \subseteq A^{\circ-}$. Hence, by using the increasingness of the operation $-\circ$ and Theorems 32, 31, and 26, we can see that

$$A^{-\circ} \subseteq A^{\circ--\circ} \subseteq A^{\circ-\circ} \subseteq A^{\circ-}.$$

Therefore, by Remark 56, we also have $A \in \mathscr{A}_{\mathscr{R}}$.

Theorem 83 *We have*

(1) $\mathscr{A}_{\mathscr{R}} \cap \mathscr{T}^{p}_{\mathscr{R}} \subseteq \mathscr{T}^{s}_{\mathscr{R}}$;
(2) $\mathscr{A}_{\mathscr{R}} \cap \mathscr{T}^{p}_{\mathscr{R}} = \mathscr{A}_{\mathscr{R}} \cap \mathscr{T}^{a}_{\mathscr{R}}$; *(3)* $\mathscr{A}_{\mathscr{R}} \cap \mathscr{T}^{s}_{\mathscr{R}} = \mathscr{A}_{\mathscr{R}} \cap \mathscr{T}^{b}_{\mathscr{R}}$.

Proof If $A \in \mathscr{A}_{\mathscr{R}} \cap \mathscr{T}^{p}_{\mathscr{R}}$, then $A \in \mathscr{T}^{p}_{\mathscr{R}}$ and $A \in \mathscr{A}_{\mathscr{R}}$. Hence, by using Definition 15 and Remark 56, we can see that $A \subseteq A^{-\circ} \subseteq A^{\circ-}$. Thus, by Remark 56, we also have $A \in \mathscr{T}^{s}_{\mathscr{R}}$. Therefore, inclusion (1) is true.

By Remark 56, it is clear that, for any $A \subseteq X$, the following assertions are equivalent:

(a) $A \in \mathscr{A}_{\mathscr{R}}$; (b) $A^{-\circ} = A^{-\circ} \cap A^{\circ-}$; (c) $A^{\circ-} = A^{-\circ} \cup A^{\circ-}$.

Hence, by Definition 15, it is clear that, for any $A \in \mathscr{A}_{\mathscr{R}}$, we have

$$A \in \mathscr{T}^{p}_{\mathscr{R}} \iff A \subseteq A^{-\circ} \iff A \subseteq A^{-\circ} \cap A^{\circ-} \iff A \in \mathscr{T}^{a}_{\mathscr{R}},$$

and quite similarly

$$A \in \mathscr{T}^{s}_{\mathscr{R}} \iff A \subseteq A^{\circ-} \iff A \subseteq A^{-\circ} \cup A^{\circ-} \iff A \in \mathscr{T}^{b}_{\mathscr{R}}.$$

Therefore, equalities (2) and (3) are also true.

Remark 59 By using Theorem 56 and assertion (3), we can also see that

$$\mathscr{A}_{\mathscr{R}} \cap \mathscr{T}^{p}_{\mathscr{R}} \subseteq \mathscr{A}_{\mathscr{R}} \cap (\mathscr{T}^{s}_{\mathscr{R}} \cup \mathscr{T}^{p}_{\mathscr{R}}) \subseteq \mathscr{A}_{\mathscr{R}} \cap \mathscr{T}^{b}_{\mathscr{R}} = \mathscr{A}_{\mathscr{R}} \cap \mathscr{T}^{s}_{\mathscr{R}} \subseteq \mathscr{T}^{s}_{\mathscr{R}}.$$

Therefore, assertion (3) is somewhat stronger than assertion (1).

20 Some Further Theorems on the Family $\mathscr{A}_{\mathscr{R}}$

The following theorem shows that, in contrast to $\mathscr{T}_{\mathscr{R}}$ and $\mathscr{T}^{\kappa}_{\mathscr{R}}$, the family $\mathscr{A}_{\mathscr{R}}$ is closed under elementwise complementation.

Theorem 84 *For any $A \in \mathscr{A}_{\mathscr{R}}$, we have $A^{c} \in \mathscr{A}_{\mathscr{R}}$.*

Proof By using Remark 56 and Theorem 2, we can see that

$$\begin{aligned} A \in \mathscr{A}_{\mathscr{R}} &\implies A^{-\circ} \subseteq A^{\circ-} \implies A^{-c-c} \subseteq A^{\circ c \circ c} \\ &\implies A^{\circ c \circ} \subseteq A^{-c-} \implies A^{c-\circ} \subseteq A^{c\circ-} \implies A^{c} \in \mathscr{A}_{\mathscr{R}}. \end{aligned}$$

Remark 60 By using the practical, but ambiguous notation

$$\mathscr{A}^{c} = [\mathscr{A}]^{c} = \mathscr{C}_{X}[\mathscr{A}] = \{\mathscr{C}_{X}(A) : \quad A \in \mathscr{A}\},$$

the above theorem can be reformulated in the instructive form that $\mathscr{A}^{c}_{\mathscr{R}} = \mathscr{A}_{\mathscr{R}}$.

From our former Theorems 79 and 81–83, by using Theorem 84, we can immediately derive the following four theorems.

Theorem 85 *We have*

$$\mathscr{N}_{\mathscr{R}} \cup \mathscr{N}_{\mathscr{R}}^{c} \subseteq \mathscr{A}_{\mathscr{R}}.$$

Theorem 86 *If $\mathscr{R}$ is a reflexive relator on X, then*

$$\mathscr{T}_{\mathscr{R}} \cup \mathscr{F}_{\mathscr{R}} \subseteq \mathscr{A}_{\mathscr{R}}.$$

Remark 61 If $\mathscr{R}$ is a reflexive relator on X, then we can also state that $\mathscr{T}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}} \subseteq \mathscr{A}_{\mathscr{R}}(-, \circ) \cap \mathscr{A}_{\mathscr{R}}(\circ, -)$.

Namely, if $A \in \mathscr{T}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}}$, then by using Definition 3 and Theorem 26 we can see that

$$A^{\circ -} = A^{-} = A = A^{\circ} = A^{-\circ}.$$

Therefore, $A \in \mathscr{A}_{\mathscr{R}}(\circ, -)$, and thus $\mathscr{T}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}} \subseteq \mathscr{A}_{\mathscr{R}}(\circ, -)$ also holds.

Theorem 87 *If $\mathscr{R}$ is a topological relator on X, then*

$$\mathscr{T}_{\mathscr{R}}^{s} \cup \mathscr{F}_{\mathscr{R}}^{s} \subseteq \mathscr{A}_{\mathscr{R}}.$$

Theorem 88 *We have*

(1) $\mathscr{A}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}}^{p} \subseteq \mathscr{F}_{\mathscr{R}}^{s}$;
(2) $\mathscr{A}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}}^{p} = \mathscr{A}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}}^{a}$; *(3)* $\mathscr{A}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}}^{s} = \mathscr{A}_{\mathscr{R}} \cap \mathscr{F}_{\mathscr{R}}^{b}$.

Now, analogously to Theorem 84, we can also prove the following

Theorem 89 *If $\mathscr{R}$ is a topological relator on X, then for any $A \in \mathscr{A}_{\mathscr{R}}$ we have*

(1) $\mathrm{cl}_{\mathscr{R}}(A) \in \mathscr{A}_{\mathscr{R}}$; *(2)* $\mathrm{int}_{\mathscr{R}}(A) \in \mathscr{A}_{\mathscr{R}}$.

Proof If $A \in \mathscr{A}_{\mathscr{R}}$, then by Remark 56 we have

$$A^{-\circ} \subseteq A^{\circ -}.$$

Moreover, by Theorems 32, 31 and 26, we also have

$$A^{--} \subseteq A^{-} \qquad \text{and} \qquad A \subseteq A^{-}.$$

Hence, by using the increasingness of the operations $\circ$ and $\circ -$, we can infer that

$$A^{--\circ} \subseteq A^{-\circ} \subseteq A^{\circ -} \subseteq A^{-\circ -}.$$

Therefore, by the definition of $\mathscr{A}_{\mathscr{R}}$, we also have $A^{-} \in \mathscr{A}_{\mathscr{R}}$.

Assertion (2) can now be easily derived from (1) by using Theorems 2 and 84. Namely, by Theorem 2, we have $A^\circ = A^{c-c}$.

Remark 62 Assertions (1) and (2) can be reformulated in the form that

$$\mathscr{A}_{\mathscr{R}}^{-} \cup \mathscr{A}_{\mathscr{R}}^{\circ} \subseteq \mathscr{A}_{\mathscr{R}}.$$

The following theorem shows that, analogously to the families $\mathscr{T}_{\mathscr{R}}^{\alpha}$ and $\mathscr{T}_{\mathscr{R}}^{r}$, the family $\mathscr{A}_{\mathscr{R}}$ may also be frequently closed under pairwise intersections.

Theorem 90 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then for any $A, B \in \mathscr{A}_{\mathscr{R}}$ we have $A \cap B \in \mathscr{A}_{\mathscr{R}}$.*

Proof By Remark 56, we have

$$A^{-\circ} \subseteq A^{\circ-} \qquad \text{and} \qquad B^{-\circ} \subseteq B^{\circ-}.$$

Hence, by using the increasingness of the operations—and $\circ$, Theorems 32 and 31, Corollary 18 and Theorem 42, we can see that

$$\begin{aligned}(A \cap B)^{-\circ} &\subseteq A^{-\circ} \cap B^{-\circ} \subseteq A^{\circ-} \cap B^{-\circ} \subseteq \left(A^{\circ} \cap B^{-\circ}\right)^{-} \\ &\subseteq \left(A^{\circ} \cap B^{\circ-}\right)^{-} \subseteq \left(A^{\circ} \cap B^{\circ}\right)^{--} \subseteq \left(A^{\circ} \cap B^{\circ}\right)^{-} = (A \cap B)^{\circ-}.\end{aligned}$$

Thus, by Remark 56, we also have $A \cap B \in \mathscr{A}_{\mathscr{R}}$.

Now, as a useful consequence of Theorems 84 and 90, we can also state

Corollary 26 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then for any $A, B \in \mathscr{A}_{\mathscr{R}}$ we have (1) $A \setminus B \in \mathscr{A}_{\mathscr{R}}$; (2) $A \cup B \in \mathscr{A}_{\mathscr{R}}$; (3) $A \Delta B \in \mathscr{A}_{\mathscr{R}}$.*

Proof To prove these, recall that

$$A \setminus B = A \cap B^{c}, \qquad A \cup B = (A^{c} \cap B^{c})^{c}, \qquad A \Delta B = (A \setminus B) \cup (B \setminus A).$$

Remark 63 Thus, if $\mathscr{R}$ is a topologically filtered, topological relator on X, then $\mathscr{A}_{\mathscr{R}}$ forms an *algebra* of subsets of X.

The countable infinite set $\mathbb{Q}$ of all rational numbers, used also by Levine [48] and Császár [15], shows that $\mathscr{A}_{\mathscr{R}}$ is not in general a *σ-algebra*.

Example 4 If X and $\mathscr{R}$ are as in Example 3 and $A_r = \{r\}$ for all $r \in \mathbb{Q}$, then $A_r \in \mathscr{A}_{\mathscr{R}}$ for all $r \in \mathbb{Q}$ such that $\bigcup_{r \in \mathbb{Q}} A_r = \mathbb{Q} \notin \mathscr{A}_{\mathscr{R}}$.

To check this, note that $A_r \in \mathscr{F}_{\mathscr{R}}$, and thus by Theorem 86 we also have $A^r \in \mathscr{A}_{\mathscr{R}}$ for all $r \in \mathbb{Q}$. However, $\mathbb{Q}^{-\circ} = \mathbb{R}^{\circ} = \mathbb{R}$, but $\mathbb{Q}^{\circ-} = \emptyset^{-} = \emptyset$. Therefore, by Remark 56, we have $\mathbb{Q} \notin \mathscr{A}_{\mathscr{R}}$.

21 Characterizations of the Family $\mathscr{A}_{\mathscr{R}}$

By using the corresponding definitions, one can easily establish the following

Theorem 91 *For any $A \subseteq X$, the following assertions are equivalent:*

(1) $A \in \mathscr{A}_{\mathscr{R}}$;
(2) if $x \in X$ *and there exists* $R \in \mathscr{R}$ *such that for any* $y \in R(x)$ *and* $S \in \mathscr{R}$ *we have* $S(y) \cap A \neq \emptyset$, *then for every* $U \in \mathscr{R}$ *there exist* $z \in U(x)$ *and* $V \in \mathscr{R}$ *such that* $V(z) \subseteq A$.

However, this intrinsic characterization of the family $\mathscr{A}_{\mathscr{R}}$ cannot, certainly, be used to prove our former and subsequent theorems on $\mathscr{A}_{\mathscr{R}}$.

Therefore, analogously to the first part of [15, Theorem 2.3] of Császár, we shall also prove a more particular, but much deeper characterization theorem.

Theorem 92 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then for any $A \subseteq X$ the following assertions are equivalent:*

(1) $A \in \mathscr{A}_{\mathscr{R}}$;
(2) there exist $V \in \mathscr{T}_{\mathscr{R}}$ *and* $B \in \mathscr{N}_{\mathscr{R}}$ *such that* $A = V \cup B$;
(3) there exist $V \in \mathscr{T}^{s}_{\mathscr{R}}$ *and* $B \in \mathscr{N}_{\mathscr{R}}$ *such that* $A = V \cup B$.

Proof If (1) holds, then by Remark 56 we have $A^{-\circ} \subseteq A^{\circ-}$. Moreover, by Theorems 32 and 26, we can also note that $A^{\circ} \subseteq A$, and thus

$$A = A^{\circ} \cup (A \setminus A^{\circ}).$$

Therefore, by taking

$$V = A^{\circ} \qquad \text{and} \qquad B = A \setminus A^{\circ}$$

we can state that $A = V \cup B$ with $V \cap B = \emptyset$. Moreover, by Theorem 32 and 31, we can also note that $V = A^{\circ} \in \mathscr{T}_{\mathscr{R}}$, and thus $V \subseteq V^{\circ}$.

On the other hand, by using Corollary 2, Theorem 42 and the inclusions

$$V^{-} = A^{\circ-} \supseteq A^{-\circ} \supseteq B^{-\circ},$$

we can see that

$$\begin{aligned} B = A \setminus V \implies V \cap B = \emptyset \implies V \cap B^{-} = \emptyset \\ \implies (V \cap B^{-})^{\circ} = \emptyset \implies V^{\circ} \cap B^{-\circ} = \emptyset \implies V \cap B^{-\circ} = \emptyset \\ \implies V^{-} \cap B^{-\circ} = \emptyset \implies B^{-\circ} = \emptyset \implies B \in \mathscr{N}_{\mathscr{R}}. \end{aligned}$$

Therefore, (2) also holds.

Now, since the implication (2) $\Longrightarrow$ (3) is immediate from the inclusion $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}^s_{\mathscr{R}}$, we need only show that (3) also implies (1).

For this, note that if (3) holds, then by Theorems 82 and 79 we also have $V \in \mathscr{A}_{\mathscr{R}}$ and $B \in \mathscr{A}_{\mathscr{R}}$. Hence, by Corollary 26, we can already infer that $A = V \cup B \in \mathscr{A}_{\mathscr{R}}$.

Remark 64 From the inclusions $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}^{\alpha}_{\mathscr{R}} \subseteq \mathscr{T}^{a}_{\mathscr{R}} \subseteq \mathscr{T}^s_{\mathscr{R}}$, valid for any reflexive relator $\mathscr{R}$ on X, it clear that in the above theorem we may write $\mathscr{T}^{\alpha}_{\mathscr{R}}$ and $\mathscr{T}^{a}_{\mathscr{R}}$ instead of $\mathscr{T}^s_{\mathscr{R}}$.

Moreover, if $\mathscr{R}$ is a reflexive relator on X, then from assertion (2) of Theorem 92, by using the inclusion $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}^p_{\mathscr{R}}$, we can also infer that there exist $V \in \mathscr{T}^p_{\mathscr{R}}$ and $B \in \mathscr{N}_{\mathscr{R}}$ such that $A = V \cup B$.

However, this statement cannot certainly be used to obtain assertion (1) of Theorem 92. Namely, if $\mathscr{R}$ is as in Theorem 92 and $V \in \mathscr{T}^p_{\mathscr{R}}$, then by Theorem 61 we can only state that there exist $W \in \mathscr{T}_{\mathscr{R}}$ and $C \in \mathscr{D}_{\mathscr{R}}$ such that $V = W \cap C$. And, C need not belong to $\mathscr{A}_{\mathscr{R}}$.

Namely, if $\mathscr{R}$ is a nonvoid relator on X and in addition to $C \in \mathscr{D}_{\mathscr{R}}$ we also have $C \in \mathscr{A}_{\mathscr{R}}$, then $X = X^{\circ} = C^{-\circ} \subseteq C^{\circ -}$. Therefore, $X = C^{\circ -}$, and thus $C^{\circ} \in \mathscr{D}_{\mathscr{R}}$ which usually does not hold.

Now, analogously to [9, Lemma 1.3] of Chattopadhyay and Roy, we can also prove the following

Theorem 93 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then the following assertions are equivalent:*

(1) $\mathscr{A}_{\mathscr{R}} = \mathscr{T}_{\mathscr{R}}$; (2) $\mathscr{A}_{\mathscr{R}} \subseteq \mathscr{T}_{\mathscr{R}}$;
(3) $\mathscr{T}_{\mathscr{R}} = \mathscr{F}_{\mathscr{R}}$; (4) $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{F}_{\mathscr{R}}$; (5) $\mathscr{F}_{\mathscr{R}} \subseteq \mathscr{T}_{\mathscr{R}}$.

Proof By Theorem 81, it is clear that assertions (1) and (2) are equivalent even if $\mathscr{R}$ is assumed only to be reflexive. Moreover, by using Theorem 11, we can easily see that assertions (3), (4) and (5) are always equivalent. Therefore, it is sufficient to prove only the equivalence of (2) and (4).

For this, note that if (4) does not hold, then there exists $V \in \mathscr{T}_{\mathscr{R}}$ such that $V \notin \mathscr{F}_{\mathscr{R}}$. Hence, by using Theorem 11, we can infer that $V^c \in \mathscr{F}_{\mathscr{R}}$ and $V^c \notin \mathscr{T}_{\mathscr{R}}$. Now, by Theorem 86, we can also state that $V^c \in \mathscr{A}_{\mathscr{R}}$. Thus, assertion (2) does not also hold. Therefore, (2) implies (4) even if $\mathscr{R}$ is assumed to be only reflexive.

On the other hand, if $A \in \mathscr{A}_{\mathscr{R}}$, then by Theorem 92 there exist $V \in \mathscr{T}_{\mathscr{R}}$ and $B \in \mathscr{N}_{\mathscr{R}}$ such that

$$A = V \cup B.$$

Moreover, if (4) holds, then (5) also holds. Therefore, by using Theorems 32, 26 and 31 , we can also see that

$$B \subseteq B^{-} = B^{-\circ} = \emptyset,$$

and thus $B = \emptyset$. Hence, we can already infer that $A = V \cup B = V \cup \emptyset = V \in \mathscr{T}_{\mathscr{R}}$. Therefore, (2) also holds.

22 Some Further Characterizations of the Family $\mathscr{A}_{\mathscr{R}}$

Analogously to [15, Theorem 2.3] of Császár and [29, Theorem 2.2] of Ganster, Reilly and Vamanamurthy, we can also prove the following two theorems.

Theorem 94 *If $\mathscr{R}$ is a nonvoid, topologically filtered, topological relator on X, then for any $A \subseteq X$ the following assertions are equivalent:*

(1) $A \in \mathscr{A}_{\mathscr{R}}$;
(2) there exists $V \in \mathscr{T}_{\mathscr{R}}$ such that $A \Delta V \in \mathscr{N}_{\mathscr{R}}$;
(3) there exists $V \in \mathscr{T}_{\mathscr{R}}$ such that $A \setminus V \in \mathscr{N}_{\mathscr{R}}$ and $V \setminus A \in \mathscr{N}_{\mathscr{R}}$.

Proof If (1) holds, then by Theorem 92 there exist $V \in \mathscr{T}_{\mathscr{R}}$ and $B \in \mathscr{N}_{\mathscr{R}}$ such that $A = V \cup B$. Hence, we can infer that

$$A \setminus V \subseteq B \qquad \text{and} \qquad V \setminus A = \emptyset.$$

Now, since $B \in \mathscr{N}_{\mathscr{R}}$ and $\mathscr{N}_{\mathscr{R}}$ is descending, we can see that $A \setminus V \in \mathscr{N}_{\mathscr{R}}$. Moreover, since $\mathscr{R}$ is nonvoid and reflexive, we can also see that $\emptyset^{-\circ} = \emptyset^{\circ} = \emptyset$, and thus $V \setminus A = \emptyset \in \mathscr{N}_{\mathscr{R}}$. Therefore, (3) also holds.

Conversely, if (3) holds, then by defining

$$W = V \setminus (V \setminus A)^{-}$$

and using Theorems 32, 32, and 11 and Corollary 15 we can see that

$$W = V \cap (V \cap A^c)^{-c} \in \mathscr{T}_{\mathscr{R}}.$$

Moreover, by using Theorems 2, 32, and 26, we can also see that

$$W = V \cap (V \cap A^c)^{c\circ} = V \cap (V^c \cup A)^{\circ} \subseteq V \cap (V^c \cup A) = V \cap A \subseteq A,$$

and thus $A = W \cup (A \setminus W)$. Furthermore, we can also note that

$$A \setminus W = A \cap W^c = A \cap \big(V \cap (V \cap A^c)^{-c}\big)^c = A \cap \big(V^c \cup (V \cap A^c)^{-}\big) =$$
$$(A \cap V^c) \cup \big(A \cap (V \cap A^c)^{-}\big) = (A \setminus V) \cup \big(A \cap (V \setminus A)^{-}\big) \subseteq (A \setminus V) \cup (V \setminus A)^{-}.$$

Hence, by using the assumptions $A \setminus V \in \mathscr{N}_{\mathscr{R}}$ and $V \setminus A \in \mathscr{N}_{\mathscr{R}}$ and three basic properties of $\mathscr{N}_{\mathscr{R}}$ established at the end of Section 15, we can already infer that $A \setminus W \in \mathscr{N}_{\mathscr{R}}$. Therefore, by Theorem 92, assertion (1) also holds.

Now, to complete the prove it remains only to note that, because of the equality $A \Delta V = (A \setminus V) \cup (V \setminus A)$ and two basic properties of $\mathscr{N}_{\mathscr{R}}$, assertions (2) and (3) are also equivalent.

Remark 65 Note that if, $V \in \mathscr{T}_{\mathscr{R}}^{\kappa}$, with $\kappa = s, \alpha$ or a, then by Theorems 31, 11, and 69, we have $W = V \cap (V \cap A^c)^{-c} \in \mathscr{T}_{\mathscr{R}}^{\kappa}$. Therefore, by Theorem 92 and Remark 64, in Theorem 94 we can write $\mathscr{T}_{\mathscr{R}}^{\kappa}$ instead of $\mathscr{T}_{\mathscr{R}}$.

However, it is now more important to note that, by a general definition of Kuratowski [45, p. 11], Theorem 94 can be briefly reformulated by stating that $A \in \mathscr{A}_{\mathscr{R}}$ if and only if A is *congruent to a member* V *of* $\mathscr{T}_{\mathscr{R}}$ *modulo* $\mathscr{N}_{\mathscr{R}}$.

Thus, under the assumptions of Theorem 94, by [45, p. 12], we can also state that $A \in \mathscr{A}_{\mathscr{R}}$ if and only if there exist $V \in \mathscr{T}_{\mathscr{R}}$ and $B, C \in \mathscr{N}_{\mathscr{R}}$ such that $A = (V \setminus B) \cup C$. However, by Theorem 62, we have $V \setminus B \in \mathscr{T}_{\mathscr{R}}^{\alpha}$.

Theorem 95 *If $\mathscr{R}$ is a nonvoid, topologically filtered, topological relator on X, then for any $A \subseteq X$ the following assertions are equivalent:*

(1) $A \in \mathscr{A}_{\mathscr{R}}$;
(2) there exist $V \in \mathscr{T}_{\mathscr{R}}^{s}$ *and* $W \in \mathscr{F}_{\mathscr{R}}^{s}$ *such that* $A = V \cap W$;
(3) there exist $V \in \mathscr{T}_{\mathscr{R}}^{\alpha}$ *and* $W \in \mathscr{F}_{\mathscr{R}}^{\alpha}$ *such that* $A = V \cap W$.

Proof Because of the inclusion $\mathscr{T}_{\mathscr{R}}^{\alpha} \subseteq \mathscr{T}_{\mathscr{R}}^{s}$ and its consequence $\mathscr{F}_{\mathscr{R}}^{\alpha} \subseteq \mathscr{F}_{\mathscr{R}}^{s}$, it is clear that (3) implies (2).

Moreover, if (2) holds, then by Definition 15 and Theorem 72, we have

$$V \subseteq V^{\circ -} \qquad \text{and} \qquad W^{-\circ} \subseteq W.$$

Hence, by using the inclusions $A \subseteq V$ and $A \subseteq W$, and the corresponding properties of the operations—and $\circ$, we can infer that

$$A^{-\circ} \subseteq V^{-\circ} \subseteq V^{\circ --\circ} = V^{\circ -\circ} \subseteq V^{\circ -}$$

and

$$A^{-\circ} \subseteq W^{-\circ} \subseteq W, \qquad \text{andthus} \qquad A^{-\circ} = A^{-\circ\circ} \subseteq W^{\circ}.$$

Now, by using Corollary 18 and Theorem 42, we can already see that

$$A^{-\circ} \subseteq V^{\circ -} \cap W^{\circ} \subseteq (V^{\circ} \cap W^{\circ})^{-} = (V \cap W)^{\circ -} = A^{\circ -}.$$

Thus, by Remark 56, assertion (1) also hols.

Finally, if (1) holds, then by Theorem 84 we also have $A^c \in \mathscr{A}_{\mathscr{R}}$. Thus, by Theorem 92, there exist $B \in \mathscr{N}_{\mathscr{R}}$ and $V \in \mathscr{T}_{\mathscr{R}}$ such that $A^c = B \cap V$, and thus

$$A = B^c \cap V^c.$$

Moreover, by using Theorems 62 and 57, we can see that $B^c \in \mathscr{T}^{\alpha}_{\mathscr{R}}$ and $V \in \mathscr{T}^{\alpha}_{\mathscr{R}}$, and hence also $V^c \in \mathscr{F}^{\alpha}_{\mathscr{R}}$. Therefore, assertion (3) also holds.

Remark 66 If $\mathscr{R}$ is nonvoid, topologically filtered topological relator on X, then necessary and sufficient conditions for a subset A of X, in order that $A^{-\circ} = A^{\circ-}$, i.e., $A \in \mathscr{A}(-, \circ) \cap \mathscr{A}(\circ, -)$ could hold, can be derived from the results of Levine [48], Choda and Matoba [10] and Chapman [7].

23 Lower and Upper Nearness Relations for Sets

If $A, B \subseteq X$, then A is said to be *near* to B, with respect to the relator $\mathscr{R}$, if $A \in \mathrm{Cl}_{\mathscr{R}}(B)$. That is, $R[A] \cap B \neq \emptyset$ for all $R \in \mathscr{R}$.

Note that if $A \cap \mathrm{cl}_{\mathscr{R}}(B) \neq \emptyset$, then by Theorem 3 we also have $A \in \mathrm{Cl}_{\mathscr{R}}(B)$. Therefore, A is near to B with respect to $\mathscr{R}$.

Moreover, if in particular $\mathscr{R}$ is nonvoid and topologically fine, then by Theorem 19 the converse of the above implication is also true.

Therefore, for some particular purposes, the relation $\mathrm{cl}_{\mathscr{R}}$ can also be naturally used to define a reasonable nearness relation for sets.

However, our main motivation for the subsequent definition has mainly come from the famous observations of Levine [49] and Corson and Michael [11] that for a subset A of a topological space $X(\mathscr{T})$ we can state that;

(1) A is preopen if and only if there exists $V \in \mathscr{T}$ such that $A \subseteq V \subseteq A^{-}$;
(2) A is semi-open if and only if there exists $V \in \mathscr{T}$ such that $V \subseteq A \subseteq V^{-}$.

Now, by an observation of Andrijević [3], we can also state that A is β-open if and only if there exists a preopen subset V of $X(\mathscr{T})$ such that $A \subseteq V \subseteq A^{-}$. Therefore, for the approximations of sets, we may also use generalized open sets.

The above observations strongly suggest that, analogously to the definition of the big lower and upper bound relations $\mathrm{Lb}_{\mathscr{R}}$ and $\mathrm{Ub}_{\mathscr{R}}$ mentioned in Remark 13, we may also naturally introduce the following

Definition 20 If $A, B \subseteq X$, such that

$$A \subseteq B \subseteq \mathrm{cl}_{\mathscr{R}}(A),$$

then we shall write

$$A \in \mathrm{Ln}_{\mathscr{R}}(B) \qquad \text{and} \qquad B \in \mathrm{Un}_{\mathscr{R}}(A).$$

Moreover, in this case, we shall say that A is near to B from below and B is near to A from above with respect to the relator $\mathscr{R}$.

Remark 67 Thus, the relations $\mathrm{Ln}_{\mathscr{R}}$ and $\mathrm{Un}_{\mathscr{R}}$ are also not independent of each other. Namely, by the above definition, evidently have $\mathrm{Un}_{\mathscr{R}} = \mathrm{Ln}^{-1}_{\mathscr{R}}$.

Therefore, in the sequel, we shall only establish some basic properties of the relation $\mathrm{Ln}_{\mathscr{R}}$.

Theorem 96 *For any $A, B \subseteq X$, the following assertions are equivalent:*

(1) $A \in \mathrm{Ln}_{\mathscr{R}}(B)$;
(2) $A \subseteq B$ *and for each* $x \in B$ *and* $R \in \mathscr{R}$ *we have* $R(x) \cap A \neq \emptyset$.

Proof By the corresponding definitions, we have

$$A \in \mathrm{Ln}_{\mathscr{R}}(B) \iff A \subseteq B \subseteq A^{-}$$

and

$$B \subseteq A^{-} \iff \forall x \in B : \ x \in A^{-} \iff \forall x \in B : \ \forall R \in \mathscr{R} : \ R(x) \cap A \neq \emptyset.$$

Therefore, assertions (1) and (2) are also equivalent.

Remark 68 In principle, this intrinsic characterization of the relation $\mathrm{Ln}_{\mathscr{R}}$ can be used to establish every possible properties of $\mathrm{Ln}_{\mathscr{R}}$.

However, in most of the forthcoming proofs, it will be more convenient to use Definition 20 and the basic properties of the relation $\mathrm{cl}_{\mathscr{R}}$.

Theorem 97 *For any $A, B \subseteq X$, the following assertions are equivalent:*

(1) $A^{c} \in \mathrm{Ln}_{\mathscr{R}}\left(B^{c}\right)$; *(2)* $\mathrm{int}_{\mathscr{R}}(A) \subseteq B \subseteq A$.

Proof If (1) holds, then by Definition 20 we have $A^{c} \subseteq B^{c} \subseteq A^{c-}$. Hence, by using Theorem 2, we can infer that

$$A^{\circ} = A^{c-c} \subseteq B \subseteq A.$$

Thus, (2) also holds. The converse implication (2) $\Longrightarrow$ (1) can be proved quite similarly, by reversing the above argument.

Remark 69 Analogously to Remark 2, assertion (1) can also be expressed in the more instructive form that $A \in (\mathrm{Ln}_{\mathscr{R}} \circ \mathscr{C}_X)^{c}(B)$.

Theorem 98 *If $\mathscr{R}$ is a topological relator on X, then for any $A, B \subseteq X$, the following assertions are equivalent:*

(1) $A \in \mathrm{Ln}_{\mathscr{R}}(B)$; *(2)* $A \subseteq B$ *and* $\mathrm{cl}_{\mathscr{R}}(A) = \mathrm{cl}_{\mathscr{R}}(B)$.

Proof If (2) holds, then by Theorems 32 and 26, we can see that

$$A \subseteq B \subseteq B^{-} = A^{-}.$$

Therefore, by Definition 20, assertion (1) also holds even if $\mathscr{R}$ is assumed to be only reflexive.

Conversely if (1) holds, then by Definition 20 we have $A \subseteq B \subseteq A^{-}$. Hence, by using the increasingness of – and Theorem 31, we can infer that

$$A^{-} \subseteq B^{-} \subseteq A^{--} \subseteq A^{-}.$$

Therefore, $A^{-} = B^{-}$, and thus (2) also holds even if $\mathscr{R}$ is assumed to be only quasi-topological.

24 Some Set-Theoretic Properties of the Relation $\mathrm{Ln}_{\mathscr{R}}$

To prove the following six theorems, it will be convenient to use Theorem 96.

Theorem 99 *For any $A \subseteq X$, the following assertions are equivalent: (1) $A \in \mathrm{Ln}_{\mathscr{R}}(\emptyset)$; (2) $A = \emptyset$.*

Proof By Theorem 96, assertion (1) holds if and only if

(a) $A \subseteq \emptyset$; (b) $\forall\, x \in \emptyset : \quad \forall\, R \in \mathscr{R} : \quad R(x) \cap A \neq \emptyset$.
Hence, since (a) is equivalent to (2), and (b) automatically holds, it is clear that (1) and (2) are equivalent.

Theorem 100 *For any $B \subseteq X$, the following assertions are equivalent:*

(1) $\emptyset \in \mathrm{Ln}_{\mathscr{R}}(B)$; (2) either $B = \emptyset$ or $\mathscr{R} = \emptyset$.

Proof By Theorem 96, assertion (1) holds if and only if

(a) $\emptyset \subseteq B$; (b) $\forall\, x \in B : \quad \forall\, R \in \mathscr{R} : \quad R(x) \cap \emptyset \neq \emptyset$.
Hence, since (a) automatically holds, and (b) can only hold if and only if (2) holds, it is clear that (1) and (2) are equivalent.

Theorem 101 *For any $a \in X$ and $B \subseteq X$, the following assertions are equivalent:*

(1) $\{a\} \in \mathrm{Ln}_{\mathscr{R}}(B)$;
(2) $a \in B$ and for any $x \in B$ and $R \in \mathscr{R}$ we have $a \in R(x)$.

Proof By Theorem 96, assertion (1) holds if and only if

(a) $\{a\} \subseteq B$; (b) $\forall\, x \in B : \quad \forall\, R \in \mathscr{R} : \quad R(x) \cap \{a\} \neq \emptyset$.
Hence, since (a) is equivalent to $a \in B$, and $R(x) \cap \{a\} \neq \emptyset$ is equivalent to $a \in R(x)$, it is clear that (1) and (2) are equivalent.

Theorem 102 *For any $A \subseteq X$ and $b \in X$, the following assertions are equivalent:*

(1) $A \in \mathrm{Ln}_{\mathscr{R}}(\{b\})$;
(2) either $A = \emptyset$ and $\mathscr{R} = \emptyset$, or $A = \{b\}$ and for any $R \in \mathscr{R}$ we have $b \in R(b)$.

Proof By Theorem 96, assertion (1) holds if and only if

(a) $A \subseteq \{b\}$; (b) $\forall\, x \in \{b\} : \quad \forall\, R \in \mathscr{R} : \quad R(x) \cap A \neq \emptyset$.
Hence, since (a) can hold if and only if either $A = \emptyset$ or $A = \{b\}$, and moreover $x \in \{b\}$ is equivalent to $x = b$, and $R(b) \cap A \neq \emptyset$ can hold if and only if $A = \{b\}$ and $b \in R(b)$, it is clear that (1) and (2) are equivalent.

Theorem 103 *For any $A \subseteq X$, the following assertions are equivalent:*

(1) $A \in \mathrm{Ln}_{\mathscr{R}}(X)$;
(2) for any $x \in X$ and $R \in \mathscr{R}$ we have $R(x) \cap A \neq \emptyset$.

Proof By Theorem 96, assertion (1) holds if and only if

(a) $A \subseteq X$; (b) $\forall\, x \in X : \quad \forall\, R \in \mathscr{R} : \quad R(x) \cap A \neq \emptyset$.
Hence, since (a) automatically holds, and (b) coincides with (2), it is clear that (1) and (2) are equivalent.

Theorem 104 *For any $B \subseteq X$, the following assertions are equivalent:*

(1) $X \in \mathrm{Ln}_{\mathscr{R}}(B)$; (2) $B = X$ and $\mathscr{R}$ is non-partial.

Proof By Theorem 96, assertion (1) holds if and only if

(a) $X \subseteq B$; (b) $\forall\, x \in B : \quad \forall\, R \in \mathscr{R} : \quad R(x) \cap X \neq \emptyset$.
Hence, since (a) is equivalent to $B = X$, and $R(x) \cap X = R(x)$, it is clear that now (b) means only that for any $x \in X$ and $R \in \mathscr{R}$ we have $R(x) \neq \emptyset$, i.e., $\mathscr{R}$ is non-partial. Therefore, (1) and (2) are equivalent.

To prove the following two theorems, we shall again use Definition 20 and some basic properties of the relation $\mathrm{cl}_{\mathscr{R}}$.

Theorem 105 *If $(A_i)_{i \in I}$ and $(B_i)_{i \in I}$ are families of subsets of X such that $A_i \in \mathrm{Ln}_{\mathscr{R}}(B_i)$ for all $i \in I$, then*

$$\bigcup_{i \in I} A_i \in \mathrm{Ln}_{\mathscr{R}}\Big(\bigcup_{i \in I} B_i\Big).$$

Proof By Definition 20, we have $A_i \subseteq B_i \subseteq A_i^-$ for all $i \in I$. Hence, by using the increasingness of the corresponding operations, we can infer that

$$\bigcup_{i \in I} A_i \subseteq \bigcup_{i \in I} B_i \subseteq \bigcup_{i \in I} A_i^- \subseteq \Big(\bigcup_{i \in I} A_i\Big)^-.$$

Therefore, by Definition 20, the required assertion is also true.

Theorem 106 *If $\mathscr{R}$ is a topologically filtered relator on X and $A, B \subseteq X$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then for any $V \in \mathscr{T}_{\mathscr{R}}$ we have*

$$A \cap V \in \mathrm{Ln}_{\mathscr{R}}(B \cap V).$$

Proof By Definition 20, we have $A \subseteq B \subseteq A^{-}$. Hence, by using Corollary 18, we can infer that

$$A \cap V \subseteq B \cap V \subseteq A^{-} \cap V \subseteq (A \cap V)^{-}.$$

Thus, by Definition 20, the required assertion is also true.

The next simple example shows that the condition $V \in \mathscr{T}_{\mathscr{R}}$ of the above theorem cannot be omitted or replaced with $V \in \mathscr{F}_{\mathscr{R}}$.

Example 5 If X and $\mathscr{R}$ are as in Example 3, then for the sets

$$A = [0,\ 1[\,, \qquad B = [0,\ 1] \qquad \text{and} \qquad V = \{1\},$$

we have $A \in \mathrm{Ln}_{\mathscr{R}}(B)$ and $A \cap V \notin \mathrm{Ln}_{\mathscr{R}}(B \cap V)$.

25 Some Topological Properties of the Relation $\mathbf{Ln}_{\mathscr{R}}$

Theorem 107 *If $A, B \subseteq X$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$ and Φ is an increasing relation on $\mathscr{P}(X)$ to X, then*

(1) $\Phi(A) \in \mathrm{Ln}_{\mathscr{R}}\big(\Phi(B)\big)$ *if* $A \in \mathscr{A}\big(\mathrm{cl}_{\mathscr{R}}, \Phi\big)$;
(2) $\Phi(B) \in \mathrm{Ln}_{\mathscr{R}}\big(\Phi\big(\mathrm{cl}_{\mathscr{R}}(A)\big)\big)$ *if* $B \in \mathscr{A}\big(\mathrm{cl}_{\mathscr{R}}, \Phi\big)$.

Proof By Definition 20 and the increasingness of $-$, we have

$$A \subseteq B \subseteq A^{-} \subseteq B^{-}.$$

Hence, by using the increasingness of Φ, we can infer that

$$\Phi(A) \subseteq \Phi(B) \subseteq \Phi(A^{-}) \subseteq \Phi(B^{-}).$$

Moreover, if $A \in \mathscr{A}\big(-,\ \Phi\big)$, then by Definition 18 we have

$$\Phi(A^{-}) \subseteq \Phi(A)^{-}.$$

Hence, we can that see that

$$\Phi(A) \subseteq \Phi(B) \subseteq \Phi(A)^{-}.$$

Thus, by Definition 20, we have $\Phi(A) \in \mathrm{Ln}_{\mathscr{R}}\big(\Phi(B)\big)$.

While, if $B \in \mathscr{A}\big(-,\ \Phi\big)$, then by Definition 18 we have

$$\Phi(B^{-}) \subseteq \Phi(B)^{-}.$$

Hence, we can see that

$$\Phi(B) \subseteq \Phi(A^{-}) \subseteq \Phi(B)^{-}$$

Thus, by Definition 20, we have $\Phi(B) \in \mathrm{Ln}_{\mathscr{R}}\left(\Phi(A^{-})\right)$.

Remark 70 Note that if in particular $\Phi(A) = A$ for all $A \subseteq X$, or $\Phi(A) = \mathrm{cl}_{\mathscr{R}}(A)$ for all $A \subseteq X$, then Φ is increasing, and by Definition 18 we have $A\left(\mathrm{cl}_{\mathscr{R}}, \Phi\right) = \mathscr{P}(X)$.

Therefore, as an immediate consequence of the above theorem, we can state

Corollary 27 *If $A, B \subseteq X$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$ then*

(1) $B \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right)$;
(2) $\mathrm{cl}_{\mathscr{R}}(A) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(B)\right)$; *(3)* $\mathrm{cl}_{\mathscr{R}}(B) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right)\right)$.

Remark 71 Note that if in particular $\Phi(A) = \mathrm{int}_{\mathscr{R}}(A)$ for all $A \subseteq X$, then Φ is increasing, and by Definition 19 we have $\mathscr{A}\left(\mathrm{cl}_{\mathscr{R}}, \Phi\right) = \mathscr{A}_{\mathscr{R}}$.

Therefore, as an immediate consequence of Theorem 107, we can also state

Corollary 28 *If $A, B \subseteq X$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then*

(1) $\mathrm{int}_{\mathscr{R}}(A) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right)$ *if* $A \in \mathscr{A}_{\mathscr{R}}$;
(2) $\mathrm{int}_{\mathscr{R}}(B) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right)\right)$ *if* $B \in \mathscr{A}_{\mathscr{R}}$.

From the above two corollaries, we can easily derive the following two corollaries.

Corollary 29 *If $A \in \mathscr{A}_{\mathscr{R}}$ and $B \subseteq X$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then*

(1) $\mathrm{int}_{\mathscr{R}}(B) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(A)\right)\right)$;
(2) $\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(A)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right)\right)$;
(3) $\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(A)\right)\right)\right)$.

Proof By Corollary 28, we have

$$A^{\circ} \in \mathrm{Ln}_{\mathscr{R}}(B^{\circ}).$$

Hence, by using Corollary 27, we can infer that
$B^{\circ} \in \mathrm{Ln}_{\mathscr{R}}(A^{\circ-})$, $A^{\circ-} \in \mathrm{Ln}_{\mathscr{R}}(B^{\circ-})$, $B^{\circ-} \in \mathrm{Ln}_{\mathscr{R}}(A^{\circ--})$.
Therefore, assertions (1), (2) and (3) are true.

Corollary 30 *If $A \subseteq X$ and $B \in \mathscr{A}_{\mathscr{R}}$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then*

(1) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right)\right)$;
(2) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(B)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}\left(cl_{\mathscr{R}}(A)\right)\right)\right)$;
(3) $\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right)\right)\right)$.

Proof By Corollary 28, we have

$$B^{\circ} \in \mathrm{Ln}_{\mathscr{R}}(A^{-\circ}).$$

Hence, by using Corollary 27, we can infer that
$A^{-\circ} \in \mathrm{Ln}_{\mathscr{R}}(B^{\circ-})$, $B^{\circ-} \in \mathrm{Ln}_{\mathscr{R}}(A^{-\circ-})$, $A^{-\circ-} \in \mathrm{Ln}_{\mathscr{R}}(B^{\circ--})$.

Therefore, assertions (1), (2) and (3) are true.

Remark 72 Note that if $\mathscr{R}$ is a topological relator on X, then by Theorems 32, 26, and 31, we have $A^{--} = A^{-}$ for all $A \subseteq X$.

Therefore, in this particular case,assertions (3) in Corollaries 27, 29, and 30 can be simplified by writing $\mathrm{cl}_{\mathscr{R}}$ instead of $\mathrm{cl}_{\mathscr{R}}\ \mathrm{cl}_{\mathscr{R}}$.

26 Some Further Topological Properties of the Relation $\mathrm{Ln}_{\mathscr{R}}$

From the above corollaries, by using the results of Section 20, we can easily derive several further topological properties of the relation $\mathrm{Ln}_{\mathscr{R}}$.

For instance, by Corollaries 28 and 29 and Theorem 85, we can at once state the following two theorems.

Theorem 108 *If $A \in \mathscr{N}_{\mathscr{R}} \cup \mathscr{N}_{\mathscr{R}}^{c}$ and $B \subseteq X$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then*

(1) $\mathrm{int}_{\mathscr{R}}(A) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right)$;
(2) $\mathrm{int}_{\mathscr{R}}(B) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(A)\right)\right)$;
(3) $\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(A)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right)\right)$;
(4) $\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(A)\right)\right)\right)$.

Theorem 109 *If $A \subseteq X$ and $B \in \mathscr{N}_{\mathscr{R}} \cup \mathscr{N}_{\mathscr{R}}^{c}$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then*

(1) $\mathrm{int}_{\mathscr{R}}(B) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right)\right)$;
(2) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right)\right)$;
(3) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(B)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right)\right)\right)$;
(4) $\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right)\right)\right)$.

Remark 73 Note that if in particular $\mathscr{R}$ is reflexive (resp. topological), then in the above two theorems, by Theorem 86 (resp. 87), we may write $\mathscr{T}_{\mathscr{R}} \cup \mathscr{F}_{\mathscr{R}}$ (resp. $\mathscr{T}_{\mathscr{R}}^{s} \cup \mathscr{F}_{\mathscr{R}}^{s}$) instead of $\mathscr{N}_{\mathscr{R}} \cup \mathscr{N}_{\mathscr{R}}^{c}$.

Moreover, if $\mathscr{R}$ is topological, then assertions (3) in the above two theorems can also be simplified by writing $\mathrm{cl}_{\mathscr{R}}$ instead of $\mathrm{cl}_{\mathscr{R}}\ \mathrm{cl}_{\mathscr{R}}$.

More importantly, concerning topological relators, we can also prove the following two theorems.

Theorem 110 *If $\mathscr{R}$ is a topological relator on X, $A \in \mathscr{A}$ and $B \subseteq X$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then*

(1) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(B)\right)\right)$;
(2) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(B)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(\mathrm{cl}_{\mathscr{R}}(A))\right)\right)$;

(3) $\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right) \in \operatorname{Ln}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(B)\right)\right)\right)$;
(4) $\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right)\right) \in \operatorname{Ln}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}(\operatorname{cl}_{\mathscr{R}}(B))\right)\right)$;
(5) $\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(B)\right)\right) \in \operatorname{Ln}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}(\operatorname{cl}_{\mathscr{R}}(A))\right)\right)$.

Proof By Corollary 27 and Theorem 89, we have

$$A^{-} \in \operatorname{Ln}_{\mathscr{R}}(B^{-}) \qquad \text{and} \qquad A^{-} \in \mathscr{A}_{\mathscr{R}}.$$

Hence, by using Corollary 28 and 29, we can infer that

$$A^{-\circ} \in \operatorname{Ln}_{\mathscr{R}}(B^{-\circ})$$

and

$$B^{-\circ} \in \operatorname{Ln}_{\mathscr{R}}(A^{-\circ-}); \qquad A^{-\circ-} \in \operatorname{Ln}_{\mathscr{R}}(B^{-\circ-}); \qquad B^{-\circ-} \in \operatorname{Ln}_{\mathscr{R}}(A^{-\circ-}).$$

Thus, assertions (1), (2), (4) and (5) are true.

Moreover, by Corollary 27, we now also have $B \in \operatorname{Ln}_{\mathscr{R}}(A^{-})$. Hence, by using Corollary 30, we can infer that $A^{-\circ} \in \operatorname{Ln}_{\mathscr{R}}(B^{-\circ-})$. Thus, assertion (3) is also true.

Theorem 111 *If $\mathscr{R}$ is a topological relator on X, $A \subseteq X$ and $B \in \mathscr{A}_{\mathscr{R}}$ such that $A \in \operatorname{Ln}_{\mathscr{R}}(B)$, then*

(1) $\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(B)\right) \in \operatorname{Ln}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right)\right)$;
(2) $\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(B)\right) \in \operatorname{Ln}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}(\operatorname{cl}_{\mathscr{R}}(A))\right)\right)$;
(3) $\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right) \in \operatorname{Ln}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(B)\right)\right)\right)$;
(4) $\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(A)\right)\right) \in \operatorname{Ln}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}(\operatorname{cl}_{\mathscr{R}}(B))\right)\right)$;
(5) $\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}(B)\right)\right) \in \operatorname{Ln}_{\mathscr{R}}\left(\operatorname{cl}_{\mathscr{R}}\left(\operatorname{int}_{\mathscr{R}}(\operatorname{cl}_{\mathscr{R}}(A))\right)\right)$.

Proof By Corollary 27 and Theorem 89, we have

$$A^{-} \in \operatorname{Ln}_{\mathscr{R}}(B^{-}) \qquad \text{and} \qquad B^{-} \in \mathscr{A}_{\mathscr{R}}.$$

Hence, by using Corollary 28 and 30, we can infer that

$$B^{-\circ} \in \operatorname{Ln}_{\mathscr{R}}(A^{-\circ})$$

and

$$A^{-\circ} \in \operatorname{Ln}_{\mathscr{R}}(B^{-\circ-}); \qquad B^{-\circ} \in \operatorname{Ln}_{\mathscr{R}}(A^{-\circ-}); \qquad A^{-\circ-} \in \operatorname{Ln}_{\mathscr{R}}(B^{-\circ-}).$$

Thus, assertions (1), (3), (2) and (4) are true.

Moreover, by Corollary 27, we now also have $B^{-} \in \operatorname{Ln}_{\mathscr{R}}(A^{-})$. Hence, by using Corollary 29, we can infer that $B^{-\circ-} \in \operatorname{Ln}_{\mathscr{R}}(A^{-\circ-})$. Thus, assertion (5) is also true.

27 Some Relation-Theoretic Properties of the Relation $\mathrm{Ln}_{\mathscr{R}}$

By using Definition 20, we can also easily establish the following two theorems.

Theorem 112 *If $\mathscr{R}$ is a reflexive relator on X, then $\mathrm{Ln}_{\mathscr{R}}$ is a reflexive relation on $\mathscr{P}(X)$.*

Proof Namely, if $A \subseteq X$, then we trivially have $A \subseteq A$. Moreover, by Theorem 26, we also have $A \subseteq A^{-}$. Thus, by Definition 20, we have $A \in \mathrm{Ln}_{\mathscr{R}}(A)$.

Remark 74 Actually, by the corresponding definitions and Theorem 26, the converse of the above theorem is also true.

For, if the relation $\mathrm{Ln}_{\mathscr{R}}$ is reflexive, then for any $A \subseteq X$ we have $A \in \mathrm{Ln}_{\mathscr{R}}(A)$, and thus $A \subseteq A^{-}$. Therefore, by Theorem 26, the relator $\mathscr{R}$ is reflexive.

Theorem 113 *If $\mathscr{R}$ is a quasi-topological relator on X, then $\mathrm{Ln}_{\mathscr{R}}$ is a transitive relation on $\mathscr{P}(X)$.*

Proof If $A, B, C \subseteq X$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$ and $B \in \mathrm{Ln}_{\mathscr{R}}(C)$, then by Definition 20 we have

$$A \subseteq B \subseteq A^{-} \qquad \text{and} \qquad B \subseteq C \subseteq B^{-}.$$

Moreover, by using Theorem 31 and the increasingness of $-$, we can see that

$$B^{-} \subseteq A^{--} \subseteq A^{-}.$$

Hence, we can already see that $A \subseteq C \subseteq A^{-}$, and thus $A \in \mathrm{Ln}_{\mathscr{R}}(C)$ also holds.

Thus, since the relation $\mathrm{Ln}_{\mathscr{R}}$ is always antisymmetric, we can also state

Corollary 31 *If $\mathscr{R}$ is a topological relator on X, then $\mathrm{Ln}_{\mathscr{R}}$ is a partial order relation on $\mathscr{P}(X)$.*

Remark 75 Note that this corollary can also be immediately derived from Theorem 98.

Now, from Theorem 107 and its corollaries, by using Theorem 113, we can easily derive the following five theorems.

Theorem 114 *If $\mathscr{R}$ is a quasi-topological relator on X and $A, B \in \mathscr{A}\left(\mathrm{cl}_{\mathscr{R}}, \Phi\right)$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then for any increasing relation on $\mathscr{P}(X)$ to X, we have*

$$\Phi(A) \in \mathrm{Ln}_{\mathscr{R}}\left(\Phi\left(\mathrm{cl}_{\mathscr{R}}(A)\right)\right).$$

Proof By Theorem 107, we have both

$$\Phi(A) \in \mathrm{Ln}_{\mathscr{R}}\left(\Phi(B)\right) \qquad \text{and} \qquad \Phi(B) \in \mathrm{Ln}_{\mathscr{R}}\left(\Phi\left(\mathrm{cl}_{\mathscr{R}}(A)\right)\right).$$

Moreover, by Theorem 113, the relation $\mathrm{Ln}_{\mathscr{R}}$ is transitive. Hence, it is clear that the required assertion is true.

Theorem 115 *If $\mathscr{R}$ is a quasi-topological relator on X and $A, B \subseteq X$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$ then*

(1) $B \in \mathrm{Ln}_{\mathscr{R}}(\mathrm{cl}_{\mathscr{R}}(B))$; *(2)* $A \in \mathrm{Ln}_{\mathscr{R}}(\mathrm{cl}_{\mathscr{R}}(B))$;
(3) $\mathrm{cl}_{\mathscr{R}}(A) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right)\right)$.

Proof Assertion (3) is the $\Phi = \mathrm{cl}_{\mathscr{R}}$ particular case of Theorem 114. Assertion (1) can be derived from Corollary 27 by using Theorem 113. While, assertion (2) is an immediate consequence of the assumption $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, assertion (1) and Theorem 113.

The following three theorems can be derived quite similarly from Corollaries 28, 29, and 30 by using Theorem 113.

Theorem 116 *If $\mathscr{R}$ is a quasi-topological relator on X and $A \in \mathscr{A}_{\mathscr{R}}$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$ for some $B \in \mathscr{A}_{\mathscr{R}}$, then*

$$\mathrm{int}_{\mathscr{R}}(A) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right)\right).$$

Theorem 117 *If $\mathscr{R}$ is a quasi-topological relator on X and $A \in \mathscr{A}_{\mathscr{R}}$ and $B \subseteq X$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then*

(1) $\mathrm{int}_{\mathscr{R}}(B) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right)\right)$;
(2) $\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(A)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(A)\right)\right)\right)$.

Theorem 118 *If $\mathscr{R}$ is a quasi-topological relator on X and $A \subseteq X$ and $B \in \mathscr{A}_{\mathscr{R}}$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then*

(1) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}\left(cl_{\mathscr{R}}(A)\right)\right)\right)$;
(2) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(B)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(B)\right)\right)\right)$.

Moreover, by using Theorem 113, from the results of Section 26, we can also derive some similar theorems.

For instance, from Theorems 110 and 111, by using Theorem 113, we can immediately derive the following two theorems.

Theorem 119 *If $\mathscr{R}$ is a topological relator on X and $A \in \mathscr{A}_{\mathscr{R}}$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$ for some $B \subseteq X$, then*

$$\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(\mathrm{cl}_{\mathscr{R}}(A))\right)\right).$$

Theorem 120 *If $\mathscr{R}$ is a topological relator on X, $A \subseteq X$ and $B \in \mathscr{A}_{\mathscr{R}}$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$, then*

(1) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(\mathrm{cl}_{\mathscr{R}}(A))\right)\right)$;
(2) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(B)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(\mathrm{cl}_{\mathscr{R}}(B))\right)\right)$.

From the latter two theorems, by using Theorem 113, we can immediately derive

Theorem 121 *If $\mathscr{R}$ is a topological relator on X and $A, B \subseteq X$ such that such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$ and either $A \in \mathscr{A}_{\mathscr{R}}$ or $B \in \mathscr{A}_{\mathscr{R}}$, then*

(1) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(A)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(\mathrm{cl}_{\mathscr{R}}(A))\right)\right)$;
(2) $\mathrm{int}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}(B)\right) \in \mathrm{Ln}_{\mathscr{R}}\left(\mathrm{cl}_{\mathscr{R}}\left(\mathrm{int}_{\mathscr{R}}(\mathrm{cl}_{\mathscr{R}}(B))\right)\right)$.

28 Lower and Upper Nearness Closures of Families of Sets

By using our former nearness relations and a very particular case of the induced topological closure $\mathrm{cl}_{\mathscr{R}}$, we may naturally introduce

Definition 21 For any $\mathscr{A} \subseteq \mathscr{P}(X)$, the families

$$\mathscr{A}^{\ell} = \mathscr{A}^{\ell_{\mathscr{R}}} = \mathrm{cl}_{\mathrm{Ln}_{\mathscr{R}}}(\mathscr{A}) \quad \text{and} \quad \mathscr{A}^{u} = \mathscr{A}^{u_{\mathscr{R}}} = \mathrm{cl}_{\mathrm{Un}_{\mathscr{R}}}(\mathscr{A})$$

will be called the *lower and upper nearness closures* of the family $\mathscr{A}$ with respect to the relator $\mathscr{R}$, respectively.

Thus, by Theorem 5, Remark 67 and Definition 1, we can at once state the following two theorems.

Theorem 122 *For any $\mathscr{A} \subseteq \mathscr{P}(X)$, we have*

(1) $\mathscr{A}^{\ell} = \mathrm{Ln}_{\mathscr{R}}^{-1}[\,\mathscr{A}\,]$; *(2)* $\mathscr{A}^{u} = \mathrm{Ln}_{\mathscr{R}}[\,\mathscr{A}\,]$.

Theorem 123 *For any $A \subseteq X$ and $\mathscr{A} \subseteq \mathscr{P}(X)$, we have*

(1) $A \in \mathscr{A}^{\ell}$ *if and only if* $\mathscr{A} \cap \mathrm{Ln}_{\mathscr{R}}(A) \neq \emptyset$;
(2) $A \in \mathscr{A}^{u}$ *if and only if* $\mathscr{A} \cap \mathrm{Ln}_{\mathscr{R}}^{-1}(A) \neq \emptyset$.

Hence, by the corresponding definitions, it is clear that we also have the following

Theorem 124 *For any $A \subseteq X$ and $\mathscr{A} \subseteq \mathscr{P}(X)$, we have*

(1) $A \in \mathscr{A}^{\ell}$ *if and only if there exists* $V \in \mathscr{A}$ *such that* $V \in \mathrm{Ln}_{\mathscr{R}}(A)$;
(2) $A \in \mathscr{A}^{u}$ *if and only if there exists* $V \in \mathscr{A}$ *such that* $A \in \mathrm{Ln}_{\mathscr{R}}(V)$.

From this theorem, by using Definition 20 and Theorem 96, we can immediately derive the following two theorems.

Theorem 125 *For any $A \subseteq X$ and $\mathscr{A} \subseteq \mathscr{P}(X)$, we have*

(1) $A \in \mathscr{A}^{\ell}$ *if and only if there exists* $V \in \mathscr{A}$ *such that* $V \subseteq A \subseteq \mathrm{cl}_{\mathscr{R}}(V)$;
(2) $A \in \mathscr{A}^{u}$ *if and only if there exists* $V \in \mathscr{A}$ *such that* $A \subseteq V \subseteq \mathrm{cl}_{\mathscr{R}}(A)$.

Theorem 126 *For any $A \subseteq X$ and $\mathscr{A} \subseteq \mathscr{P}(X)$, we have*

(1) $A \in \mathscr{A}^{\ell}$ if and only if there exists $V \in \mathscr{A}$ such that $V \subseteq A$, and for any $x \in A$ and $R \in \mathscr{R}$ we have $R(x) \cap V \neq \emptyset$;
(2) $A \in \mathscr{A}^{u}$ if and only if there exists $V \in \mathscr{A}$ such that $A \subseteq V$, and for any $x \in V$ and $R \in \mathscr{R}$ we have $R(x) \cap A \neq \emptyset$.

Proof To prove the "only if part" of (2), note that if $A \in \mathscr{A}^{u}$, then by Theorem 124 there exists $V \in \mathscr{A}$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(V)$. Hence, by Theorem 96, we can see that $A \subseteq V$ and for any $x \in V$ and $R \in \mathscr{R}$ we have $R(x) \cap A \neq \emptyset$.

Remark 76 The most important particular cases of Definition 21 are when $\mathscr{A} = \tau_{\mathscr{S}}$ or $\mathscr{T}_{\mathscr{S}}$ for some nonvoid relator $\mathscr{S}$ on X. That is, $\mathscr{A}$ is a minimal structure or a generalized topology on X. (See [85].)

In our former papers [67, 68], the members of the families $\mathscr{T}_{\mathscr{R}}^{\ell}$, $\mathscr{T}_{\mathscr{R}}^{u}$, $\mathscr{T}_{\mathscr{R}}^{su}$ and $\mathscr{T}_{\mathscr{R}}^{p\ell}$ have been called the *topologically quasi-open, pseudo-open, γ-open and δ-open subsets* of the relator space $X(\mathscr{R})$, respectively.

29 Some Set-Theoretic Properties of the Families $\mathscr{A}^{\ell}$ and $\mathscr{A}^{u}$

By using Theorems 124 and 99–104, we can easily prove the following three theorems.

Theorem 127 *For any $A \subseteq X$ and $\mathscr{A} \subseteq \mathscr{P}(X)$ we have*

(1) $\emptyset \in \mathscr{A}^{\ell}$ if and only if $\emptyset \in \mathscr{A}$;
(2) $\emptyset \in \mathscr{A}^{u}$ if and only if either $\emptyset \in \mathscr{A}$, or $\mathscr{A} \neq \emptyset$ and $\mathscr{R} = \emptyset$.

Proof By Theorems 124 and 99, we have

$$\emptyset \in \mathscr{A}^{\ell} \iff \exists\, V \in \mathscr{A} :\ V \in \mathrm{Ln}_{\mathscr{R}}(\emptyset) \iff \exists\, V \in \mathscr{A} :\ V = \emptyset \iff \emptyset \in \mathscr{A}.$$

While, by Theorems 124 and 100, we have

$$\emptyset \in \mathscr{A}^{u} \iff \exists\, V \in \mathscr{A} :\ \emptyset \in \mathrm{Ln}_{\mathscr{R}}(V) \iff$$
$$\exists\, V \in \mathscr{A} :\ V = \emptyset \text{ or } \mathscr{R} = \emptyset \iff \emptyset \in \mathscr{A} \text{ or } (\mathscr{A} \neq \emptyset,\ \mathscr{R} = \emptyset).$$

Theorem 128 *For any $a \in X$ and $\mathscr{A} \subseteq \mathscr{P}(X)$ we have*

(1) $\{a\} \in \mathscr{A}^{\ell}$ if $\{a\} \in \mathscr{A}$ and for any $R \in \mathscr{R}$ we have $a \in R(a)$;
(2) $\{a\} \in \mathscr{A}^{u}$ if and only if there exists $V \in \mathscr{A}$ such that $a \in V$ and for any $x \in V$ and $R \in \mathscr{R}$ we have $a \in R(x)$.

Proof By Theorems 124 and 101, we have

$$\{a\} \in \mathscr{A}^u \iff \exists\, V \in \mathscr{A} :\ \{a\} \in \mathrm{Ln}_{\mathscr{R}}(V) \iff$$
$$\exists\, V \in \mathscr{A} :\ \ a \in V\,,\ \ \forall\, x \in V :\ \forall\, R \in \mathscr{R} :\ \ a \in R\,(x).$$

While, by Theorems 124 and 102, we have

$$\{a\} \in \mathscr{A}^\ell \iff \exists\, V \in \mathscr{A} :\ \ V \in \mathrm{Ln}_{\mathscr{R}}(\{a\}) \iff$$
$$\exists\, V \in \mathscr{A} :\ \ (\,V = \emptyset,\ \mathscr{R} = \emptyset\,) \text{ or } (\,V = \{a\}\,,\ \forall\, R \in \mathscr{R} :\ \ a \in R\,(a)\,).$$

Theorem 129 *For any $A \subseteq X$ and $\mathscr{A} \subseteq \mathscr{P}(X)$ we have*

(1) $X \in \mathscr{A}^\ell$ if $X \in \mathscr{A}$ and $\mathscr{R}$ non-partial;
(2) $X \in \mathscr{A}^u$ if and only if $X \in \mathscr{A}$ and $\mathscr{R}$ non-partial.

Proof By Theorems 124 and 104, we have

$$X \in \mathscr{A}^u \iff \exists\, V \in \mathscr{A} :\ X \in \mathrm{Ln}_{\mathscr{R}}(V) \iff \exists\, V \in \mathscr{A} :\ V = \emptyset\,,\ \mathscr{R}\text{isnon} - \text{partial}.$$

While, by Theorems 124 and 103, we have

$$X \in \mathscr{A}^\ell \iff \exists\, V \in \mathscr{A} :\ V \in \mathrm{Ln}_{\mathscr{R}}(X) \iff$$
$$\exists\, V \in \mathscr{A} :\ \forall\, x \in X :\ \forall\, R \in \mathscr{R} :\ \ R\,(x) \cap V \neq \emptyset.$$

From Theorems 127 and 129, we can obtain the following

Corollary 32 *If $\mathscr{R}$ is a non-partial relator on X and $\mathscr{A}$ is a minimal structure on X, then $\mathscr{A}^\ell$ and $\mathscr{A}^u$ are also minimal structures on X.*

In addition to this corollary, by using Theorems 124 and 105, we can also prove

Theorem 130 *If $\mathscr{A} \subseteq \mathscr{P}(X)$ such that $\mathscr{A}$ is closed under arbitrary unions, then $\mathscr{A}^\ell$ and $\mathscr{A}^u$ are also closed under arbitrary unions.*

Proof If $A_i \in \mathscr{A}^\ell$ for all $i \in I$, then by Theorem 124 for each $i \in I$ there exists $V_i \in \mathscr{A}$ such that $V_i \in \mathrm{Ln}_{\mathscr{R}}(A_i)$. Hence, by using the assumed property of $\mathscr{A}$ and Theorem 105, we can infer that

$$\bigcup_{i \in I} A_i \in \mathscr{A} \qquad \text{and} \qquad \bigcup_{i \in I} A_i \in \mathrm{Ln}_{\mathscr{R}}\Big(\bigcup_{i \in I} B_i\Big).$$

Therefore, by Theorem 124, we also have $\bigcup_{i \in I} A_i \in \mathscr{A}^\ell$.

This proves the first statement of the theorem. The second statement can be proved quite similarly.

Now, as and immediate consequence of Corollary 32 and Theorem 130, we can also state

Corollary 33 *If $\mathscr{R}$ is a non-partial relator on X and $\mathscr{A}$ is a generalized topology on X, then $\mathscr{A}^\ell$ and $\mathscr{A}^u$ are also generalized topologies on X.*

In this respect, it is also worth proving the following

Theorem 131 *If $\mathscr{A}$ is a stack on X, then*

(1) $\mathscr{A}^\ell \subseteq \mathscr{A}$; *(2)* $\mathscr{A}^{u-} \subseteq \mathscr{A}$.

Proof For instance, if $A \in \mathscr{A}^u$, then by Theorem 125, there exists $V \in \mathscr{A}$ such that $A \subseteq V \subseteq A^-$. Thus, since $\mathscr{A}$ is ascending, $A^- \in \mathscr{A}$ also holds. Therefore, assertion (2) is true.

Moreover, by using Theorems 124 and 97, we can also prove

Theorem 132 *For any $A \subseteq X$ and $\mathscr{A} \subseteq \mathscr{P}(X)$, we have*

(1) $A \in \mathscr{A}^{lc}$ *if and only if there exists* $W \in \mathscr{A}^c$ *such that* $\operatorname{int}_{\mathscr{R}}(W) \subseteq A \subseteq W$;
(2) $A \in \mathscr{A}^{uc}$ *if and only if there exists* $W \in \mathscr{A}^c$ *such that* $\operatorname{int}_{\mathscr{R}}(A) \subseteq W \subseteq A$.

Proof If $A \in \mathscr{A}^{lc}$, then $A^c \in \mathscr{A}^l$. Thus, by Theorem 124, there exists $V \in \mathscr{A}$ such that $V \in \operatorname{Ln}_{\mathscr{R}}(A^c)$. Hence, by taking $W = V^c$, we can see that $W \in \mathscr{A}^c$ such that $W^c \in \operatorname{Ln}_{\mathscr{R}}(A^c)$. Therefore, by Theorem 97, we can also state that $W^\circ \subseteq A \subseteq W$.

Thus, we have proved the "only if part" of (1). The if part of (1) and assertion (2) can be proved quite similarly.

30 Intersection Properties of the Families $\mathscr{A}^\ell$ and $\mathscr{A}^u$

Theorem 133 *If $\mathscr{R}$ is a topologically filtered relator on X, and moreover $U \in \mathscr{T}_{\mathscr{R}}$ and $\mathscr{A} \subseteq \mathscr{P}(X)$ such that $U \cap V \in \mathscr{A}$ for all $V \in \mathscr{A}$, then*

(1) $U \cap A \in \mathscr{A}^\ell$ *for all* $A \in \mathscr{A}^\ell$; *(2)* $U \cap A \in \mathscr{A}^u$ *for all* $A \in \mathscr{A}^u$.

Proof If $A \in \mathscr{A}^\ell$, then by Theorem 124 there exists $V \in \mathscr{A}$ such that $V \in \operatorname{Ln}_{\mathscr{R}}(A)$. Hence, by assumption, we can infer that $U \cap V \in \mathscr{A}$. Moreover, by using Theorem 106, we can see that

$$U \cap V \in \operatorname{Ln}_{\mathscr{R}}(U \cap A).$$

Thus, by Theorem 124, we also have $U \cap A \in \mathscr{A}^\ell$. Therefore, assertion (1) is true.

While, if $A \in \mathscr{A}^u$, then by Theorem 124 there exists $V \in \mathscr{A}$ such that $A \in \operatorname{Ln}_{\mathscr{R}}(V)$. Hence, by assumption, we can infer that $U \cap V \in \mathscr{A}$. Moreover, by using Theorem 106, we can see that

$$U \cap A \in \operatorname{Ln}_{\mathscr{R}}(U \cap V).$$

Thus, by Theorem 124, we also have $U \cap A \in \mathscr{A}^u$. Therefore, assertion (2) is also true.

Repeated applications of this theorem give the following

Corollary 34 *If $\mathscr{R}$ is a topologically filtered relator on X, and moreover $U \in \mathscr{T}_{\mathscr{R}}$ and $\mathscr{A} \subseteq \mathscr{P}(X)$ such that $U \cap V \in \mathscr{A}$ for all $V \in \mathscr{A}$, then $U \cap A \in \mathscr{A}^\kappa$ for all $A \in \mathscr{A}^\kappa$ with $\kappa = \ell\ell,\ uu,\ \ell u$ and $u\ell$.*

Proof By Theorem 133, for instance we have $U \cap A \in \mathscr{A}^\ell$ for all $A \in \mathscr{A}^\ell$. Hence, by applying Theorem 133, to the family $\mathscr{A}^\ell$ instead of $\mathscr{A}$, for instance we can infer that $U \cap A \in \mathscr{A}^{\ell\ell}$ for all $A \in \mathscr{A}^{\ell\ell}$.

From Theorem 133 and Corollary 34, by Corollary 15, it is clear that in particular we also have

Corollary 35 *If $\mathscr{R}$ is a topologically filtered relator on X and $U \in \mathscr{T}_{\mathscr{R}}$, then $U \cap A \in \mathscr{T}_{\mathscr{R}}^\kappa$ for all $A \in \mathscr{T}^\kappa$ with $\kappa = \ell,\ u,\ \ell\ell,\ uu,\ \ell u$ and $u\ell$.*

Proof By Corollary 15, we have $U \cap V \in \mathscr{T}_{\mathscr{R}}$ for all $V \in \mathscr{T}_{\mathscr{R}}$. Thus, by Theorem 133, for instance we have $U \cap A \in \mathscr{T}_{\mathscr{R}}^\ell$ for all $A \in \mathscr{T}_{\mathscr{R}}^\ell$.

Moreover, from Theorem 133, by using Theorem 69, we can derive

Corollary 36 *If $\mathscr{R}$ is a topologically filtered relator on X and $U \in \mathscr{T}_{\mathscr{R}}$, then $U \cap A \in \mathscr{T}_{\mathscr{R}}^{\kappa\nu}$ for all $A \in \mathscr{T}_{\mathscr{R}}^{\kappa\mu}$ with $\kappa = s,\ p,\ \alpha, \beta,\ a,\ b$ and $\mu = \ell,\ u$.*

Proof By Theorem 69, for instance we have $U \cap A \in \mathscr{T}_{\mathscr{R}}^s$ for all $A \in \mathscr{T}_{\mathscr{R}}^s$. Thus, by Theorem 133, for instance we have $U \cap A \in \mathscr{T}_{\mathscr{R}}^{s\ell}$ for all $A \in \mathscr{T}_{\mathscr{R}}^{s\ell}$.

Remark 77 If $\mathscr{R}$ is a nonvoid, topologically filtered, topological relator on X, then by Theorems 70, 71, and 144, we can see that

$$\mathscr{T}_{\mathscr{R}}^\alpha = \left\{ A \subseteq X : \quad \forall\, B \in \mathscr{T}_{\mathscr{R}}^\ell : \quad A \cap B \in \mathscr{T}_{\mathscr{R}}^\ell \right\}.$$

This fact can also be used to prove that now $\mathscr{T}_{\mathscr{R}}^\alpha$ is an ordinary topology on X.

Other classes of generalized topologically open sets do not, in general, form topologies. For instance, the following example shows that the family $\mathscr{T}_{\mathscr{R}}^p$ need not be a topology.

Example 6 If X and $\mathscr{R}$ are as in Example 3 and

$$A = \mathbb{Q} \qquad \text{and} \qquad B = \{1\} \cup \mathbb{Q}^c\,,$$

then $A, B \in \mathscr{T}_{\mathscr{R}}^p$, but $A \cap B \notin \mathscr{T}_{\mathscr{R}}^p$.

To check this, recall that families of all rational and irrational numbers are dense in $\mathbb{R}$. Therefore,

$$A^{-\circ} = \mathbb{R}^\circ = \mathbb{R} = X$$

and quite similarly $B^{-\circ} = X$. Thus, in particular $A, B \in \mathscr{T}^{p}_{\mathscr{R}}$. However,

$$\big(A \cap B\big)^{-\circ} = \{1\}^{-\circ} = \{1\}^{\circ} = \emptyset,$$

and thus $A \cap B \notin \mathscr{T}^{p}_{\mathscr{R}}$.

Remark 78 In this respect, it is also worth mentioning that if $\mathscr{R}$ is a topologically filtered, topological relator on X, then by Theorem 78 the family $\mathscr{T}^{r}_{\mathscr{R}}$ is closed under pairwise intersections.

However, the family $\mathscr{T}^{r}_{\mathscr{R}}$ does not also form a topology. Namely, in contrast to the various families of generalized topological open sets, it is not, in general, closed even under pairwise unions.

31 Some Algebraic and Topological Properties of the Operations ℓ and u

By using Theorems 122, 112, and 113, we can easily establish the following basic properties of the operations ℓ and u introduced in Definition 21.

Theorem 134 *The operations ℓ and u are union-preserving.*

Thus, in particular, we can also state

Corollary 37 *The operations ℓ and u are increasing.*

Theorem 135 *If $\mathscr{R}$ is a reflexive relator on X, then the operations ℓ and u are expansive.*

Hence, by using Theorem 131, we can derive

Corollary 38 *If $\mathscr{R}$ is a reflexive relator on X and $\mathscr{A}$ is a stack on X, then $\mathscr{A}^{\ell} = \mathscr{A}$.*

Theorem 136 *If $\mathscr{R}$ is a quasi-topological relator on X, then the operations ℓ and u are upper quasi-idempotent.*

Proof By Theorem 113, $\mathrm{Ln}_{\mathscr{R}}$ is a transitive relation on $\mathscr{P}(X)$. Hence, it is clear that $\mathrm{Ln}^{-1}_{\mathscr{R}}$ is also a transitive relation on $\mathscr{P}(X)$. Therefore, $\mathrm{Ln}^{-1}_{\mathscr{R}} \circ \mathrm{Ln}^{-1}_{\mathscr{R}} \subseteq \mathrm{Ln}^{-1}_{\mathscr{R}}$. Hence, by using Theorem 122, we can already see that

$$\mathscr{A}^{\ell\ell} = \mathrm{Ln}^{-1}_{\mathscr{R}}\big[\,\mathrm{Ln}^{-1}_{\mathscr{R}}[\,\mathscr{A}\,]\,\big] = \big(\,\mathrm{Ln}^{-1}_{\mathscr{R}} \circ \mathrm{Ln}^{-1}_{\mathscr{R}}\,\big)[\,\mathscr{A}\,] \subseteq \mathrm{Ln}^{-1}_{\mathscr{R}}[\,\mathscr{A}\,] = \mathscr{A}^{\ell}$$

for all $\mathscr{A} \subseteq \mathscr{P}(X)$. Thus, the operation ℓ is quasi-idempotent. The corresponding assertion for u can be proved even more easily.

Now, as an immediate consequence of the above three theorems, we can state

Corollary 39 *If $\mathscr{R}$ is a topological relator on X, then ℓ and u are union-preserving closure operations on $\mathscr{P}(X)$.*

By using Theorem 124, Corollaries 27, and 28 and Theorem 115, we can prove the following three theorems.

Theorem 137 *For any $\mathscr{A} \subseteq \mathscr{P}(X)$, we have*

(1) $\mathscr{A}^{\ell} \subseteq \mathscr{A}^{-u}$; (2) $\mathscr{A}^{u-} \subseteq \mathscr{A}^{\ell}$;
(3) $\mathscr{A}^{\ell-} \subseteq \mathscr{A}^{-\ell}$; (4) $\mathscr{A}^{u-} \subseteq \mathscr{A}^{-u}$.

Proof If $A \in \mathscr{A}^{\ell}$, then by Theorem 124 there exists $V \in \mathscr{A}$ such that $V \in \mathrm{Ln}_{\mathscr{R}}(A)$. Hence, by using Corollary 27, we can infer that $A \in \mathrm{Ln}_{\mathscr{R}}(V^{-})$. Now, since $V^{-} \in \mathscr{A}^{-}$, by Theorem 124 we can see that $A \in \mathscr{A}^{-u}$. Therefore, $\mathscr{A}^{\ell} \subseteq \mathscr{A}^{-u}$.

While, if $A \in \mathscr{A}^{u}$, then by Theorem 124 there exists $V \in \mathscr{A}$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(V)$. Hence, by using Corollary 27, we can infer that $A^{-} \in \mathrm{Ln}_{\mathscr{R}}(V^{-})$. Now, since $V^{-} \in \mathscr{A}^{-}$, by Theorem 124 we can see that $A^{-} \in \mathscr{A}^{-u}$. Therefore, $\mathscr{A}^{u-} \subseteq \mathscr{A}^{-u}$ also holds.

Thus, we have proved assertions (1) and (4). The proof of assertions (2) and (3) are quite similar.

Theorem 138 *We have*

(1) $\mathscr{A}_{\mathscr{R}}^{\ell\circ} \subseteq \mathscr{A}_{\mathscr{R}}^{\circ\ell}$; *(2)* $\mathscr{A}_{\mathscr{R}}^{u-\circ} \subseteq \mathscr{A}_{\mathscr{R}}^{\circ\ell}$.

Proof If $A \in \mathscr{A}_{\mathscr{R}}^{\ell}$, then by Theorem 124 there exists $V \in \mathscr{A}_{\mathscr{R}}$ such that $V \in \mathrm{Ln}_{\mathscr{R}}(A)$. Hence, by using Corollary 28 we can infer that $V^{\circ} \in \mathrm{Ln}_{\mathscr{R}}(A^{\circ})$. Now, since $V^{\circ} \in \mathscr{A}_{\mathscr{R}}^{\circ}$, by Theorem 124 we can see that $A^{\circ} \in \mathscr{A}_{\mathscr{R}}^{\circ\ell}$. Therefore, $\mathscr{A}_{\mathscr{R}}^{\ell\circ} \subseteq \mathscr{A}_{\mathscr{R}}^{\circ\ell}$.

While, if $A \in \mathscr{A}_{\mathscr{R}}^{u}$, then by Theorem 124 there exists $V \in \mathscr{A}_{\mathscr{R}}$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(V)$. Hence, by using Corollary 28, we can infer that $V^{\circ} \in \mathrm{Ln}_{\mathscr{R}}(A^{-\circ})$. Now, since $V^{\circ} \in \mathscr{A}_{\mathscr{R}}^{\circ}$, by Theorem 124 we can see that $A^{-\circ} \in \mathscr{A}_{\mathscr{R}}^{\circ\ell}$. Therefore, $\mathscr{A}_{\mathscr{R}}^{u-\circ} \subseteq \mathscr{A}_{\mathscr{R}}^{\circ\ell}$ also holds.

Theorem 139 *If $\mathscr{R}$ is a quasi-topological relator on X, then for any $\mathscr{A} \subseteq \mathscr{P}(X)$ we have $\mathscr{A}^{\ell-} \subseteq \mathscr{A}^{\ell}$.*

Proof If $A \in \mathscr{A}^{\ell}$, then by Theorem 124 there exists $V \in \mathscr{A}$ such that $V \in \mathrm{Ln}_{\mathscr{R}}(A)$. Hence, by using Theorem 115, we can infer that $V \in \mathrm{Ln}_{\mathscr{R}}(A^{-})$. Now, by Theorem 124, we can see that $A^{-} \in \mathscr{A}^{\ell}$. Therefore, $\mathscr{A}^{\ell-} \subseteq \mathscr{A}^{\ell}$.

Theorem 140 *If $\mathscr{R}$ is a quasi-topological relator on X, then for any $\mathscr{A} \subseteq \mathscr{P}(X)$ we have*

(1) $\mathscr{A}^{\ell u} \subseteq \mathscr{A}^{-u}$; *(2)* $\mathscr{A}^{\ell u-} \subseteq \mathscr{A}^{\ell}$;
(3) $\mathscr{A}^{u\ell} \subseteq \mathscr{A}^{-u}$; *(4)* $\mathscr{A}^{u\ell-} \subseteq \mathscr{A}^{\ell}$.

Proof If $A \in \mathscr{A}^{\ell u}$, then by Theorem 124 there exists $B \in \mathscr{A}^{\ell}$ such that $A \in \mathrm{Ln}_{\mathscr{R}}(B)$. Moreover, also by Theorem 124, there exists $V \in \mathscr{A}$ such that $V \in \mathrm{Ln}_{\mathscr{R}}(B)$. Hence, by using Corollary 27, we can infer that $B \in \mathrm{Ln}_{\mathscr{R}}\left(V^{-}\right)$. Now, by using Theorem 113, we can see that $A \in \mathrm{Ln}_{\mathscr{R}}\left(V^{-}\right)$. Thus, since $V^{-} \in \mathscr{A}^{-}$, by using Theorem 124, we can see that $A^{-} \in \mathscr{A}^{-u}$. Therefore, $\mathscr{A}^{\ell u} \subseteq \mathscr{A}^{-u}$.

While, if $A \in \mathscr{A}^{u\ell}$, then by Theorem 124 there exists $B \in \mathscr{A}^{u}$ such that $B \in \mathrm{Ln}_{\mathscr{R}}(A)$. Moreover, also by Theorem 124, there exists $V \in \mathscr{A}$ such that $B \in \mathrm{Ln}_{\mathscr{R}}(V)$. Hence, by using Corollary 27, we can infer that

$$V \in \mathrm{Ln}_{\mathscr{R}}(B^{-}) \qquad \text{and} \qquad B^{-} \in \mathrm{Ln}_{\mathscr{R}}(A^{-}).$$

Now, by using Theorem 113, we can see that $V \in \mathrm{Ln}_{\mathscr{R}}\left(A^{-}\right)$. Thus, by using Theorem 124, we can see that $A^{-} \in \mathscr{A}^{-\ell}$. Therefore, $\mathscr{A}^{u\ell} \subseteq \mathscr{A}^{\ell}$.

Thus, we have proved assertions (1) and (4). The proofs of assertions (2) and (3) are quite similar.

32 Nearness Closures of the Families $\mathscr{T}_{\mathscr{R}}$, $\mathscr{T}^{s}_{\mathscr{R}}$ and $\mathscr{T}^{p}_{\mathscr{R}}$

The importance of Definition 21 lies mainly in the following theorems.

Theorem 141 *We have*

(1) $\mathscr{T}^{\ell}_{\mathscr{R}} \subseteq \mathscr{T}^{s}_{\mathscr{R}}$; *(2)* $\mathscr{T}^{u}_{\mathscr{R}} \subseteq \mathscr{T}^{p}_{\mathscr{R}}$.

Proof If $A \in \mathscr{T}^{\ell}_{\mathscr{R}}$, then by Theorem 125 there exists $V \in \mathscr{T}_{\mathscr{R}}$ such that

$$V \subseteq A \subseteq V^{-}.$$

Hence, by using the definition of $\mathscr{T}_{\mathscr{R}}$ and the increasingness of $\circ$, we can infer that $V \subseteq V^{\circ} \subseteq A^{\circ}$. Now, by using the increasingness of $-$, we can also see that

$$A \subseteq V^{-} \subseteq A^{\circ -}.$$

Therefore, by Definition 15, we also have $A \in \mathscr{T}^{s}_{\mathscr{R}}$.

While, if $A \in \mathscr{T}^{u}_{\mathscr{R}}$, then by Theorem 125, there exists $V \in \mathscr{T}_{\mathscr{R}}$ such that

$$A \subseteq V \subseteq A^{-}.$$

Hence, by using the definition of $\mathscr{T}_{\mathscr{R}}$ and the increasingness of $\circ$, we can infer that

$$A \subseteq V \subseteq V^{\circ} \subseteq A^{- \circ}.$$

Therefore, by Definition 15, we also have $A \in \mathscr{T}^{p}_{\mathscr{R}}$.

Theorem 142 *If $\mathscr{R}$ is a reflexive relator on X, then*

(1) $\mathscr{T}_{\mathscr{R}}^{\ell} \subseteq \mathscr{T}_{\mathscr{R}}^{s\ell}$; *(2)* $\mathscr{T}_{\mathscr{R}}^{u} \subseteq \mathscr{T}_{\mathscr{R}}^{pu}$.

Proof By Theorems 135 and 141 and Corollary 37, we have

$$\mathscr{T}_{\mathscr{R}}^{\ell} \subseteq \mathscr{T}_{\mathscr{R}}^{\ell\ell} \subseteq \mathscr{T}_{\mathscr{R}}^{s\ell} \qquad \text{and} \qquad \mathscr{T}_{\mathscr{R}}^{u} \subseteq \mathscr{T}_{\mathscr{R}}^{uu} \subseteq \mathscr{T}_{\mathscr{R}}^{pu}.$$

Theorem 143 *If $\mathscr{R}$ is a quasi-topological relator on X, then*

(1) $\mathscr{T}_{\mathscr{R}}^{s\ell} \subseteq \mathscr{T}_{\mathscr{R}}^{s}$; *(2)* $\mathscr{T}_{\mathscr{R}}^{pu} \subseteq \mathscr{T}_{\mathscr{R}}^{p}$.

Proof If $A \in \mathscr{T}_{\mathscr{R}}^{s\ell}$, then by Theorem 125 there exists $V \in \mathscr{T}_{\mathscr{R}}^{s}$ such that

$$V \subseteq A \subseteq V^{-}.$$

Moreover, by Definition 15, we have $V \subseteq V^{\circ -}$. Hence, by using the increasingness of–and Theorem 31, we can infer that

$$V^{-} \subseteq V^{\circ --} \subseteq V^{\circ -}.$$

On the other hand, because of $V \subseteq A$ and the increasingness of $\circ -$, we also have $V^{\circ -} \subseteq A^{\circ -}$. Therefore, $V^{-} \subseteq A^{\circ -}$ also holds. Hence, because of $A \subseteq V^{-}$, we can already see that $A \subseteq A^{\circ -}$. Thus, by Definition 15, we also have $A \in \mathscr{T}_{\mathscr{R}}^{s}$. Therefore, assertion (1) is true.

While, if $A \in \mathscr{T}_{\mathscr{R}}^{pu}$, then by Theorem 125 there exists $V \in \mathscr{T}_{\mathscr{R}}^{p}$ such that

$$A \subseteq V \subseteq A^{-}.$$

Moreover, by Definition 15, we have $A \subseteq A^{-\circ}$. Hence, by using that $B \subseteq A$, we can see that $B \subseteq A^{-\circ}$. Moreover, from the inclusion $A \subseteq B^{-}$, by using the increasingness of–and Theorem 31, we can infer that

$$A^{-} \subseteq B^{--} \subseteq B^{-}.$$

Hence, by using the increasingness of $\circ$, we can infer that $A^{-\circ} \subseteq B^{-\circ}$. Therefore, because of $B \subseteq A^{-\circ}$, we also have $B \subseteq B^{-\circ}$. Hence, by Definition 15, we can see that $B \in \mathscr{T}_{\mathscr{R}}^{p}$ also holds. Therefore, assertion (2) is also true.

Theorem 144 *If $\mathscr{R}$ is a topological relator on X, then*

(1) $\mathscr{T}_{\mathscr{R}}^{\ell} = \mathscr{T}_{\mathscr{R}}^{s}$; *(2)* $\mathscr{T}_{\mathscr{R}}^{u} = \mathscr{T}_{\mathscr{R}}^{p}$.

Proof If $A \in \mathscr{T}_{\mathscr{R}}^{s}$, then by Definition 15 we have $A \subseteq A^{\circ -}$. Hence, by taking $V = A^{\circ}$, we get $A \subseteq V^{-}$. Moreover, by using Theorems 32, 26, and 31, we can see that

$$V = A^\circ \subseteq A \qquad \text{and} \qquad V = A^\circ \in \mathscr{T}_{\mathscr{R}}.$$

Thus, by Theorem 125, we also have $A \in \mathscr{T}^{\ell}_{\mathscr{R}}$. This proves that $\mathscr{T}^{s}_{\mathscr{R}} \subseteq \mathscr{T}^{\ell}_{\mathscr{R}}$.

While, if $A \in \mathscr{T}^{p}_{\mathscr{R}}$, then by Definition 15 we have $A \subseteq A^{-\circ}$. Hence, by taking $V = A^{-\circ}$, we get $A \subseteq V$. Moreover, by using Theorems 32, 26, and 31, we can see that

$$V = A^{-\circ} \subseteq A^{-} \qquad \text{and} \qquad V = A^{-\circ} \in \mathscr{T}_{\mathscr{R}}.$$

Thus, by Theorem 125, we also have $A \in \mathscr{T}^{u}_{\mathscr{R}}$. This proves that $\mathscr{T}^{p}_{\mathscr{R}} \subseteq \mathscr{T}^{u}_{\mathscr{R}}$.

Now, by Theorem 141, we can see that assertions (1) and (2) are also true.

From Theorems 142 and 143, by using Theorem 144, we can immediately derive

Corollary 40 *If $\mathscr{R}$ is a topological relator on X, then*

(1) $\mathscr{T}^{s\ell}_{\mathscr{R}} = \mathscr{T}^{s}_{\mathscr{R}}$; *(2)* $\mathscr{T}^{pu}_{\mathscr{R}} = \mathscr{T}^{p}_{\mathscr{R}}$.

Remark 79 Note that this corollary can also be immediately derived from Theorem 144 by using Corollary 39.

33 Some Further Theorems on the Families $\mathscr{T}^{s}_{\mathscr{R}}$ and $\mathscr{T}^{p}_{\mathscr{R}}$

In addition to Corollary 40, we can also prove the following

Theorem 145 *If $\mathscr{R}$ is a topological relator, then*

(1) $\mathscr{T}_{\mathscr{R}} = \mathscr{T}^{s\circ}_{\mathscr{R}}$; *(2)* $\mathscr{T}_{\mathscr{R}} = \mathscr{T}^{p\circ}_{\mathscr{R}}$.

Proof By Theorem 32, $\mathscr{R}$ is reflexive and quasi-topological. Thus, if in particular $V \in \mathscr{T}_{\mathscr{R}}$, then by Theorem 57 we also have $V \in \mathscr{T}^{s}_{\mathscr{R}}$.

Moreover, by Definition 3 and Theorem 26, we also have $V \subseteq V^\circ$ and $V^\circ \subseteq V$, and thus also $V = V^\circ$. Therefore, $V \in \mathscr{T}^{s\circ}_{\mathscr{R}}$, and thus $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}^{s\circ}_{\mathscr{R}}$.

On the other hand, by Theorem 31 we have $A^\circ \in \mathscr{T}_{\mathscr{R}}$ for all $A \subseteq X$. Thus, $\mathscr{P}(X)^\circ \subseteq \mathscr{T}_{\mathscr{R}}$, and thus in particular $\mathscr{T}^{s\circ}_{\mathscr{R}} \subseteq \mathscr{T}_{\mathscr{R}}$ also holds.

Therefore, assertion (1) is true. The proof of assertion (2) is quite similar.

However, it is now more important to note that, by using Corollary 40, we can prove the following theorems.

Theorem 146 *If $\mathscr{R}$ is a topological relator on X, then*

(1) $\mathscr{T}^{s-}_{\mathscr{R}} \subseteq \mathscr{T}^{s}_{\mathscr{R}}$; *(2)* $A^- \in \mathscr{T}^{p}_{\mathscr{R}}$ *implies* $A \in \mathscr{T}^{p}_{\mathscr{R}}$ *for all* $A \subseteq X$.

Proof If $A \subseteq X$, then by Theorems 32 and 31, we have $A \subseteq A^- \subseteq A^-$. Hence, if $A \in \mathscr{T}^{s}_{\mathscr{R}}$, then by using Theorem 125 we can see that $A^- \in \mathscr{T}^{s\ell}_{\mathscr{R}}$. Thus, by Corollary 40, we also have $A^- \in \mathscr{T}^{s}_{\mathscr{R}}$. Therefore, assertion (1) is true.

On the other hand, if $A \subseteq X$ such that $A^- \in \mathscr{T}^p_{\mathscr{R}}$, then from the inclusions $A \subseteq A^- \subseteq A^-$ by using Theorem 125 we can see that $A \in \mathscr{T}^{pu}_{\mathscr{R}}$. Thus, by Corollary 40, we also have $A \in \mathscr{T}^p_{\mathscr{R}}$. Therefore, assertion (2) is also true.

From this theorem, by using Theorem 57, we can immediately derive

Corollary 41 *If $\mathscr{R}$ is a topological relator on X, then*

(1) $\mathscr{T}^-_{\mathscr{R}} \subseteq \mathscr{T}^s_{\mathscr{R}}$; *(2)* $A^- \in \mathscr{T}_{\mathscr{R}}$ *implies* $A \in \mathscr{T}^p_{\mathscr{R}}$ *for all* $A \subseteq X$.

Proof Namely, by Theorem 32 and 57, we have $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}^s_{\mathscr{R}}$ and $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}^p_{\mathscr{R}}$. Hence, by Theorem 146, it is clear that in particular assertions are also true.

Remark 80 If $\mathscr{R}$ is a topological relator on X, then in particular we also have $\mathscr{P}(X)^{\circ -} \subseteq \mathscr{T}^s_{\mathscr{R}}$.

Namely, by Theorems 32 and 31, we have $\mathscr{P}(X)^\circ \subseteq \mathscr{T}_{\mathscr{R}}$. Hence, by using Corollary 41, we can infer that $\mathscr{P}(X)^{\circ -} \subseteq \mathscr{T}^-_{\mathscr{R}} \subseteq \mathscr{T}^s_{\mathscr{R}}$.

Theorem 147 *If $\mathscr{R}$ is a topological relator on X, then $\mathscr{A} = \mathscr{T}^s_{\mathscr{R}}$ is the smallest subset of $\mathscr{P}(X)$ such that*

(1) $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{A}$; *(2)* $\mathscr{A}^\ell \subseteq \mathscr{A}$.

Proof If $\mathscr{A} = \mathscr{T}^s_{\mathscr{R}}$, then from Theorems 57 and 141, we can see that (1) and (2) hold.

While, if $\mathscr{A} \subseteq \mathscr{P}(X)$ such that (1) and (2) hold, then by using Theorem 144 and Corollary 37 we can see that

$$\mathscr{T}^s_{\mathscr{R}} = \mathscr{T}^\ell_{\mathscr{R}} \subseteq \mathscr{A}^\ell \subseteq \mathscr{A}.$$

Therefore, the stated minimality property of $\mathscr{T}^s_{\mathscr{R}}$ is also true.

Remark 81 Assertion (1) of Theorem 144 showed that, for a topological relator $\mathscr{R}$ on X, the family $\mathscr{T}^s_{\mathscr{R}}$ is the l-closure of $\mathscr{T}_{\mathscr{R}}$ in $\mathscr{P}(X)$.

While, assertion (1) of Theorem 147 shows that, for a topological relator $\mathscr{R}$ on X, the family $\mathscr{T}^s_{\mathscr{R}}$ is the smallest ł-closed subset of $\mathscr{P}(X)$ containing $\mathscr{T}_{\mathscr{R}}$.

Theorem 148 *If $\mathscr{R}$ is a topological relator on X, then $\mathscr{A} = \mathscr{T}^p_{\mathscr{R}}$ is the smallest subset of $\mathscr{P}(X)$ such that*

(1) $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{A}$; *(2)* $\mathscr{A}^u \subseteq \mathscr{A}$.

Proof If $\mathscr{A} = \mathscr{T}^p_{\mathscr{R}}$, then from Theorems 57 and 141, we can see that (1) and (2) hold.

While, if $\mathscr{A} \subseteq \mathscr{P}(X)$ such that (1) and (2) hold, then by using Theorem 144 and Corollary 37 we can see that

$$\mathscr{T}^p_{\mathscr{R}} = \mathscr{T}^u_{\mathscr{R}} \subseteq \mathscr{A}^u \subseteq \mathscr{A}.$$

Therefore, the stated minimality property of $\mathscr{T}^p_{\mathscr{R}}$ is also true.

Remark 82 In the above theorems, by Theorem 144, we may also write $\mathscr{T}_{\mathscr{R}}^{\ell}$ and $\mathscr{T}_{\mathscr{R}}^{u}$ instead of $\mathscr{T}_{\mathscr{R}}^{s}$ and $\mathscr{T}_{\mathscr{R}}^{p}$, respectively.

34 Some Basic Properties of the Families $\mathscr{T}_{\mathscr{R}}^{su}$ and $\mathscr{T}_{\mathscr{R}}^{p\ell}$

The following theorems will show that the families $\mathscr{T}_{\mathscr{R}}^{su}$ and $\mathscr{T}_{\mathscr{R}}^{p\ell}$ are usually more important than the families $\mathscr{T}_{\mathscr{R}}^{sl}$ and $\mathscr{T}_{\mathscr{R}}^{pu}$ considered in Section 32.

Theorem 149 *We have*

$$\mathscr{T}_{\mathscr{R}}^{su} \cup \mathscr{T}_{\mathscr{R}}^{p\ell} \subseteq \mathscr{T}_{\mathscr{R}}^{\beta}.$$

Proof If $A \in \mathscr{T}_{\mathscr{R}}^{su}$, then by Theorem 125 there exists $V \in \mathscr{T}_{\mathscr{R}}^{s}$ such that

$$A \subseteq V \subseteq A^{-}.$$

Hence, by using Definition 15 and the increasingness of $\circ -$, we can infer that

$$A \subseteq V \subseteq V^{\circ -} \subseteq A^{-\circ -}.$$

Thus, by Definition 15, we also have $A \in \mathscr{T}_{\mathscr{R}}^{\beta}$. Therefore, $\mathscr{T}_{\mathscr{R}}^{su} \subseteq \mathscr{T}_{\mathscr{R}}^{\beta}$.

While, if $A \in \mathscr{T}_{\mathscr{R}}^{p\ell}$, then by Theorem 125 there exists $V \in \mathscr{T}_{\mathscr{R}}^{p}$ such that

$$V \subseteq A \subseteq V^{-}.$$

Hence, by using Definition 15 and the increasingness of—and $- \circ -$, we can infer that

$$A \subseteq V^{-} \subseteq V^{-\circ -} \subseteq A^{-\circ -}.$$

Thus, by Definition 15, we also have $A \in \mathscr{T}_{\mathscr{R}}^{\beta}$. Therefore, $\mathscr{T}_{\mathscr{R}}^{p\ell} \subseteq \mathscr{T}_{\mathscr{R}}^{\beta}$.

From this theorem, by using Theorem 141 and Corollary 37, we can derive

Corollary 42 *We have* $\mathscr{T}_{\mathscr{R}}^{\ell u} \cup \mathscr{T}_{\mathscr{R}}^{u\ell} \subseteq \mathscr{T}_{\mathscr{R}}^{\beta}$.

Theorem 150 *If $\mathscr{R}$ is a reflexive relator on X, then*

(1) $\mathscr{T}_{\mathscr{R}}^{s} \cup \mathscr{T}_{\mathscr{R}}^{u} \subseteq \mathscr{T}_{\mathscr{R}}^{su}$; *(2)* $\mathscr{T}_{\mathscr{R}}^{p} \cup \mathscr{T}_{\mathscr{R}}^{\ell} \subseteq \mathscr{T}_{\mathscr{R}}^{p\ell}$.

Proof By Theorem 135, we have $\mathscr{T}_{\mathscr{R}}^{s} \subseteq \mathscr{T}_{\mathscr{R}}^{su}$. Moreover, by Theorem 57, we have $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}_{\mathscr{R}}^{s}$. Hence, by using Corollary 37, we can infer that $\mathscr{T}_{\mathscr{R}}^{u} \subseteq \mathscr{T}_{\mathscr{R}}^{su}$.

Therefore, assertion (1) is true. Assertion (2) can be proved quite similarly.

From this theorem, by using Theorem 141, we can derive

Corollary 43 *If $\mathscr{R}$ is a reflexive relator on X, then* $\mathscr{T}_{\mathscr{R}}^{\ell} \cup \mathscr{T}_{\mathscr{R}}^{u} \subseteq \mathscr{T}_{\mathscr{R}}^{su} \cap \mathscr{T}_{\mathscr{R}}^{p\ell}$.

However, this is of no particular importance since we can now also prove a stronger statement.

Theorem 151 *If $\mathscr{R}$ is a reflexive relator on X, then*

$$\mathscr{T}_{\mathscr{R}}^{\ell} \cup \mathscr{T}_{\mathscr{R}}^{u} \subseteq \mathscr{T}_{\mathscr{R}}^{\ell u} \cap \mathscr{T}_{\mathscr{R}}^{u\ell}.$$

Proof By Theorem 135, we have $\mathscr{T}_{\mathscr{R}}^{\ell} \subseteq \mathscr{T}_{\mathscr{R}}^{\ell u}$ and $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}_{\mathscr{R}}^{\ell}$. Hence, by using Corollary 37, we can infer that $\mathscr{T}_{\mathscr{R}}^{u} \subseteq \mathscr{T}_{\mathscr{R}}^{\ell u}$.

Therefore, $\mathscr{T}_{\mathscr{R}}^{\ell} \cup \mathscr{T}_{\mathscr{R}}^{u} \subseteq \mathscr{T}_{\mathscr{R}}^{\ell u}$. The inclusion $\mathscr{T}_{\mathscr{R}}^{\ell} \cup \mathscr{T}_{\mathscr{R}}^{u} \subseteq \mathscr{T}_{\mathscr{R}}^{u\ell}$ can be proved quite similarly.

From this theorem, by using Theorem 135, we can derive

Corollary 44 *If $\mathscr{R}$ is a reflexive relator on X, then* $\mathscr{T}_{\mathscr{R}} \subseteq \mathscr{T}_{\mathscr{R}}^{\ell u} \cap \mathscr{T}_{\mathscr{R}}^{u\ell}$.

Moreover, by using Theorem 135, we can also prove

Theorem 152 *If $\mathscr{R}$ is a reflexive relator on X, then*

$$\mathscr{T}_{\mathscr{R}}^{a} \subseteq \mathscr{T}_{\mathscr{R}}^{su} \cap \mathscr{T}_{\mathscr{R}}^{p\ell}.$$

Proof By Theorems 56 and 135, we have

$$\mathscr{T}_{\mathscr{R}}^{a} = \mathscr{T}_{\mathscr{R}}^{s} \cap \mathscr{T}_{\mathscr{R}}^{p} \subseteq \mathscr{T}_{\mathscr{R}}^{su} \cap \mathscr{T}_{\mathscr{R}}^{p\ell}.$$

However, it is now more important to note that, in addition to Theorem 149, we can also prove the following two theorems.

Theorem 153 *If $\mathscr{R}$ is a topological relator on X, then* $\mathscr{T}_{\mathscr{R}}^{su} = \mathscr{T}_{\mathscr{R}}^{\beta}$.

Proof By Theorem 149, we always have $\mathscr{T}_{\mathscr{R}}^{su} \subseteq \mathscr{T}_{\mathscr{R}}^{\beta}$. Therefore, we need only prove that now $\mathscr{T}_{\mathscr{R}}^{\beta} \subseteq \mathscr{T}_{\mathscr{R}}^{su}$ also holds.

For this, note that if $A \in \mathscr{T}_{\mathscr{R}}^{\beta}$, then by Theorem 63 we have also $A^{-} \in \mathscr{T}_{\mathscr{R}}^{s}$. Hence, by defining $V = A^{-}$, we can note that $V \in \mathscr{T}_{\mathscr{R}}^{s}$ such that $V \subseteq A^{-}$. Moreover, by Theorems 32 and 26, we can see that $A \subseteq A^{-} = V$ is also true. Therefore, by Theorem 125, we also have $A \in \mathscr{T}_{\mathscr{R}}^{u}$.

Theorem 154 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then* $\mathscr{T}_{\mathscr{R}}^{p\ell} = \mathscr{T}_{\mathscr{R}}^{\beta}$.

Proof By Theorem 149, we always have $\mathscr{T}_{\mathscr{R}}^{p\ell} \subseteq \mathscr{T}_{\mathscr{R}}^{\beta}$. Therefore, we need only prove that now $\mathscr{T}_{\mathscr{R}}^{\beta} \subseteq \mathscr{T}_{\mathscr{R}}^{p\ell}$ also holds.

For this, note that if $A \in \mathscr{T}_{\mathscr{R}}^{\beta}$, then by Theorem 63 we have $A^{-} \in \mathscr{T}_{\mathscr{R}}^{s}$. Hence, by using Theorems 32, 26, 31, and 59, we can infer that

$$A^{-} = A^{--} = A^{-\circ -}.$$

Now, by defining $V = A \cap A^{-\circ}$, we can note that $V \subseteq A$. Moreover, by using Corollary 18 and Theorems 32, 31, and 26, we can see that

$$V^{-} = \left(A \cap A^{-\circ}\right)^{-} \supseteq A^{-} \cap A^{-\circ} = A^{-\circ}.$$

Hence, by using the increasingness of $\circ$ and Theorems 32, 31, and 26, we can infer that

$$V^{-\circ} \supseteq A^{-\circ\circ} = A^{-\circ} \supseteq A \cap A^{-\circ} = V.$$

Thus, by Definition 15, we also have $V \in \mathscr{T}_{\mathscr{R}}^{p}$.

Moreover, quite similarly, we can now also note that

$$A \subseteq A^{-} = A^{-\circ -} \subseteq V^{--} = V^{-}.$$

Therefore, by Theorem 125, we also have $A \in \mathscr{T}_{\mathscr{R}}^{p\ell}$.

From Theorems 153 and 154, by using Corollary 39, we can immediately derive the following two corollaries.

Corollary 45 *If $\mathscr{R}$ is a topological relator on X, then $\mathscr{T}_{\mathscr{R}}^{\beta u} = \mathscr{T}_{\mathscr{R}}^{\beta}$.*

Corollary 46 *If $\mathscr{R}$ is a topologically filtered, topological relator on X, then $\mathscr{T}_{\mathscr{R}}^{\beta\ell} = \mathscr{T}_{\mathscr{R}}^{\beta}$.*

35 Some Further Theorems on the Closure Operation $\triangle$

By Theorems 17 and 16, we evidently have the following

Theorem 155 *For any $\Diamond = *, \#$ and $\wedge$, we have*

(1) $\mathscr{A}_{\mathscr{R}} = \mathscr{A}_{\mathscr{R}^{\Diamond}}$,
(2) $\mathrm{Ln}_{\mathscr{R}} = \mathrm{Ln}_{\mathscr{R}^{\Diamond}}$ *and* $\mathrm{Un}_{\mathscr{R}} = \mathrm{Un}_{\mathscr{R}^{\Diamond}}$;
(3) $\ell_{\mathscr{R}} = \ell_{\mathscr{R}^{\Diamond}}$ *and* $u_{\mathscr{R}} = u_{\mathscr{R}^{\Diamond}}$.

Proof By Theorem 17, we have $\mathscr{R}^{\wedge} = \mathscr{R}^{\Diamond\wedge}$. Hence, by using Theorem 16, we can infer that

$$\mathrm{cl}_{\mathscr{R}} = \mathrm{cl}_{\mathscr{R}^{\Diamond}} \qquad \text{and} \qquad \mathrm{int}_{\mathscr{R}} = \mathrm{int}_{\mathscr{R}^{\Diamond}}.$$

Thus, by Definitions 19, 20, and 21, the required assertions are also true.

Moreover, analogously to the results of [67, Sections 33–35], we can also easily prove the following theorems.

Theorem 156 *If $\mathscr{R}$ is nonvoid, non-partial relator on X, then*

$$\mathscr{E}_{\mathscr{R}} \cup \mathscr{E}_{\mathscr{R}}^{c} \subseteq \mathscr{A}_{\mathscr{R}^{\vartriangle}}.$$

Proof By Theorem 28, we have

$$\emptyset \notin \mathscr{E}_{\mathscr{R}} \qquad \text{and} \qquad X \in \mathscr{D}_{\mathscr{R}}.$$

Therefore, if $A \in \mathscr{E}_{\mathscr{R}}$, then by Corollary 6 we have

$$\mathrm{cl}_{\mathscr{R}^{\vartriangle}}\left(\mathrm{int}_{\mathscr{R}^{\vartriangle}}(A)\right) = \mathrm{cl}_{\mathscr{R}^{\vartriangle}}(X) = X.$$

Thus, $\mathrm{int}_{\mathscr{R}^{\vartriangle}}\left(\mathrm{cl}_{\mathscr{R}^{\vartriangle}}(A)\right) \subseteq \mathrm{cl}_{\mathscr{R}^{\vartriangle}}\left(\mathrm{int}_{\mathscr{R}^{\vartriangle}}(A)\right)$, i.e., $A \in \mathscr{A}_{\mathscr{R}^{\vartriangle}}$ trivially holds.

While, if $A \in \mathscr{E}_{\mathscr{R}}^{c}$, i.e., $A^{c} \in \mathscr{E}_{\mathscr{R}}$, then by Theorem 7 we have $A \notin \mathscr{D}_{\mathscr{R}}$. Therefore, by Corollary 6, we have

$$\mathrm{int}_{\mathscr{R}^{\vartriangle}}\left(\mathrm{cl}_{\mathscr{R}^{\vartriangle}}(A)\right) = \mathrm{int}_{\mathscr{R}^{\vartriangle}}(\emptyset) = \emptyset.$$

Thus, $\mathrm{int}_{\mathscr{R}^{\vartriangle}}\left(\mathrm{cl}_{\mathscr{R}^{\vartriangle}}(A)\right) \subseteq \mathrm{cl}_{\mathscr{R}^{\vartriangle}}\left(\mathrm{int}_{\mathscr{R}^{\vartriangle}}(A)\right)$, i.e., $A \in \mathscr{A}_{\mathscr{R}^{\vartriangle}}$ trivially holds.

Theorem 157 *If $\mathscr{R}$ is non-degenerated relator on X, then*

$$\mathscr{D}_{\mathscr{R}} \cap \mathscr{A}_{\mathscr{R}^{\vartriangle}} \subseteq \mathscr{E}_{\mathscr{R}}.$$

Proof By Definition 10 and Theorem 30, we have

$$X \neq \emptyset, \qquad \mathscr{R} \neq \emptyset \qquad \text{and} \qquad \emptyset \notin \mathscr{D}_{\mathscr{R}}, \qquad X \in \mathscr{E}_{\mathscr{R}}.$$

Thus, if $A \in \mathscr{D}_{\mathscr{R}}$, then by Corollary 6 we have

$$\mathrm{int}_{\mathscr{R}^{\vartriangle}}\left(\mathrm{cl}_{\mathscr{R}^{\vartriangle}}(A)\right) = \mathrm{int}_{\mathscr{R}^{\vartriangle}}(X) = X.$$

Now, if $A \in \mathscr{A}_{\mathscr{R}^{\vartriangle}}$, i.e., $\mathrm{int}_{\mathscr{R}^{\vartriangle}}\left(\mathrm{cl}_{\mathscr{R}^{\vartriangle}}(A)\right) \subseteq \mathrm{cl}_{\mathscr{R}^{\vartriangle}}\left(\mathrm{int}_{\mathscr{R}^{\vartriangle}}(A)\right)$, then we also have

$$\mathrm{cl}_{\mathscr{R}^{\vartriangle}}\left(\mathrm{int}_{\mathscr{R}^{\vartriangle}}(A)\right) = X \neq \emptyset.$$

Hence, by using Corollary 6, we can infer that $\mathrm{int}_{\mathscr{R}^{\vartriangle}}(A) \in \mathscr{D}_{\mathscr{R}}$. Thus, since $\emptyset \notin \mathscr{D}_{\mathscr{R}}$, we can also state that $\mathrm{int}_{\mathscr{R}^{\vartriangle}}(A) \neq \emptyset$. Hence, by using Corollary 6, we can already infer that $A \in \mathscr{E}_{\mathscr{R}}$.

Theorem 158 *If $\mathscr{R}$ is a nonvoid relator on X, then for any $A \subseteq X$ we have*

(1) $\mathrm{Un}_{\mathscr{R}^\triangle}(A) = \mathscr{P}^{-1}(A)$ *if* $A \in \mathscr{D}_{\mathscr{R}}$;
(2) $\mathrm{Un}_{\mathscr{R}^\triangle}(A) = \emptyset$ *if* $A \notin \mathscr{D}_{\mathscr{R}} \setminus \{\emptyset\}$; *(3)* $\mathscr{U}_{\mathscr{R}^\triangle}(\emptyset) = \{\emptyset\}$ *if* $\emptyset \notin \mathscr{D}_{\mathscr{R}}$.

Proof If $A \in \mathscr{D}_{\mathscr{R}}$, then by Definition 20 and Corollary 6, for any $B \subseteq X$ we have

$$B \in \mathrm{Un}_{\mathscr{R}^\triangle}(A) \iff A \in \mathrm{Ln}_{\mathscr{R}^\triangle}(B) \iff A \subseteq B \subseteq \mathrm{cl}_{\mathscr{R}^\triangle}(A)$$
$$\iff A \subseteq B \subseteq X \iff A \in \mathscr{P}(B) \iff B \in \mathscr{P}^{-1}(A).$$

Therefore, assertion (1) is true.

While, if $A \notin \mathscr{D}_{\mathscr{R}}$, then by Definition 20 and Corollary 6, for any $B \subseteq X$ we have

$$B \in \mathrm{Un}_{\mathscr{R}^\triangle}(A) \iff A \in \mathrm{Ln}_{\mathscr{R}^\triangle}(B)$$
$$\iff A \subseteq B \subseteq \mathrm{cl}_{\mathscr{R}^\triangle}(A) \iff A \subseteq B \subseteq \emptyset \iff A = B = \emptyset.$$

Hence, it is clear that assertions (2) and (3) are also true.

From this theorem, by using Theorem 30, we can immediately derive

Corollary 47 *If $\mathscr{R}$ is a non-degenerated relator on X, then for any $A \subseteq X$ we have*

(1) $\mathrm{Un}_{\mathscr{R}^\triangle}(A) = \emptyset$ *if* $A \notin \mathscr{D}_{\mathscr{R}}$; *(2)* $\mathscr{U}_{\mathscr{R}^\triangle}(A) = \mathscr{P}^{-1}(A)$ *if* $A \in \mathscr{D}_{\mathscr{R}}$.

Hence, by using Remark 67 and Theorem 122, we can easily derive

Theorem 159 *If $\mathscr{R}$ is a non-degenerated relator on X, then for any $\mathscr{A} \subseteq \mathscr{P}(X)$ we have*

$$\mathscr{A}^{\ell_{\mathscr{R}^\triangle}} = \bigcup_{A \in \mathscr{A} \cap \mathscr{D}_{\mathscr{R}}} \mathscr{P}^{-1}(A).$$

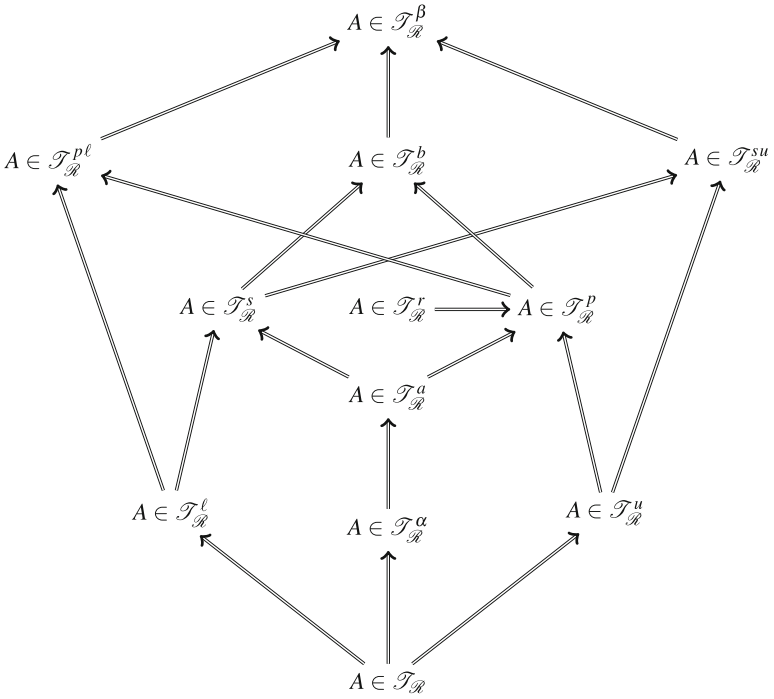
Proof By Remark 67, Theorem 122 and Corollary 47, we can see that

$$\mathscr{A}^{\ell_{\mathscr{R}^\triangle}} = \mathrm{Ln}^{-1}_{\mathscr{R}^\triangle}[\mathscr{A}] = \mathrm{Un}_{\mathscr{R}^\triangle}[\mathscr{A}] = \bigcup_{A \in \mathscr{A}} \mathrm{Un}_{\mathscr{R}^\triangle}(A) = \bigcup_{A \in \mathscr{A} \cap \mathscr{D}_{\mathscr{R}}} \mathscr{P}^{-1}(A).$$

36 An Illustrating Diagram and Two Related Examples

The following diagram and the subsequent two examples have been constructed by Muwafaq Mahdi Salih, a PhD student of the second author from the University of Duhok, Kurdistan Region, Iraq.

Diagram 1 For a reflexive relator $\mathscr{R}$ on X, the following implications hold:



Remark 83 By our former theorems, nine from the above implications do not require the relator $\mathscr{R}$ to be reflexive.

Moreover, the following two examples established in [67] show that seventeen implications in Diagram 1 are not reversible.

Example 7 If $X = \{1, 2, 3\}$ and R is a relation on X such that

$$R(1) = \{1, 2\} \qquad \text{and} \qquad R(2) = R(3) = X,$$

then $\mathscr{R} = \{R\}$ is a reflexive relator on X such that:

(1) $\mathscr{T}_{\mathscr{R}} = \mathscr{T}_{\mathscr{R}}^{r} = \mathscr{T}_{\mathscr{R}}^{\ell} = \{\emptyset, X\}$;
(2) $\mathscr{T}_{\mathscr{R}}^{s} = \mathscr{T}_{\mathscr{R}}^{a} = \mathscr{T}_{\mathscr{R}}^{\alpha} = \{\emptyset, \{1, 2\}, X\}$;
(3) $\mathscr{T}_{\mathscr{R}}^{u} = \mathscr{T}_{\mathscr{R}}^{p} = \mathscr{T}_{\mathscr{R}}^{b} = \mathscr{T}_{\mathscr{R}}^{\beta} = \mathscr{T}_{\mathscr{R}}^{su} = \mathscr{T}_{\mathscr{R}}^{p\ell} = \mathscr{P}(X) \setminus \{\{3\}\}$.

Example 8 If $X = \{1, 2, 3, 4\}$ and R_1 and R_2 are relations on X such that

$$R_1(1) = R_1(2) = \{1, 2, 3\}, \qquad R_1(3) = R_1(4) = \{1, 3, 4\};$$

$$R_2(1) = \{1, 2, 3\}, \qquad R_2(2) = \{1, 2\}, \qquad R_2(3) = R_2(4) = \{3, 4\};$$

then $\mathscr{R} = \{R_1,\ R_2\}$ is a reflexive relator on X such that:

(1) $\mathscr{T}_{\mathscr{R}} = \mathscr{T}_{\mathscr{R}}^{\alpha} = \{\emptyset,\ \{3,4\},\ X\}$;
(2) $\mathscr{T}_{\mathscr{R}}^{r} = \mathscr{T}_{\mathscr{R}} \cup \{\{2\}\}$;
(3) $\mathscr{T}_{\mathscr{R}}^{\ell} = \mathscr{T}_{\mathscr{R}}^{a} = \mathscr{T}_{\mathscr{R}} \cup \{\{1,3,4\}\}$;
(4) $\mathscr{T}_{\mathscr{R}}^{s} = \mathscr{T}_{\mathscr{R}}^{\ell} \cup \{\{1,2\}\}$;
(5) $\mathscr{T}_{\mathscr{R}}^{p} = \mathscr{P}(X) \setminus \{\{1\},\ \{1,2\}\}$;
(6) $\mathscr{T}_{\mathscr{R}}^{u} = \mathscr{T}_{\mathscr{R}}^{p} \setminus \{\{2\}\}$;
(7) $\mathscr{T}_{\mathscr{R}}^{b} = \mathscr{T}_{\mathscr{R}}^{p\ell} = \mathscr{P}(X) \setminus \{\{1\}\}$;
(8) $\mathscr{T}_{\mathscr{R}}^{\beta} = \mathscr{T}_{\mathscr{R}}^{su} = \mathscr{P}(X)$.

Remark 84 Unfortunately, the above two examples cannot be used to show that the implication $A \in \mathscr{T}_{\mathscr{R}}^{su} \implies A \in \mathscr{T}_{\mathscr{R}}^{\beta}$ is also not reversible.

Note that, by Theorem 149, this implication does not also require the relator $\mathscr{R}$ to be reflexive. Moreover, if $\mathscr{R}$ is topological, then by Theorem 153 the reverse implication is also true.

Note 1 Some of the results of this paper can be generalized according to the ideas of Gargouri and Rezgui [31] and the second author [96].

References

1. M.E. Abd El-Monsef, S.N. El-Deeb, R.A. Mahmoud, β-open sets and β-continuous mappings. Bull. Fac. Sci. Assiut Univ. **12**, 77–90 (1983)
2. P. Alexandroff, Zur Begründung der n-dimensionalen mengentheorischen Topologie. Math. Ann. **94**, 296–308 (1925)
3. D. Andrijević, Semi-preopen sets. Mat. Vesnik **38**, 24–32 (1986)
4. D. Andrijević, On b-open sets. Mat. Vesnik **48**, 59–64 (1996)
5. N. Biswas, On some mappings in topological spaces. Bull. Cal. Math. Soc. **61**, 127–135 (1969)
6. N. Bourbaki, *General Topology, Chapters 1–4* (Springer, Berlin, 1989)
7. T.A. Chapman, A further note on closure and interior operators. Am. Math. Monthly **69**, 524–529 (1962)
8. Ch. Chattopadhyay, Ch. Bandyopadhyay, On structure of δ-sets. Bull. Calcutta Math. Soc. **83**, 281–290 (1991)
9. Ch. Chattopadhyay, U.K. Roy, δ-sets, irresolvable and resolvable spaces. Math. Slovaca **42**, 371–378 (1992)
10. H. Choda, K. Matoba, On a theorem of Levine. Proc. Jpn. Acad. **37**, 462–463 (1961)
11. H.H. Corson, E. Michael, Metrizability of countable unions. Ill. J. Math. **8**, 351–360 (1964)
12. Á. Császár, *Foundations of General Topology* (Pergamon Press, London, 1963)
13. Á. Császár, Generalized open sets. Acta Math. Hungar. **75**, 65–87 (1997)
14. Á. Császár, On the γ-interior and γ-closure of a set. Acta Math. Hungar. **80**, 89–93 (1998)
15. Á. Császár, γ-quasi-open sets. Stud. Sci. Math. Hungar. **38**, 171–176 (2001)
16. Á. Császár, Remarks on γ-quasi-open sets. Stud. Sci. Math. Hungar. **39**, 137–141 (2002)
17. Á. Császár, Further remarks on the formula for the γ-interior. Acta Math. Hungar. **113**, 325–332 (2006)
18. Á. Császár, Remarks on quasi-topologies. Acta Math. Hungar. **119**, 197–200 (2008)
19. B.A. Davey, H.A. Priestley, *Introduction to Lattices and Order* (Cambridge University Press, Cambridge, 2002)
20. A.S. Davis, Indexed systems of neighbordoods for general topological spaces. Am. Math. Monthly **68**, 886–893 (1961)
21. K. Dlaska, N. Ergun, M. Ganster, On the topology generated by semi-regular sets. Indian J. Pure Appl. Math. **25**, 1163–1170 (1994)

22. J. Dontchev, Survey on preopen sets. Meetings on Topological Spaces, Theory and Applications, Yatsushiro College of Technology, Kumamoto, Japan (1998), 18 pp.
23. Z. Duszyński, T. Noiri, Semi-open, semi-closed sets and semi-continuity of functions. Math. Pannon. **23**, 195–200 (2012)
24. V.A. Efremovič, The geometry of proximity. Mat. Sb. **31**, 189–200 (1952) (Russian)
25. V.A. Efremović, A.S. Švarc, A new definition of uniform spaces. Metrization of proximity spaces. Dokl. Acad. Nauk. SSSR **89**, 393–396 (1953) (Russian)
26. N. Elez, O. Papaz, The new operators in topological spaces. Math. Moravica **17**, 63–68 (2013)
27. P. Fletcher, W.F. Lindgren, *Quasi-Uniform Spaces* (Marcel Dekker, New York, 1982)
28. M. Ganster, Preopen sets and resolvable spaces. Kyungpook J. **27**, 135–143 (1987)
29. M. Ganster, I.L. Reilly, M.K. Vamanamurthy, Remarks on locally closed sets. Math. Pannon. **3**, 107–113 (1992)
30. B. Ganter, R. Wille, *Formal Concept Analysis* (Springer, Berlin, 1999)
31. R. Gargouri, A. Rezgui, A unification of weakening of open and closed subsets in a topological spaces. Bull. Malays. Math. Sci. Soc. **40**, 1219–1230 (2017)
32. S. Givant, P. Halmos, *Introduction to Boolean Algebras* (Springer, Berlin, 2009)
33. T. Glavosits, Generated preorders and equivalences. Acta Acad. Paed. Agrienses, Sect. Math. **29**, 95–103 (2002)
34. W.H. Gottschalk, Intersection and closure. Proc. Am. Math. Soc. **4**, 470–473 (1953)
35. W. Hunsaker, W. Lindgren, Construction of quasi-uniformities. Math. Ann. **188**, 39–42 (1970)
36. D.H. Hyers, On the stability of the linear functional equation. Proc. Nat. Acad. Sci. U.S.A **27**, 222–224 (1941)
37. Y. Isomichi, New concept in the theory of topological spaces–Supercondensed set, subcondensed set, and condensed set. Pac. J. Math. **38**, 657–668 (1971)
38. Y.B. Jun, S.W. Jeong, H.j. Lee, J.W. Lee, Applications of pre-open sets. Appl. Gen. Top. **9**, 213–228 (2008)
39. S.-M. Jung, Interiors and closure of sets and applications. Int. J. Pure Math. **3**, 41–45 (2016)
40. S.-M. Jung, D. Nam, Some properties of interior and closure in general topology. Mathematics **7**, 624 (2019)
41. J.L. Kelley, *General Topology* (Van Nostrand Reinhold Company, New York, 1955)
42. H. Kenyon, Two theorems on relations. Trans. Am. Math. Soc. **107**, 1–9 (1963)
43. V.L. Kljushin, Al bayati J.H. Hussein, On simply-open sets. Vestnik UDC **3**, 34–38 (2011). (Russian)
44. K. Kuratowski, Sur l'opération $\overline{A}$ de l'analysis situs. Fund. Math. **3**, 182–199 (1922) (An English translation: On the operation $\overline{A}$ in analysis situs, prepared by M. Bowron in 2010, is available on the Internet)
45. K. Kuratowski, *Topology I* (Academic Press, New York, 1966)
46. J. Kurdics, A note on connection properties. Acta Math. Acad. Paedagog. Nyházi. **12**, 57–59 (1990)
47. J. Kurdics, Á. Száz, Well-chainedness characterizations of connected relators. Math. Pannon. **4**, 37–45 (1993)
48. N. Levine, On the commutivity of the closure and interior operators in topological spaces. Am. Math. Montly **68**, 474–477 (1961)
49. N. Levine, Semi-open sets and semi-continuity in topological spaces. Am. Math. Monthly **70**, 36–41 (1963)
50. N. Levine, Some remarks on the closure operator in topological spaces. Am. Math. Monthly **70**, 553 (1963)
51. N. Levine, On uniformities generated by equivalence relations. Rend. Circ. Mat. Palermo **18**, 62–70 (1969)
52. N. Levine, On Pervin's quasi uniformity. Math. J. Okayama Univ. **14**, 97–102 (1970)
53. J. Mala, Relators generating the same generalized topology. Acta Math. Hungar. **60**, 291–297 (1992)
54. J. Mala, Á. Száz, Properly topologically conjugated relators. Pure Math. Appl. Ser. B **3**, 119–136 (1992)

55. J. Mala, Á. Száz, Modifications of relators. Acta Math. Hungar. **77**, 69–81 (1997)
56. A.S. Mashhour, M.E. Abd El-Monsef, S.N. El-Deeb, On precontinuous and weak precontinuous mappings. Proc. Math. Phys. Soc. Egypt **53**, 47–53 (1982)
57. S.A. Naimpally, B.D. Warrack, *Proximity Spaces* (Cambridge University Press, Cambridge, 1970)
58. H. Nakano, K. Nakano, Connector theory. Pac. J. Math. **56**, 195–213 (1975)
59. A.A. Nasef, R. Mareay, More on simplly open sets and its applications. South Asian J. Math. **5**, 100–108. (2015)
60. A.A. Nasef, R. Mareay, Ideals and some applications of simply open sets. J. Adv. Math. **13**, 7264–7271 (2017)
61. A. Neubrunnová, On transfinite sequences of certain types of functions. Acta Fac. Rer. Natur. Univ. Commun. Math. **30**, 121–126 (1975)
62. O. Njåstad, On some classes of nearly open sets. Pac. J. Math. **15**, 195–213 (1965)
63. G. Pataki, Supplementary notes to the theory of simple relators. Radovi Mat. **9**, 101–118 (1999)
64. G. Pataki, On the extensions, refinements and modifications of relators. Math. Balk. **15**, 155–186 (2001)
65. G. Pataki, Á. Száz, A unified treatment of well-chainedness and connectedness properties. Acta Math. Acad. Paedagog. Nyházi. (N.S.) **19**, 101–165 (2003)
66. W.J. Pervin, Quasi-uniformization of topological spaces. Math. Ann. **147**, 316–317 (1962)
67. Th.M. Rassias, M. Salih, Á. Száz, Characterizations of generalized topologically open sets in relator spaces, in *Recent Trends on Pure and Applied Mathematics, Special Issue of the Montes Taurus*, ed. by G.V. Milovanovic, Thm. M. Rassias, Y. Simsek. J. Pure Appl. Math., Dedicated to Professor Hari Mohan Srivastava on the occasion of his 80th Birthday, Montes Taurus J. Pure Appl. Math. **3**, 39–94 (2021)
68. Th.M. Rassias, M. Salih, Á. Száz, Set-theoretic properties of generalized topologically open sets in relator spaces, in *Mathematical Analysis in Interdisciplinary Research*, ed. by I.N. Parasidis, E. Providas, Th.M. Rassias, to appear
69. M. Salih, Á. Száz, Generalizations of some ordinary and extreme connectedness properties of topological spaces to relator spaces. Electron. Res. Arch. **28**, 471–548 (2020)
70. P. Sivagami, Remarks on γ-interior. Acta Math. Hungar. **119**, 81–94 (2008)
71. Yu.M. Smirnov, On proximity spaces. Math. Sb. **31**, 543–574 (1952) (Russian)
72. M.H. Stone, Application of the theory of Boolean rings to general topology. Trans. Am. Math. Soc. **41**, 374–481 (1937)
73. Á. Száz, Basic tools and mild continuities in relator spaces. Acta Math. Hungar. **50**, 177–201 (1987)
74. Á. Száz, Directed, topological and transitive relators. Publ. Math. Debrecen **35**, 179–196 (1988)
75. Á. Száz, Relators, Nets and Integrals. Unfinished doctoral thesis, Debrecen (1991), 126 pp.
76. Á. Száz, Structures derivable from relators. Singularité **3**, 14–30 (1992)
77. Á. Száz, Refinements of relators. Tech. Rep., Inst. Math., Univ. Debrecen, vol. 76 (1993), 19 pp.
78. Á. Száz, Cauchy nets and completeness in relator spaces. Colloq. Math. Soc. János Bolyai **55**, 479–489 (1993)
79. Á. Száz, Neighbourhood relators. Bolyai Soc. Math. Stud. **4**, 449–465 (1995)
80. Á. Száz, Uniformly, proximally and topologically compact relators. Math. Pannon. **8**, 103–116 (1997)
81. Á. Száz, Somewhat continuity in a unified framework for continuities of relations. Tatra Mt. Math. Publ. **24**, 41–56 (2002)
82. Á. Száz, Upper and lower bounds in relator spaces. Serdica Math. J. **29**, 239–270 (2003)
83. Á. Száz, Rare and meager sets in relator spaces. Tatra Mt. Math. Publ. **28**, 75–95 (2004)
84. Á. Száz, Galois-type connections on power sets and their applications to relators. Tech. Rep., Inst. Math., Univ. Debrecen 2005/2 (2005), 38 pp.
85. Á. Száz, Minimal structures, generalized topologies, and ascending systems should not be studied without generalized uniformities. Filomat **21**, 87–97 (2007)

86. Á. Száz, Galois type connections and closure operations on preordered sets. Acta Math. Univ. Comenian. (N.S.) **78**, 1–21 (2009)
87. Á. Száz, Inclusions for compositions and box products of relations. J. Int. Math. Virt. Inst. **3**, 97–125 (2013)
88. Á. Száz, A particular Galois connection between relations and set functions. Acta Univ. Sapientiae, Math. **6**, 73–91 (2014)
89. Á. Száz, Generalizations of Galois and Pataki connections to relator spaces. J. Int. Math. Virtual Inst. **4**, 43–75 (2014)
90. Á. Száz, Basic tools, increasing functions, and closure operations in generalized ordered sets, in *Contributions in Mathematics and Engineering*, ed. by P.M. Pardalos, Th.M. Rassias. In Honor of Constantion Caratheodory (Springer, Berlin, 2016), pp. 551–616
91. Á. Száz, Four general continuity properties, for pairs of functions, relations and relators, whose particular cases could be investigated by hundreds of mathematicians. Tech. Rep., Inst. Math., Univ. Debrecen, 2017/1 (2017), 17 pp.
92. Á. Száz, The closure-interior Galois connection and its applications to relational equations and inclusions. J. Int. Math. Virt. Inst. **8**, 181–224 (2018)
93. Á. Száz, Corelations are more powerful tools than relations, in *Applications of Nonlinear Analysis*, ed. by Th.M. Rassias. Optimization and Its Applications, vol. 134 (Springer, Berlin, 2018), pp. 711–779
94. Á. Száz, Relationships between inclusions for relations and inequalities for corelations. Math. Pannon. **26**, 15–31 (2018)
95. Á. Száz, Galois and Pataki connections on generalized ordered sets. Earthline J. Math. Sci. **2**, 283–323 (2019)
96. Á. Száz, Birelator spaces are natural generalizations of not only bitopological spaces, but also ideal topological spaces, in *Mathematical Analysis and Applications, Springer Optimization and Its Applications*, ed. by Th.M. Rassias, P.M. Pardalos, vol. 154 (Springer, Switzerland, 2019), pp. 543–586
97. W.J. Thron, *Topological Structures* (Holt, Rinehart and Winston, New York, 1966)
98. H. Tietze, Beiträge zur allgemeinen Topologie I. Axiome für verschiedene Fassungen des Umgebungsbegriffs. Math. Ann. **88**, 290–312 (1923)
99. A. Weil, Sur les espaces á structure uniforme et sur la topologie générale. Actual. Sci. Ind., vol. 551 (Herman and Cie, Paris 1937)

Graphical Mean Curvature Flow

Andreas Savas-Halilaj

Abstract In this survey article, we discuss recent developments on the mean curvature flow of graphical submanifolds, generated by smooth maps between Riemannian manifolds. We will see interesting applications of this technique, in the understanding of the homotopy type of maps between manifolds.[1][2]

1 Introduction

Let $f\colon M \to N$ be a smooth map between two manifolds M and N. It is a fundamental problem to find *canonical representatives* in the homotopy class of f. By a canonical representative is usually meant a map in the homotopy class of the given map f which is a critical point of a suitable functional. In the mid-1960s, Eells and Sampson [34] introduced the *harmonic maps* as critical points of the energy density, to attack the aforementioned problem.

One possible approach to construct harmonic maps is via the *harmonic map heat flow*. If M is compact and N is negatively curved, in [34] Eells and Sampson were able to prove long-time existence and convergence of the flow, showing that under these assumptions one finds harmonic representatives in a given homotopy class. In general, one can neither expect long-time existence nor convergence of this flow. For example, the situation is very complicated in the case of maps between spheres.

[1] These notes are based partly on a series of lectures delivered by the author at the Chern Institute of Mathematics held in Tianjin-China in November 2019.

[2] The author would like to acknowledge support by the General Secretariat for Research and Technology (GSRT) and the Hellenic Foundation for Research and Innovation (HFRI) Grant No:133.

A. Savas-Halilaj (✉)
Department of Mathematics, University of Ioannina, Ioannina, Greece
e-mail: ansavas@uoi.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_20

There is another important functional that we may consider in the space of smooth maps. Given a map $f : M \to N$ between Riemannian manifolds, let us denote its *graph* in the product space $M \times N$ by

$$\Gamma(f) = \{(x, f(x)) \in M \times N : x \in M\}.$$

Following the terminology introduced by Schoen [79], a map whose graph is minimal submanifold is called *minimal map*. Therefore, minimal maps are critical points of the volume functional.

In this survey, among others, we will discuss deformation of graphical submanifolds via the *mean curvature flow*. Before stating the problems that we would like to deal with, let us provide some basic facts and definitions. Let M be a smooth m-dimensional manifold, $T > 0$ a positive number and $F : M \times [0, T) \to P$ a smooth time-dependent family of immersions of M into a Riemannian manifold P. We say that F evolves in time under the *mean curvature flow* if it satisfies the evolution equation

$$dF(\partial_t)(x, t) = H(x, t)$$

for any $(x, t) \in M \times [0, T)$, where $H(x, t)$ stands for the mean curvature vector at the point x of the immersion $F(\cdot, t) : M \to P$. It is a well-known fact that if M is compact and $F_0 : M \to P$ is an immersion, then the initial value problem for the mean curvature flow admits a unique smooth solution on a maximal time interval $[0, T_{\max})$, where $0 < T_{\max} \leq \infty$. Suppose now that P is the product manifold $M \times N$ and F_0 is the graph of a map $f : M \to N$. Notice that long as the submanifolds deformed under mean curvature flow remain graphical, one obtains a smooth family of maps which belong to the homotopy class of the map f. In the case of long-time existence and convergence of the flow, we obtain a smooth homotopy from f to a minimal map.

The first result regarding evolutions by mean curvature of graphical submanifolds is due to Ecker and Huisken [33]. They proved long-time existence of entire graphical hypersurfaces in $\mathbb{R}^{n+1}$. Moreover, Ecker and Huisken proved convergence to a flat subspace, if the growth rate at infinity of the initial graphical submanifold is linear. On the other hand, in higher codimensions, the complexity of the normal bundle makes the situation more complicated. Results analogous to that of Ecker and Huisken are not available any more without further assumptions. However, the ideas developed in the paper of Ecker and Huisken opened a new era for the study of the mean curvature flow of submanifolds in Riemannian manifolds of arbitrary codimension; see for example [12, 13, 16–18, 60–62, 64, 75, 77, 78, 85, 87–91, 94–96, 98–100].

This new deformation of maps between Riemannian manifolds via the mean curvature flow has been used in order to have a better understanding of the relation between the *k-dilation* Dil_k and the homotopy type of maps. In order to be precise, let us recall at first the following definition:

Definition 1 Let $f : M \to N$ be a map between two Riemannian manifolds. We say that $Dil_k(f) \leq \alpha$ if f maps each k-dimensional submanifold $\Sigma \subset M$ to an image with k-dimensional volume at most $\alpha \cdot \mathcal{H}^k(\Sigma)$, where $\mathcal{H}^k(\Sigma)$ stands for the k-dimensional Hausdorff measure of Σ. In particular, we say that f is *area decreasing* if $Dil_2(f) \leq 1$, *strictly area decreasing* if $Dil_2(f) = 1$, and *area preserving* if $Dil_2(f) = 1$.

Roughly speaking, the k-dilation measures how much the map $f : M \to N$ contracts k-dimensional volumes. Gromov in [38] realized that there is a close relationship between the 1-dilation of a map and its homotopy type. For instance, he proved that if f is a map from $\mathbb{S}^m$ to $\mathbb{S}^m$, then its degree is at most $Dil_1^m(f)$ and this bound is sharp up to a constant factor. Motivated by this result, in [40, 41] Gromov proposed the following:

Problem 1 Let $f : \mathbb{S}^m \to \mathbb{S}^n$ be a smooth map between euclidean spheres. Is there a number $\varepsilon(k, m, n)$ such that if $Dil_k(f) < \varepsilon$ would imply that f is null-homotopic?

Tsui and Wang in [91] proved using the mean curvature flow that smooth strictly area decreasing maps $f : \mathbb{S}^m \to \mathbb{S}^n$ can be smoothly deformed to a constant map. Guth [42] proved this result cannot be extended in the case of maps with k-dilation strictly less than 1, if $k \geq 3$. The result of Tsui and Wang was generalized by Lee and Lee in [60]. In the matter of fact, they proved that any strictly area decreasing map between compact Riemannian manifolds M and N whose sectional curvatures are bounded by $\sec_M \geq \sigma_1$ and $\sigma_2 \geq \sec_N$, where σ_1, σ_2 are two real constants such that $\sigma_1 \geq \sigma_2 > 0$ or $\sigma_1 > 0 \geq \sigma_2$, is homotopic by mean curvature flow to a constant map. We would like to point out here that the curvature assumptions can be relaxed to

$$\sec_M > -\sigma \quad \text{and} \quad \mathrm{Ric}_M \geq (m-1)\sigma \geq (m-1)\sec_N,$$

where σ is a positive constant number, as it was shown in [75] by Savas-Halilaj and Smoczyk.

In the case of a smooth area decreasing map $f : M \to N$ between two compact Riemann surfaces M and N of the same constant sectional curvature σ, we have a complete picture of the behaviour of the mean curvature flow. *It turns out that, under the mean curvature flow, such a map either instantly becomes strictly area decreasing or it was and remains an area preserving map. Moreover, the mean curvature flow preserves the graphical property, exists for all time, and converges to a minimal surface Σ_∞ of the product $M \times N$. Additionally:*

(I) *If the evolved graphs are generated by strictly area decreasing maps then:*

  (a) *If $\sigma > 0$, then Σ_∞ is the graph of a constant map.*
  (b) *If $\sigma = 0$, then Σ_∞ is the graph of an affine minimal map.*

(II) *If the evolved graphs are generated by area preserving maps then:*

(a) *If* $\sigma > 0$, *then* Σ_∞ *is the graph of an isometry.*
(b) *If* $\sigma = 0$, *then* Σ_∞ *is the graph of an affine minimal diffeomorphism.*

The first steps in the proof of the above result were made in the seminal works of Smoczyk [85] and Wang [94, 95, 100], where the area preserving case was investigated. The strictly area decreasing case was first treated by Tsui and Wang [91], in the positive case, and completed recently by Savas-Halilaj and Smoczyk in [78]. The primary goal of this survey is to present a unified proof of this result, based in [78].

From the results of Wang [94, 95, 100], we get another proof of Smale's Theorem [84] which says that any diffeomorphism of $\mathbb{S}^2$ can be smoothly deformed into an isometry. Let us mention here that, according to a deep theorem of Hatcher [49], any diffeomorphism of $\mathbb{S}^3$ can be deformed into an isometry of $\mathbb{S}^3$. Such a result is not expected for spheres of dimension greater or equal than 4; see for example [28]. However, the following problem is challenging:

Problem 2 Let $f : \mathbb{S}^m \to \mathbb{S}^m$, $m \geq 4$, be a smooth diffeomorphism. Under which conditions f can be smoothly deformed into an isometry of the sphere?

Another interesting problem is the investigation of the symplectomorphism group of the complex projective space $\mathbb{CP}^m$. Gromov [39] proved that the bi-holomorphic group of $\mathbb{CP}^2$ is a deformation retract of its symplectomorphism group. A natural problem is to determine whether a similar result holds for $\mathbb{CP}^m$ with $m \geq 3$. In the matter of fact, the following problem is still open:

Problem 3 Let $f : \mathbb{CP}^m \to \mathbb{CP}^m$, $m \geq 3$, be a smooth symplectomorphism. Is it true that the mean curvature flow deforms the graph $\Gamma(f)$ of f smoothly to the graph $\Gamma(g)$ of bi-holomorphic map $g : \mathbb{CP}^m \to \mathbb{CP}^m$? Is it true that any minimal symplectomorphism $f : \mathbb{CP}^m \to \mathbb{CP}^m$ is a bi-holomorphic isometry?

Medoš and Wang in [64] made some contribution by giving an affirmative answer to the above problem under the additional assumption that the singular values of the differential of the symplectomorphism are close to 1.

The paper is structured as follows. In Section 2 we set up the notation and recall basic facts from submanifold geometry. In Section 3 we discuss minimal submanifolds in euclidean spaces. We introduce the generalized Gauss map and prove the Ruh-Vilms Theorem. Section 4 describes the class of graphical submanifolds and review some Bernstein-type theorems. Section 5 is devoted to the maximum principle for scalar and systems of PDEs. In Section 6, we introduce the mean curvature flow, prove short-time existence, and derive various basic evolution equations. Section 7 describes how to built smooth singularity models for the mean curvature flow. Section 8 combines results from the previous sections to prove our main result.

2 Riemannian Submanifolds

In this section we set up the notation and recall some basic facts from submanifold geometry. We closely follow the exposition in [5, 29, 55, 59, 92, 102].

2.1 Notation and Conventions

Let M be a m-dimensional manifold and (E, π, M) a vector bundle of rank k over M. We often denote the bundles only by its total space E. The fiber of E at a point $x \in M$ is denoted by E_x, the tangent space of M at a point $x \in M$ will be denoted by $T_x M$ and the space of sections of E is denoted by $\Gamma(E)$. For the tangent bundle of M, we use the symbol TM. Sections of the tangent bundle are called *vector fields* and usually $\Gamma(TM)$ is denoted by $\mathfrak{X}(M)$. A smooth map $T : E \to V$ between two vector bundles E and V over M which maps the fiber E_x linearly to V_x, for any $x \in M$ is called *bundle morphism*. If additionally, T is bijective we call it *bundle isomorphism*.

Definition 2 A (*linear*) *connection* on a vector bundle E is a map $\nabla^E \colon \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$, written $\nabla^E(v, \phi) = \nabla^E_v \phi$, satisfying the properties:

(a) For any $v_1, v_2 \in \mathfrak{X}(M)$ and $\phi \in \Gamma(E)$, it holds

$$\nabla^E_{v_1+v_2}\phi = \nabla^E_{v_1}\phi + \nabla^E_{v_2}\phi.$$

(b) For any $v \in \mathfrak{X}(M)$, $f \in C^\infty(M)$ and $\phi \in \Gamma(E)$, it holds

$$\nabla^E_{fv}\phi = f\nabla^E_v\phi.$$

(c) For any $v \in \mathfrak{X}(M)$, $f \in C^\infty(M)$ and $\phi_1, \phi_2 \in \Gamma(E)$, it holds

$$\nabla^E_v(\phi_1 + \phi_2) = \nabla^E_v\phi_1 + \nabla^E_v\phi_2.$$

(d) For any $v \in \mathfrak{X}(M)$, $\phi \in \Gamma(E)$ and $f \in C^\infty(M)$, it holds

$$\nabla^E_v(f\phi) = (vf)\phi + f\nabla^E_v\phi.$$

For any $\phi \in \Gamma(E)$ and $x_0 \in M$, the value $\nabla^E\phi_x|_{x_0}$ of the quantity $\nabla^E_v\phi$ at $x_0 \in M$ depends only on the value of v at x_0 and on the restriction of ϕ along a curve passing through x_0 with speed v. If $\phi_1, \phi_2 \in \Gamma(E)$ coincide on a neighbourhood of $x_0 \in M$, then

$$\nabla^E_{v_1}\phi_1|_{x_0} = \nabla^E_{v_2}\phi_2|_{x_0},$$

for any pair of vector fields $v_1, v_2 \in \mathfrak{X}(M)$ with $v_1|_{x_0} = v_2|_{x_0}$.

Definition 3 A section $\phi \in \Gamma(E)$ is said to be *parallel* with respect to ∇^E if, for any vector field v on M, it holds $\nabla^E_v \phi = 0$.

We can define higher derivatives of sections of a vector bundle over a manifold M whose tangent bundle TM is equipped with a connection.

Definition 4 Suppose that M is a smooth manifold and E a vector bundle over M. Let ∇^M be a connection of TM and ∇^E a connection of E. For any pair $v_1, v_2 \in \mathfrak{X}(M)$, the map $\nabla^2_{v_1,v_2} : \Gamma(E) \to \Gamma(E)$, given by

$$\nabla^2_{v_1,v_2}\phi = \nabla^E_{v_1}\nabla^E_{v_2}\phi - \nabla^E_{\nabla^M_{v_1} v_2}\phi,$$

is called the *second covariant derivative* of ϕ, with respect to the directions v_1 and v_2. By coupling the connections ∇^M and ∇^E, one may define, the k-th derivative ∇^k of a section ϕ in $\Gamma(E)$.

To each connection there is associated an important operator, which measures the non commutativity of the covariant derivatiation.

Definition 5 The operator $R^E : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$, defined by

$$R^E(v_1, v_2, \phi) = \nabla^2_{v_1,v_2}\phi - \nabla^2_{v_2,v_1}\phi,$$

for any $v_1, v_2 \in \mathfrak{X}(M)$ and $\phi \in \Gamma(E)$, is called the *curvature operator* of ∇^E.

Now let us turn our attention to vector bundles equipped with a Riemannian metric structure.

Definition 6 A *Riemannian metric* on a vector bundle E of rank k over the manifold M is a smooth map $g_E : \Gamma(E) \times \Gamma(E) \to C^\infty(M)$, such that its restriction to the fibers is a positive definite inner product.

Definition 7 A connection ∇^E is called *compatible with the Riemannian metric* g_E or *metric compatible* if it satisfies

$$v g_E(\phi_1, \phi_2) = g_E\big(\nabla^E_v \phi_1, \phi_2\big) + g_E\big(\phi_1, \nabla^E_v \phi_2\big),$$

for any $v \in \mathfrak{X}(M)$ and $\phi_1, \phi_2 \in \Gamma(E)$. A vector bundle E endowed with these structures is called *Riemannian vector bundle endowed with a compatible linear connection*.

We say that a set of sections $\{\phi_1, \dots, \phi_k\}$ forms an *orthonormal frame*, with respect to g_E if and only if $g_E(\phi_i, \phi_j) = \delta_{ij}$, for any $i, j \in \{1, \dots, k\}$. In particular, around any point x_0 of M it is possible to find a local orthonormal frame $\{\phi_1, \dots, \phi_k\}$ such that

$$\nabla_v \phi_i|_{x_0} = 0$$

for any tangent vector v. Such frames are called *normal* or *geodesic frames*.

Let us restrict ourselves at the tangent bundle TM of M. Given a Riemannian metric g on M, there is a unique connection ∇, referred as the *Levi-Civita connection*, which is compatible with the Riemannian metric. More precisely, ∇ is given by the *Koszul formula*

$$\begin{aligned} 2g\big(\nabla_{v_1} v_2, v_3\big) = {} & v_1\big(g(v_2, v_3)\big) + v_2\big(g(v_1, v_3)\big) - v_3\big(g(v_1, v_2)\big) \\ & + g\big([v_1, v_2], v_3\big) - g\big([v_1, v_3], v_2\big) - g\big([v_2, v_3], v_1\big), \end{aligned}$$

for all $v_1, v_2, v_3 \in \mathfrak{X}(M)$. The Levi-Civita also satisfy

$$\nabla_{v_1} v_2 - \nabla_{v_2} v_1 = [v_1, v_2],$$

for any $v_1, v_2 \in \mathfrak{X}(M)$.

Denote by R the curvature operator with respect to the connection ∇. Combining R with g we obtain a (4,0)-tensor which, for simplicity, we again denote with the letter R. More precisely,

$$R(v_1, v_2, v_3, v_4) = -g(R(v_1, v_2, v_3), v_4),$$

for any $v_1, v_2, v_3, v_4 \in \mathfrak{X}(M)$. If v_1, v_2 are linearly independent vectors, then

$$\sec(v_1, v_2) = \frac{R(v_1, v_2, v_1, v_2)}{g(v_1, v_1)g(v_2, v_2) - g(v_1, v_2)^2},$$

is called the *sectional curvature* of the plane spanned by the vectors v_1 and v_2. By contracting the operator R with g we obtain the *Ricci operator* Ric and the *scalar curvature* scal, i.e.,

$$\mathrm{Ric}(v_1, v_2) = \sum_{i=1}^{m} R(v_1, e_i, v_2, e_i) \quad \text{and} \quad \mathrm{scal} = \sum_{i=1}^{m} \mathrm{Ric}(e_i, e_i),$$

where $v \in \mathfrak{X}(M)$ and $\{e_1, \ldots, e_m\}$ is a local orthonormal frame on M.

Remark 1 One can use the operations of Linear Algebra to produce new vector bundles from given ones. For example, if E and V are vector bundles over a manifold M, then E^*, $E \times V$, $E \otimes V$, $E \oplus V$, $\mathrm{Hom}(E; V)$, $\Lambda^r(V)$ and $\mathrm{Sym}^r(V)$ gives rise to new bundles over M. If M is endowed with a Riemannian metric, then this metric and its Levi-Civita connections extends in a natural way to all the aforementioned bundles; for more details see [59] or [102].

2.2 The Pull-back Bundle

Let M and N be two manifolds, (E, π, N) is a vector bundle of rank k over N and suppose that $f\colon M \to N$ is a smooth map. The map f induces a new vector bundle of rank k over M. Indeed, take as total space

$$f^*E = \big\{(x, \xi) : x \in M,\ \xi \in E_{f(x)}\big\}$$

and as projection the map $\pi_f : f^*E \to M$ given by $\pi_f(x, \xi) = x$. The space $\Gamma(f^*E)$ contains all sections of E with base point at $f(M)$ and inherit naturally a vector space structure from $E_{f(x)}$, given by

$$(x, \xi) + (x, \eta) = (x, \xi + \eta) \quad \text{and} \quad \lambda(x, \xi) = (x, \lambda\xi).$$

The triple (f^*E, π_f, M) carries the structure of a vector bundle over M. This bundle is called the *pull-back* or the *induced by f vector bundle* on M.

Suppose that h is a Riemannian metric on E and ∇^E is a metric compatible connection. The map f induces a connection ∇^{f^*E} on the pull-back bundle which is defined as follows: Let $\{\vartheta_1, \dots, \vartheta_k\}$ be a local orthonormal frame field of E in a neighbourhood of the point $f(x) \in N$. Then, any section $\phi \in \Gamma\big(f^*E\big)$ can be written in the form

$$\phi|_x = \Big(x, \sum\nolimits_{\alpha=1}^{k} \phi^\alpha(x)\vartheta_\alpha|_{f(x)}\Big) \cong \sum\nolimits_{\alpha=1}^{k} \phi^\alpha(x)\vartheta_\alpha|_{f(x)},$$

where ϕ^α, $\alpha \in \{1, \dots, k\}$, are the *components* of ϕ with respect to the given orthonormal frame field. Define now the induced connection by

$$\nabla^{f^*E}_v \phi|_x = \sum_{\alpha=1}^{k} \big(v\phi^\alpha\big)\,\vartheta_\alpha|_{f(x)} + \sum_{\alpha=1}^{k} \phi^\alpha \nabla^E_{df(v)}\vartheta_\alpha|_{f(x)},$$

for $x \in M$ and $v \in T_xM$. One can easily show that the curvature operator R^{f^*E} of ∇^{f^*E} is given by

$$R^{f^*E}\big(v_1, v_2, \phi|_x\big) = R^E\big(df(v_1), df(v_2), \phi|_x\big),$$

for any $x \in M$, $v_1, v_2 \in T_xM$ and $\phi \in E|_{f(x)}$.

Let us discuss the case where $f : (M, g, \nabla^g) \to (N, h, \nabla^h)$ is a map between Riemannian manifolds. The restriction of h on f^*TN, induces a Riemannian metric on f^*TN, which is compatible with the pull-back connection, that is

$$vh(\phi_1, \phi_2) = h\big(\nabla^{f^*TN}_v \phi_1, \phi_2\big) + h\big(\phi_1, \nabla^{f^*TN}_v \phi_2\big).$$

Moreover, for $v_1, v_2 \in \mathfrak{X}(M)$, it holds

$$\nabla^{f^*TN}_{v_1} df(v_2) - \nabla^{f^*TN}_{v_2} df(v_1) = df\big([v_1, v_2]\big).$$

Definition 8 The *Hessian* of a map $f : (M, g, \nabla^g) \to (N, h, \nabla^h)$ is defined to be the symmetric tensor $B : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \Gamma(f^*E)$ given by

$$B(v_1, v_2) = \nabla^{f^*TN}_{v_1} df(v_2) - df(\nabla^g_{v_1} v_2),$$

for any $v_1, v_2 \in \mathfrak{X}(M)$. The trace of B with respect to g is denoted by $\Delta_{g,h} f$ and is called the *Laplacian* of f. If the Laplacian of f is zero, then f is called a *harmonic map.*

2.3 The Second Fundamental Form

Consider Riemannian manifolds (M, g, ∇^g) and (N, h, ∇^h) of dimension m and n, respectively, with $m \leq n$. A map $f : M \to N$ is called an *isometric immersion* if and only if $f^*h = g$. For simplicity, we often denote both metrics g and h by $\langle \cdot, \cdot \rangle$. At every $x \in M$, we have the orthogonal decomposition

$$T_{f(x)}N = df_x(T_xM) \oplus N_{f(x)}M,$$

where $N_{f(x)}M$ is the orthogonal complement of $df_x(T_xM)$ with respect to h. The union NM of all normal spaces form a vector bundle of rank $n - m$ over M which is called the *normal bundle*. According to the above decomposition, any section $v \in \Gamma(f^*TN)$ can be decomposed in a unique way in the form

$$v = v^\top + v^\perp,$$

where $v^\top$ is the *tangent* and $v^\perp$ is the *normal* part of v along the submanifold. A well known fact in submanifold theory is that

$$\big(\nabla^{f^*TN}_{v_1} df(v_2)\big)^\top = df\big(\nabla^g_{v_1} v_2\big), \tag{1}$$

for any $v_1, v_2 \in \mathfrak{X}(M)$. In submanifold theory, the Hessian of f is denoted by the letter A, i.e., we have

$$A(v_1, v_2) = \nabla^{f^*TN}_{v_1} df(v_2) - df(\nabla^g_{v_1} v_2).$$

The tensor A is called the *second fundamental form* of f. If ξ is a normal vector, then the tensor A^ξ given by

$$A^\xi(v_1, v_2) = \langle A(v_1, v_2), \xi \rangle,$$

for any tangent vectors v_1, v_2, is called *shape operator with respect to* ξ. The *Weingarten operator* A_ξ associated with ξ is defined by

$$\langle A_\xi v_1, v_2\rangle = A^\xi(v_1, v_2) = \langle A(v_1, v_2), \xi\rangle.$$

The Laplacian of f or, equivalently, the trace of A with respect to g is called the (*unormalized*) *mean curvature* and is denoted by the letter H, that is

$$H = \operatorname{trace}_g A.$$

Definition 9 A submanifold with zero mean curvature is called *minimal.*

The restriction of h on NM gives rise to a Riemannian metric on the normal bundle. Moreover, the restriction of ∇^h on NM induces a connection $\nabla^\perp$ on NM which is compatible with the metric; i.e., just define

$$\nabla^\perp_v \xi = \big(\nabla^N_v \xi\big)^\perp.$$

The curvature tensor of the normal bundle is denoted by $R^\perp$ and is given by

$$R^\perp(v_1, v_2, \xi) = \nabla^\perp_{v_1}\nabla^\perp_{v_2}\xi - \nabla^\perp_{v_2}\nabla^\perp_{v_1}\xi - \nabla^\perp_{[v_1,v_2]}\xi.$$

As usual, we can form from $R^\perp$ a $C^\infty(M)$-valued tensor which we denote again by $R^\perp$, that is

$$R^\perp(v_1, v_2, \xi, \eta) = -\langle R^\perp(v_1, v_2, \xi), \eta\rangle.$$

The Riemann curvature operator R of M, the curvature operator $\widetilde{R}$ of N and the normal curvature $R^\perp$ are related to the second fundamental form A through the *Gauss-Codazzi-Ricci equations*:

(a) **Gauss equation:**

$$\begin{aligned} R(v_1, v_2, v_3, v_4) = {} & \widetilde{R}\big(df(v_1), df(v_2), df(v_3), df(v_4)\big) \\ & + \langle A(v_1, v_3), A(v_2, v_4)\rangle - \langle A(v_2, v_3), A(v_1, v_4)\rangle. \end{aligned}$$

(b) **Codazzi equation:**

$$\big(\nabla^\perp_{v_1} A\big)(v_2, v_3) - \big(\nabla^\perp_{v_2} A\big)(v_1, v_3) = \big(\widetilde{R}\big(df(v_1)df(v_2), df(v_3)\big)\big)^\perp.$$

(c) **Ricci equation:**

$$R^\perp(v_1, v_2, \xi, \eta) = \widetilde{R}\big(df(v_1), df(v_2), \xi, \eta\big)$$

$$+\sum_k \big(A^\xi(v_1,e_k)A^\eta(v_2,e_k) - A^\eta(v_1,e_k)A^\xi(v_2,e_k)\big),$$

where $v_1, v_2, v_3, v_4 \in \mathfrak{X}(M)$, $\xi, \eta \in NM$ and $\{e_1, \ldots, e_m\}$ is a local orthonormal frame field with respect to g.

2.4 Local Representations

Let $f : (M, g) \to (N, h)$ be a smooth map between Riemannian manifolds. For computational reasons, we need expressions for components of various tensorial quantities. We can express coordinates with respect to local charts or with respect to orthonormal frames.

Let discuss at first the notation with respect to a local coordinate system. Choose a chart (U, φ) around a point $x \in M$ and a chart (V, ψ) around $f(x) \in N$. Assume that $\varphi : U \to \mathbb{R}^m$ is represented as $\varphi = (x_1, \ldots, x_m)$ and suppose that $\psi : V \to \mathbb{R}^n$ is represented as $\psi = (y_1, \ldots, y_n)$. We use Latin indices to describe quantities on M and Greek indices for quantities on N. From the charts φ and ψ, we obtain for f the local expression expression

$$\psi \circ f \circ \varphi^{-1} = (f^1, \ldots, f^n),$$

where

$$f^\alpha = y^\alpha \circ f \circ \varphi^{-1}.$$

Denote now the basic vector fields associated with the charts (U, φ) and (V, ψ) by $\{\partial_{x_i}, \ldots, \partial_{x_m}\}$ and $\{\partial_{y_1}, \ldots, \partial_{y_n}\}$, respectively. Moreover, denote their corresponding dual forms by $\{dx_1, \ldots, dx_m\}$ and $\{dy_1, \ldots, dy_n\}$. With respect to these conventions, the Riemannian metrics g and h can be written in the form

$$g = \sum_{i,j} g_{ij} dx_i \otimes dx_j \quad \text{and} \quad h = \sum_{\alpha,\beta} h_{\alpha\beta} dy_\alpha \otimes dy_\beta.$$

The *Christoffel symbols* Γ_{ij}^k of the metric g, are defined by the formula

$$\nabla^g_{\partial_{x_i}} \partial_{x_j} = \sum_k \Gamma_{ij}^k \partial_{x_k}$$

and they can be expressed in terms of the metric as

$$\Gamma_{ij}^k = \frac{1}{2} \sum_l g^{kl}\big(- \partial_{x_l} g_{ij} + \partial_{x_i} g_{jl} + \partial_{x_j} g_{li}\big),$$

where g^{ij} are the components of the inverse of the matrix of the metric g, with respect to the basis $\{\partial_{x_1}, \dots, \partial_{x_m}\}$. Similarly, are defined the Christoffel symbols $\Gamma^{\alpha}_{\beta\gamma}$ of h. The differential of the map f and the pull-back via f of the metric h are given by

$$df = \sum_{\alpha} f^{\alpha}_{x_i} \partial_{y_\alpha} \otimes dx_i \quad \text{and} \quad f^*h = \sum_{\alpha,\beta} h_{\alpha\beta} f^{\alpha}_{x_i} f^{\beta}_{x_j}.$$

By a straightforward computation, we see that the Hessian B of f can be represented in the form

$$B(\partial_{x_i}, \partial_{x_j}) = \sum_{\alpha} B^{\alpha}_{ij} \partial_{y_\alpha} = \sum_{\alpha} \big(f^{\alpha}_{x_i x_j} - \sum_{k} \Gamma^{k}_{ij} f^{\alpha}_{x_k} + \sum_{\beta,\gamma} \Gamma^{\alpha}_{\beta\gamma} f^{\beta}_{x_i} f^{\gamma}_{x_j} \big) \partial_{y_\alpha}.$$

Suppose now that $f : M \to N$ is an isometric immersion. Then, the second fundamental form A and the mean curvature H are represented, respectively, as

$$A(\partial_{x_i}, \partial_{x_j}) = \sum_{\alpha} A^{\alpha}_{ij} \partial_{y_\alpha} = \sum_{\alpha} \big(f^{\alpha}_{x_i x_j} - \sum_{k} \Gamma^{k}_{ij} f^{\alpha}_{x_k} + \sum_{\beta,\gamma} \Gamma^{\alpha}_{\beta\gamma} f^{\beta}_{x_i} f^{\gamma}_{x_j} \big) \partial_{y_\alpha}$$

and

$$\begin{aligned} H &= \sum_{\alpha} H^{\alpha} \partial_{y_\alpha} = \sum_{i,j,\alpha} g^{ij} A^{\alpha}_{ij} \partial_{y_\alpha} \\ &= \sum_{i,j} g^{ij} \big(f^{\alpha}_{x_i x_j} - \sum_{k} \Gamma^{k}_{ij} f^{\alpha}_{x_k} + \sum_{\beta,\gamma} \Gamma^{\alpha}_{\beta\gamma} f^{\beta}_{x_i} f^{\gamma}_{x_j} \big) \partial_{y_\alpha}. \end{aligned} \tag{2}$$

Let us discuss now expressions of tensors in local orthonormal frames. Let $\{e_1, \dots, e_m\}$ be a local orthonormal frame of the tangent bundle and $\{\xi_{m+1}, \dots, \xi_n\}$ a local orthonormal frame of the normal bundle. Here we use Latin indices for components on the tangent bundle and Greek indices for components on the normal bundle. For example, we write:

$$\begin{aligned} A^{\alpha}_{ij} &= \langle A(e_i, e_j), \xi_\alpha \rangle = \langle A_{ij}, \xi_\alpha \rangle, \\ \widetilde{R}_{ijkl} &= \widetilde{R}\big(df(e_i), df(e_j), df(e_k), df(e_l)\big), \\ \widetilde{R}_{ij\alpha\beta} &= \widetilde{R}\big(df(e_i), df(e_j), \xi_\alpha, \xi_\beta\big). \end{aligned}$$

Now the Gauss-Codazzi-Ricci equations can be written as:

(a) **Gauss equation:**

$$R_{ijkl} = \widetilde{R}_{ijkl} + \sum_{\alpha} \big(A^{\alpha}_{ik} A^{\alpha}_{jl} - A^{\alpha}_{jk} A^{\alpha}_{il} \big). \tag{3}$$

(b) **Codazzi equation:**

$$\left(\nabla^{\perp}_{e_i} A\right)^{\alpha}_{jk} - \left(\nabla^{\perp}_{e_j} A\right)^{\alpha}_{ik} = -\sum_{\alpha} \widetilde{R}_{ijk\alpha}. \tag{4}$$

(c) **Ricci equation:**

$$R^{\perp}_{ij\alpha\beta} = \widetilde{R}_{ij\alpha\beta} + \sum_{k} \left(A^{\alpha}_{ik} A^{\beta}_{jk} - A^{\beta}_{ik} A^{\alpha}_{jk}\right). \tag{5}$$

3 Minimal Submanifolds

The theory of minimal submanifolds is one of the most active subjects of differential geometry. There is a vast of literature, but here we will present rather basic facts concerning higher codimensional minimal submanifolds. For more details we refer to [21, 22, 70].

3.1 The Gauss Map of a Minimal Submanifold

One of the most important objects in the submanifold geometry is the Gauss map. For codimension one oriented submanifolds in the euclidean space, the *Gauss map* associates to every point of the hypersurface its oriented unit normal vector. This concept can be generalized to higher codimensional oriented submanifolds. Let $f : M \to \mathbb{R}^n$ be an isometric immersion of a m-dimensional oriented Riemannian manifold M into the euclidean space. The image $df(T_x M)$, can be taken after a suitable parallel displacement in $\mathbb{R}^n$, into a point $\mathcal{G}(x)$ of the *oriented Grassmann space* $\mathbb{G}_+(m, n)$ of m-dimensional oriented subspaces of $\mathbb{R}^n$. The map $\mathcal{G} : M \to \mathbb{G}_+(m, n)$ defined in this way, is called the *generalized Gauss map*.

There is a natural way to visualize the Grassmann space $\mathbb{G}_+(m, n)$. Let us denote by $\Lambda^m(\mathbb{R}^n)$ the dual space of all alternative multilinear forms of degree m. Elements of $\Lambda^m(\mathbb{R}^n)$ are called *m-vectors.* Hence, given vectors $v_1, \ldots, v_m$ on $\mathbb{R}^n$, the *exterior product* $v_1 \wedge \cdots \wedge v_m$ is the linear map which on an alternating form Ω of degree m takes the value

$$(v_1 \wedge \cdots \wedge v_m)(\Omega) = \Omega(v_1, \ldots, v_m).$$

The exterior product is linear in each variable separately. Interchanging two elements the sign of the product changes and if two variables are the same the exterior product vanishes. An m-vector ξ is called *simple* or *decomposable* if it can be written as a single wedge product of vectors, that is

$$\xi = v_1 \wedge \cdots \wedge v_m.$$

Note that are m-vectors which are not simple. Using standard techniques from Linear Algebra one can verify that the exterior product $v_1 \wedge \cdots \wedge v_m$ is zero if and only if the vectors are linearly dependent. Moreover, if $\{e_1, \dots, e_n\}$ consists a basis for $\mathbb{R}^n$, then the m-vectors

$$\{e_{i_1} \wedge \cdots \wedge e_{i_m} : 1 \leq i_1 < \cdots < i_m \leq n\}$$

consists a basis of $\Lambda^m(\mathbb{R}^n)$. Therefore, the dimension of the vector space of m-vectors is

$$\dim \Lambda^m(\mathbb{R}^n) = \binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

Each simple vector represents a unique m-dimensional subspace of $\mathbb{R}^n$. Moreover, if ξ and η are simple vectors representing the same subspace, then there exists a non-zero real number such that $\xi = \lambda\eta$. Therefore, there is an obvious equivalence relation on the space of simple vectors such that the space of equivalence classes is to an one to one correspondence with the space of m-dimensional subspaces of $\mathbb{R}^n$. Additionally, we can consider the following relation on the set of non-zero simple m-vectors: ξ *and* η *are called equivalent if and only if* $\xi = \lambda\eta$ *for some positive number* λ. Denote by $[\xi]$ the class containing all simple m-vectors that are equivalent to ξ. The equivalence classes now obtained are called *oriented m-dimensional subspaces* of $\mathbb{R}^n$.

We can equip $\Lambda^m(\mathbb{R}^n)$ with a natural inner product, which for simplicity we denote again by $\langle \cdot\,, \cdot \rangle$. Indeed, define

$$\langle v_1 \wedge \cdots \wedge v_m, w_1 \wedge \cdots \wedge w_m \rangle = \det\big(\langle v_i, w_j \rangle\big)_{1 \leq i,j \leq m}$$

on simple m-vectors and then extend linearly. Moreover, if $\{e_1, \dots, e_n\}$ is an orthonormal basis of $\mathbb{R}^n$ then, the m-vectors

$$\{e_{i_1} \wedge \cdots \wedge e_{i_m} : 1 \leq i_1 < \cdots < i_m \leq n\}$$

consists an orthonormal basis for the exterior power $\Lambda^m(\mathbb{R}^n)$. Moreover, it turns out that for vectors $v_1, \dots, v_m$ in $\mathbb{R}^n$, the norm

$$|v_1 \wedge \cdots \wedge v_m|$$

gives the *m-volume of the parallelepiped* spanned by these vectors.

We can equip $\mathbb{G}_+(m, n)$ with a natural differentiable structure. For every m-dimensional subspace V_0 of $\mathbb{G}_+(m, n)$, consider the open neighbourhood $U(V_0)$ of oriented m-dimensional subspaces whose orthogonal projection into V_0 is one-to-one. Let $\{e_1, \dots, e_m\}$ be an orthonormal base spanning V_0 and $\{\eta_{m+1}, \dots, \eta_n\}$ an orthonormal base spanning its orthogonal complement $V_0^\perp$ in $\mathbb{R}^n$. Then, we may parametrize $U(V_0)$ via $\xi : \mathbb{R}^{m(n-m)} \to U(V_0)$ given by

$$(x_{1m+1},\dots,x_{i\alpha},\dots,x_{mn}) \to \xi(x_{1m+1},\dots,x_{i\alpha},\dots,x_{mn}) \tag{6}$$
$$= \big(e_1+\sum\nolimits_\alpha x_{1\alpha}\eta_\alpha\big)\wedge\cdots\wedge\big(e_m+\sum\nolimits_\alpha x_{m\alpha}\eta_\alpha\big).$$

Two charts $U(V_i), U(V_j)$ with distinct V_i, V_j are analytically compatible.

Definition 10 The map $\Psi : \mathbb{G}_+(m,n) \to \mathbb{S}^{\binom{n}{m}-1}$ given by

$$\Psi\big([v_1\wedge\cdots\wedge v_m]\big) = \frac{v_1\wedge\cdots\wedge v_m}{|v_1\wedge\cdots\wedge v_m|}$$

is called the *Plücker embedding*. We regard the Grassmann space $\mathbb{G}_+(m,n)$ as a Riemannian manifold with the induced by Ψ Riemannian metric.

Theorem 1 *The Plücker embedding is minimal.*

Proof Fix a m-dimensional linear space $V_0 \in \mathbb{G}_+(m,n)$ and consider the parametrization $\xi : \mathbb{R}^{m(n-m)} \to U(V_0) \subset \mathbb{G}_+(m,n)$ described in (6). Now

$$\Psi = W\xi = \frac{\big(e_1+\sum_\alpha x_{1\alpha}\eta_\alpha\big)\wedge\cdots\wedge\big(e_n+\sum_\alpha x_{m\alpha}\eta_\alpha\big)}{\big|\big(e_1+\sum_\alpha x_{1\alpha}\eta_\alpha\big)\wedge\cdots\wedge\big(e_n+\sum_\alpha x_{m\alpha}\eta_\alpha\big)\big|},$$

where the index α run from $m+1$ to n and

$$W = \frac{1}{|\xi|} = \frac{1}{\big|\big(e_1+\sum_\alpha x_{1\alpha}\eta_\alpha\big)\wedge\cdots\wedge\big(e_n+\sum_\alpha x_{m\alpha}\eta_\alpha\big)\big|}.$$

Note that

$$\xi_{x_{i\alpha}} = \big(e_1+\sum\nolimits_\alpha x_{1\alpha}\eta_\alpha\big)\wedge\cdots\wedge\big(e_{i-1}+\sum\nolimits_\alpha x_{i-1\alpha}\eta_\alpha\big)\wedge\eta_\alpha$$
$$\wedge\big(e_{i+1}+\sum\nolimits_\alpha x_{i+1\alpha}\eta_\alpha\big)\wedge\cdots\wedge\big(e_m+\sum\nolimits_\alpha x_{m\alpha}\eta_\alpha\big)$$

and

$$\xi_{x_{i\alpha}x_{j\beta}} = \big(e_1+\sum\nolimits_\alpha x_{1\alpha}\eta_\alpha\big)\wedge\cdots\wedge\big(e_{i-1}+\sum\nolimits_\alpha x_{i-1\alpha}\eta_\alpha\big)\wedge\eta_\alpha$$
$$\wedge\big(e_{i+1}+\sum\nolimits_\alpha x_{i+1\alpha}\eta_\alpha\big)\wedge\cdots\wedge\big(e_{j-1}+\sum\nolimits_\alpha x_{j-1\alpha}\eta_\alpha\big)\wedge\eta_\beta$$
$$\wedge\big(e_{j+1}+\sum\nolimits_\alpha x_{j+1\alpha}\eta_\alpha\big)\wedge\cdots\wedge\big(e_m+\sum\nolimits_\alpha x_{m\alpha}\eta_\alpha\big)\big(1-\delta_{ij}\big),$$

where $i, j \in \{1,\dots,m\}$ and $\alpha,\beta \in \{m+1,\dots,n\}$. In particular,

$$\xi_{x_{i\alpha}}(0) = e_1\wedge\cdots\wedge e_{i-1}\wedge\eta_\alpha\wedge e_{i+1}\wedge\cdots\wedge e_m \tag{7}$$

and

$$\xi_{x_{i\alpha}x_{j\beta}}(0) = e_1 \wedge \cdots \wedge e_{i-1} \wedge \eta_\alpha \wedge e_{i+1} \wedge \cdots \wedge e_{j-1} \wedge \eta_\beta \wedge e_{j+1} \wedge \cdots \wedge e_m(1-\delta_{ij}). \tag{8}$$

Additionally,

$$W_{x_{i\alpha}} = -W^3\langle \xi, \xi_{x_{i\alpha}}\rangle$$

and

$$W_{x_{i\alpha}x_{j\beta}} = -3W^5\langle \xi, \xi_{x_{i\alpha}}\rangle\langle \xi, \xi_{x_{j\beta}}\rangle - W^3\langle \xi_{x_{i\alpha}}, \xi_{x_{j\beta}}\rangle - W^3\langle \xi, \xi_{x_{i\alpha}x_{j\beta}}\rangle.$$

Moreover,

$$W(0) = 1, \quad W_{x_{i\alpha}}(0) = 0 \quad \text{and} \quad W_{x_{i\alpha}x_{j\beta}}(0) = -\delta_{ij}\delta_{\alpha\beta}. \tag{9}$$

From (7) and (9) we see that

$$\Psi_{x_{i\alpha}}(0) = e_1 \wedge \cdots \wedge e_{i-1} \wedge \eta_\alpha \wedge e_{i+1} \wedge \cdots \wedge e_m.$$

Hence, the vectors

$$\{\partial_{x_{1m+1}}|_0, \ldots, \partial_{x_{i\alpha}}|_0, \ldots, \partial_{x_{mn}}|_0\}$$

form an orthonormal basis of $T_{V_0}\mathbb{G}_+(m,n)$ with respect to the induced by Ψ Riemannian metric. Moreover, from (7), (8), and (9) we deduce that

$$\Psi_{x_{i\alpha}x_{j\beta}}(0) = -\delta_{ij}\delta_{\alpha\beta}\Psi(0) + \xi_{x_{i\alpha}x_{j\beta}}(0).$$

According to (8), the second fundamental form A of the Plücker embedding at the point V_0 is equal to

$$\begin{aligned} A(\partial_{x_{i\alpha}}, \partial_{x_{j\beta}}) = {} & e_1 \wedge \cdots \wedge e_{i-1} \wedge \eta_\alpha \wedge e_{i+1} \\ & \wedge \cdots \wedge e_{j-1} \wedge \eta_\beta \wedge e_{j+1} \wedge \cdots \wedge e_m(1-\delta_{ij}) \end{aligned} \tag{10}$$

and, in particular,

$$A(\partial_{x_{i\alpha}}, \partial_{x_{i\alpha}}) = 0 \tag{11}$$

for any $i \in \{1, \ldots, m\}$ and $\alpha \in \{m+1, \ldots, n\}$. Thus, the mean curvature H of the embedding Ψ at V_0 is given by

$$H(V_0) = \sum_{i,\alpha} A(\partial_{x_{i\alpha}}, \partial_{x_{i\alpha}}) = 0.$$

Consequently, Ψ gives rise to a minimal submanifold of the sphere. □

In 1970, Ruh and Vilms [74] obtained an important link between minimality of a submanifold and the harmonicity of its generalized Gauss map. More precisely, the following result holds:

Theorem 2 *Let $f : M \to \mathbb{R}^n$ be a minimal isometric immersion. Then, the generalized Gauss map $\mathcal{G}$ of f is a harmonic map.*

Proof Consider the map

$$F = \Psi \circ \mathcal{G} : M \to \mathbb{S}^{\binom{n}{m}-1} \subset \Lambda^m(\mathbb{R}^n)$$

where Ψ is the Plücker embedding. From the composition formula, we have

$$B_F(v_1, v_2) = d\Psi\big(B_{\mathcal{G}}(v_1, v_2)\big) + A_\Psi\big(d\mathcal{G}(v_1), d\mathcal{G}(v_2)\big) \tag{12}$$

for any $v_1, v_2 \in \mathfrak{X}(M)$, where B_F and $B_{\mathcal{G}}$ are the Hessians of F and $\mathcal{G}$ and A_Ψ the second fundamental form of Ψ, respectively. Fix a local orthonormal frame field $\{e_1, \ldots, e_m\}$ defined on an open neighbourhood U of M and a local orthonormal frame $\{\eta_{m+1}, \ldots, \eta_n\}$ in the normal bundle of the immersion. Note that since f is isometric immersion, for any $x \in U$, we have

$$F(x) = df_x(e_1) \wedge df_x(e_2) \wedge \cdots \wedge df_x(e_m).$$

Fix now a point $x_0 \in U$ and for simplicity suppose that the frame $\{e_1, \ldots, e_m\}$ is normal at x_0. By straightforward computations we see that at x_0 we have

$$dF(e_j) = A(e_j, e_1) \wedge \cdots \wedge df(e_m) + \cdots + df(e_1) \wedge \cdots \wedge A(e_j, e_m),$$

where A is the second fundamental form of f. Hence, in view of (7), we obtain that the differential of $\mathcal{G}$ at x_0 is equal to

$$d\mathcal{G}(e_j) = \sum_\alpha A^\alpha_{1j}\eta_\alpha \wedge e_2 \wedge \cdots \wedge e_m + \cdots + e_1 \wedge e_2 \wedge \cdots \wedge \sum_\alpha A^\alpha_{mj}\eta_\alpha.$$

Recall that, from the Codazzi equations (4), we have at x_0 that

$$\big(\nabla^{f^*T\mathbb{R}^n}_{e_j} A\big)_{kl} = \big(\nabla^\perp_{e_k} A\big)_{jl} - \sum_{i,\alpha} A^\alpha_{kl} A^\alpha_{ij} df(e_i),$$

for any $j, k, l \in \{1, \ldots, m\}$. Differentiating dF and estimating at x_0 we get

$$\begin{aligned}\nabla^{F^*T\Lambda^m(\mathbb{R}^n)}_{e_j} dF(e_j) = &-|A|^2 F + \big(\nabla^\perp_{e_1} A\big)_{jj} \wedge df(e_2) \wedge \cdots \wedge df(e_m)\\ &+ df(e_1) \wedge \big(\nabla^\perp_{e_2} A\big)_{jj} \wedge \cdots \wedge df(e_m) + \cdots + df(e_1) \wedge \cdots \wedge \big(\nabla^\perp_{e_m} A\big)_{jj}\\ &+ A_{1j} \wedge A_{2j} \wedge \cdots \wedge df(e_m) + \cdots + df(e_1) \wedge \cdots \wedge A_{jm-1} \wedge A_{jm}.\end{aligned}$$

Therefore, at x_0 we have the identity

$$\begin{aligned} B_F(e_j, e_j) &= \left(\nabla^{\perp}_{e_1} A\right)_{jj} \wedge df(e_2) \wedge \cdots \wedge df(e_m) \\ &+ df(e_1) \wedge \left(\nabla^{\perp}_{e_2} A\right)_{jj} \wedge \cdots \wedge df(e_m) + \cdots + df(e_1) \wedge \cdots \wedge \left(\nabla^{\perp}_{e_m} A\right)_{jj} \\ &+ A_{1j} \wedge A_{2j} \wedge \cdots \wedge df(e_m) + \cdots + df(e_1) \wedge \cdots \wedge A_{jm-1} \wedge A_{jm}. \end{aligned}$$

Summing up and using the minimality of f, we see that at x_0 it holds

$$\begin{aligned} \operatorname{trace}_g B_F = & \sum_{j,\alpha} A^{\alpha}_{1j} \eta_\alpha \wedge \sum_{j,\beta} A^{\beta}_{2j} \eta_\beta \wedge df(e_3) \wedge \cdots \wedge df(e_m) \\ & + \sum_{j,\alpha} A^{\alpha}_{1j} \eta_\alpha \wedge df(e_2) \wedge \sum_{j,\beta} A^{\beta}_{3j} \eta_\beta \wedge \cdots \wedge df(e_m) \\ & \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ & + df(e_1) \wedge df(e_2) \wedge \cdots \wedge \sum_{j,\alpha} A^{\alpha}_{m-1\,j} \eta_\alpha \wedge \sum_{j,\alpha} A^{\alpha}_{mj} \eta_\beta. \end{aligned}$$

Hence, bearing in mind equation (10) which gives the formula for the second fundamental form of Ψ it follows that at x_0 we have

$$\operatorname{trace}_g B_F = \sum\nolimits_j A_\Psi\left(d\mathcal{G}(e_j), d\mathcal{G}(e_j)\right).$$

Combining the last equality with (12) we get the desired result. □

3.2 Weierstrass Representations

For two dimensional surfaces in the euclidean space, there is a link between minimality and holomorphicity through a general formula which express a simply-connected minimal surface in terms of complex functions with certain properties. In particular, the following result holds:

Theorem 3 *Let $\varphi_1, \ldots, \varphi_n : U \subset \mathbb{C} \to \mathbb{C}$ be holomorphic functions, where U is simply connected, such that $\varphi_1^2 + \cdots + \varphi_n^2 = 0$ and $|\varphi_1|^2 + \cdots + |\varphi_n|^2 > 0$. Then, the map*

$$F(z) = \left(Re \int_{z_0}^{z} \varphi_1(\zeta) d\zeta, \ldots, Re \int_{z_0}^{z} \varphi_n(\zeta) d\zeta \right), \quad z \in U,$$

where $z_0 \in U$, gives rise to minimal surface in $\mathbb{R}^n$. Conversely, every minimal surface in $\mathbb{R}^n$, at least locally, can be described in this form.

The idea to obtain such a parametrization is the following: Let $f : M \to \mathbb{R}^n$ be a minimal immersion of a 2-manifold. Choose a local isothermal system of coordinates ($U \subset \mathbb{C}, z = x + iy$), where U is simply connected; see [54]. Then, the induced by f metric g on M has the form $g = E|dz|^2$, where E is a smooth positive function. Moreover, in these coordinates, the Laplace–Beltrami operator Δ with respect to g is expressed by

$$\Delta = E^{-1}(\partial_x \partial_x + \partial_y \partial_y).$$

With respect to such parameters, minimality is equivalent to harmonicity. Consider now the map $\varphi = (\varphi_1, \ldots, \varphi_n) : U \to \mathbb{C}^n$, $\varphi = f_x - if_y$. One can readily check that φ is holomorphic and its components satisfy

$$\varphi_1^2 + \cdots + \varphi_n^2 = 0 \quad \text{and} \quad |\varphi_1|^2 + \cdots + |\varphi_n|^2 = 2E > 0.$$

By fixing a point $z_0 \in U$ it is clear that, up to a parallel transport,

$$f(z) = Re \int_{z_0}^{z} \varphi(\zeta) d\zeta, \quad z \in U.$$

The map φ has also a very important geometric interpretation. At first we observe that the variety

$$Q_{n-2} = \left\{ [z_1, \ldots, z_n] \in \mathbb{CP}^{n-1} : z_1^2 + \cdots + z_n^2 = 0 \right\}$$

is diffeomorphic with $\mathbb{G}_+(2, n)$. To see this, consider a 2-plane $\Pi \subset \mathbb{R}^n$ that is spanned by $u \wedge v$, where the vectors u and v satisfy $|u| - |v| = 0$ and $\langle u, v \rangle = 0$. Then, vector $w = u + iv$ belongs to Q_{n-2}. Hence, to each oriented 2-plane we associate a point in Q_{n-2}. In fact, this correspondence actually is a diffeomorphism. Consequently, the map $\overline{\varphi} : U \to Q_{n-2}$, $\overline{\varphi} = f_x + if_y$, is exactly the generalized Gauss map of the minimal surface.

Let M be a manifold of dimension $2m$ endowed with a Riemannian metric g and a metric connection ∇. An *almost complex structure* on M is by definition a bundle isomorphism $J : TM \to TM$ satisfying $J \circ J = -I$. The pair (M, J) is called an *almost complex manifold*. If J is an isometry with respect to g and parallel with respect to ∇, then the triple (M, g, J) is called *Kähler manifold*. In this case, the 2-form Ω given by

$$\Omega(v_1, v_2) = g\big(Jv_1, v_2\big),$$

where $v_1, v_2 \in \mathfrak{X}(M)$, is closed and is called the *Kähler form*. A smooth map $f : (M, J_M) \to (N, J_N)$ between Kähler manifolds is called *holomorphic* if $df \circ J_M = J_N \circ df$ and *anti-holomorphic* if $df \circ J_M = -J_N \circ df$. If the map f is a holomorphic or anti-holomorphic isometric immersion, then $f(M)$ will be called an *immersed complex submanifold* of N. Such immersions are automatically minimal.

With the terminology we just introduced and the discussion above, we can now state the following result which was originally proved by Chern [20].

Theorem 4 *An oriented surface of the euclidean space is minimal if and only if its generalized Gauss map is anti-homolomorphic.*

We will present now another interesting category of submanifolds, the so called Lagrangian submanifolds.

Definition 11 Let M^m be a Riemannian manifold, (N^{2m}, g_N, Ω) be a Kähler manifold and $f\colon M^m \to N^{2m}$ an isometric immersion. The immersion f will be called *Lagrangian* if and only if $f^*\Omega = 0$.

Let us conclude this section with the following parametrization of minimal Lagrangian surfaces in $\mathbb{R}^4$; see Chen and Morvan [19] and Aiyama [1, 2].

Theorem 5 *Suppose that $f, g : U \to \mathbb{C}$ are two holomorphic maps defined in a simply connected domain U of the complex plane satisfying $|f_z|^2 + |g_z|^2 > 0$. Then the map*

$$F = \frac{e^{i\beta/2}}{\sqrt{2}}(f - i\overline{g}, g + i\overline{f}),$$

where β is a real number, is a minimal conformal Lagrangian immersion in $\mathbb{C}^2$. The generalized Gauss map $\mathcal{G}$ takes values in $\mathbb{S}^2 \times \{(e^{i\beta}, 0)\} \simeq \mathbb{C} \cup \{\infty\}$ and is given by the formula

$$\mathcal{G} = f_z/g_z.$$

Conversely, every minimal Lagrangian immersion $f : M \to \mathbb{C}^2$ can be, at least locally, parametrized as above.

4 Scalar and Vectorial Maximum Principles

The maximum principle is one of the most useful tools employed in the study of PDEs. All maximum principles rely on the following elementary result of calculus: *Suppose that Ω is an open, bounded domain of $\mathbb{R}^m$ and let $u : \overline{\Omega} \to \mathbb{R}$ be a continuous function which is C^2-smooth in Ω. If u attains its maximum at interior point x_0, then*

$$\nabla u(x_0) = 0 \quad \textit{and} \quad \nabla^2 u(x_0) \leq 0.$$

As an immediate consequence of this fact is that any continuous and C^2-smooth up to the boundary strictly convex function must attain its maximum at the boundary of Ω. In the matter of fact, one can show a little bit more: *Any continuous and C^2-*

smooth up to the boundary weakly convex function either attain its maximum at the boundary of Ω or otherwise is constant. The above principle holds for a large class of solutions of partial differential inequalities.

4.1 Hopf's Maximum Principles

Suppose that Ω is a bounded, open and connected domain of $\mathbb{R}^m$. We wish to study operators $\mathcal{L}\colon C^2(\Omega) \to C^0(\Omega)$ of the form

$$\mathcal{L} = \sum_{i=1}^{m} a_{ij}\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{m} b_i \partial_{x_i}, \tag{13}$$

where here $a_{ij} = a_{ij}$, $b_j\colon \Omega \to \mathbb{R}$, $i, j \in \{1, \ldots, m\}$, are uniformly bounded functions and ∂_{x_i}, $i \in \{1, \ldots, m\}$, the partial derivatives with respect the cartesian coordinates of $\mathbb{R}^m$. The symmetric matrix $\mathcal{A}$ with coefficients the functions a^{ij} is called the *representative matrix* of $\mathcal{L}$.

Definition 12 The operator $\mathcal{L}$ is called *elliptic* if the matrix $\mathcal{A}$ is positive at each point of Ω. Moreover, $\mathcal{L}$ is called *uniformly elliptic* if the smallest eigenvalue of its matrix $\mathcal{A}$ is a function which is bounded away from zero.

Theorem 6 (Hopf's Strong Maximum Principle) *Let $\Omega \subset \mathbb{R}^m$ be an open, connected and bounded domain. Suppose that $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ is a solution of the differential inequality*

$$\mathcal{L}u + hu \geq 0,$$

where $\mathcal{L}$ is an uniformly elliptic differential operator with uniformly bounded coefficients and h a continuous function on $\overline{\Omega}$.

(a) Suppose that $h = 0$ and that u attains its maximum at an interior point of Ω. Then, u is constant.
(b) Suppose that $h \leq 0$ and that u attains a non-negative maximum at an interior point of Ω. Then, u is constant.

For the proof see [35] or [72].

4.2 Maximum Principles for Systems

We would like to have a form of the maximum principle that is applicable for sections in vector bundles. To generalize, first note that Hopf's maximum principle for functions can be re-formulated as follows:

Let Ω be an open subset of $\mathbb{R}^m$ and $u : \Omega \subset \mathbb{R}^m \to [a, b]$ a C^2-smooth function satisfying the uniformly elliptic differential equation

$$\sum_{i,j=1}^{m} a_{ij} u_{x_i x_j} + \sum_{i=1}^{m} b_i u_{x_i} = 0.$$

If a point of Ω is mapped into a boundary point of $[a, b]$, then any point of Ω is mapped into the boundary.

From this point of view of the statement of Hopf's maximum principle, one can guess how the generalization of the maximum principle for vector valued maps should be. The interval is replaced by a convex set K and the statement reads:
Let $\Omega \subset \mathbb{R}^m$ be open, $K \subset \mathbb{R}^n$ closed convex and $u : \Omega \to K$ a C^2-smooth vector valued map satisfying the uniformly elliptic differential system

$$\sum_{i,j=1}^{m} a_{ij} u_{x_i x_j} + \sum_{i=1}^{m} b_i u_{x_i} = 0.$$

If a point of Ω is mapped into a boundary point of K then every point is mapped into the boundary.

4.2.1 Convexity and Distance Functions

A crucial role in the proof of the vectorial maximum principle plays the geometry of the (signed) distance function from the boundary of a convex set. In this subsection, we review the basic definitions of the geometry of convex sets in euclidean space such as supporting half-spaces, tangent cones, normal vectors and distance functions.

Definition 13 A subset K of $\mathbb{R}^n$ is called *convex* if for any pair of points $z, w \in K$, the segment

$$\mathcal{E}_{z,w} = \{tz + (1-t)w \in \mathbb{R}^n : t \in [0, 1]\}$$

is contained in K. The set K is said to be *strictly convex*, if for any pair $z, w \in K$ the segment $\mathcal{E}_{z,w}$ belongs to the interior of K.

A convex set $K \subset \mathbb{R}^n$ may have non-smooth boundary. It is a well-known fact in Convex Geometry that the boundary ∂K is a continuous hypersurface of $\mathbb{R}^n$. In fact, according to a result of Reidemeister [73], the boundary ∂K is Lipschitz continuous and so almost everywhere differentiable. In particular, there is no well-defined tangent or normal space of K in the classical sense. However, there is a way to generalize these notions for convex subsets of $\mathbb{R}^n$.

Definition 14 Let K be a closed convex subset of the euclidean space $\mathbb{R}^n$. A *supporting half-space* of the set K is a closed half-space of $\mathbb{R}^n$ which contains K and has points of K on its boundary. A supporting hyperplane of K is a hyperplane which is the boundary of a *supporting half-space* of K. The *tangent cone* $C_{y_0}K$ of K at $y_0 \in \partial K$ is defined as the intersection of all supporting half-spaces of K that contain y_0.

Definition 15 Let $K \subset \mathbb{R}^n$ be a closed convex subset and $y_0 \in \partial K$. Then:

(a) A non-zero vector ξ is called *normal vector* of ∂K at y_0, if ξ is normal to a supporting hyperplane of K passing through y_0. This normal vector is called *inward pointing*, if it points into the half-space containing K.
(b) A vector η is called *inward pointing* at $y_0 \in \partial K$, if

$$\langle \xi, \eta \rangle \geq 0$$

for any inward pointing normal vector ξ at y_0.

Let $K \subset \mathbb{R}^n$ be a closed convex set and $d : \mathbb{R}^n \to \mathbb{R}$ the function given by

$$d(z) = \begin{cases} +\operatorname{dist}(z, \partial K), \text{ if } z \in K, \\ \\ -\operatorname{dist}(z, \partial K), \text{ if } z \notin K. \end{cases}$$

Note that for each $x \in \mathbb{R}^n$ there is at least one point $y \in \partial K$ such that

$$\operatorname{dist}(z, \partial K) = |y - z|.$$

Moreover, the function d is Lipschitz continuous. For a better understanding of the properties of d, let us suppose that ∂K is C^2-smooth. Denote by ξ the inward pointing unit normal vector field along ∂K and by the A the corresponding Weingarten operator. Because K is convex, from Hadamard's Theorem, A is non-negative definite. In particular, K is strictly convex if and only if A is positive definite; see for example [29]. Fix a point $y_0 \in \partial K$. In an open neighbourhood $U \subset \mathbb{R}^n$ of y_0, the part $U \cap \partial K$ can be parametrized via an embedding $f : \Omega = U \cap T_{y_0}K \to \mathbb{R}^n$, which assigns to each point of Ω the height of ∂K from its tangent plane at y_0. Recall from multi-variable calculus that the distance of any point $z \in K^0$ to ∂K is realized as the intersection of a straight line passing through z and meeting ∂K orthogonally. Hence, the *level set*

$$K_t = \{z \in K : d(z) = t\},$$

of d is parametrized locally via the map $f_t : \Omega \to \mathbb{R}^n$ given by

$$f_t = f + t\xi.$$

Proposition 1 *There exists a positive real number ε, such that f_t is an immersion for all $t \in (-\varepsilon, \varepsilon)$. Moreover, the unit normal along f_t coincide with the unit normal ξ of f. Additionally, the Weingarten operator A_t of f_t is related to the Weingarten operator A of f by the formula*

$$A_t = (I - tA)^{-1} \circ A.$$

In particular, K_t is strictly convex if and only if ∂K is strictly convex.

Proof We have

$$df_t = df + td\xi = df \circ (I - tA).$$

Hence, ξ is a unit normal vector field along f_t. Therefore,

$$-df \circ A = d\xi = -df_t \circ A_t = -df \circ (I - tA) \circ A_t$$

and so

$$A = (I - tA) \circ A_t.$$

From the above formula we deduce that there exists a positive constant ε such that K_t is convex for all $t \in (-\varepsilon, \varepsilon)$. In addition, if ∂K is strictly convex, the level sets close to the boundary are also strictly convex. □

Proposition 2 *Let K be a closed and convex set in $\mathbb{R}^n$.*

(a) *For any $y_0 \in \partial K$ there exists a neighbourhood $U \subset \mathbb{R}^n$ containing y_0, such that d is C^2-smooth function on $U \cap K^0$.*
(b) *Let v, w tangent vectors of $K_{d(z)}$ at $z \in K^0$. Then, the Hessian $\nabla^2 d$ of d satisfies*

$$\nabla^2 d(v, w) = -\frac{A(v, w)}{1 - d(z)A(v, w)},$$

where A is the shape operator of ∂K associated to the inward pointing unit normal, and

$$\nabla^2 d(v, \xi) = 0.$$

Proof Parametrize, locally, the boundary ∂K as the image of an embedding $f : \Omega \subset \mathbb{R}^{n-1} \to \mathbb{R}^n$. Define the map $F : \Omega \times \mathbb{R} \subset \mathbb{R}^n \to \mathbb{R}^n$, given by

$$F(x, t) = f(x) + t\xi(x).$$

Then, $dF_{(y_0,t)}(\partial_t) = \xi_{y_0}$ and $dF_{(y_0,t)}(v) = df_{x_0}(v - tAv)$, for any index $i \in \{1, \ldots, n-1\}$ and $v \in T_{y_0}\Omega$. From the inverse mapping theorem, there exists

an open subset $D \subset \Omega$ and a positive number ε such that the map F is a C^2-diffeomorphism for any $(x, t) \in D \times (-\varepsilon, \varepsilon)$. Hence, under a change of coordinates, d may be regarded as a C^2-smooth function defined on $D \times (-\varepsilon, \varepsilon)$. In the matter of fact, in these coordinates, we have

$$d(x, t) = \langle F(x, t) - f(x), \xi(x) \rangle = t.$$

Therefore, $\nabla d_{(x,t)} = \xi_x$. Because $|\nabla d| = 1$, we deduce that $\nabla^2 d$ vanishes on the normal bundle of any level set K_t. Moreover, $\nabla^2 d = -A_t$ on the tangent space of any level set K_t. Now the desired result follows from Proposition 1. This completes the proof. □

4.2.2 Weinberger's Maximum Principle

Weinberger [101] established a strong maximum principle for C^2-smooth maps $u : \Omega \subset \mathbb{R}^m \to K \subset \mathbb{R}^n$ with values in a closed convex set $K \subset \mathbb{R}^n$, whose boundary ∂K satisfies some regularity conditions that he called "*slab conditions*". Inspired by the ideas of Weinberger, Wang [93] gave a geometric proof of the strong maximum principle, in the case where the boundary ∂K of K is of class C^2. The idea of Wang was to apply Hopf's maximum principle to the function $d \circ u : \Omega \to \mathbb{R}$, whose value at x is equal to the distance of $u(x)$ from the boundary ∂K. Later, Evans [36] removed all additional regularity requirements on the boundary of K.

Theorem 7 (Weinberger-Evans) *Let K be a closed, convex set of $\mathbb{R}^n$ and $u : \Omega \subset \mathbb{R}^m \to K \subset \mathbb{R}^n$, $u = (u^1, \ldots, u^n)$, a solution of the uniformly elliptic system of partial differential equations*

$$\mathcal{L}u(x) + \Psi(x, u(x)) = 0, \quad x \in \Omega,$$

where Ω is a connected open domain of $\mathbb{R}^m$ and $\Psi : \Omega \times \mathbb{R}^n \to \mathbb{R}^n$ a continuous map that is Lipschitz continuous in the second variable. Suppose further that Ψ is pointing into K.

(a) *If there is a point $x_0 \in \Omega$ such that $u(x_0) \in \partial K$, then $u(x) \in \partial K$ for any point $x \in \Omega$.*
(b) *Assume additionally that the boundary ∂K is strictly convex. If there is a point $x_0 \in \Omega$ such that $u(x_0) \in \partial K$, then u is constant.*

Proof Let us give the proof in the case where the boundary of K is smooth of class C^2, following the ideas in [93]. Consider the function $f = d \circ u : \Omega \to \mathbb{R}$. We compute,

$$f_{x_i} = \sum\nolimits_{\alpha=1}^{n} d_{u^\alpha} u^\alpha_{x_i}, \tag{14}$$

and

$$f_{x_i x_j} = \sum_{\alpha=1}^{n} d_{u^\alpha} u^\alpha_{x_i x_j} + \sum_{\alpha,\beta=1}^{n} d_{u^\alpha u^\beta} u^\alpha_{x_i} u^\beta_{x_j}.$$

Consider now the uniformly differential operator $\widetilde{\mathcal{L}}$ given by

$$\widetilde{\mathcal{L}} = \mathcal{L} - \sum_{i=1}^{m} b_i \partial_{x_i}.$$

By a straightforward computation, we get

$$\widetilde{\mathcal{L}} f = \sum_{\alpha=1}^{n} d_{u^\alpha} \sum_{i,j=1}^{m} a_{ij} u^\alpha_{x_i x_j} + \sum_{i,j=1}^{m} a_{ij} \sum_{\alpha,\beta=1}^{n} d_{u^\alpha u^\beta} u^\alpha_{x_i} u^\beta_{x_j}. \quad (15)$$

Denote the first sum in the right hand side of (15) by I and the second sum by II. Observe at first that

$$\mathrm{I}(x) = -\langle \nabla d_{u(x)}, \Psi(x, u(x)) \rangle.$$

We restrict our selves in a sufficiently small neighbourhood $U \subset \mathbb{R}^n$ around $u(x_0)$ and in an neighbourhood V of x_0 such that $u(V) \subset U$. For each $x \in V$, denote by $\widehat{u}(x)$ the unique point on ∂K with the property

$$f(x) = d(u(x)) = |u(x) - \widehat{u}(x)|.$$

Recall that the integral curves ∇d are straight lines perpendicular to each level set of d. Thus, $\nabla d_{u(x)} = \nabla d_{\widehat{u}(x)}$. Since Ψ is inward pointing, we get that

$$\langle \nabla d_{u(x)}, \Psi(x, \widehat{u}(x)) \rangle = \langle \nabla d_{\widehat{u}(x)}, \Psi(x, \widehat{u}(x)) \rangle \geq 0.$$

Therefore, exploiting the Lipschitz property of Ψ, we get that

$$\begin{aligned}
\mathrm{I}(x) &= -\langle \nabla d_{u(x)}, \Psi(x, u(x)) - \Psi(x, \widehat{u}(x)) + \Psi(x, \widehat{u}(x)) \rangle \\
&= -\langle \nabla d_{u(x)}, \Psi(x, u(x)) - \Psi(x, \widehat{u}(x)) \rangle - \langle \nabla d_{u(x)}, \Psi(x, \widehat{u}(x)) \rangle \\
&\leq |\nabla d_{u(x)}| \cdot |\Psi(x, u(x)) - \Psi(x, \widehat{u}(x))| \\
&\leq h(x) |u(x) - \widehat{u}(x)| \\
&= h(x) f(x),
\end{aligned}$$

where h is a non-negative bounded function. Recall from Proposition 1 that U is foliated by level sets of d. Thus, we can decompose u_{x_i} in the form

$$u_{x_i} = u^\top_{x_i} + u^\perp_{x_i}$$

where $(\cdot)^\top$ denotes the orthogonal projection into the tangent space and $(\cdot)^\top$ the orthogonal projection into the normal space of the foliation. Bearing in mind the conclusions of Proposition 2, we deduce that

$$\sum\nolimits_{\alpha,\beta=1}^{n} d_{u^\alpha u^\beta} u^\alpha_{x_i} u^\beta_{x_j} = \nabla^2 d\big(u^\top_{x_i}, u^\top_{x_j}\big) = \frac{A\big(u^\top_{x_i}, u^\top_{x_j}\big)}{1 - f(x) A\big(u^\top_{x_i}, u^\top_{x_j}\big)}.$$

Since, A is non-negative definite and $\mathcal{A} = (a_{ij})$ is positive definite, we deduce that

$$\mathrm{II} = \mathrm{trace}\left(\mathcal{A} \cdot \nabla^2 d\right) \leq 0.$$

In addition, for any x such that $u(x) \in \partial K$, we have that

$$\mathrm{II}(x) = \begin{cases} \text{strictly negative}, & \text{if } \partial K \text{ is strictly convex in close to } u(x), \\ \text{zero} & \text{if } \partial K \text{ is flat in a neighbourhood of } u(x). \end{cases} \tag{16}$$

Putting everything together, we get

$$\widetilde{\mathcal{L}} f(x) - h(x) f(x) \leq 0.$$

Since $f \geq 0$ and there exists a point x_0 such that $f(x_0) = 0$, from Hopf's strong maximum principle we deduce that $f \equiv 0$. This implies now that all the values of u lie in the boundary of K. Moreover, going back to the original equation for f, we see that $\mathrm{II} \equiv 0$. Consequently, if ∂K is strictly convex, from (16) it follows that u must be constant. This completes the proof. □

4.3 Maximum Principles for Bundles

To state the maximum principle for sections in vector bundles, we must introduce an appropriate notion of convexity for subsets of vector bundles. Let us recall at first the following definition of Hamilton [45]:

Definition 16 (Hamilton) Suppose that E is a vector bundle over the manifold M and let K be a closed subset of E.

(a) The set K is called *fiber-convex* or *convex in the fiber*, if for each $x \in M$, the set $K_x = K \cap E_x$ is a convex subset of the fiber E_x.
(b) The set K is called *invariant under parallel transport*, if for every smooth curve $\gamma : [0, b] \to M$ and any vector $v \in K_{\gamma(0)}$, the unique parallel section $v(t) \in E_{\gamma(t)}$, $t \in [0, b]$, along $\gamma(t)$ with $v(0) = v$, is contained in K.
(c) A *fiberwise map* $\Psi : E \to E$ is a map such that $\pi \circ \Psi = \pi$, where π denotes the bundle projection. We say that a fiberwise map Ψ points into K (or is inward pointing), if for any $x \in M$ and any $\vartheta \in \partial K_x$ the vector $\Psi(\vartheta)$ belongs to the tangent cone $C_\vartheta K_x$ of K_x at ϑ.

Let E be a Riemannian vector bundle over a manifold M equipped with a metric compatible connection. We consider uniformly elliptic operators $\mathcal{L}$ that are given locally by

$$\mathcal{L} = \sum_{i,j=1}^{m} a_{ij} \nabla^2_{e_i,e_j} + \sum_{j=1}^{m} b_j \nabla_{e_j}, \tag{E}$$

where $\{e_1, \dots, e_m\}$ is a local orthonormal frame of M, $\mathcal{A} = (a_{ij})$ a symmetric, uniformly positive definite tensor and $b = \sum_{i=1}^{m} b_i e_i$ is a smooth vector field.

For the proof of the maximum principle, we will use a result due to Böhm and Wilking [9].

Lemma 1 *Let M be a Riemannian manifold and E a Riemannian vector bundle over M equipped with a metric compatible connection. Let $K \subset E$ be a closed and fiber-convex subset which is invariant under parallel transport. If ϕ is a smooth section with values in K then, for any $x \in M$ and $v \in T_x M$, the Hessian*

$$\nabla^2_{v,v}\phi = \nabla_v \nabla_v \phi - \nabla_{\nabla_v v}\phi$$

belongs into the tangent cone of K_x at the point $\phi|_x$.

Proof It suffices to prove the result in the case where there exists a point x_0 which is mapped via ϕ in the boundary of K, since otherwise the result is trivially true. Consider a unit vector $v \in T_{x_0}M$ and an normal coordinate system $(x_1, \dots, x_m)$ in an open neighbourhood U around a point x_0 such that $\partial_{x_1}\big|_{x_0} = v$. Moreover, pick a basis $\{\phi_1|_{x_0}, \dots, \phi_k|_{x_0}\}$ of E_{x_0} and extend it into a local geodesic orthonormal frame field. Then,

$$\phi = u_1\phi_1 + \cdots + u_k\phi_k,$$

where the components $u_i : U \to \mathbb{R}$, $i \in \{1, \dots, k\}$, are smooth functions. A simple computation shows

$$\nabla^2_{v,v}\phi|_{x_0} = \nabla_{\partial_{x_1}} \nabla_{\partial_{x_1}} \phi|_{x_0} - \nabla_{\nabla_{\partial_{x_1}} \partial_{x_1}} \phi|_{x_0} = \sum_{i=1}^{k} (\partial_{x_1} \partial_{x_1} u_i)(x_0) \phi_i|_{x_0}$$

$$= \sum_{i=1}^{k} (u_i \circ \gamma)''(0) \phi_i|_{x_0},$$

where $\gamma : (-\varepsilon, \varepsilon) \to U$ is a length minimizing geodesic such that

$$\gamma(0) = x_0 \quad \text{and} \quad \gamma'(0) = \partial_{x_1}\big|_{x_0}.$$

Define now the set

$$\mathcal{K} = \left\{(y_1, \ldots, y_k) \in \mathbb{R}^k : \sum\nolimits_{i=1}^{k} y_i \phi_i|_{x_0} \in K_{x_0}\right\}.$$

Clearly $\mathcal{K}$ is a closed and convex subset of $\mathbb{R}^k$. Since $\phi \in K$ and K is invariant under parallel transport, we deduce that the curve $\sigma : (-\varepsilon, \varepsilon) \to \mathbb{R}^k$, given by

$$\sigma = (u_1 \circ \gamma, \ldots, u_k \circ \gamma),$$

lies in $\mathcal{K}$. It suffices to prove that $\sigma''(0)$ points into $\mathcal{K}$. Indeed, because $\mathcal{K}$ is convex, for any unit inward pointing normal ξ of $\mathcal{K}$ at $\sigma(0)$, we have

$$g(t) = \langle \xi, \sigma(t) - \sigma(0) \rangle \geq 0,$$

for any $t \in (-\varepsilon, \varepsilon)$. Because g attains its minimum at $t = 0$, from standard calculus we get that $g''(0) \geq 0$, which implies $\langle \sigma''(0), \xi \rangle \geq 0$. This completes the proof. □

Remark 2 According to the above result, it follows that if ϕ is a section lying in a set satisfying the conditions of Lemma 1 and $\mathcal{L}$ is a uniformly elliptic operator of second order, then section $\mathcal{L}\phi$ always points into K.

Theorem 8 (Strong Elliptic Maximum Principle) *Suppose that M is a Riemannian manifold (without boundary) and E a vector bundle of rank k over M equipped with a Riemannian metric g_E and a metric compatible connection. Let K be a closed fiber-convex subset of the bundle E that is invariant under parallel transport and $\phi \in \Gamma(E)$, $\phi : M \to K$, a smooth section such that*

$$\mathcal{L}\phi + \Psi(\phi) = 0,$$

where $\mathcal{L}$ is a uniformly elliptic operator of second order of the form given in (E) and Ψ is a smooth fiberwise map that points into K. If there exists a point $x_0 \in M$ such that $\phi|_{x_0} \in \partial K_{x_0}$, then $\phi|_x \in \partial K_x$ for any point $x \in M$. If, additionally, in a neighbourhood of $\phi|_{x_0}$ the set K_{x_0} is strictly convex and the boundary ∂K_{x_0} is C^2-smooth, then ϕ is a parallel section.

Proof We follow the exposition in [76]. Let $\{\phi_1, \ldots, \phi_k\}$ be a geodesic orthonormal frame field defined in a neighbourhood U around $x_0 \in M$. Hence, $\phi = u_1\phi_1 + \cdots + u_k\phi_k$, where $u_i : U \to \mathbb{R}$, $i \in \{1, \ldots, k\}$, are smooth functions. With respect to this frame we have

$$\begin{aligned}\mathcal{L}\phi &= \sum\nolimits_{i=1}^{k} \big(\mathcal{L}u_i + \big(\text{gradient terms of } u_i\big) + \sum\nolimits_{j=1}^{k} u_j \, g_E(\mathcal{L}\phi_j, \phi_i)\big)\phi_i \\ &= -\sum\nolimits_{i=1}^{k} g_E(\Psi(\phi), \phi_i)\phi_i.\end{aligned}$$

Therefore, the map $u : U \to \mathbb{R}^k$, $u = (u_1, \dots, u_k)$, satisfies a uniformly elliptic system of second order of the form

$$\widetilde{\mathcal{L}} u + \Phi(u) = 0, \tag{17}$$

where $\Phi : \mathbb{R}^k \to \mathbb{R}^k$, $\Phi = (\Phi_1, \dots, \Phi_k)$, is given by

$$\Phi_i(u) = \mathrm{g}_E\big(\Psi\big(\sum\nolimits_{j=1}^k u_j \phi_j\big) + \sum\nolimits_{j=1}^k u_j \mathcal{L}\phi_j \, , \phi_i\big), \tag{18}$$

for any $i \in \{1, \dots, k\}$. Consider now the convex set

$$\mathcal{K} = \big\{(y_1, \dots, y_k) \in \mathbb{R}^k : \sum\nolimits_{i=1}^k y_i \phi_i|_{x_0} \in K_{x_0}\big\}.$$

Claim 1: *For any point $x \in U$ we have $u(x) \in \mathcal{K}$.*

Indeed, fix a point $x \in U$ and let $\gamma : [0, 1] \to U$ be the geodesic curve joining the points x and x_0. Denote by θ the parallel section which is obtained by the parallel transport of $\phi|_x$ along the geodesic γ. Then,

$$\theta|_{\gamma(t)} = \sum\nolimits_{i=1}^k y_i \, \phi_i|_{\gamma(t)},$$

where $y_i : [0, 1] \to \mathbb{R}$, $i \in \{1, \dots, k\}$, are smooth functions. Because, θ and ϕ_i, $i \in \{1, \dots, k\}$, are parallel along γ, it follows that

$$0 = \nabla^{\gamma^* E}_{\partial_t} \theta = \sum\nolimits_{i=1}^k y_i'(t) \phi_i|_{\gamma(t)}.$$

Hence, $y_i(t) = y_i(0) = u_i(x)$, for any $t \in [0, 1]$ and $i \in \{1, \dots, k\}$. Therefore,

$$\theta|_{\gamma(1)} = \theta|_{x_0} = \sum\nolimits_{i=1}^k u_i(x) \phi_i|_{x_0}.$$

Since by our assumptions K is invariant under parallel transport, it follows that $\theta|_{x_0} \in K_{x_0}$. Hence, $u(U) \subset \mathcal{K}$ and this proves Claim 1.

Claim 2: *For any $y \in \partial\mathcal{K}$ the vector $\Phi(y)$ defined in (18) points into $\mathcal{K}$ at y.*

First note that the boundary of each slice K_x is invariant under parallel transport. From (18) we deduce that it suffices to prove that both terms appearing on the right hand side of (18) point into $\mathcal{K}$. The first term points into $\mathcal{K}$ by assumption on Ψ. The second term is inward pointing due to Lemma 1. This completes the proof of Claim 2.

Observe now that the solution of the uniformly second order elliptic partial differential system (17) satisfies all the assumptions of Theorem 7. Therefore, because $u(x_0) \in \partial\mathcal{K}$ it follows that $u(U)$ is contained in the boundary $\partial\mathcal{K}$ of $\mathcal{K}$. Consequently, $\phi|_x \in \partial K$ for any $x \in U$. Since M is connected, we deduce that

$\phi(M) \subset \partial K$. Note, that if $\mathcal{K}$ is additionally strictly convex at $u(x_0)$, then the map u is constant. This implies that

$$\phi|_x = \sum\nolimits_{i=1}^{k} u_i(x_0)\phi_i|_x$$

for any $x \in U$. Thus, ϕ is a parallel section taking all its values in ∂K. □

4.3.1 Maximum Principles for Symmetric Tensors

Let (E, g_E) be a Riemannian vector bundle over a manifold M. For any $\phi \in \mathrm{Sym}(E^* \otimes E^*)$, a real number λ is called *eigenvalue* of ϕ with respect to g_E at the point $x \in M$, if there exists a non-zero vector $v \in E_x$, such that

$$\phi(v, w) = \lambda g_E(v, w),$$

for any $w \in E_x$. The linear subspace $\mathrm{Eig}_{\lambda,\phi}(x)$ of E_x given by

$$\mathrm{Eig}_{\lambda,\phi}(x) = \{v \in E_x : \phi(v, w) = \lambda g_E(v, w), \text{ for any } w \in E_x\},$$

is called the *eigenspace* of λ at x. Since ϕ is symmetric, it admits k real eigenvalues $\lambda_1(x), \ldots, \lambda_k(x)$ at each point $x \in M$. We will always arrange the eigenvalues such that $\lambda_1(x) \leq \cdots \leq \lambda_k(x)$. If $\lambda_1(x) \geq 0$ (resp. > 0) we say that ϕ is *non-negatively (resp. positively) definite* at x.

Before stating the main results, let us recall the following definition due to Hamilton [44].

Definition 17 A fiberwise map $\Psi : \mathrm{Sym}(E^* \otimes E^*) \to \mathrm{Sym}(E^* \otimes E^*)$ is said to satisfy the *null-eigenvector condition*, if whenever ϑ is a non-negative symmetric 2-tensor at a point $x \in M$ and if $v \in T_xM$ is a null-eigenvector of ϑ, then $\Psi(\vartheta)(v, v) \geq 0$.

The next theorem consists the elliptic analogue of the maximum principle of Hamilton [45]. More precisely:

Theorem 9 *Let (M, g) be a Riemannian manifold (without boundary) and suppose that (E, g_E) is a Riemannian vector bundle over M equipped with a metric connection. Assume that $\phi \in \mathrm{Sym}(E^* \otimes E^*)$ is non-negative definite and satisfies*

$$\mathcal{L}\phi + \Psi(\phi) = 0,$$

where Ψ is a smooth fiberwise map satisfying the null-eigenvector condition. If there is a point of M where ϕ has a zero eigenvalue, then ϕ must have a zero eigenvalue everywhere.

Proof Denote by K the set of non-negative definite symmetric 2-tensors, i.e.,

$$K = \{\vartheta \in \mathrm{Sym}(E^* \otimes E^*) : \vartheta \geq 0\}.$$

Each set K_x is a closed and convex. Then,

$$\partial K_x = \{\vartheta \in K_x : \text{exists nonzero } v \in T_x M \text{ such that } \vartheta(v, \cdot) = 0\}.$$

The tangent cone of K_x at a point $\vartheta \in \partial K$ is given by

$$C_\vartheta K_x = \{\phi \in \mathrm{Sym}(E_x^* \otimes E_x^*) : \phi(v, v) \geq 0, \text{ for all } v \in \mathrm{Eig}_{0,\vartheta}(x)\}.$$

Claim 1. *The set K is invariant under parallel translation.*

Let $\gamma : [0, 1] \to M$ be a geodesic, P_t the parallel transport operator of vectors along γ and Π_t the parallel transport operator of 2-tensors along the curve γ. Consider $\vartheta \in K_{\gamma(0)}$. Then, for any $v \in T_{\gamma(0)}M$, we have

$$\partial_t \left\{ \left(\Pi_t \vartheta\right)(P_t v, P_t v) \right\} = \left(\nabla_{\partial_t} \Pi_t \vartheta\right)(P_t v, P_t v) + 2\Pi_t \theta \left(\nabla_{\partial_t} P_t v, P_t v\right) = 0.$$

Therefore, for any vector $v \in T_{\gamma(0)}M$, it holds $\left(\Pi_t \vartheta\right)(P_t v, P_t v) = \vartheta(v, v)$. Consequently, for any $w \in T_{\gamma(t)}M$, we obtain that

$$\left(\Pi_t \vartheta\right)(w, w) = \vartheta(P_t^{-1} w, P_t^{-1} w) \geq 0.$$

This proves the claim.

Claim 2. *Let $\Psi : \mathrm{Sym}(E^* \otimes E^*) \to \mathrm{Sym}(E^* \otimes E^*)$ be a smooth fiberwise map satisfying the null-eigenvector condition. Then, for any $x \in M$ and $\vartheta \in \partial K$, the vector $\Psi(x, \vartheta)$ points into K.*

Indeed, let $\vartheta \in \partial K_{x_0}$. Then Ψ points inwards of K_{x_0} at ϑ if and only if

$$\langle v^* \otimes v^*, \Psi(x, \vartheta) \rangle = \Psi(x, \vartheta)(v, v) \geq 0,$$

for any x in M and null-eigenvector $v \in T_{x_0}M$ of ϕ.

This complete the proof. □

4.3.2 A Second Derivative Test for Symmetric 2-tensors

Theorem 10 *Let (M, g) be a Riemannian manifold (without boundary) and (E, g_E) a Riemannian vector bundle of rank k over the manifold M equipped with a metric connection ∇. Suppose that $\phi \in \mathrm{Sym}(E^* \otimes E^*)$ is a smooth symmetric 2-tensor. If the biggest eigenvalue λ_k of ϕ admits a local maximum λ at an interior point $x_0 \in M$, then*

$$(\nabla \phi)(v, v) = 0 \quad \text{and} \quad (\mathcal{L}\phi)(v, v) \leq 0,$$

for all $v \in \mathrm{Eig}_{\lambda,\phi}(x_0)$ *and for all uniformly elliptic second order operators* $\mathcal{L}$.

Remark 3 The above theorem is due to Hamilton [44]. Replacing ϕ by $-\phi$ in Theorem 10, we get a similar result for the smallest eigenvalue λ_1 of ϕ.

Proof Let $v \in \mathrm{Eig}_{\lambda,\phi}(x_0)$ be a unit vector and $V \in \Gamma(E)$ such that $V|_{x_0} = v$ and $\nabla V|_{x_0} = 0$. Define the symmetric 2-tensor S given by $S = \phi - \lambda \mathrm{g}_E$. From our assumptions, S is non-positive definite in a small neighbourhood of x_0. Moreover, the biggest eigenvalue of S at x_0 equals 0. Consider the smooth function $f : M \to \mathbb{R}$, given by $f(x) = S(V|_x, V|_x)$. The function f is non-positive in the same neighbourhood around x_0 and attains a local maximum at x_0. In particular, $f(x_0) = 0$, $df(x_0) = 0$ and $(\mathcal{L}f)(x_0) \leq 0$. Consider a local orthonormal frame $\{e_1, \ldots, e_m\}$ with respect to g defined in a neighbourhood of the point x_0. A simple calculation yields

$$df(e_i) = \left(\nabla_{e_i} S\right)(V, V) + 2S\left(\nabla_{e_i} V, V\right).$$

Taking into account that g_E is parallel, we deduce that

$$0 = (\nabla f)(x_0) = (\nabla S)(v, v) = (\nabla \phi)(v, v).$$

Furthermore,

$$\begin{aligned}\nabla^2_{e_i,e_j} f = {} & (\nabla^2_{e_i,e_j} S)(V, V) + 2S(V, \nabla^2_{e_i,e_j} V) \\ & +2\left(\nabla_{e_i} S\right)(\nabla_{e_j} V, V) + 2\left(\nabla_{e_j} S\right)(\nabla_{e_i} V, V) \\ & +2S(\nabla_{e_i} V, \nabla_{e_j} V).\end{aligned}$$

Bearing in mind the definition of S and using the fact that g_E is parallel with respect to ∇, we obtain

$$\begin{aligned}\mathcal{L}f &= (\mathcal{L}\phi)(V, V) + 2S(V, \mathcal{L}V) \\ &\quad +2\sum\nolimits_{i,j=1}^{m} a_{ij}\left\{S(\nabla_{e_i} V, \nabla_{e_j} V) + 2(\nabla_{e_i} S)(\nabla_{e_j} V, V)\right\} \\ &= (\mathcal{L}\phi)(V, V) + 2S(V, \mathcal{L}V) \\ &\quad +2\sum\nolimits_{i,j=1}^{m} a_{ij}\left\{S(\nabla_{e_i} V, \nabla_{e_j} V) + 2(\nabla_{e_i} S)(\nabla_{e_j} V, V)\right\}.\end{aligned}$$

Estimating at x_0 and taking into account that $V|_{x_0} = v$ is a null eigenvector of S at x_0, we get

$$0 \geq (\mathcal{L}f)(x_0) = (\mathcal{L}\phi)(v, v).$$

This completes the proof. □

5 Graphical Submanifolds

5.1 Definitions

Let (M, g_M) and (N, g_N) be Riemannian manifolds of dimension m and n, respectively. The induced metric on $M \times N$ will be denoted by $\mathrm{g}_{M\times N} = g_M \times g_N$. We often denote the product metric also by $\langle \cdot, \cdot \rangle$. A smooth map $f : M \to N$ defines an embedding $F : M \to M \times N$, given by $F(x) = (x, f(x))$, for any $x \in M$. The *graph* of f is defined to be the submanifold

$$\Gamma(f) = F(M) = \{(x, f(x)) \in M \times N : x \in M\}.$$

Since F is an embedding, it induces another Riemannian metric $g = F^*\mathrm{g}_{M\times N}$ on M. The two natural projections $\pi_M : M \times N \to M$ and $\pi_N : M \times N \to N$ are submersions, that is they are smooth and have maximal rank. Note that the tangent bundle of the product manifold $M \times N$, splits as a direct sum

$$T(M \times N) = TM \oplus TN.$$

The four metrics g_M, g_N, $\mathrm{g}_{M\times N}$ and g are related by

$$\mathrm{g}_{M\times N} = \pi_M^* g_M + \pi_N^* g_N \quad \text{and} \quad g = F^*\mathrm{g}_{M\times N} = g_M + f^* g_N\,. \tag{19}$$

The Levi-Civita connection $\nabla^{\mathrm{g}_{M\times N}}$ associated to the Riemannian metric $\mathrm{g}_{M\times N}$ on $M \times N$ is related to the Levi-Civita connections ∇^{g_M} on (M, g_M) and ∇^{g_N} on (N, g_N) by

$$\nabla^{\mathrm{g}_{M\times N}} = \pi_M^* \nabla^{g_M} \oplus \pi_N^* \nabla^{g_N}\,.$$

The corresponding curvature operator $\widetilde{R}$ on the product $M \times N$ is related to the curvature operators on (M, g_M) and R_N on (N, g_N) by

$$\widetilde{R} = \pi_M^* R_M \oplus \pi_N^* R_N.$$

The map $f : M \to N$ is called *minimal* if $\Gamma(f) \subset M \times N$ is minimal.

5.2 Singular Value Decomposition

For any fixed point $x \in M$, let $\lambda_1^2(x) \leq \cdots \leq \lambda_m^2(x)$ be the eigenvalues of f^*g_N with respect to g_M. The corresponding values $\lambda_i \geq 0$, $i \in \{1, \ldots, m\}$, are usually called *singular values* of the differential df of f at the point x. Let $r = r(x) = \operatorname{rank} df(x)$. Then, $r \leq \min\{m, n\}$ and $\lambda_1(x) = \cdots = \lambda_{m-r}(x) = 0$.

It is well-known that the singular values can be used to define the so-called *singular decomposition* of df. At the point x, consider an orthonormal basis $\{\alpha_1, \ldots, \alpha_{m-r}; \alpha_{m-r+1}, \ldots, \alpha_m\}$ with respect to g_M which diagonalizes $f^* g_N$. Moreover, at $f(x)$ consider an orthonormal basis $\{\beta_1, \ldots, \beta_{n-r}; \beta_{n-r+1}, \ldots, \beta_n\}$ with respect to g_N such that, for any $i \in \{m-r+1, \ldots, m\}$,

$$df(\alpha_i) = \lambda_i(x)\beta_{n-m+i}.$$

It is well-known fact that, with the above ordering, the singular values give rise to continuous functions. In the matter of fact, they are even smooth on an open and dense subset of M. In particular, they are smooth on open subsets where the corresponding multiplicities are constant and the corresponding eigenspaces are smooth distributions; see [83].

We may define a special basis for the tangent and the normal space of the graph in terms of the singular values. The vectors

$$e_i = \begin{cases} \alpha_i & , 1 \leq i \leq m-r, \\ \dfrac{\alpha_i}{\sqrt{1+\lambda_i^2(x)}} & , m-r+1 \leq i \leq m, \end{cases} \tag{20}$$

form an orthonormal basis with respect to the metric $g_{M\times N}$ of the tangent space $dF(T_x M)$ of the graph $\Gamma(f)$ at x. Moreover, the vectors

$$\xi_i = \begin{cases} \beta_i & , 1 \leq i \leq n-r, \\ \dfrac{-\lambda_{i+m-n}(x)\alpha_{i+m-n} \oplus \beta_i}{\sqrt{1+\lambda_{i+m-n}^2(x)}} & , n-r+1 \leq i \leq n, \end{cases} \tag{21}$$

give an orthonormal basis with respect to $g_{M\times N}$ of the normal space $N_x M$.

5.3 Length and Area Decreasing Maps

Let (M, g_M) and (N, g_N) be two Riemannian manifolds of dimensions m and n respectively. For any smooth map $f : M \to N$ its differential df induces a map $\Lambda^k df : \Lambda^k TM \to \Lambda^k TN$ given by

$$\left(\Lambda^k df\right)(v_1, \ldots, v_k) = df(v_1) \wedge \cdots \wedge df(v_k),$$

for any smooth vector fields $v_1, \ldots, v_k \in TM$. The map $\Lambda^k df$ is called the *k-Jacobian* of f. The *supremum norm* or the *k-dilation* $|\Lambda^k df|(x)$ of the map f at a point $x \in M$ is defined as the supremum of

$$\sqrt{\det\left((f^*g_N(v_i, v_j))_{1\leq i,j\leq k}\right)}$$

when $\{v_1, \dots, v_m\}$ runs over all orthonormal bases of T_xM.

Definition 18 A smooth map $f : (M, g_M) \to (N, g_N)$ between Riemannian manifolds is called (*weakly*) *k-volume decreasing* if $|\Lambda^k df| \leq 1$, *strictly k-volume decreasing* if $|\Lambda^k df| < 1$ and *k-volume preserving* if $|\Lambda^k df| = 1$. For $k = 1$ we use the term *length* instead of 1-volume and if $k = 2$ we use the term *area* instead of 2-volume.

There is a way to express the length and area decreasing property of a map in terms of positivity of symmetric tensors. Define on M the symmetric 2-tensors $S_{M\times N}$ and S given by

$$S_{M\times N} = \pi_M^* g_M - \pi_N^* g_N \quad \text{and} \quad S = F^* S_{M\times N} = g_M - f^* g_N.$$

With respect to the basis of the singular value decomposition, we have

$$S_{M\times N}(e_i, e_j) = \frac{1-\lambda_i^2}{1+\lambda_i^2}\delta_{ij}, \quad 1 \leq i, j \leq m. \tag{22}$$

Hence, the eigenvalues $\mu_1, \mu_2, \dots, \mu_m$ of S with respect to g, are

$$\mu_1 = \frac{1-\lambda_m^2}{1+\lambda_m^2} \leq \cdots \leq \mu_m = \frac{1-\lambda_1^2}{1+\lambda_1^2}.$$

Hence, f is length decreasing if $S \geq 0$. Additionally let us mention that

$$S_{M\times N}(\xi_i, \xi_j) = \begin{cases} -\delta_{ij} & , \ 1 \leq i \leq n-r, \\ -\dfrac{1-\lambda_{i+m-n}^2}{1+\lambda_{i+m-n}^2}\delta_{ij} & , \ n-r+1 \leq i \leq n. \end{cases} \tag{23}$$

and

$$S_{M\times N}(e_{m-r+i}, \xi_{n-r+j}) = -\frac{2\lambda_{m-r+i}}{1+\lambda_{m-r+i}^2}\delta_{ij}, \quad 1 \leq i, j \leq r. \tag{24}$$

Observe now that, for any pair of indices $i, j \in \{1, \dots, m\}$, we have

$$\mu_i + \mu_j = \frac{1-\lambda_i^2}{1+\lambda_i^2} + \frac{1-\lambda_j^2}{1+\lambda_j^2} = \frac{2(1-\lambda_i^2\lambda_j^2)}{(1+\lambda_i^2)(1+\lambda_j^2)}.$$

Hence, the map is strictly area decreasing, if and only if the tensor S is *strictly* 2-*positive*, i.e., the sum of the two smallest eigenvalues is positive. The 2-positivity of a tensor $T \in \mathrm{Sym}(T^*M \otimes T^*M)$ can be expressed as the positivity of another tensor $T^{[2]} \in \mathrm{Sym}(\Lambda^2 T^*M \otimes \Lambda^2 T^*M)$. Indeed, let P and Q be two symmetric 2-tensors. Then, the *Kulkarni-Nomizu product* $P \owedge Q$ given by

$$\begin{aligned}(P \owedge Q)(v_1 \wedge w_1, v_2 \wedge w_2) = {} & P(v_1, v_2)Q(w_1, w_2) + P(w_1, w_2)Q(v_1, v_2) \\ & - P(w_1, v_2)Q(v_1, w_2) - P(v_1, w_2)Q(w_1, v_2)\end{aligned}$$

is an element of the vector bundle $\mathrm{Sym}(\Lambda^2 T^*M \otimes \Lambda^2 T^*M)$. Now, to every element $T \in \mathrm{Sym}(T^*M \otimes T^*M)$ let us assign an element $T^{[2]}$ of the bundle $\mathrm{Sym}(\Lambda^2 T^*M \otimes \Lambda^2 T^*M)$, by setting

$$T^{[2]} = T \owedge g.$$

We point out that the Riemannian metric G of $\Lambda^2 TM$ is given by

$$G = \tfrac{1}{2} g \owedge g = \tfrac{1}{2} g^{[2]}.$$

The relation between the eigenvalues of the tensor T and the eigenvalues of $T^{[2]}$ is explained in the following lemma:

Lemma 2 *Let T be a symmetric 2-tensor with eigenvalues $\mu_1 \leq \cdots \leq \mu_m$ and corresponding eigenvectors $\{v_1, \ldots, v_m\}$ with respect to the metric g. Then the eigenvalues of the symmetric 2-tensor $T^{[2]}$ with respect to G are $\mu_i + \mu_j$, for any $1 \leq i < j \leq m$, with corresponding eigenvectors $v_i \wedge v_j$, for any $1 \leq i < j \leq m$.*

5.4 Minimal Graphs in the Euclidean Space

Let us discuss the case of graphs generated by smooth maps $f : \mathbb{R}^m \to \mathbb{R}^n$. The induced metric g on the graph is given in local coordinates in the form

$$g_{ij} = \delta_{ij} + \sum_{i,j=1}^{m} \sum_{\alpha=1}^{n} f^{\alpha}_{x_i} f^{\alpha}_{x_j}.$$

As usual, the components of the inverse matrix of the induced metric g are denoted by g^{ij}. It is not difficult to show that $\Gamma(f)$ is minimal if and only if the components of

$$f = (f^1, \ldots, f^m)$$

satisfy the following system of differential equations

$$\sum_{i,j=1}^{m} g^{ij} f^{\alpha}_{x_i x_j} = 0. \tag{MSE}$$

The equation is known in the literature as the *minimal surface equation*. For graphical hypersurfaces, that is for graphs generated by functions smooth $f : U \subset \mathbb{R}^m \to \mathbb{R}$, minimality is expressed by the equation

$$\operatorname{div}\left(\frac{\operatorname{grad} f}{\sqrt{1 + |\operatorname{grad} f|^2}}\right) = 0.$$

There is a long history of attempts to study entire solutions of the minimal surface equation. Bernstein [8] proved that the only entire minimal graphs in the $\mathbb{R}^3$ are planes. However, there was a gap in the original proof of Bernstein which was fixed 40 years later; see [51, 65]. In the meantime, several complex analysis proofs have been obtained; for more details see the surveys of Osserman [70, 71].

It was conjectured for a long time that the theorem of Bernstein holds in any dimension for graphical hypersurfaces. For $m = 3$, its validity was proved by De Giorgi [30], for $n = 4$ by Almgren [3] and for $m = 5, 6, 7$ by Simons [80]. It was a big surprise when Bombieri, De Giorgi and Giusti [10] proved that, for $m \geq 8$, there are entire solutions of the minimal surface equation other than the affine ones.

In higher codimensions, the situation is more complicated. There are plenty of non-flat entire minimal graphs. For example, the graph of an entire holomorphic map $f : \mathbb{C}^m \to \mathbb{C}^n$ is minimal. Moreover, Osserman [70] has constructed examples of complete minimal two-dimensional graphs in $\mathbb{R}^4$, which are not complex analytic with respect to any orthogonal complex structure on $\mathbb{R}^4$. For instance, the graph $\Gamma(f)$ over the map $f : \mathbb{R}^2 \to \mathbb{R}^2$, given by

$$f(x, y) = \left(e^{\frac{x}{2}} - 3e^{-\frac{x}{2}}\right)\left(\cos\frac{y}{2}, -\sin\frac{y}{2}\right)$$

for any $(x, y) \in \mathbb{R}^2$, is such an example. Now the obvious questions became:

Question 1 If entire solutions of the minimal surface equation need not be linear, do they have any other distinguishing characteristics? What additional restrictions on entire solutions would guarantee linearity in all dimensions?

The first result in this direction was obtained by Osserman [69] for two-dimensional graphs, generated by maps $f : \mathbb{R}^2 \to \mathbb{R}^n$. He proved that if the differential df of the map f is bounded, then must be a plane. In fact, he proved the following more general theorem:

Theorem 11 *Suppose that Σ is a complete, oriented minimal surface (not necessarily graphical) in the euclidean space $\mathbb{R}^n$. Assume that the Gauss map of Σ omits an open neighbourhood in the Grassmannian. Then, Σ is flat.*

Let us restrict now in two dimensional graphs in $\mathbb{R}^4$, i.e., minimal graphs generated by maps $f = (f_1, f_2)$. For such graphs, Simon [81] proved that if one component of f have bounded gradient, then f is affine. Later on, Schoen [79] obtained a Bernstein-type result by imposing the assumption that $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a diffeomorphism. Moreover, Ni [68] has derived a result of Bernstein type under the assumption that f is an area-preserving map. In this case, area preserving is equivalent with the condition $|\det(df)| = 1$. The function $\mathrm{Jac}(f) = \det(df)$ is called the *Jacobian determinant* of f. All these result were generalized by Hasanis, Savas-Halilaj and Vlachos in [47, 48], just by assuming that $\mathrm{Jac}(f)$ is bounded. In fact, the following result is shown:

Theorem 12 *Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be an entire solution of the minimal surface equation. Assume that $\Gamma(f)$ is not a plane. Then, $\Gamma(f)$ is a complex analytic curve if and only if the Jacobian determinant $\mathrm{Jac}(f)$ of f does not take every real value. In particular if $\Gamma(f)$ is a complex analytic curve, then:*

(a) *The Jacobian determinant $\mathrm{Jac}(f)$ takes every real value in $(0, +\infty)$ or in $[0, +\infty)$ if f is holomorphic.*
(b) *The Jacobian determinant $\mathrm{Jac}(f)$ takes every real value in $(-\infty, 0)$ or in $(-\infty, 0]$, if f is anti-holomorphic.*

All these proofs use strongly the fact that the Gauss map of a minimal surface in the euclidean space is anti-holomorphic.

The first Bernstein-type theorem which was valid for arbitrary dimension and codimension is due to Hildebrandt, Jost, and Widman [50]. They obtained such a result under the assumption of a certain quantitative bound for the slope, that is a bound on the norm of the differential of the generating map.

Let us describe here briefly their technique. Note at first that a bound on the differential of the map forces the Gauss map of the graph to lie in a bounded region of the Grassmannian manifold. In particular, the first step is to determine which bounds on the differential will force the Gauss map to have its range in a sufficiently small convex subset of the Grassmannian. The second step is to find a convex function defined on the convex set, which contains the Gauss image of the graph, and to compose it with the Gauss map. By Theorem 2 of Ruh and Vilms, the Gauss map is harmonic. Consequently, the composition of the Gauss map with the convex function will give rise to a subharmonic function defined on the graph. The third step is to show that this particular subharmonic function is constant and the Gauss map is parallel. Of course, there are many difficulties to overcome to run this program. The first problem is the complexity of the Grassmannian manifolds. For example, it is not so easy to identify which are the convex subsets of the Grassmannian and their corresponding convex supporting functions. One way is to consider distance balls. In fact, Hildebrandt, Jost, and Widman [50] identified the largest ball in the Grassmannian manifold on which the square of the distance function is convex. Another major difficulty is that an entire euclidean minimal graph is complete and non-compact. Consequently, the standard maximum principle cannot be applied directly. Let us mention here that the original assumption on the slope was obtained by Hildebrandt, Jost, and Widman in [50] was

$$E(f) = \sqrt{\det(I + df^t df)} \leq \beta_0 < \cos^{-p}\left(\frac{\pi}{2\sqrt{2\,p}}\right)$$

where β_0 is a constant and $p = \min\{m, n\}$. Over the years, the bound on $E(f)$ was improved. Recently, Jost, Xin, and Yang [56] proved the following:

Theorem 13 *Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be an entire solution of the minimal surface equation. Suppose that there exists a number β_0 such that*

$$\beta_0 < \begin{cases} 3, & \text{if} \quad n \geq 2, \\ \infty, & \text{if} \quad n = 1, \end{cases}$$

and

$$E(f) = \sqrt{\det(I + df^t df)} \leq \beta_0.$$

Then $\Gamma(f)$ is an affine subspace of $\mathbb{R}^m \times \mathbb{R}^n$.

Remark 4 For codimension one graphs, the above theorem was first obtained by Moser [67].

Question 2 Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be an entire solution of the minimal surface equation such that

$$E(f) = \sqrt{\det(I + df^t df)} < 9.$$

Is it true that $\Gamma(f)$ is an affine subspace of $\mathbb{R}^m \times \mathbb{R}^n$?

Remark 5 The number 9 in the above conjecture should be the sharp bound. The reason is that there are examples of Lipschitz minimal maps constructed by Lawson and Osserman [58] with $E(f) = 9$; see also [37]. These examples are generated from the map $f : \mathbb{C}^2 - \{0\} = \mathbb{R}^4 - \{0\} \to \mathbb{R} \times \mathbb{C} = \mathbb{R}^3$ given by

$$f(x) = \frac{\sqrt{5}}{2}|x|\mathcal{H}\left(\frac{x}{|x|}\right),$$

where $\mathcal{H} : \mathbb{C}^2 \to \mathbb{R} \times \mathbb{C}$ is the Hopf-map $\mathcal{H}(z, w) = (|z|^2 - |w|^2, 2z\overline{w})$.

Let us conclude this section by mentioning some results in special situations. The first one is due to Fischer-Colbrie [37] and it says that a 3-dimensional complete minimal graph with bounded differential is totally geodesic. In the matter of fact, the following holds:

Theorem 14 *Let $f : \mathbb{R}^3 \to \mathbb{R}^n$ be an entire solution of the minimal surface equation. If $|df|$ is uniformly bounded, then $\Gamma(f)$ is flat.*

In codimension two, no specific bound on the differential of f is needed. Recently, Assimos and Jost [6] obtained the following interesting theorem:

Theorem 15 *Let $f : \mathbb{R}^m \to \mathbb{R}^2$ be an entire solution of the minimal surface equation. Suppose that there exists a number β_0 such that*

$$E(f) = \sqrt{\det(I + df^t df)} \leq \beta_0.$$

Then $\Gamma(f)$ is an affine subspace of $\mathbb{R}^m \times \mathbb{R}^2$.

The next result we would like to mention is due to Wang [97]. He obtained the following theorem for strictly area decreasing minimal graphs.

Theorem 16 *Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be an entire solution of the minimal surface equation. Suppose that there exists numbers $\delta_1 \in (0, 1)$ and $\delta_2 > 0$ such that $|\Lambda^2 df| \leq 1 - \delta_1$ and $E(f) \leq \delta_2$. Then $\Gamma(f)$ is an affine subspace of $\mathbb{R}^m \times \mathbb{R}^n$.*

Remark 6 The above cannot be extended for k-volume decreasing minimal maps with $k > 2$. For example, consider $f : \mathbb{C}^2 = \mathbb{R}^4 \to \mathbb{C}^2 = \mathbb{R}^4$, given by

$$f(z, w) = (\beta_0 z + h(w), w),$$

where $z, w \in \mathbb{C}$, $h : \mathbb{C} \to \mathbb{C}$ is a non-affine holomorphic map and β_0 a real number. Observe that the graph $\Gamma(f)$ is a non-flat minimal submanifold of $\mathbb{R}^8$ and $|\Lambda^4 df| = |\beta_0|$. Consequently, there exists an abundance of non-flat minimal graphs in the euclidean space with arbitrary small 4-Jacobian.

6 Mean Curvature Flow

In this section, we introduce the notion of the mean curvature flow. Later, we will examine how various geometric quantities evolve under the mean curvature flow. Suppose that M is a manifold of dimension m, let $T > 0$ be a real number and $F : M \times [0, T) \to N$ a smooth time-dependent family of immersions of M into a Riemannian manifold N of dimension n. We follow the exposition in [5, 32, 63, 86].

Definition 19 Let N be a Riemannian manifold. We say that a family of immersions $F\colon M \times [0, T) \to N$ evolves by *mean curvature flow* (MCF for short) with initial data the immersion $F_0\colon M \to N$ if it satisfies the initial value problem

$$\begin{cases} dF_{(x,t)}(\partial_t) = H(F(x, t)) \\ F(x, 0) = F_0(x) \end{cases},$$

for any $(x, t) \in M \times [0, T)$, where $H\big(F(x, t)\big)$ denotes the mean curvature vector of the immersion $F(\cdot, t) : M \to N$ at the point $x \in M$.

6.1 Basic Facts for Systems of Parabolic PDEs

In this section we recall basic facts about solvability of Cauchy problems; for more details see [57].

6.1.1 Differential Operators

Let M be a smooth manifold equipped with a Riemannian metric g whose associated Levi-Civita connection is ∇^M. Suppose that E_1 and E_2 are two vector bundles over M and assume that E_1 is equipped with a Riemannian metric h and a compatible connection ∇^{E_1}. As in Definition 4, from the connections ∇^M and ∇^{E_1}, one can form the k-th derivative ∇^k of a section $\phi \in \Gamma(E_1)$.

Definition 20 A map $P\colon \Gamma(E_1) \to \Gamma(E_2)$ of the form

$$(P\phi)(x) = Q\big(x, \nabla^1\phi(x), \dots, \nabla^k\phi(x)\big) \in (E_2)_x,$$

where Q is smooth in all its variables, will be called *differential operator of order* k. In the case where P is $\mathbb{R}$-linear, we say that P is a *linear differential operator of order* k. Otherwise, we say that P is non-linear.

Suppose that $P\colon \Gamma(E_1) \to \Gamma(E_2)$ is a linear differential operator of degree k. Then, in index notation, it can be written in the form

$$P\phi = \sum_{i_1,\dots,i_k} A^{i_1\cdots i_k} \nabla^k_{\partial_{x_{i_1}}\cdots\partial_{x_{i_k}}} \phi + \cdots + \sum_{i_1} A^{i_1} \nabla^1_{\partial_{x_{i_1}}} \phi + A^0\phi,$$

where for each $x \in M$, $A^{i_1\cdots i_k}(x)\colon (E_1)_x \to (E_2)_x$ is linear map. These maps are called the *coefficients* of the linear operator P.

Definition 21 Let $P\colon \Gamma(E_1) \to \Gamma(E_2)$ be a linear differential operator of order k, let x be a point in M and $\zeta = \sum_i \zeta_i dx_i \in T^*_xM$. The linear map $\sigma_\zeta(P;x)\colon (E_1)_x \to (E_2)_x$, given by

$$\sigma_\zeta(P;x)\phi = \sum_{i_1,\dots,i_k} \zeta_{i_1}\cdots\zeta_{i_k} A^{i_1\dots i_k}\phi|_x,$$

is called the *principal symbol of the operator P at the point x and in the direction* ζ. In particular, the operator P is called *elliptic* if its principal symbol is an isomorphism, for every point x and every non-zero direction ζ.

Definition 22 The *differential* or the *linearization* of P at ϕ_0, if it exists, is defined to be the linear map $DP|_{\phi_0}\colon \Gamma(E_1) \to \Gamma(E_2)$, given by the expression

$$DP|_{\phi_0}(\psi) = \lim_{s\to 0} \frac{P(\phi_0 + s\psi) - P(\phi_0)}{s},$$

for any $\psi \in \Gamma(E_1)$.

Definition 23 Let $P\colon \Gamma(E_1) \to \Gamma(E_2)$ be a differential operator of order k. We say that P is *elliptic, undetermined elliptic or overdetermined elliptic* if its linearization is so.

Example 1 Let $f\colon M \to N$ be a smooth map between manifolds endowed with Riemannian metrics g_M and g_N, respectively, and consider the operator $\Delta_{g_M,g_N}\colon C^\infty(M) \to C^\infty(M)$, given by

$$\Delta_{g_M,g_N} f = \operatorname{tr}_{g_M} B,$$

where B stands for the Hessian of f. In local coordinates, we have

$$\Delta_{g_M,g_N} f = \sum_{i,j,\alpha} g_M^{ij}\big(f^\alpha_{x_i x_j} - \sum_k \Gamma^k_{ij} f^\alpha_{x_k} + \sum_{\gamma,\delta} \Gamma^\alpha_{\gamma\delta} f^\gamma_{x_i} f^\delta_{x_j}\big)\partial_{y_\alpha}.$$

The linearization of $\Delta_{g_M,g_N} f$ is

$$\begin{aligned} D\Delta_{g_M,g_N}|_f(G) &= \lim_{s\to\infty} \frac{\Delta_{g_M,g_N}(f+sG) - \Delta_{g_M,g_N}(f)}{s} \\ &= \sum_{i,j} g_M^{ij} G^\alpha_{x_i x_j}\partial_{y_\alpha} + \text{lower order terms.} \end{aligned}$$

Hence, for any

$$\zeta = (\zeta_1, \ldots, \zeta_m) \quad \text{and} \quad \phi = (\phi_1, \ldots, \phi_n)$$

we have

$$\sigma_\zeta(D\Delta_{g_M,g_N}, x)\phi = \sum\nolimits_{i,j} g_M^{ij}\zeta_i\zeta_j\phi|_x = |\zeta|_g^2\phi|_x,$$

Consequently, the Laplacian operator Δ_{g_M,g_n} is elliptic.

6.1.2 Time-Dependent Vector Bundles

Suppose that $I \subset \mathbb{R}$ is an open interval and let $\{g(t)\}_{t\in I}$ be a smooth family of Riemannian metrics on a manifold M. This means that for any $(x,t) \in M \times I$ we have an inner product $g_{(x,t)}$ on T_xM. We can regard $\{g(t)\}_{t\in I}$ as a metric g acting on the *spatial tangent bundle* $\mathcal{H}$, defined by

$$\mathcal{H} = \{v \in T(M \times \mathbb{R}) : d\pi_2(v) = 0\},$$

where $\pi_2 : M \times I \to I$ is given by $\pi_2(x,t) = t$. Note that each $g(t)$ is a metric on $\mathcal{H}$ since $\mathcal{H}_{(x,t)}$ is isomorphic to T_xM via π_2. We can even extend g into a metric on $M \times I$, with respect to which we have the orthogonal decomposition

$$T(M \times I) = \mathcal{H} \oplus \mathbb{R}\partial_t.$$

Since $\mathcal{H}$ is a subbundle of $T(M \times I)$, any section of $\mathcal{H}$ is also a section of $T(M \times I)$. Sections of $\Gamma(\mathcal{H})$ are called *spatial vector fields*. There is a natural connection ∇ on $M \times I$. Namely, define ∇ by

$$\nabla_v w = \nabla_v^{g(t)} w,\ \nabla_v \partial_t = 0,\ \nabla_{\partial t}\partial_t = 0 \text{ and } \nabla_{\partial_t} v = [\partial_t, v], \tag{25}$$

for any $v, w \in \Gamma(\mathcal{H})$, where $\nabla^{g(t)}$ stand for the Levi-Civita connection of $g(t)$. One can readily check that ∇ is compatible with g, i.e.,

$$vg(w_1, w_2) = g(\nabla_v w_1, w_2) + g(w_1, \nabla_v w_2),$$

for any $v \in \mathfrak{X}(M \times \mathbb{R})$ and $w_1, w_2 \in \Gamma(\mathcal{H})$. Moreover, for any $w_1, w_2 \in \Gamma(\mathcal{H})$,

$$\nabla_{w_1} w_2 - \nabla_{w_2} w_1 = [w_1, w_2].$$

The situation we discussed above occurs, when we have a family of immersions $F : M \times I \to N$. In this case, F^*h gives a family of metrics on M. Endowing $M \times I$ with the connection ∇, we have for any $v \in \Gamma(\mathcal{H})$ that

$$\nabla_{\partial_t}^{F^*TN} dF(v) - \nabla_v^{F^*TN} dF(\partial_t) = dF([\partial_t, v]) = dF(\nabla_{\partial_t} v).$$

6.1.3 Parabolic Differential Equations

Let M be a manifold equipped with a family of metrics $\{g(t)\}_{[0,T)}$. Denote by $\{\nabla^{g(t)}\}_{t\in[0,T)}$ the corresponding Levi-Civita connections. Let E_1 and E_2 be vector bundles over M and assume that E_1 is equipped with a fixed time independent metric h and connections $\{\nabla(t)\}_{t\in[0,T)}$ that are compatible with h, i.e.,

$$vh(\phi_1, \phi_2) = h\big(\nabla(t)_v\phi_1, \phi_2\big) + h\big(\phi_1, \nabla(t)_v\phi_2\big),$$

for any tangent vector v, sections $\phi_1, \phi_2 \in \Gamma(E)$ and any time $t \in [0, T)$.

As in Definition 4, by coupling $\nabla(t)$ with $\nabla^{g(t)}$ we obtain repeated covariant derivatives $\nabla^k(t)$ acting on sections of E_1. Suppose now that $\{\phi(t)\}_{t\in[0,T)}$ is a smooth time-dependent family of sections of E_1, where smooth means that for any fixed $(x,t) \in M \times [0,T)$, the time-derivative

$$(\nabla_{\partial_t}\phi)(x,t) = \lim_{h\to 0} \frac{\phi(x, t+h) - \phi(x,t)}{h}$$

exists. Hence, $\{\nabla_{\partial_t}\phi\}_{t\in[0,T)}$ is another one parameter family of sections on E_1. We are interested now in expressions of the form:

$$(\nabla_{\partial_t}\phi)(x,t) = (P\phi)(x,t) = Q\big(x,t,\nabla^1(t)\phi(x,t),\ldots,\nabla^k(t)\phi(x,t)\big), \tag{26}$$

where now $P : \Gamma(E_1) \to \Gamma(E_2)$ is a time-dependent differentiable operator of order k. If for each fixed t the operator P is linear elliptic, we say that (26) is a *linear parabolic differential equation*. We say that (26) represents a *non-linear parabolic differential equation* if and only if, for any $\phi \in \Gamma(E_1)$, its linearization is parabolic.

Theorem 17 *If the differential operator P is parabolic at $\phi_0 \in \Gamma(E_1)$, then there exist a $T > 0$ and a smooth family $\phi(t) \in \Gamma(E_1)$, for $t \in [0,T]$, such that there exists a unique smooth solution for the initial value problem*

$$\begin{cases} \nabla_{\partial_t}\phi = P\phi, \\ \phi(0) = \phi_0. \end{cases}$$

for $t \in [0,T]$, where T depends on the initial data ϕ_0.

We close this section with an application of this general theory.

Definition 24 Let (M, g_M) and (N, g_N) be Riemannian manifolds. We say that a family of smooth maps $F\colon M \times [0,T) \to N$ evolves by *(harmonic) heat flow*, with initial data $F_0\colon M \to N$, if it satisfies the initial value problem

$$\begin{cases} \nabla_{\partial_t} dF = dF(\partial_t) = \Delta_{g_M,g_N} F, \\ F(\cdot,0) = F_0. \end{cases} \tag{27}$$

Theorem 18 *Let (M, g_M) be a compact Riemannian manifold and suppose that $F_0\colon (M, g_M) \to (N, g_N)$ is a smooth map into a Riemannian manifold (N, g_N). Then, (27) admits a unique, smooth solution on a maximal time interval $[0, T_{\max})$, where $0 < T_{\max} \leq \infty$.*

Proof We already computed that for $\zeta \in T^*M$, we have

$$\sigma_\zeta(D\Delta_{g_M,g_N}, x) = |\zeta|_g^2 I.$$

Hence, the parabolic theory can be used to ensure short-time existence. □

6.2 Short-time Existence of the Mean Curvature Flow

A supposed solution F of MCF can be represented in local coordinates as

$$F(x_1,\ldots,x_m,t) = \big(F^1(x_1,\ldots,x_m,t),\ldots,F^n(x_1,\ldots,x_m,t)\big).$$

Then, from (2) we have

$$H = \sum_{i,j,\alpha} g^{ij}\big(F^{\alpha}_{x_i x_j} - \sum_{k} \Gamma^{k}_{ij} F^{\alpha}_{x_k} + \sum_{\gamma,\delta} \Gamma^{\alpha}_{\gamma\delta} F^{\gamma}_{x_i} F^{\delta}_{x_j}\big)\partial_{y_\alpha},$$

where

$$g_{ij} = \sum_{\alpha,\beta} h_{\alpha\beta} F^{\alpha}_{x_i} F^{\beta}_{x_j} \quad \text{and} \quad \Gamma^{k}_{ij} = \frac{1}{2}\sum_{l} g^{kl}(\partial_{x_i} g_{jl} + \partial_{x_j} g_{il} - \partial_{x_l} g_{ij}).$$

Note that g is the induced metric and it depends on F. Hence,

$$\partial_{x_i} g_{jl} = \sum_{\beta,\gamma} \big(h_{\beta\gamma} F^{\gamma}_{x_i x_j} F^{\beta}_{x_l} + h_{\beta\gamma} F^{\gamma}_{x_j} F^{\beta}_{x_i x_l}\big) + \text{lower order terms}$$

and consequently

$$\Gamma^{k}_{ij} = \sum_{l,\beta,\gamma} g^{kl} h_{\beta\gamma} F^{\gamma}_{x_l} F^{\beta}_{x_i x_j} + \text{lower order terms}. \tag{28}$$

Combining the formula (28) with equation (2), we obtain

$$H = \sum_{i,j,\alpha,\beta} g^{ij}\big(\delta_{\alpha\beta} - \sum_{k,l,\gamma} g^{kl} h_{\beta\gamma} F^{\alpha}_{x_k} F^{\gamma}_{x_l}\big) F^{\beta}_{x_i x_j} \partial_{y_\alpha} + \text{lower order terms}.$$

By a straightforward computation, we get

$$\begin{aligned} DH|_F(G) &= \lim_{s\to 0} \frac{H(F+sG) - H(F)}{s} \\ &= \sum_{i,j,\alpha,\beta} g^{ij}\big(\delta_{\alpha\beta} - \sum_{k,l,\gamma} g^{kl} h_{\beta\gamma} F^{\alpha}_{x_k} F^{\gamma}_{x_l}\big) G^{\beta}_{x_i x_j} \partial_{y_\alpha} + \text{lower order terms}. \end{aligned}$$

Denote by π_{TM} and π_{NM} the projections of F^*TN onto $dF(T_xM)$ and NM, respectively. Then, for any $\phi = \sum_\alpha \phi_\alpha \partial_{y_\alpha} \in \Gamma(F^*TN)$, we have

$$\pi_{NM}(\phi) = \phi - \pi_{TM}(\phi) = \sum_{\alpha,\beta} \big(\delta_{\alpha\beta} - \sum_{k,l,\gamma} g^{kl} h_{\beta\gamma} F^{\alpha}_{x_k} F^{\beta}_{x_l}\big)\phi_\beta \partial_{y_\alpha}.$$

Therefore, the principal symbol is given by

$$\begin{aligned} \sigma_\zeta(DH; x)\phi &= \sum\nolimits_{i,j} g^{ij}\zeta_i\zeta_j \sum\nolimits_{\alpha,\beta} \big(\delta_{\alpha\beta} - \sum\nolimits_{k,l,\gamma} g^{kl} h_{\beta\gamma} F^{\alpha}_{x_k} F^{\beta}_{x_l}\big)\phi_\beta \partial_{y_\alpha} \\ &= |\zeta|^2_g \pi_{NM}(\phi|_x). \end{aligned}$$

Observe that the principal symbol is zero for tangent directions. Thus, MCF is degenerate and we cannot obtain information from the standard theory about short-time existence. Short-time existence and uniqueness of MCF was originally proven using results of Hamilton [43, 44] based on the Nash-Moser iteration method. We present a proof adapting a variant of the DeTurck's trick which was first used in Ricci flow [31]; see also [7, 63, 92].

Theorem 19 (Invariance Under Tangential Variations) *Suppose that* $F\colon M \times [0,T) \to N$ *is a family of immersions satisfying the system of PDEs*

$$\begin{cases} dF_{(x,t)}(\partial_t) = H\big(F(x,t)\big) + dF_{(x,t)}\big(V(x,t)\big), \\ F(x,0) = F_0(x), \end{cases} \tag{29}$$

where $(x,t) \in M \times [0,T)$*, the manifold* M *is compact and* V *is a time-dependent family of smooth vector fields. Then, there exists a unique family of diffeomorphisms* $\psi\colon M \times [0,T) \to M$*, such that the map* $\widehat{F}\colon M \times [0,T) \to N$ *given by* $\widehat{F}(x,t) = F\big(\psi(x,t),t\big)$*, is a solution of*

$$\begin{cases} d\widehat{F}_{(x,t)}(\partial_t) = H\big(\widehat{F}(x,t)\big), \\ \widehat{F}(x,0) = F_0\big(\psi(x,0)\big). \end{cases}$$

Conversely, if $F\colon M \times [0,T) \to N$ *is a solution of the mean curvature flow and* $\psi\colon M \times [0,T) \to M$ *is a family of diffeomorphisms, then* $\widehat{F}\colon M \times [0,T) \to N$ *satisfies a system of the form* (29)*.*

Proof Consider for the moment an arbitrary family a time-dependent of diffeomorphisms $\psi\colon M \times [0,T) \to M$ and define $\widehat{F}\colon M \times [0,T) \to N$ given by $\widehat{F}(x,t) = F\big(\psi(x,t),t\big)$, for $(x,t) \in M \times [0,T)$. From the chain rule, we have

$$d\widehat{F}_{(x,t)}(\partial_t) = H\big(\widehat{F}(x,t)\big) + dF_{(\psi(x,t),t)}\big(V(\psi(x,t),t) + d\psi_{(x,t)}(\partial_t)\big),$$

for any $(x,t) \in M \times [0,T)$. Hence, it suffices to find a one-parameter family of diffeomorphisms $\psi\colon M \times [0,T) \to M$ solving the initial value problem

$$\begin{cases} d\psi_{(x,t)}(\partial_t) = -V\big(\psi(x,t),t\big), \\ \psi(x,0) = I, \end{cases}$$

for any $(x,t) \in M \times [0,T)$, where $I : M \to M$ is the identity map. By Picard-Lindelöf theorem there exists a unique smooth solution of the above initial value problem. Moreover, because the initial data is the identity, taking $T > 0$ small enough we can assume that for any $t \in [0,T]$ the map $\psi(\cdot\,,t) : M \to M$ is a diffeomorphism. The converse is straightforward. □

Theorem 20 (Short-time Existence) *Let M be a compact Riemannian manifold and $F_0 \colon M \to N$ an immersion into a Riemannian manifold N. Then, the mean curvature flow with initial data the immersion F_0 admits a smooth solution on a maximal time interval $[0, T_{\max})$, where $0 < T_{\max} \leq \infty$.*

Proof The idea is to modify MCF by adding some tangential component in order to make it parabolic. Suppose that $F : M \times [0, T_{\max}) \to N$ solves MCF. Fix a Riemannian metric $\widehat{g}$ on M, denote its Levi-Civita connection by $\widehat{\nabla}$ and consider the vector field V_{DT} on M given by

$$V_{DT} = \operatorname{tr}_g(\nabla - \widehat{\nabla}). \tag{30}$$

Note that in local coordinates, V_{DT} has the form

$$V_{DT} = \sum_{i,j,k} g^{ij}(\Gamma_{ij}^k - \widehat{\Gamma}_{ij}^k)\partial_{x_k},$$

where Γ_{ij}^k and $\widehat{\Gamma}_{ij}^k$ are the Christoffel symbols of the connections ∇ and $\widehat{\nabla}$, respectively. Consider now the initial value problem,

$$\begin{cases} dF(\partial_t) = H + dF(V_{DT}) \\ F(\cdot, 0) = F_0 \end{cases}, \tag{31}$$

The first equation of (31) in local coordinates takes the form

$$F_t = \sum_{i,j,\alpha} g^{ij}\big(F^\alpha_{x_i x_j} - \sum_k \widehat{\Gamma}_{ij}^k F^\alpha_{x_k} + \sum_{\gamma,\delta} \Gamma^\alpha_{\gamma\delta} F^\gamma_{x_i} F^\delta_{x_j}\big)\partial_{y_\alpha}.$$

Since $\widehat{\Gamma}_{ij}^k$ does not depend on time, the principal symbol of (31) is

$$\sigma_\zeta\big(D(H + V_{DT}), \cdot\big) = |\zeta|^2 I.$$

Hence (31) is parabolic and has a unique solution. According to Theorem 19, from a solution of (31) we obtain a solution of the mean curvature flow. □

Definition 25 Let $F \colon M \times [0, T) \to N$ be a solution of MCF. Fix a metric $\widehat{g}$ and consider the vector field V_{DT}. The modified flow (31) is called *DeTurck mean curvature flow*.

Lemma 3 *The vector field V_{DT} defined in (30) is minus the Laplacian of the identity map $I \colon (M, g) \to (M, \widehat{g})$.*

Proof The Hessian B of the map I is given by

$$B(v_1, v_2) = \widehat{\nabla}_{dI(v_1)} dI(v_2) - dI(\nabla_{v_1} v_2) = \widehat{\nabla}_{v_1} v_2 - \nabla_{v_1} v_2,$$

for any $v_1, v_2 \in \mathfrak{X}(M)$. Hence, $\Delta_{g,\widehat{g}} I = -V_{DT}$. This completes the proof. □

Theorem 21 (Uniqueness) *Let M be a compact Riemannian manifold and $F_0 \colon M \to N$ an immersion into a Riemannian manifold N. Then, the solution of MCF, with initial data the immersion $F_0 \colon M \to N$, is unique up to diffeomorphisms.*

Proof Suppose that $\widetilde{F} \colon M \times [0, T_{\max}) \to N$ is the maximal solution of MCF, with initial data the given immersion F_0, and denote the induced metrics by $\widetilde{g}$. As in the existence part, fix a metric $\widehat{g}$ and denote by $\widehat{\nabla}$ its associated Levi-Civita connection. Consider the initial value problem

$$\begin{cases} d\phi(\partial_t) = \Delta_{\widetilde{g},\widehat{g}}\phi \\ \phi(\cdot, 0) = I \end{cases}.$$

Observe that the above problem is a parabolic and thus its solution gives rise to a unique one parameter family of diffeomorphisms $\phi : M \times [0, \varepsilon) \to M$, for at least some short time $\varepsilon > 0$. Denote by $\psi \colon M \times [0, \varepsilon) \to M$ the one parameter family of diffeomorphisms with the property that, for each t, the map $\psi(\cdot\,, t)$ is the inverse of $\phi(\cdot\,, t)$, i.e.,

$$\psi(\phi(x, t), t) = x = \phi(\psi(x, t), t)$$

for any (x, t) in space-time. From the chain rule, we have

$$d\psi_{(\phi(x,t),t)}(\partial_t) = -d\psi_{(\phi(x,t),t)}\big((\Delta_{\widetilde{g},\widehat{g}}\phi)(x)\big). \tag{32}$$

Define the map $F \colon M \times [0, \varepsilon) \to N$ given by $F(x, t) = \widetilde{F}(\psi(x, t), t)$, for any $(x, t) \in M \times [0, T_{\max})$. The induced time-dependent metric on M is $g = \psi^*\widetilde{g}$. Moreover, the map F satisfies the evolution equation

$$F_t = H + d\widetilde{F}(W), \tag{33}$$

where for any point (x, t) in space-time, we have

$$W(\psi(x, t), t) = d\psi_{(x,t)}(\partial_t).$$

Taking into account (32) and the composition formula for the Laplacian (see for example [24, page 116, equation (2.56)]), we have

$$W(\psi(x, t), t) = d\psi_{(x,t)}(V_{DT}(x)), \tag{34}$$

for any $(x, t) \in M \times [0, \varepsilon)$. From (33) and (34), we see that F satisfies the DeTurck mean curvature flow

$$dF(\partial_t) = H + dF(V_{DT}),$$

with initial data the immersion $F_0 : M \to N$.

Suppose now that $\widetilde{F}_1, \widetilde{F}_2 \colon M \times [0, T_{\max}) \to N$ are two solutions of the mean curvature flow, with the same initial condition $F_0 \colon M \to N$. As before fix a metric $\widehat{g}$ on M and denote by $\widetilde{g}_1$ and $\widetilde{g}_2$ the induced time-dependent metrics on M by $\widetilde{F}_1$ and $\widetilde{F}_2$, respectively. Denote by

$$\phi^1 \colon M \times [0, \varepsilon) \to N \quad \text{and} \quad \phi^2 \colon M \times [0, \varepsilon) \to N$$

the one-parameter family of diffeomorphisms solving the initial value problem

$$\begin{cases} d\eta(\partial_t) = \Delta_{\widetilde{g}_i, \widehat{g}}\eta, \\ \eta(\cdot, 0) = I. \end{cases}$$

Then, as we verified above, the maps

$$F_i : M \times [0, \varepsilon) \to N, \quad i \in \{1, 2\},$$

satisfy

$$\widetilde{F}_i(x, t) = F_i\big(\phi^i(x, t), t\big),$$

for any $(x, t) \in M \times [0, \varepsilon)$, form solutions of the DeTurck mean curvature flow, with common initial data the immersion $F_0 \colon M \to N$. Since the DeTurck mean curvature flow is parabolic, it follows that its solution is unique. □

6.3 Parabolic Maximum Principles

In this subsection, we state the weak and strong version of the parabolic maximum principle for scalar functions obeying a diffusion-reaction equation on a manifold equipped with a smooth time-dependent family of Riemannian metrics. Then we also present Hamilton's version [44, 45] of the parabolic maximum principle for arbitrary sections of a vector bundle; for detailed proofs see also the excellent monograph [26].

6.3.1 Scalar Parabolic Maximum Principle

Suppose that M is a smooth manifold, possibly with boundary ∂M, and $\{g(t)\}_{t\in[0,T)}$ a smooth family of Riemannian metrics. We will consider the second order time-dependent operator $\mathcal{L}$ given by

$$\mathcal{L}u = \Delta_{g(t)}u + g(t)(X, \nabla^{g(t)}u) \tag{P}$$

where

$$u \in C^2\big(M \times (0,T)\big) \cup C^0\big(\overline{M} \times [0,T]\big).$$

Note that, for each fixed time, $\mathcal{L}$ is an elliptic operator.

Theorem 22 (Comparison Principle) *Suppose that M is a compact, without boundary, manifold equipped with a smooth family $\{g(t)\}_{t\in[0,T)}$ of Riemannian metrics and $u\colon M \times [0,T) \to \mathbb{R}$ a C^2-smooth function, which satisfies the differential inequality*

$$\partial_t u - \mathcal{L}u \leq \Psi(u,t),$$

where $\mathcal{L}$ is the (time-dependent) operator defined in (P) and $\Psi : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ a smooth map. Let φ be the solution of the associated ODE

$$\begin{cases} \varphi'(t) = \Psi(\varphi(t),t), \\ \varphi(0) = \max_{x\in M} u(x,0). \end{cases}$$

Then, the solution u of the differential inequality is bounded from above by the solution φ of the ODE, that is $u(x,t) \leq \varphi(t)$, for every $(x,t) \in M \times [0,T)$.

As in the elliptic case, there exists a criterion which forces a solution of a parabolic differential inequality to be constant.

Theorem 23 (Strong Maximum Principle) *Suppose that M is a smooth manifold, possibly with boundary, equipped with a smooth family $\{g(t)\}_{t\in[0,T)}$ of Riemannian metrics. Let $u \in C^2\big(M \times (0,T)\big) \cup C^0\big(\overline{M} \times [0,T]\big)$ be a solution of*

$$\partial_t u - \mathcal{L}u + c\,u \leq 0$$

where c is a non-negative constant.

(a) *If $c = 0$ and u attains a maximum at point $(x_0,t_0) \in M \times (0,T)$ then u is constant on $M \times [0,t_0]$.*
(b) *If $c < 0$ and the function u attains a non-negative maximum at a point $(x_0,t_0) \in M \times (0,T)$, then u is constant on $M \times [0,t_0]$.*

By reversing both inequalities we obtain the corresponding minimum version of the comparison and strong principle; for the proofs see [5, 35] or [72].

6.3.2 Vectorial Parabolic Maximum Principle

Let M be a smooth manifold, possibly with boundary ∂M, equipped with a smooth family of metrics $\{g(t)\}_{t\in[0,T)}$ and associated Levi-Civita connections $\nabla^{g(t)}$. Let E be a vector bundle over M equipped with a time-independent metric h and a

family $\{\nabla(t)\}_{t\in[0,T)}$ of connections that are compatible with h. The *time-dependent Laplacian* acting on smooth sections of E is defined by

$$\Delta(t)\phi = \sum_{i=1}^{m} \left(\nabla(t)_{v_i}\nabla(t)_{v_i}\phi - \nabla(t)_{\nabla^{g(t)}_{v_i} v_i}\phi\right)$$

where $\{v_1, \dots, v_m\}$ is an orthonormal basis of $g(t)$.

Following the same lines as in the elliptic case, we can derive Weinberger-Hamilton's versions of the parabolic maximum principle.

Theorem 24 (Weak Vectorial Maximum Principle) *Suppose that M is a compact manifold, possibly with boundary ∂K, equipped with a smooth family $\{g(t)\}_{t\in[0,T)}$ of Riemannian metrics. Let E be a vector bundle over M endowed with time independent bundle metric h and a family $\{\nabla(t)\}_{t\in[0,T)}$ of connections that are compatible with h. Let K be a closed fiber-convex subset of E that is invariant under parallel transport with respect to each connection $\nabla(t)$, $t \in [0, T)$, and let $\{\phi(t)\}_{t\in[0,T)}$ be a smooth family of sections such that*

$$\nabla_{\partial_t}\phi - \Delta(t)\phi = \nabla(t)_X\phi + \Psi(\phi)$$

where X is a smooth time dependent vector field and Ψ is a smooth fiberwise map that points into K. If $\phi_{(x,t)} \in K$ for any (x, t) in the parabolic boundary of $M \times [0, T)$, i.e., for any $(x, t) \in \big(M \times \{0\}\big) \cup \big(\partial M \times [0, T)\big)$, then $\phi_{(x,t)} \in K$ for any $(x, t) \in M \times [0, T)$.

Theorem 25 (Strong Vectorial Maximum Principle) *Suppose that M is a smooth, not necessarily compact, manifold equipped with a smooth family $\{g(t)\}_{t\in[0,T)}$ of Riemannian metrics. Moreover, let E be a vector bundle over M endowed with time independent metric h and a family $\{\nabla(t)\}_{t\in[0,T)}$ of connections that are compatible with h. Assume that K is a closed fiber-convex subset of the vector bundle E that is invariant under parallel transport with respect to each connection $\nabla(t)$, $t \in [0, T)$, and let $\{\phi(t)\}_{t\in[0,T)}$ be a smooth family of sections such that*

$$\nabla_{\partial_t}\phi - \Delta(t)\phi = \nabla(t)_X\phi + \Psi(\phi)$$

where X is a smooth time dependent vector field and Ψ is a smooth fiberwise map that points into K. If there exists a point $(x_0, t_0) \in M \times (0, T)$ such that $\phi_{(x_0,t_0)} \in \partial K$, then $\phi_{(x,t)} \in \partial K$ for any $(x, t) \in M \times [0, t_0]$.

Let us describe now the parabolic maximum principle in the special case where as vector bundle we consider the space of symmetric 2-tensors.

Theorem 26 *Let M be a compact manifold equipped with a smooth family $\{g(t)\}_{t\in[0,T)}$ of Riemannian metrics. Suppose that $\{\phi(t)\}_{t\in[0,T)}$ is smooth family of symmetric 2-tensors on M such that*

$$\nabla_{\partial_t}\phi - \Delta(t)\phi = \nabla(t)_X\phi + \Psi(\phi)$$

where $\Psi : \mathrm{Sym}(T^*M \otimes T^*M) \to \mathrm{Sym}(T^*M \otimes T^*M)$ *is a smooth fiberwise map satisfying the null-eigenvector condition and* X *is a smooth time dependent vector field. If* $\phi(0) \geq 0$*, then* $\phi(t) \geq 0$ *for all* $t \in [0, T)$*. Additionally, if there is a point* $(x_0, t_0) \in M \times (0, T)$ *where* $\phi(t_0)$ *has a zero eigenvalue then* $\phi(t)$ *has a zero eigenvalue for any* $t \in (0, t_0)$*.*

6.4 Evolution Equations

We will compute the evolution of some important quantities. In order to simplify the notation, we omit upper or lower indices on connections and Laplacians which identify the corresponding bundles where they are defined. Most of these computations can be found in [4, 75–78, 86, 94, 96].

Lemma 4 *Suppose that* $F : M \times [0, T) \to N$ *is a solution of the mean curvature flow. Then, the following facts are true:*

(a) *The induced metrics* g *evolve in time under the equation*

$$\big(\nabla_{\partial_t} g\big)(v_1, v_2) = -2\langle H, A(v_1, v_2)\rangle = -2A^H(v_1, v_2),$$

for any $v_1, v_2 \in \mathfrak{X}(M)$*.*

(b) *The induced volume form* Ω *on* (M, g) *evolves according to the equation*

$$\nabla_{\partial_t}\Omega = -|H|^2\Omega.$$

Moreover, the volume of the evolved submanifolds satisfy

$$\partial_t Vol = -\int_M |H|^2\Omega.$$

(c) *There exists a local smooth time-dependent tangent orthonormal frame field and a local smooth time-dependent orthonormal frame field along the normal bundle of the evolving submanifolds.*

Proof

(a) Let $v_1, \ldots, v_m$ be time-independent tangent vector fields. Keeping in mind the notation introduced in Section 6.1.2, we have

$$\nabla_{\partial_t} dF(v_i) = \nabla_{v_i} dF(\partial_t) + dF\big([\partial_t, v_i]\big) = \nabla_{v_i} H,$$

for any $i \in \{1, \ldots, m\}$. Therefore, for any $i, j \in \{1, \ldots, m\}$, we deduce that

$$\begin{aligned}(\nabla_{\partial_t} g)(v_i, v_j) &= \partial_t(g(v_i, v_j)) - g(\nabla_{\partial_t} v_i, v_j) - g(v_i, \nabla_{\partial_t} v_j)\\ &= \partial_t \langle dF(v_i), dF(v_j)\rangle = \langle \nabla_{v_i} H, dF(v_j)\rangle + \langle \nabla_{v_j} H, dF(v_i)\rangle\\ &= -\langle H, \nabla_{v_i} dF(v_j)\rangle - \langle H, \nabla_{v_j} dF(v_i)\rangle\\ &= -2\langle H, A(v_i, v_j)\rangle.\end{aligned}$$

(b) We compute

$$\begin{aligned}\partial_t \sqrt{\det g_{ij}} &= \sum_{k,l} \frac{\left(g^{kl}\partial_t g_{kl}\right)\det g_{ij}}{2\sqrt{\det g_{ij}}} = -\sum_{k,l}\langle H, g^{kl}A_{kl}\rangle\sqrt{\det g_{ij}}\\ &= -|H|^2\sqrt{\det g_{ij}}.\end{aligned}$$

(c) The associated adjoint operator $P\colon (TM, g) \to (TM, g)$ of A^H satisfies

$$A^H(v_1, v_2) = g(Pv_1, v_2) = g(v_1, Pv_2), \tag{35}$$

for any $v_1, v_2 \in \mathfrak{X}(M)$. Consider now the family of bundle isomorphism $U(t)$: $\big(TM, g(0)\big) \to \big(TM, g(t)\big)$, given as the solution of the initial value problem

$$\begin{cases} \nabla_{\partial_t} U(t) = P \circ U(t),\\ U(0) = I. \end{cases} \tag{36}$$

By a straightforward computation, we can show that $U^*(t)g(t) = g(0)$. Hence, if $\{e_1(0), \ldots, e_m(0)\}$ is a local orthonormal frame with respect to $g(0)$, then $\{e_1(t) = U(t)e_1(0), \ldots, e_m(t) = U(t)e_m(0)\}$ is a local orthonormal frame of $g(t)$. By taking the complement of $\{e_1, \ldots, e_m\}$, we get a time-dependent frame field on the normal bundles of the evolving submanifolds.

□

Lemma 5 *The time-derivative of the second fundamental form is given by*

$$\big(\nabla^{\perp}_{\partial_t} A\big)^{\alpha}_{ij} = \big(\nabla^{\perp 2} H\big)^{\alpha}_{ij} - \sum\nolimits_{k,\beta} H^{\beta} A^{\beta}_{jk} A^{\alpha}_{ik} - \sum\nolimits_{\beta} H^{\beta} \widetilde{R}_{\beta ij\alpha},$$

where the indices are with respect to a local orthonormal frame.

Proof Suppose that $\{e_1, \ldots, e_m; \xi_{m+1}, \ldots, \xi_n\}$ is a local adapted orthonormal frame field around a fixed point (x_0, t_0). Recall that

$$\nabla_{\partial_t}\partial_t = 0, \quad \nabla_{e_i}\partial_t = 0 \quad \text{and} \quad [\partial_t, e_i] = \nabla_{\partial_t} e_i = \sum\nolimits_{j,\beta} H^{\beta} A^{\beta}_{ij} e_j. \tag{37}$$

In order to simplify the computations, we may assume that $\{e_1, \ldots, e_m\}$ is normal frame at (x_0, t_0). Under these considerations, we have that at (x_0, t_0)

$$\begin{aligned}(\nabla_{\partial_t} A)_{ij} &= \nabla_{\partial_t}\nabla_{e_i} dF(e_j) - \nabla_{\partial_t} dF(\nabla_{e_i} e_j) - A(\nabla_{\partial_t} e_i, e_j) - A(e_i, \nabla_{\partial_t} e_j)\\ &= \nabla_{e_i}\nabla_{\partial_t} dF(e_j) + \widetilde{R}\big(H, dF(e_i), dF(e_j)\big) + \nabla_{\nabla_{\partial_t} e_i} dF(e_j)\\ &\quad -dF\big(\nabla_{\partial_t}\nabla_{e_i} e_j\big) - A(\nabla_{\partial_t} e_i, e_j) - A(e_i, \nabla_{\partial_t} e_j).\end{aligned}$$

Hence,

$$\begin{aligned}(\nabla_{\partial_t} A)_{ij} &= \nabla_{e_i}\big(\nabla_{e_j} H + dF(\nabla_{\partial_t} e_j)\big) + \widetilde{R}\big(H, dF(e_i), dF(e_j)\big)\\ &\quad +\nabla_{\nabla_{\partial_t} e_i} dF(e_j) - dF\big(\nabla_{\partial_t}\nabla_{e_i} e_j\big) - A(\nabla_{\partial_t} e_i, e_j) - A(e_i, \nabla_{\partial_t} e_j)\\ &= \nabla^2_{e_i,e_j} H + \widetilde{R}\big(H, dF(e_i), dF(e_j)\big) + \nabla_{e_i} dF(\nabla_{\partial_t} e_j)\\ &\quad +\nabla_{\nabla_{\partial_t} e_i} dF(e_j) - dF\big(\nabla_{\partial_t}\nabla_{e_i} e_j\big) - A(\nabla_{\partial_t} e_i, e_j) - A(e_i, \nabla_{\partial_t} e_j)\\ &= \nabla^2_{e_i,e_j} H + \widetilde{R}\big(H, dF(e_i), dF(e_j)\big) + \nabla_{e_i} dF(\nabla_{\partial_t} e_j)\\ &\quad +\nabla_{\nabla_{\partial_t} e_i} dF(e_j) - dF\big(\nabla_{\partial_t}\nabla_{e_i} e_j\big) - A(\nabla_{\partial_t} e_i, e_j) - A(e_i, \nabla_{\partial_t} e_j)\end{aligned}$$

and so

$$(\nabla_{\partial_t} A)_{ij} = \nabla^2_{e_i,e_j} H + \widetilde{R}\big(H, dF(e_i), dF(e_j)\big) - dF\big(R^{\nabla}(\partial_t, e_i, e_j)\big)$$

where R^{∇} is the curvature operator of ∇ on $T(M\times(0,T))$. Consequently, at (x_0, t_0) we have

$$\begin{aligned}(\nabla^{\perp}_{\partial_t} A)_{ij} &= \sum\nolimits_{\alpha} \langle (\nabla^{\perp}_{\partial_t} A)_{ij}, \xi_\alpha\rangle \xi_\alpha = \sum\nolimits_{\alpha} \langle (\nabla_{\partial_t} A)_{ij}, \xi_\alpha\rangle \xi_\alpha\\ &= \sum\nolimits_{\alpha} \langle \nabla_{e_i}\nabla_{e_j} H, \xi_\alpha\rangle \xi_\alpha + \sum\nolimits_{\alpha,\beta} H^\beta \widetilde{R}_{\beta ij\alpha}\xi_\alpha.\end{aligned}$$

On the other hand,

$$\begin{aligned}\langle \nabla_{e_i}\nabla_{e_j} H, \xi_\alpha\rangle &= \langle \nabla^{\perp}_{e_i}\big(\nabla^{\perp}_{e_j} H + \sum\nolimits_k \langle \nabla_{e_j} H, dF(e_k)\rangle dF(e_k)\big), \xi_\alpha\rangle\\ &= (\nabla^{2\perp} H)^{\alpha}_{ij} - \sum\nolimits_{k,\beta} H^\beta A^{\beta}_{jk} A^{\alpha}_{ik}.\end{aligned}$$

Combining the last two equalities we obtain the result. □

Lemma 6 *The mean curvature H evolves in time under the equation*

$$(\nabla^{\perp}_{\partial_t} H)^{\alpha} = (\Delta^{\perp} H)^{\alpha} - \sum_{i,\beta} H^\beta \widetilde{R}_{\beta ii\alpha} + \sum_{i,j,\beta} H^\beta A^{\beta}_{ij} A^{\alpha}_{ij}.$$

Moreover,

$$\partial_t |H|^2 = \Delta |H|^2 - 2|\nabla^\perp H|^2 + 2|A^H|^2 - 2\sum_{i,\alpha,\beta} H^\alpha H^\beta \widetilde{R}_{\alpha i i \beta},$$

where the indices are with respect to a local orthonormal frame.

Proof Let $(x_0, t_0) \in M \times (0, T)$ and $\{e_1, \dots, e_m; \xi_{m+1}, \dots, \xi_n\}$ be a local orthonormal frame field around of (x_0, t_0). From (37) and Lemma 5, we have

$$\begin{aligned}\left(\nabla^\perp_{\partial_t} H\right)^\alpha &= \sum_i \left(\nabla^\perp_{\partial_t} A_{ii}\right)^\alpha = \sum_i (\nabla^\perp_{\partial_t} A)^\alpha_{ii} + 2\sum_i A^\alpha(\nabla_{\partial_t} e_i, e_i) \\ &= (\Delta^\perp H)^\alpha + \sum_{i,\beta} H^\beta \widetilde{R}_{\beta i i \alpha} - \sum_{i,j,\beta} H^\beta A^\beta_{ij} A^\alpha_{ij} + 2\sum_{i,j,\beta} H^\beta A^\beta_{ij} A^\alpha_{ij},\end{aligned}$$

from where we deduce the evolution equation for H. Moreover

$$\begin{aligned}\partial_t |H|^2 &= \partial_t \langle H, H\rangle = 2\langle \nabla^\perp_{\partial_t} H, H\rangle = \sum_\alpha (\nabla^\perp_{\partial_t} H)^\alpha H^\alpha \\ &= 2\sum_\alpha (\Delta H)^\alpha H^\alpha - 2\sum_{i,\alpha,\beta} H^\alpha H^\beta \widetilde{R}_{\alpha i i \beta} + 2\sum_{i,j,\alpha,\beta} H^\alpha H^\beta A^\alpha_{ij} A^\beta_{ij}.\end{aligned}$$

On the other hand

$$\sum_\alpha \Delta (H^\alpha)^2 = 2\sum_\alpha (\Delta H)^\alpha H^\alpha + 2\sum_\alpha |\nabla H^\alpha|^2.$$

Combining the last two identities we obtain the desired identity. □

6.5 Evolution Equations of Parallel Forms

Let $F : M \times [0, T) \to N$ be a solution of the mean curvature flow and suppose that Φ is a parallel k-tensor on N. Then, the pullback via F of Φ gives rise to a time-dependent k-form on M. For example, the volume form of N is such a tensor. As we will see in the next section, interesting situations occurs when N is a Riemannian product $N_1 \times N_2$ and we consider the volume forms Ω_1 and Ω_2 of N_1 and N_2, respectively.

In the next lemmata, we will compute how these pullback tensors evolve under the mean curvature flow.

Lemma 7 *The covariant derivative of the tensor $F^*\Phi$ is given by*

$$(\nabla_{e_s} F^*\Phi)_{i_1 \dots i_k} = \sum_\alpha \left(A^\alpha_{s i_1} \Phi_{\alpha i_2 \dots i_k} + \dots + A^\alpha_{s i_m} \Phi_{i_1 \dots i_{m-1} \alpha}\right),$$

for any adapted orthonormal frame field $\{e_1, \dots, e_m; \xi_{m+1}, \dots, \xi_{n-m}\}$.

Proof As usually let us suppose that $\{e_1, \ldots, e_m\}$ is a normal frame at a fixed point (x_0, t_0) in space-time. By a direct computation, we get that at (x_0, t_0) we have

$$\begin{aligned}(\nabla_{e_s} F^*\Phi)_{i_1\ldots i_k} &= e_s \Phi\big(dF(e_{i_1}), \ldots, dF(e_{i_m})\big)\\ &= \Phi\big(\nabla_{e_s} dF(e_{i_1}), \ldots, dF(e_{i_m})\big) + \cdots + \Phi\big(dF(e_{i_1}), \ldots, \nabla_{e_s} dF(e_{i_m})\big)\\ &= \Phi\big(A(e_s, e_{i_1}), \ldots, dF(e_{i_m})\big) + \cdots + \Phi\big(dF(e_{i_1}), \ldots, A(e_s, e_{i_m})\big)\\ &= \sum_\alpha \big(A^\alpha_{si_1}\Phi_{\alpha i_2\ldots i_k} + \cdots + A^\alpha_{si_m}\Phi_{i_1\ldots i_{m-1}\alpha}\big).\end{aligned}$$

This completes the proof. □

By a direct computation we can derive the expression for the Laplacian of the pullback of a parallel k-tensor on N.

Lemma 8 *The Laplacian of the k-tensor $F^*\Phi$ is given by*

$$\begin{aligned}(\Delta F^*\Phi)_{i_1\ldots i_m} &= \sum_\alpha (\nabla^\perp_{e_{i_1}} H)^\alpha \Phi_{\alpha i_2\ldots i_m} + \cdots + \sum_\alpha (\nabla^\perp_{e_{i_m}} H)^\alpha \Phi_{i_1\ldots i_{m-1}\alpha}\\ &\quad + 2\sum_{k,\alpha,\beta} A^\alpha_{ki_1} A^\beta_{ki_2}\Phi_{\alpha\beta i_2\ldots i_m} + \cdots + 2\sum_{k,\alpha,\beta} A^\alpha_{ki_{m-1}} A^\beta_{ki_m}\Phi_{i_1\ldots\alpha\beta}\\ &\quad - \sum_{k,l,\alpha}\big(A^\alpha_{ki_1}A^\alpha_{kl}\Phi_{li_2\ldots i_m} + \cdots + A^\alpha_{ki_m}A^\alpha_{kl}\Phi_{i_1\ldots i_{m-1}l}\big)\\ &\quad - \sum_{k,\alpha}\big(\widetilde{R}_{k\alpha ki_1}\Phi_{\alpha i_2\ldots i_m} + \cdots + \widetilde{R}_{k\alpha ki_m}\Phi_{i_1\ldots i_{m-1}\alpha}\big),\end{aligned}$$

for any adapted orthonormal frame field $\{e_1, \ldots, e_m; \xi_{m+1}, \ldots, \xi_{n-m}\}$.

Proof Let $\{e_1, \ldots, e_m; \xi_{m+1}, \ldots, \xi_{n-m}\}$ be an adapted normal frame at the point (x_0, t_0) in space-time. We compute,

$$\begin{aligned}\big(\nabla_{e_k}\nabla_{e_k} F^*\Phi\big)_{i_1\ldots i_m} &= e_k\big(\Phi(A_{ki_1}, \ldots, dF(e_{i_m})) + \cdots + \Phi(dF(e_{i_1}), \ldots, A_{ki_m})\big)\\ &= \Phi((\nabla_{e_k}A)_{ki_1}, \ldots, dF(e_{i_m})) + \cdots + \Phi(dF(e_{i_1}), \ldots, (\nabla_{e_k}A)_{ki_m})\\ &\quad + 2\Phi\big(A_{ki_1}, A_{ki_2}, \ldots, dF(e_{i_m})\big) + \cdots + 2\Phi\big(dF(e_{i_1}), \ldots, A_{ki_{m-1}}, A_{ki_m}\big)\\ &= \Phi((\nabla^\perp_{e_k}A)_{ki_1}, \ldots, dF(e_{i_m})) + \cdots + \Phi(dF(e_{i_1}), \ldots, (\nabla^\perp_{e_k}A)_{ki_m})\\ &\quad + 2\Phi\big(A_{ki_1}, A_{ki_2}, \ldots, dF(e_{i_m})\big) + \cdots + 2\Phi\big(dF(e_{i_1}), \ldots, A_{ki_{m-1}}, A_{ki_m}\big)\\ &\quad - \sum\nolimits_l \langle A_{ki_1}, A_{kl}\rangle F^*\Phi(e_l, \ldots e_{i_m}) - \cdots - \sum\nolimits_l \langle A_{ki_m}, A_{kl}\rangle F^*\Phi(e_{i_1}\ldots, e_l).\end{aligned}$$

Summing over k and using the Codazzi equation (4), we get the result. □

Lemma 9 *Suppose that $F\colon M\times[0,T)\to N$ is a solution of the mean curvature flow and let Φ be a parallel m-form on N. Then, $u=*(F^*\Phi)$, where $*$ is the Hodge star operator with respect to the induced Riemannian metric g, evolves in time under the equation*

$$\begin{aligned}\partial_t u-\Delta u=&-2\sum_{k,\alpha,\beta}A^\alpha_{k1}A^\beta_{k2}\Phi_{\alpha\beta 2\dots m}-\dots-2\sum_{k,\alpha,\beta}A^\alpha_{km-1}A^\beta_{km}\Phi_{1\dots\alpha\beta}\\&+\sum_{k,l,\alpha}\big(A^\alpha_{k1}A^\alpha_{kl}\Phi_{l2\dots m}+\dots+A^\alpha_{km}A^\alpha_{kl}\Phi_{1\dots m-1l}\big)\\&+\sum_{k,\alpha}\big(\widetilde{R}_{k\alpha k1}\Phi_{\alpha 2\dots m}+\dots+\widetilde{R}_{k\alpha km}\Phi_{1\dots m-1\alpha}\big),\end{aligned}$$

for any adapted orthonormal frame field $\{e_1,\dots,e_m;\xi_{m+1},\dots,\xi_{n-m}\}$.

Proof Let us make our computations again, with respect to a time-dependent orthonormal frame field as in Lemma 4. We compute,

$$\begin{aligned}\partial_t u&=\partial_t\big((F^*\Phi)(e_1,\dots,e_m)\big)\\&=\Phi\big(\nabla_{\partial_t}dF(e_1),\dots,dF(e_m)\big)+\dots+\Phi\big(dF(e_1),\dots,\nabla_{\partial_t}dF(e_m)\big).\end{aligned}$$

Taking into account the formulas (37), we have

$$\begin{aligned}\nabla_{\partial_t}dF(e_i)&=\nabla_{e_i}dF(\partial_t)+dF(\nabla_{\partial_t}e_i)=\nabla_{e_i}H+\sum_{k,\beta}H^\beta A^\beta_{ik}dF(e_k)\\&=\nabla^\perp_{e_i}H,\end{aligned}$$

for any $i\in\{1,\dots,m\}$. Hence, putting everything together, we deduce that

$$\partial_t u=\Phi\big(\nabla^\perp_{e_1}H,\dots,dF(e_m)\big)+\dots+\Phi\big(dF(e_1),\dots,\nabla^\perp_{e_m}H\big).$$

Combining with Lemma 8 we obtain the result. □

7 Formation of Singularities Under Mean Curvature Flow

In this section, we present how one can build smooth singularity models for the mean curvature flow by rescaling properly around points, where the second fundamental form attains its maximum. The proof relies heavily on a compactness theorem of Cheeger-Gromov-Taylor [14] for pointed Riemannian manifolds and on the standard compactness theorem for immersions; see for example [27].

7.1 Characterization of the Maximal Time of Existence

In the following theorem, we give a characterization of the maximal time of solutions of the mean curvature flow. Its proof has been done by Huisken in [52, 53] and is based on the parabolic maximum principle. The key observation is that all higher derivatives $\nabla^k A$, $k \in \mathbb{N}$, of the second fundamental tensor are uniformly bounded, once A is uniformly bounded. More precisely, the following result holds:

Theorem 27 *Let M be a compact manifold and let $F_0 : M \to N$ a smooth immersion into a complete Riemannian manifold N. Then, the maximal time $T_{\max}$ of the solution of the mean curvature flow, with initial data F_0, is finite if and only if*

$$\limsup_{t \to T_{\max}} \left(\max_{M \times [0,t]} |A| \right) = \infty.$$

An immediate consequence of the above result is the following theorem.

Theorem 28 *Let M be a compact manifold and $F : M \to [0, T_{\max}) \to N$ a solution of the mean curvature flow on a maximal time interval in a complete Riemannian manifold N. If the norm $|A|$ of the second fundamental form is uniformly bounded, then the maximal time of solution of the flow is infinite.*

Remark 7 When the target space N is compact and the maximal time of solution of the flow is infinite, due to a deep result of Simon [82], it follows that the flow converges smoothly and uniformly to a minimal submanifold. However, long-time existence does not automatically imply convergence. For instance, start with a latitude circle $\mathbb{S}^1$ on a complete surface of revolution that does not admit closed embedded curves as geodesics. Then the flow with initial that particular circle will run forever, but it will not converge.

Remark 8 Due to a recent result of Cooper [27], it is not necessary to have boundedness on the full norm of the second fundamental form in order to get long-time existence of the flow. In the matter of fact, he showed that uniform boundedness of the second fundamental form only in the direction of the mean curvature also leads to long-time existence.

7.2 Cheeger-Gromov Compactness for Metrics

Let us recall here the basic notions and definitions. For more details, see [5, 25] and [66]. We closely follow the exposition in [78].

Definition 26 Let (E, π, Σ) be a vector bundle endowed with a Riemannian metric g and a metric connection ∇ and suppose that $\{\xi_k\}_{k \in \mathbb{N}}$ is a sequence of sections of E. Let U be an open subset of Σ with compact closure $\bar{U}$ in Σ. Fix a natural number $p \geq 0$. We say that $\{\xi_k\}_{k \in \mathbb{N}}$ converges *C^p-smoothly* to $\xi_\infty \in \Gamma(E|_{\overline{U}})$, if for every $\varepsilon > 0$, there exists $k_0 = k_0(\varepsilon)$, such that

$$\sup_{0\leq\alpha\leq p}\sup_{x\in\bar{U}}|\nabla^{\alpha}(\xi_k-\xi_\infty)|<\varepsilon$$

where $k\geq k_0$. We say that $\{\xi_k\}_{k\in\mathbb{N}}$ C^∞-*smoothly converges* to $\xi_\infty\in\Gamma(E|_{\bar{U}})$ if $\{\xi_k\}_{k\in\mathbb{N}}$ converges in C^p to $\xi_\infty\in\Gamma(E|_{\overline{U}})$, for any $p\geq 0$.

Definition 27 Let (E,π,Σ) be a vector bundle endowed with a Riemannian metric g and a metric connection ∇. Let $\{U_n\}_{n\in\mathbb{N}}$ be an exhaustion of Σ and $\{\xi_k\}_{k\in\mathbb{N}}$ be a sequence of sections of E defined on open sets A_k of Σ. We say that $\{\xi_k\}_{k\in\mathbb{N}}$ converges smoothly on compact sets to $\xi_\infty\in\Gamma(E)$ if:

(a) For every $n\in\mathbb{N}$ there exists k_0 such that $\overline{U}_n\subset A_k$, for all natural numbers $k\geq k_0$.
(b) The sequence $\{\xi|_{\overline{U}_k}\}_{k\geq k_0}$ converges in C^∞ to the restriction of the section ξ_∞ on $\overline{U}_n$.

In the next definitions, we recall the notion of the smooth Cheeger-Gromov convergence of sequences of Riemannian manifolds.

Definition 28 A pointed Riemannian manifold (Σ,g,x) is a Riemannian manifold (Σ,g) with a choice of origin or base point $x\in\Sigma$. If the metric g is complete, we say that (Σ,g,x) is a complete pointed Riemannian manifold.

Definition 29 We will say that a sequence $\{(\Sigma_k,g_k,x_k)\}_{k\in\mathbb{N}}$ of complete, pointed Riemannian manifolds smoothly converges in the sense of Cheeger-Gromov to a complete pointed Riemannian manifold $(\Sigma_\infty,g_\infty,x_\infty)$, if there exists:

(a) An exhaustion $\{U_k\}_{k\in\mathbb{N}}$ of Σ_∞ with $x_\infty\in U_k$, for all $k\in\mathbb{N}$.
(b) A sequence of diffeomorphisms $\Phi_k\colon U_k\to\Phi_k(U_k)\subset\Sigma_k$, with

$$\Phi_k(x_\infty)=x_k$$

and such that the sequence $\{\Phi_k^*g_k\}_{k\in\mathbb{N}}$ smoothly converges in C^∞ to g_∞ on compact sets in Σ_∞.

The sequence $\{(U_k,\Phi_k)\}_{k\in\mathbb{N}}$ is called a *family of convergence pairs* of the sequence $\{(\Sigma_k,g_k,x_k)\}_{k\in\mathbb{N}}$, with respect to the limit $(\Sigma_\infty,g_\infty,x_\infty)$.

When we say *smooth convergence*, we always mean smooth convergence in the sense of Cheeger-Gromov. The family of convergence pairs is not unique. Two such families $\{(U_k,\Phi_k)\}_{k\in\mathbb{N}}$,$\{(W_k,\Psi_k)\}_{k\in\mathbb{N}}$ are equivalent in the sense that there exists an isometry I of the limit $(\Sigma_\infty,g_\infty,x_\infty)$, such that for every compact subset K of Σ_∞, there exists a natural number k_0, such that for any natural $k\geq k_0$:

(a) The mapping $\Phi_k^{-1}\circ\Psi_k$ is well defined over K.
(b) The sequence $\{\Phi_k^{-1}\circ\Psi_k\}_{k\geq k_0}$ smoothly converges to I on K.

The limiting pointed Riemannian manifold $(\Sigma_\infty,g_\infty,x_\infty)$ of the Definition 29 is unique up to isometries.

Definition 30 Let M be a Riemannian manifold. The injectivity radius at $x \in M$ is the supremum of all values r, such that the expotential map from the unit ball $B_r(x)$ in T_xM, to the manifold M, is injective.

Definition 31 A complete Riemannian manifold (Σ, g) is said to have *bounded geometry*, if the following conditions are satisfied:

(a) For any integer $j \geq 0$, there exists a uniform positive constant C_j, such that $|\nabla^j R| \leq C_j$.
(b) The injectivity radius satisfies $inj_g(\Sigma) > 0$.

The following proposition is standard and will be useful in the proof of the long-time existence of the mean curvature flow.

Proposition 3 *Suppose (Σ, g) is a complete Riemannian manifold with bounded geometry. Suppose that $\{\alpha_k\}_{k\in\mathbb{N}}$ is an increasing sequence of real numbers that tends to $+\infty$ and let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence of points on Σ. Then, the sequence $\{(\Sigma, \alpha_k^2 g, x_k)\}_{k\in\mathbb{N}}$ smoothly subconverges to the euclidean space $(\mathbb{R}^m, g_{euc}, 0)$.*

We will use the following definition of uniformly bounded geometry for a sequence of pointed Riemannian manifolds.

Definition 32 We say that a sequence $\{(\Sigma_k, g_k, x_k)\}_{k\in\mathbb{N}}$ of complete pointed Riemannian manifolds has *uniformly bounded geometry*, if the following two conditions are satisfied:

(a) For any integer $j \geq 0$, there exists a uniform constant C_j, such that for each $k \in \mathbb{N}$ it holds $|\nabla^j R_k| \leq C_j$, where R_k is the curvature operator of the metric g_k.
(b) There exists a uniform constant c_0, such that $inj_{g_k}(\Sigma_k) \geq c_0 > 0$.

In the next result, we state the Cheeger-Gromov compactness theorem for sequences of complete pointed Riemannian manifolds. The version that we present here is due to Hamilton [46].

Theorem 29 *Let $\{(\Sigma_k, g_k, x_k)\}_{k\in\mathbb{N}}$ be a sequence of complete pointed Riemannian manifolds with uniformly bounded geometry. Then, the sequence $\{(\Sigma_k, g_k, x_k)\}_{k\in\mathbb{N}}$ subconverges smoothly to a complete pointed Riemannian manifold $(\Sigma_\infty, g_\infty, x_\infty)$.*

Remark 9 We would like to mention here that due to an estimate from Cheeger, Gromov and Taylor [14], the above compactness theorem still holds under the weaker assumption that the injectivity radius is uniformly bounded from below by a positive constant, only along the base points $\{x_k\}_{k\in\mathbb{N}}$, thereby avoiding the assumption of the uniform lower bound for $inj_{g_k}(\Sigma_k)$.

7.3 Convergence of Immersions

Definition 33 Let $F_k\colon (\Sigma_k, g_k, x_k) \to (P_k, h_k, y_k)$ be a sequence of isometric immersions, such that $F(x_k) = y_k$, for any $k \in \mathbb{N}$. We say that the sequence $\{F_k\}_{k\in\mathbb{N}}$ *converges smoothly* to an isometric immersion

$$F_\infty\colon (\Sigma_\infty, g_\infty, x_\infty) \to (P_\infty, h_\infty, y_\infty)$$

if the following conditions are satisfied:

(a) The sequence $\{(\Sigma_k, g_k, x_k)\}_{k\in\mathbb{N}}$ smoothly converges to $(\Sigma_\infty, g_\infty, x_\infty)$.
(b) The sequence $\{(P_k, h_k, y_k)\}_{k\in\mathbb{N}}$ smoothly converges to $(P_\infty, h_\infty, y_\infty)$.
(c) If $\{(U_k, \Phi_k)\}_{k\in\mathbb{N}}$ is a family of convergence pairs of $\{(\Sigma_k, g_k, x_k)\}_{k\in\mathbb{N}}$ and $\{(W_k, \Psi_k)\}_{k\in\mathbb{N}}$ is a family of convergence pairs of $\{(P_k, h_k, y_k)\}_{k\in\mathbb{N}}$, then for each $k \in \mathbb{N}$, we have $F_k \circ \Phi_k(U_k) \subset \Psi_k(W_k)$ and $\Psi_k^{-1} \circ F \circ \Phi_k$ smoothly converges to F_∞ on compact sets.

Lemma 10 *Suppose that (P, h) is a complete Riemannian manifold with bounded geometry. Then, for any $C > 0$, there exists a positive constant $r > 0$, such that $inj_g(\Sigma) > r$, for any isometric immersion $F\colon (\Sigma, g) \to (P, h)$ such that the norm $|A_F|$ of its second fundamental form satisfies $|A_F| \leq C$.*

The last lemma and the Cheeger-Gromov compactness theorem allow us to deduce a compactness theorem in the category of sequences of immersions; see for example [27].

Theorem 30 *Let $\{(\Sigma_k, g_k, x_k)\}_{k\in\mathbb{N}}$ and $\{(P_k, h_k, y_k)\}_{k\in\mathbb{N}}$ be two sequences of complete Riemannian manifolds with dimensions m and l, respectively. Suppose that $F_k\colon (\Sigma_k, g_k, x_k) \to (P_k, h_k, y_k)$ is a family of isometric immersions, where $F_k(x_k) = y_k$. Assume that:*

(a) *Each Σ_k is compact.*
(b) *The sequence $\{(P_k, h_k, y_k)\}_{k\in\mathbb{N}}$ has uniformly bounded geometry.*
(c) *For any integer $j \geq 0$, there exists a uniform constant C_j, such that*

$$|(\nabla^{F_k})^j A_{F_k}| \leq C_j,$$

for any $k \in \mathbb{N}$. Here, A_{F_k} stands for the second fundamental form of F_k.

Then, the sequence $\{F_k\}_{k\in\mathbb{N}}$ subconverges smoothly to a complete isometric immersion $F_\infty\colon (\Sigma_\infty, g_\infty, x_\infty) \to (P_\infty, h_\infty, y_\infty)$.

7.4 Modeling the Singularities

In the following theorem, we describe a method of rescaling around points, where the second fundamental form attains its maximum.

Theorem 31 *Let Σ be a compact manifold and let $F\colon \Sigma \times [0, T_{\max}) \to (P, h)$ be a solution of mean curvature flow, where P is a Riemannian manifold with bounded geometry and $T_{\max} \leq \infty$ is the maximal time of existence of a smooth solution. Suppose that there exists a point $x_\infty \in \Sigma$ and a sequence of points $\{(x_k, t_k)\}_{k\in\mathbb{N}}$ in $\Sigma \times [0, T)$ with $\lim x_k = x_\infty$, $\lim t_k = T_{\max}$ such that*

$$a_k = \max_{M\times[0,t_k]} |A(x,t)| = |A(x_k,t_k)| \to \infty.$$

Then:

(a) *The family of maps $F_k\colon \Sigma \times [-a_k^2 t_k, 0] \to (P, a_k^2 h), k \in \mathbb{N}$, given by*

$$F_k(x,s) = F_{k,s}(x) = F(x, s/a_k^2 + t_k),$$

form a sequence of mean curvature flow solutions. The mean curvature H_{F_k} and the norm $|A_{F_k}|$ of the second fundamental form of F_k satisfy the equation

$$H_{F_k} = \frac{1}{a_k^2} H(x, s/a_k^2 + t_k) \quad \text{and} \quad |A_{F_k}(x,s)| = \frac{1}{a_k} |A(x, s/a_k^2 + t_k)|.$$

Moreover, for any $s \leq 0$ we have

$$|A_{F_k}(x,s)| \leq 1 \quad \text{and} \quad |A_{F_k}(x_k, 0)| = 1,$$

for any $k \in \mathbb{N}$.

(b) *For any fixed $s \leq 0$, the sequence $\{(\Sigma, F_{k,s}^*(a_k^2 h), x_k)\}_{k\in\mathbb{N}}$ smoothly subconverges in the Cheeger-Gromov sense to a connected complete pointed Riemannian manifold $(\Sigma_\infty, g_\infty(s), x_\infty)$, where Σ_∞ does not depend on the choice of s. Moreover, the sequence $\left\{\left(P, a_k^2 h, F_k(x_k, s)\right)\right\}_{k\in\mathbb{N}}$ smoothly subconverges in the Cheeger-Gromov sense to the standard Euclidean space $(\mathbb{R}^l, g_{euc}, 0)$.*

(c) *There is an ancient smooth solution $F_\infty\colon \Sigma_\infty \times (-\infty, 0] \to \mathbb{R}^l$ of the mean curvature flow, such that for each fixed time $s \leq 0$, the sequence $\{F_{k,s}\}_{k\in\mathbb{N}}$ smoothly subconverges in the Cheeger-Gromov sense to $F_{\infty,s}$. Additionally,*

$$|A_{F_\infty}| \leq 1 \quad \text{and} \quad |A_{F_\infty}(x_\infty, 0)| = 1.$$

(d) *If $\dim \Sigma = 2$ and $H_{F_\infty} = 0$, then the limiting Riemann surface Σ_∞ has finite total curvature. In the matter of fact, the limiting surface Σ_∞ is conformally*

diffeomorphic to a compact Riemann surface minus a finite number of points and is of parabolic type.

For the proof see [15] and [66].

8 Graphical MCF of Surfaces in Four Manifolds

Let (M, g_M) and (N, g_N) be compact Riemann surfaces. Recall that a smooth map $f : M \to N$ is called *area decreasing* if $|\Lambda^2 df| \leq 1$, where $\Lambda^2 df$ is the 2-Jacobian of f. Being area decreasing means that the map f contracts 2-dimensional regions of M. If $|\Lambda^2 df| < 1$ the map is called *strictly area decreasing* and if $|\Lambda^2 df| \equiv 1$ the map is said *area preserving*.

We will deform area decreasing maps f by evolving their corresponding graphs

$$\Gamma(f) = \big\{(x, f(x)) \in M \times N : x \in M\big\},$$

under the mean curvature flow in the Riemannian product 4-manifold

$$(M \times N, \mathrm{g}_{M\times N} = \pi_M^* g_M + \pi_N^* g_N),$$

where $\pi_M : M \times N \to M$ and $\pi_N : M \times N \to N$ are the natural projection maps.

Our goal is to give a detailed, unified proof of the following theorem, which was shown in [78, 85, 95, 100]. For the strictly area decreasing case, we closely follow the presentation in [78].

Theorem 32 *Let (M, g_M) and (N, g_N) be compact Riemann surfaces and $f : M \to N$ be a smooth area decreasing map. Suppose that the curvatures σ_M of g_M and σ_N of g_N are related by*

$$\min \sigma_M \geq \max \sigma_N .$$

Then there exists a family of smooth area decreasing maps $f_t : M \to N$, $t \in [0, \infty)$, $f_0 = f$, such that the graphs $\Gamma(f_t)$ of f_t move by mean curvature flow in $(M \times N, \mathrm{g}_{M\times N})$. Furthermore, there exist only two possible categories of initial data sets and corresponding solutions:

(I) *The curvatures σ_M and σ_N are constant and equal and the map f_0 is area preserving. In this category, each f_t is area preserving and $\Gamma(f_t)$ smoothly converges to a minimal Lagrangian graph $\Gamma(f_\infty)$ in $M \times N$, with respect to the symplectic form*

$$\Omega_{M\times N} = \pi_M^* \Omega_M \mp \pi_N^* \Omega_N,$$

depending on whether the map f_0 is orientation preserving or reversing, respectively. Here Ω_M and Ω_N are the positively oriented volume forms of M and N, respectively.

(II) *All other possible cases. In this category, for $t > 0$ each map f_t is strictly area decreasing. Moreover, depending on the sign of $\sigma = \min \sigma_M$ we have the following behavior:*

(a) *If $\sigma > 0$, then the family $\Gamma(f_t)$ smoothly converges to the graph of a constant map.*
(b) *If $\sigma = 0$, then $\Gamma(f_t)$ smoothly converges to a totally geodesic graph $\Gamma(f_\infty)$ of $M \times N$.*
(c) *If $\sigma < 0$, then $\Gamma(f_t)$ smoothly converges to a minimal surface M_∞ of the product manifold $M \times N$.*

8.1 Jacobians of the Projection Maps

Let Ω_M denote the Kähler form of the Riemann surface (M, g_M) and Ω_N the Kähler form of (N, g_N). We can extend Ω_M and Ω_N to two parallel 2-forms on the product manifold $M \times N$ by pulling them back via the projection maps π_M and π_N. That is we may define the parallel forms $\Omega_1 = \pi_M^* \Omega_M$ and $\Omega_2 = \pi_N^* \Omega_N$. Define now two smooth functions u_1 and u_2 given by

$$u_1 = *(F^* \Omega_1) = *\left\{(\pi_M \circ F)^* \Omega_M\right\} = *(I^* \Omega_M)$$

and

$$u_2 = *(F^* \Omega_2) = *\left\{(\pi_N \circ F)^* \Omega_N\right\} = *(f^* \Omega_N)$$

where here $*$ stands for the Hodge star operator with respect to the metric g. Note that u_1 is the Jacobian of the projection map from $\Gamma(f)$ to the first factor of $M \times N$ and u_2 is the Jacobian of the projection map of $\Gamma(f)$ to the second factor of $M \times N$. With respect to the basis $\{e_1, e_2; \xi_3, \xi_4\}$ of the singular decomposition, we can write

$$u_1 = \frac{1}{\sqrt{(1+\lambda^2)(1+\mu^2)}} \quad \text{and} \quad |u_2| = \frac{\lambda\mu}{\sqrt{(1+\lambda^2)(1+\mu^2)}}.$$

Another important quantity that plays a crucial role in the case of maps between equi-dimensional manifolds is the *Jacobian determinant*, i.e., the map given by

$$\mathrm{Jac}(f) = \frac{*(f^* \Omega_N)}{*(I^* \Omega_M)} = \frac{u_2}{u_1}.$$

Moreover, the difference between u_1 and $|u_2|$ measures how far f is from being area preserving. In particular:

$$\begin{aligned} u_1 - |u_2| \geq 0 \quad &\text{if and only if} \quad f \text{is area decreasing,} \\ u_1 - |u_2| > 0 \quad &\text{if and only if} \quad f \text{is strictly area decreasing,} \\ u_1 - |u_2| = 0 \quad &\text{if and only if} \quad f \text{is area preserving.} \end{aligned}$$

8.2 The Kähler Angles

There are two natural complex structures associated to the product space $(M \times N, \mathrm{g}_{M\times N})$, namely $J_1 = \pi_M^* J_M - \pi_N^* J_N$ and $J_2 = \pi_M^* J_M + \pi_N^* J_N$, where J_M and J_N are the complex structures on M and N defined by

$$\Omega_M(\cdot\,,\cdot) = g_M(J_M\,\cdot\,,\cdot) \quad \text{and} \quad \Omega_N(\cdot\,,\cdot) = g_N(J_N\,\cdot\,,\cdot).$$

Chern and Wolfson [23] introduced a function which measures the deviation of the tangent plane $dF(T_xM)$ from a complex line of the space $T_{F(x)}(M \times N)$. More precisely, if we consider $(M \times N, \mathrm{g}_{M\times N})$ as a complex manifold with respect to J_1 then its corresponding *Kähler angle* a_1 is given by the formula

$$\cos a_1 = \varphi = \mathrm{g}_{M\times N}\big(J_1 dF(v_1), dF(v_2)\big) = u_1 - u_2.$$

For our convenience we require that $a_1 \in [0, \pi]$. Note that in general a_1 is not smooth at points where $\varphi = \pm 1$. If there exists a point $x \in M$ such that $a_1(x) = 0$ then $dF(T_xM)$ is a complex line of $T_{F(x)}(M \times N)$ and x is called a *complex point* of F. If $a_1(x) = \pi$ then $dF(T_xM)$ is an anti-complex line of $T_{F(x)}(M \times N)$ and x is said *anti-complex point* of F. In the case where $a_1(x) = \pi/2$, the point x is called *Lagrangian point* of the map F. In this case $u_1 = u_2$. Similarly, if we regard the product manifold $(M \times N, \mathrm{g}_{M\times N})$ as a Kähler manifold with respect to the complex structure J_2, then its corresponding Kähler angle a_2 is defined by the formula

$$\cos a_2 = \vartheta = \mathrm{g}_{M\times N}\big(J_2 dF(v_1), dF(v_2)\big) = u_1 + u_2.$$

The graph $\Gamma(f)$ in the product Kähler manifold $(M \times N, \mathrm{g}_{M\times N}, J_i)$ is called *symplectic* with respect to the Kähler form related to J_i, if the corresponding Kähler angle satisfies $\cos a_i > 0$. Therefore a map f is strictly area decreasing if and only if its graph $\Gamma(f)$ is symplectic with respect to both Kähler forms.

8.3 Structure Equations

Around each point $x \in \Gamma(f)$ we choose an adapted local orthonormal frame $\{e_1, e_2; \xi_3, \xi_4\}$ along the graph. In this special case the *Gauss equation* reads

$$2\sigma_g = 2u_1^2\sigma_M + 2u_2^2\sigma_N + |H|^2 - |A|^2,$$

where here σ_g is the Gauss curvature of the induced metric. From the *Ricci equation* we see that the curvature σ_n of the normal bundle of $\Gamma(f)$ is given by the formula

$$\sigma_n = R^{\perp}_{1234} = u_1u_2(\sigma_M + \sigma_N) + A^3_{11}A^4_{12} - A^3_{12}A^4_{11} + A^3_{12}A^4_{22} - A^3_{22}A^4_{12}.$$

The sum of the last four terms in the above formula is equal to minus the commutator $\sigma^{\perp}$ of the matrices $A^3 = (A^3_{ij})$ and $A^4 = (A^4_{ij})$, that is

$$\sigma^{\perp} = \langle [A^3, A^4]e_1, e_2\rangle = -A^3_{11}A^4_{12} + A^3_{12}A^4_{11} - A^3_{12}A^4_{22} + A^3_{22}A^4_{12}. \tag{38}$$

In the case where $u_1 = u_2$ and $\sigma_M = \sigma = \sigma_n$, it turns out that the immersion F is Lagrangian and $\sigma_g = \sigma_n$. In this case, the following algebraic equality holds

$$\sigma^{\perp} = \frac{|A|^2 - |H|^2}{2}. \tag{39}$$

8.4 Estimates for the Jacobians and the Kähler Angles

Let us evolve now by mean curvature flow the graph $\Gamma(f)$. Denote by $T_{\max}$ the maximal time of solution of the flow and by T_Γ the time until graphical property is preserved. Of course, $0 < T_\Gamma \leq T_{\max}$. We will give here several a priori estimates for the Jacobians u_1 and u_2 and the Kähler angles. The proofs are straightforward and follow directly as special cases of the general formulas of Section 6.5.

Lemma 11 *The gradients of the functions φ, ϑ at a point $x \in M$ satisfy the equations*

$$|\nabla\varphi|^2 = \big(1-\varphi^2\big)\big((A^3_{11} + A^4_{12})^2 + (A^3_{12} + A^4_{22})^2\big),$$
$$|\nabla\vartheta|^2 = \big(1-\vartheta^2\big)\big((A^3_{11} - A^4_{12})^2 + (A^3_{12} - A^4_{22})^2\big),$$

As long the mean curvature flow remains graphical, the Jacobians u_1 and u_2 satisfy the following coupled system of parabolic equations

$$\partial_t u_1 - \Delta u_1 = |A|^2u_1 + 2\sigma^{\perp}u_2 + \sigma_M(1 - u_1^2 - u_2^2)u_1 - 2\sigma_N u_1u_2^2,$$

$$\partial_t u_2 - \Delta u_2 = |A|^2 u_2 + 2\sigma^{\perp} u_1 + \sigma_N(1 - u_1^2 - u_2^2)u_2 - 2\sigma_M u_1^2 u_2.$$

Moreover, φ and ϑ satisfy the following system of equations

$$\partial_t \varphi - \Delta\varphi = (|A|^2 - 2\sigma^{\perp})\varphi + \frac{1}{2}\big(\sigma_M(\varphi + \vartheta) + \sigma_N(\varphi - \vartheta)\big)(1 - \varphi^2),$$

$$\partial_t \vartheta - \Delta\vartheta = (|A|^2 + 2\sigma^{\perp})\vartheta + \frac{1}{2}\big(\sigma_M(\varphi + \vartheta) - \sigma_N(\varphi - \vartheta)\big)(1 - \vartheta^2).$$

Lemma 12 *Let $f : (M, g_M) \to (N, g_N)$ be an area decreasing map between compact Riemann surfaces. Suppose that the curvatures of g_M and g_N satisfy $\sigma = \min \sigma_M \geq \max \sigma_N$. Then the following statements hold.*

(a) *The conditions $\mathrm{Jac}(f) \leq 1$ or $\mathrm{Jac}(f) \geq -1$ are both preserved as long as the flow remains graphical. In particular, the area decreasing property is preserved as long as the flow remains graphical.*
(b) *If there is a point $(x_0, t_0) \in M \times (0, T_\Gamma)$ where $\mathrm{Jac}^2(f)(x_0, t_0) = 1$, then $\mathrm{Jac}^2(f) \equiv 1$ in space and time and $\sigma_M \equiv \sigma \equiv \sigma_N$.*
(c) *The flow remains graphical as long as it exists, that is $T_\Gamma = T_{\max}$.*

Proof

(a) From Lemma 11, we deduce that

$$\partial_t \varphi - \Delta\varphi = \big(|A|^2 - 2\sigma^{\perp} + \sigma_N(1 - \varphi^2)\big)\varphi + \frac{1}{2}(\sigma_M - \sigma_N)(\varphi + \vartheta)(1 - \varphi^2).$$

Note that the quantities $1 - \varphi^2$ and $\varphi + \vartheta$ are non-negative. Hence, because of our curvature assumptions, the last line of the above equality is positive. Thus, there exists a time dependent function h such that

$$\partial_t \varphi - \Delta\varphi \geq h\,\varphi.$$

From the maximum principle we deduce that φ stays non-negative in time.

(b) From the strong maximum principle it follows that if φ vanishes at a point $(x_0, t_0) \in M \times (0, T_\Gamma)$, then it vanishes identically in space and time. In this case, ϑ is positive. Going back to the evolution equation of φ, we see that σ_M and σ_N must be constant equal to σ. Similarly, we prove the results concerning ϑ.

(c) By compactness, initially, we have that $\min_{x \in M} u_1(x, 0) = \varepsilon > 0$. By continuity, the minimum of u_1 stays positive for small values of t. However, we will show that the flow remains graphical as long as it exists. As a matter of fact, we will show that

$$\min_{x \in M} u_1(x, t) > 0,$$

as long as the flow exists. Suppose to the contrary, that there exists a first time where the graphical property does not hold. This means that there exists a point (x_0, t_0) in space-time with $t_0 < T$, such that $u_1(x_0, t_0) = 0$ and $u_1(x, t) > 0$, for all $(x, t) \in M \times [0, t_0)$. Since the area decreasing property is preserved by the flow and $|A|^2$ is bounded on $M \times [0, t_0]$, there exists a constant $c(t_0) \in \mathbb{R}$, such that

$$\partial_t u_1 - \Delta u_1 \geq c(t_0) u_1,$$

for all $(x, t) \in M \times [0, t_0)$. From the parabolic maximum principle, we get $u_1(x, t) \geq e^{c(t_0)t}$, for all $(x, t) \in M \times [0, t_0)$. Passing to the limit as t approaches t_0, we obtain

$$u_1(x_0, t_0) = \lim_{t \to t_0} u_1(x_0, t) \geq e^{c(t_0)t_0} > 0,$$

which leads to a contradiction.

This completes the proof. □

From the Lemma 12 we see that, under our assumptions, the evolved maps $\{f_t\}_{t \in (0, T_{\max})}$ are either strictly area decreasing or area preserving. This fact leads us to investigate these two cases separately.

8.4.1 Strictly Area Decreasing Case

We will explore the behaviour of $\rho : M \times [0, T_{\max}) \to \mathbb{R}$ given by $\rho = \varphi \vartheta$ under the graphical mean curvature flow.

Lemma 13 *Let (M, g_M) and (N, g_N) be compact Riemann surfaces such that their curvatures σ_M and σ_N are related by $\sigma = \min \sigma_M \geq \max \sigma_N$. The following hold true:*

(a) *If $\sigma \geq 0$, then there exists a positive constant c_0 such that*

$$\rho \geq \frac{c_0 e^{\sigma t}}{\sqrt{1 + c_0^2 e^{2\sigma t}}},$$

for any (x, t) in space-time.

(b) *If $\sigma < 0$, then there exists a positive constant c_0 such that*

$$\rho \geq \frac{c_0 e^{2\sigma t}}{\sqrt{1 + c_0^2 e^{4\sigma t}}},$$

for any (x, t) in space-time.

Proof From Lemma 11 we get,

$$\partial_t \rho - \Delta\rho = 2\rho|A|^2 - 2\langle\nabla\varphi, \nabla\vartheta\rangle + 2(1-\rho)\sigma_M u_1^2 - 2(1+\rho)\sigma_N u_2^2.$$

Note that

$$\begin{aligned} -2\rho\langle\nabla\varphi, \nabla\vartheta\rangle + \frac{1}{2}|\nabla\rho|^2 &= \frac{1}{2}\big(|\nabla(\varphi\vartheta)|^2 - 4\varphi\vartheta\langle\nabla\varphi, \nabla\vartheta\rangle\big) \\ &= \frac{1}{2}\big(\varphi^2|\nabla\vartheta|^2 + \vartheta^2|\nabla\varphi|^2 - 2\varphi\vartheta\langle\nabla\varphi, \nabla\vartheta\rangle\big) \\ &\geq \frac{1}{2}\big(|\varphi\nabla\vartheta| - |\vartheta\nabla\varphi|\big)^2. \end{aligned}$$

Since by assumption $\sigma_M \geq \sigma \geq \sigma_N$, we deduce that

$$\partial_t \rho - \Delta\rho \geq -\frac{1}{2\rho}|\nabla\rho|^2 + 2\sigma\rho(1 - u_1^2 - u_2^2).$$

One can algebraically check that

$$1 - \rho^2 \leq 2(1 - u_1^2 - u_2^2) \leq 2(1-\rho^2). \tag{40}$$

(a) Suppose at first that $\sigma \geq 0$. Then

$$\partial_t \rho - \Delta\rho \geq -\frac{1}{2\rho}|\nabla\rho|^2 + \sigma\rho(1-\rho^2).$$

From the comparison maximum principle we obtain

$$\rho \geq \frac{c_0 e^{\sigma t}}{\sqrt{1 + c_0^2 e^{2\sigma t}}},$$

where c_0 is a positive constant.

(b) In the case where $\sigma < 0$, from the Equation (40) we deduce that

$$\partial_t \rho - \Delta\rho \geq -\frac{1}{2\rho}|\nabla\rho|^2 + 2\sigma\rho(1-\rho^2),$$

from where we get the desired estimate.

This completes the proof. □

8.4.2 Area Preserving Case

Suppose that the family of the graphs is generated orientation preserving by area decreasing maps. This means that φ is identically zero. In the next lemma we derive an estimate for the Kähler angle ϑ.

Lemma 14 *Suppose that M and N are compact with the same constant sectional curvature σ and that $f : M \to N$ is an area preserving map. Then, there exists a positive real number c_0 such that*

$$1 \geq \vartheta(x,t) \geq \frac{c_0 e^{\sigma t}}{\sqrt{1 + c_0^2 e^{2\sigma t}}},$$

for any point (x,t) in space-time.

Proof Since $|A|^2 + 2\sigma^{\perp} \geq 0$, from the evolution equation of ϑ, we get

$$\partial_t \vartheta - \Delta\vartheta \geq \sigma\vartheta(1 - \vartheta^2).$$

According to the parabolic maximum principle, there exist a positive real number c_0 such that

$$\vartheta(x,t) \geq \frac{c_0 e^{\sigma t}}{\sqrt{1 + c_0^2 e^{2\sigma t}}},$$

for any (x,t) in space-time. This completes the proof. □

8.5 Curvature Decay Estimates

8.5.1 Strictly Area Decreasing Case

Lemma 15 *Let $f : (M, g_M) \to (N, g_N)$ be a strictly area decreasing map. Suppose that the curvatures of M and N satisfy $\sigma = \min \sigma_M \geq \max \sigma_N$. Let $\delta : [0,T) \to \mathbb{R}$ be a positive increasing real function and τ the time dependent function given by $\tau = \log\left(\delta|H|^2 + \varepsilon\right)$, where ε is a non-negative number. Then,*

$$\begin{aligned}\partial_t \tau - \Delta\tau \leq{}& \frac{2\delta}{\delta|H|^2 + \varepsilon}|H|^2|A|^2 + \frac{\delta'}{\delta|H|^2 + \varepsilon}|H|^2 \\ &+ \frac{2\delta}{\delta|H|^2 + \varepsilon}|H|^2\sigma_M(1 - u_1^2 - u_2^2) + \frac{1}{2}|\nabla\tau|^2.\end{aligned}$$

Proof Recall from Lemma 6 that $|H|^2$ evolves in time under the equation

$$\begin{aligned}\partial_t|H|^2 - \Delta|H|^2 = 2|A^H|^2 - 2|\nabla^\perp H|^2 \\ +2\widetilde{R}(H, dF(e_1), H, dF(e_1)) + 2\widetilde{R}(H, dF(e_2), H, dF(e_2)),\end{aligned}$$

where $\{e_1, e_2\}$ is a local orthonormal frame with respect to g. Using the special frames of the singular value decomposition we see that

$$\begin{aligned}&\widetilde{R}\big(H, dF(e_1), H, dF(e_1)\big) + \widetilde{R}\big(H, dF(e_2), H, dF(e_2)\big)\\ &\quad= \sigma_M u_1^2(\lambda^2+\mu^2)|H|^2 - (\sigma_M-\sigma_N)u_1^2\big(\lambda^2(H^4)^2+\mu^2(H^3)^2\big)\\ &\quad\le \sigma_M(1-u_1^2-u_2^2)|H|^2.\end{aligned}$$

Note that from Cauchy–Schwarz inequality $|A^H| \le |A|\cdot|H|$. Moreover, observe that at points where the mean curvature vector is non-zero, from Kato's inequality, we have that

$$\big|\nabla^\perp H\big|^2 \ge \big|\nabla|H|\big|^2.$$

Consequently, at points where the norm $|H|$ of the mean curvature is not zero the following inequality holds

$$\partial_t|H|^2 - \Delta|H|^2 \le -2\big|\nabla|H|\big|^2 + 2|A|^2|H|^2 + 2\sigma_M(1-u_1^2-u_2^2)|H|^2.$$

Now let us compute the evolution equation of the function τ. We have,

$$\begin{aligned}\partial_t\tau - \Delta\tau &= \frac{\delta(\partial_t|H|^2 - \Delta|H|^2)}{\delta|H|^2+\varepsilon} + \frac{\delta^2|\nabla|H|^2|^2}{(\delta|H|^2+\varepsilon)^2} + \frac{\delta'|H|^2}{\delta|H|^2+\varepsilon}\\ &\le -\frac{2\delta}{\delta|H|^2+\varepsilon}\big|\nabla|H|\big|^2 + \frac{\delta^2}{(\delta|H|^2+\varepsilon)^2}\big|\nabla|H|^2\big|^2\\ &\quad+\frac{2\delta}{\delta|H|^2+\varepsilon}|H|^2|A|^2 + \frac{\delta'}{\delta|H|^2+\varepsilon}|H|^2\\ &\quad+\frac{2\delta}{2\delta|H|^2+\varepsilon}|H|^2\sigma_M(1-u_1^2-u_2^2).\end{aligned}$$

Note that

$$-\frac{2\delta}{\delta|H|^2+\varepsilon}\big|\nabla|H|\big|^2 + \frac{1}{2}\frac{\delta^2}{(\delta|H|^2+\varepsilon)^2}|\nabla|H|^2\big|^2 \le 0.$$

Therefore,

$$\partial_t \tau - \Delta \tau \leq \frac{1}{2}|\nabla \tau|^2 + \frac{2\delta}{\delta|H|^2+\varepsilon}|H|^2|A|^2 + \frac{\delta'}{\delta|H|^2+\varepsilon}|H|^2 + \frac{2\delta}{\delta|H|^2+\varepsilon}|H|^2\sigma_M(1-u_1^2-u_2^2),$$

and this completes the proof. □

Theorem 33 *Let $f : (M, g_M) \to (N, g_N)$ be an area decreasing map, where M and N are compact Riemann surfaces. Suppose that the curvatures of M and N satisfy $\sigma = \min \sigma_M \geq \sup \sigma_N$. Then the following statements hold:*

(a) *There exist a positive time independent constant C such that $|H|^2 \leq C$.*
(b) *If $\sigma \geq 0$, there exist a time independent constant C so that $|H|^2 \leq Ct^{-1}$.*

Proof Consider the time dependent function $\Theta = \log(\delta|H|^2 + \varepsilon) - \log \rho$, where δ is a positive increasing function. From Lemmas 6 and 13 and $|H|^2 \leq 2|A|^2$, we deduce that

$$\partial_t \Theta - \Delta \Theta \leq \frac{1}{2}\langle \nabla \Theta, \nabla \tau + \nabla \rho \rangle + \frac{\delta'|H|^2 - \varepsilon|H|^2 - 2\varepsilon\sigma(1-u_1^2-u_2^2)}{\delta|H|^2+\varepsilon}.$$

Choosing $\delta = 1$ and $\varepsilon = 0$, we obtain that

$$\partial_t \Theta - \Delta \Theta \leq \frac{1}{2}\langle \nabla \Theta, \nabla \tau + \nabla \rho \rangle.$$

From the maximum principle the norm $|H|$ remains uniformly bounded in time regardless of the sign of the constant σ. In the case where $\sigma \geq 0$, choosing $\varepsilon = 1$ and $\delta = t$, we deduce that Θ remains uniformly bounded in time which gives the desired decay estimate for H. □

8.5.2 Area Preserving Case

In the sequel, we provide a very important decay estimate due to Wang [91] for the mean curvature in the area preserving case.

Theorem 34 *Suppose that M and N are compact Riemannian manifolds with the same constant sectional curvature σ and that $f : M \to N$ is an area preserving map. Then, the following decay estimate holds:*

$$\int \frac{|H|^2}{\vartheta} \Omega \leq e^{\sigma t},$$

where Ω is the volume element of the induced metric.

Proof The idea is to compare $|H|$ with ϑ. We compute

$$\partial_t\left(\vartheta^{-1}|H|^2\right)-\Delta\left(\vartheta^{-1}|H|^2\right)=\vartheta^{-1}(\partial_t|H|^2-\Delta|H|^2)-\vartheta^{-2}|H|^2(\partial_t\vartheta-\Delta\vartheta)$$
$$+2\vartheta^{-2}\langle\nabla|H|^2,\nabla\vartheta\rangle-2\vartheta^{-3}|H|^2|\nabla\vartheta|^2.$$

But from the evolution equation of ϑ and $|H|^2$, we obtain

$$\begin{aligned}&\partial_t\left(\vartheta^{-1}|H|^2\right)-\Delta\left(\vartheta^{-1}|H|^2\right)\\&=\vartheta^{-1}\big(-2|\nabla^\perp H|^2+2\textstyle\sum_{k,\alpha,\beta}H^\alpha H^\beta\widetilde{R}_{\alpha k\beta k}+2\sum_{i,j}(A^H_{ij})^2\big)\\&-\vartheta^{-2}|H|^2\big((|A|^2+2\sigma^\perp)\vartheta+\sigma\vartheta(1-\vartheta^2)\big)+2\vartheta^{-2}\langle\nabla|H|^2,\nabla\vartheta\rangle-2\vartheta^{-3}|H|^2|\nabla\vartheta|^2.\end{aligned}\tag{41}$$

Using the Equation (39) and the formula

$$\sum_{k,\alpha,\beta}H^\alpha H^\beta\widetilde{R}_{\alpha k\beta k}=\sigma\left(1-\frac{\vartheta^2}{2}\right)|H|^2\tag{42}$$

the identity (41) becomes

$$\begin{aligned}&\partial_t\left(\vartheta^{-1}|H|^2\right)-\Delta\left(\vartheta^{-1}|H|^2\right)\\&=\vartheta^{-3}\Big(4\vartheta|H|\langle\nabla\vartheta,\nabla|H|\rangle-2|\nabla\vartheta|^2|H|^2-2\vartheta^2|\nabla^\perp H|^2\Big)\\&+\vartheta^{-1}\Big(2\textstyle\sum_{i,j}(A^H_{ij})^2-2|H|^2|A|^2+|H|^4\Big)+\sigma\vartheta^{-1}|H|^2.\end{aligned}$$

Integrating and using Stokes' theorem, we have

$$\begin{aligned}\partial_t\left(\int\vartheta^{-1}|H|^2\Omega\right)&=\int\vartheta^{-1}|H|^2\nabla_{\partial_t}\Omega\\&+2\int\vartheta^{-3}\Big(2\vartheta|H|\langle\nabla\vartheta,\nabla|H|\rangle-|\nabla\vartheta|^2|H|^2-\vartheta^2|\nabla^\perp H|^2\Big)\Omega\\&+\int\vartheta^{-1}\Big(2\textstyle\sum_{i,j}(A^H_{ij})^2-2|H|^2|A|^2+|H|^4\Big)\Omega+\sigma\int\vartheta^{-1}|H|^2\Omega.\end{aligned}$$

Using

$$\big|\nabla|H|\big|\leq\big|\nabla^\perp H\big|$$

in the first term on the right hand side of the above equation and completing the square, we have

$$2\vartheta|H|\langle\nabla\vartheta,\nabla|H|\rangle-|\nabla\vartheta|^2|H|^2-\vartheta^2\big|\nabla|H|\big|^2=-\big||H|\nabla\vartheta-\vartheta\nabla|H|\big|^2\leq 0.$$

Moreover, from Lemma 4, we have $\nabla_{\partial_t}\Omega=-|H|^2\Omega$. Also, by Cauchy–Schwarz inequality, we get

$$\sum\nolimits_{i,j}(A^H_{ij})^2\leq\sum\nolimits_{i,j}|A_{ij}|^2|H|^2=|A|^2|H|^2.$$

Therefore, putting everything together, we get

$$\partial_t\bigg(\int\vartheta^{-1}|H|^2\Omega\bigg)\leq\sigma\int\vartheta^{-1}|H|^2\Omega$$

and by integration, we obtain the result. □

8.6 Long-time Existence

We will show now that the graphical MCF exists for all times.

Theorem 35 *Let (M,g_M) and (N,g_N) be compact Riemann surfaces such that their curvatures σ_M and σ_N are related by $\sigma=\min\sigma_M\geq\max\sigma_N$. Also, let $f:M\to N$ be an area preserving map. Evolve the graph off under the mean curvature flow. Then, the norm of the second fundamental form of the evolved graphs stays uniformly bounded in time and so the graphical mean curvature flow exists for all times.*

Proof Suppose that $|A|$ is not uniformly bounded. Then, there exists a sequence $\{(x_k,t_k)\}_{k\in\mathbb{N}}$ in $M\times[0,T_{\max})$ with $\lim t_k=T_{\max}\leq\infty$, and such that

$$a_k=\max_{(x,t)\in M\times[0,t_k]}|A(x,t)|=|A(x_k,t_k)|\to\infty.$$

Let $F_k\colon M\times[-a_k^2t_k,0]\to(N,a_k^2g_N)$ be the graph of the "*rescaled map*"

$$f\colon(M,a_k^2g_M)\to(N,a_k^2g_N).$$

Claim: *The singular values are invariant under parabolic rescalings.*

Let $\{\alpha_1,\alpha_2\}$ and $\{\beta_1,\beta_2\}$ orthonormal frames of the singular value decomposition of f. Then $\{\widetilde{\alpha}_1=\alpha_1/a_k,\widetilde{\alpha}_2=\alpha_2/a_k\}$ is an orthonormal frame with respect to $a_k^2g_M$ and $\{\widetilde{\beta}_1=\beta_1/\alpha_k,\widetilde{\beta}_2=\beta_2/\alpha_k\}$ is orthonormal with respect to g_N. Therefore, the singular values of the rescaled map f are given by

$$df(\widetilde{\alpha}_1)=\frac{1}{a_k}df(\alpha_1)=\lambda\frac{\beta_1}{a_k}=\lambda\widetilde{\beta}_1$$

and

$$df(\widetilde{\alpha}_2) = \frac{1}{a_k} df(\alpha_2) = \mu \frac{\beta_2}{a_k} = \mu \widetilde{\beta}_2.$$

This completes the proof of the claim.

Thus, $\varphi_{F_k} = \varphi$ and $\vartheta_{F_k} = \vartheta$. Also, from Theorem 31(a) we have

$$H_{F_k}(x, s) = \frac{1}{a_k^2} H(x, s/a_k^2 + t_k),$$

for any $(x, s) \in M \times [-a_k^2 t_k, 0]$.

CASE 1 Suppose that the evolved graphs are generated by strictly area decreasing maps. Since from the estimate of Lemma 15 the norm $|H|$ of the mean curvature is uniformly bounded and the convergence is smooth, it follows that $F_\infty : \Sigma_\infty \to \mathbb{R}^4$ is a complete minimal immersion of parabolic type. Hence, any non-negative superharmonic function must be constant. Since the convergence is smooth, the corresponding Kähler angles φ_∞, ϑ_∞ of F_∞ with respect to the complex structures $J = (J_{\mathbb{R}^2}, -J_{\mathbb{R}^2})$ and $J_2 = (J_{\mathbb{R}^2}, J_{\mathbb{R}^2})$ of $\mathbb{R}^4$ are non-negative. As in Lemma 11 we get that

$$\Delta\varphi_\infty + \big(|A_{F_\infty}|^2 - 2\sigma^\perp_{F_\infty}\big)\varphi_\infty = 0, \tag{43}$$

$$\Delta\vartheta_\infty + \big(|A_{F_\infty}|^2 + 2\sigma^\perp_{F_\infty}\big)\vartheta_\infty = 0, \tag{44}$$

where $-\sigma^\perp_{F_\infty}$ is the normal curvature of F_∞. Moreover,

$$|\nabla\varphi_\infty|^2 = (1-\varphi_\infty^2)\Big(\big((A_{F_\infty})^3_{11} + (A_{F_\infty})^4_{12}\big)^2 + \big((A_{F_\infty})^3_{12} - (A_{F_\infty})^4_{11}\big)^2\Big), \tag{45}$$

$$|\nabla\vartheta_\infty|^2 = (1-\vartheta_\infty^2)\Big(\big((A_{F_\infty})^3_{11} - (A_{F_\infty})^4_{12}\big)^2 + \big((A_{F_\infty})^3_{12} + (A_{F_\infty})^4_{11}\big)^2\Big). \tag{46}$$

Note that from (38) one can easily derive the inequalities

$$|A_{F_\infty}|^2 \pm 2\sigma^\perp_{F_\infty} \geq 0.$$

From (43) and (44) we deduce that φ_∞ and ϑ_∞ are superharmonic and consequently they must be constants. Thus, the functions $(u_1)_\infty$ and $(u_2)_\infty$ are also constants. We will distinguish three subcases:

Sub-case A Suppose at first that $\varphi_\infty > 0$ and $\vartheta_\infty > 0$. Then from (43) and (44) we deduce that

$$|A_{F_\infty}|^2 \pm 2\sigma^\perp_{F_\infty} = 0$$

which implies that $|A_{F_\infty}| = 0$. This contradicts the fact that there is a point where $|A_{F_\infty}| = 1$.

Sub-case B Suppose that both constants φ_∞ and ϑ_∞ are zero. Then from the Equations (45) and (46) we obtain that A_{F_∞} vanishes identically, which is a again a contradiction.

Sub-case C Suppose now that only one of the constants φ_∞, ϑ_∞ is zero. Let us assume that $\varphi_\infty = 0$ and $\vartheta_\infty > 0$. The case $\varphi_\infty > 0$ and $\vartheta_\infty = 0$ is treated in a similar way. Since $\varphi_\infty = 0$, $F_\infty : \Sigma_\infty \to \mathbb{R}^4$ must be a minimal Lagrangian immersion. Note that in this case necessarily $(u_1)_\infty = (u_2)_\infty = \text{const} > 0$. Recall from Theorem 5 that the minimal Lagrangian F_∞ can be locally reparametrized in the form

$$F_\infty = \frac{1}{\sqrt{2}} e^{i\beta/2}\big(\mathcal{F}_1 - i\overline{\mathcal{F}_2}, \mathcal{F}_2 + i\overline{\mathcal{F}_1}\big),$$

where β is a constant and $\mathcal{F}_1, \mathcal{F}_2 : \mathbb{D} \subset \mathbb{C} \to \mathbb{C}$ are holomorphic functions defined in a simply connected domain $\mathbb{D}$ such that

$$|(\mathcal{F}_1)_z|^2 + |(\mathcal{F}_2)_z|^2 > 0.$$

The Gauss map of F_∞ is described by $\mathcal{G} : \mathbb{D} \to \mathbb{S}^2 = \mathbb{C} \cup \{\infty\}$ given by

$$\mathcal{G} = (\mathcal{F}_1)_z / (\mathcal{F}_2)_z.$$

Because $(u_1)_\infty = \text{const} > 0$ we get that F_∞ is the graph of an area preserving map h. Then

$$\mathcal{F}_1 = (z + i\overline{h})/2, \quad \mathcal{F}_2 = (-i\overline{z} + h)/2 \quad \text{and} \quad |h_z|^2 - |h_{\bar{z}}|^2 = 1.$$

Therefore

$$\mathcal{G} = (\mathcal{F}_1)_z / (\mathcal{F}_2)_z = (1 - ih_{\bar{z}})/h_z.$$

A straightforward computation shows that

$$|\mathcal{G}|^2 = \frac{\big|1 + i\overline{h_{\bar{z}}}\big|^2}{|h_z|^2} = \frac{1 + |h_{\bar{z}}|^2 + i\big(\overline{h_{\bar{z}}} - h_{\bar{z}}\big)}{1 + |h_{\bar{z}}|^2} = 1 + \frac{2\,\mathrm{Im}(h_{\bar{z}})}{1 + |h_{\bar{z}}|^2} \leq 2.$$

Hence, the image of $\mathcal{G}$ is contained in a bounded subset of $\mathbb{C} \cup \{\infty\}$. But then, due to Theorem 11 the immersion F_∞ must be flat, which is a contradiction.

CASE 2 Let us treat now the area preserving case. In this situation, we have that

$$\frac{|H_{F_k}|^2}{\vartheta_{F_k}} = \frac{1}{a_k^2}\frac{|H|^2}{\vartheta}.$$

We distinguish two subcases:

Sub-case A Let us suppose that $\sigma \leq 0$. Using Lemma 34, we have

$$\int \frac{|H_{F_k}|^2}{\vartheta_{F_k}}\Omega_k = \frac{1}{a_k^2}\int \frac{|H|^2}{\vartheta}\Omega \leq \frac{1}{a_k^2}e^{\sigma(s/a_k^2+t_k)} \leq \frac{1}{a_k^2}c,$$

where $c > 0$. Since the convergence is smooth, we have

$$0 = \lim_{k\to\infty}\int \frac{|H_{F_k}|^2}{\vartheta_{F_k}}\Omega = \int \lim_{k\to\infty}\frac{|H_{F_k}|^2}{\vartheta_{F_k}}\Omega = \int \frac{|H_{F_\infty}|^2}{\vartheta_\infty}\Omega.$$

Therefore, $H_{F_\infty} = 0$. Proceeding exactly in the same way as in CASE 1 we can prove that F_∞ is flat, something which leads to a contradiction.

Sub-case B Let us treat now the case $\sigma > 0$. We will show at first that $T_{\max} = \infty$. To show this, assume in contrary that $T_{\max} < +\infty$. Then,

$$\int \frac{|H|^2}{\vartheta}\Omega \leq e^{\sigma t} \leq e^{\sigma T_{\max}} < +\infty.$$

As in the previous case, we deduce that $H_{F_\infty} = 0$. Performing exactly the same procedure as above, we get a contradiction. Therefore, there is no finite time singularity and the flow exists for all times. It remains to show that $|A|^2 \leq C$, where C is time independent. Indeed, since $\lambda\mu = 1$, we obtain

$$\vartheta = \frac{2\lambda}{1+\lambda^2} \leq 1.$$

On the other hand, from Lemma 14, we have

$$1 \geq \vartheta \geq \frac{c_0 e^t}{\sqrt{1+c_0^2 e^{2t}}},$$

which tends to 1 as $t \to \infty$. Therefore, $\vartheta_\infty = 1$ and $\lambda_\infty = 1$. Therefore, f_∞ is an isometry and, thus, F_∞ must be totally geodesic. The latter implies $|A_{F_\infty}| = 0$ and this is again a contradiction.

This completes the proof. □

8.7 Convergence and Proof of Theorem 32

We are ready to prove the main theorem stated in the introduction of this section. We will show that the graphical mean curvature flow of an area preserving map converges to an isometry in the positive case, to an affine map in the zero case, and to a minimal surface in the negative case. Recall that from Theorem 35, we already know that the norm of the second fundamental form stays uniformly bounded in time. Since

$$\nabla_{\partial_t} \Omega = -\int_M |H|^2 \Omega$$

and since the graphical flow exists for all time we have that there exists a time-independent constant C, such that

$$\int_0^\infty \left(\int_M |H|^2 \Omega \right) dt \leq C.$$

Therefore, there exists a sequence $\{t_k\}_{k\in\mathbb{N}}$, such that

$$\lim_{k\to\infty} \int_M |H|^2 \Omega = 0. \tag{47}$$

From Theorem 35, the norms of the second fundamental forms of the evolving submanifolds and their derivatives are uniformly bounded in time. Since the product manifold $M \times N$ is compact, after passing to a subsequence of $\{t_k\}_{k\in\mathbb{N}}$ if necessary, we deduce that the flow subconverges smoothly to a smooth surface M_∞ of $M \times N$; see for example [11, Theorem 1.1]. From (47) M_∞ should be minimal. Due to a deep result of Simon [82], it follows that the flow converges smoothly and uniformly to a minimal surface $M_\infty \subset M \times N$. Additionally, we have the following situations:

Area Preserving Case Let us treat the case where the evolving maps are area preserving.

(a) If $\sigma > 0$, then from Lemma 14(c), we have $\vartheta \to 1$, as $t \to \infty$. Therefore, M_∞ is the graph of an isometry $f_\infty \colon M \to N$.

(b) If $\sigma = 0$, then from Lemma 14(c), we have that $\vartheta \geq c_0 > 0$. Hence, the surface M_∞ is the graph of a map $f_\infty \colon M \to N$. From Lemma 11 and the fact that $2\sigma_\infty^\perp = |A_\infty|^2$, we have

$$-\Delta \vartheta_\infty = 2|A_\infty|^2 \vartheta_\infty \geq 0.$$

By the strong maximum principle, we get $|A_\infty|^2 = 0$. Hence, M_∞ is totally geodesic.

Strictly Area Decreasing Case Assume that our maps are area decreasing.

(a) Suppose that $\sigma > 0$. In this case the flow is smoothly converging to a graphical minimal surface $M_\infty = \Gamma(f_\infty)$ of $M \times N$. Due to Theorem 13(a), the biggest singular value tends to zero as time goes to infinity. Hence, M_∞ must be totally geodesic and f_∞ is a constant map.

(b) Assume that $\sigma = 0$. As in the previous case, we have smooth convergence of the flow to a minimal graphical surface $M_\infty = \Gamma(f_\infty)$ of $M \times N$, where f_∞ is a strictly area decreasing map. Because of the formula

$$\partial_t \int_M \Omega = -\int_M |H|^2 \Omega \leq 0,$$

we obtain that

$$\int_M \Omega \leq \int_M \Omega_M = \text{constant}\,.$$

From Theorem 33(b), there is a non-negative constant C such that

$$\int_M |H|^2 \Omega \leq \frac{C}{t} \int_M \Omega \leq \frac{C}{t} \int_M \Omega.$$

Due to our assumptions we have $u_2^2 \leq u_1^2 \leq 1$ and $\min \sigma_M \geq 0 \geq \sup \sigma_N$. Moreover, recall that

$$\Omega = \sqrt{(1+\lambda^2)(1+\mu^2)}\,\Omega_M = u_1^{-1}\Omega_M.$$

From the Gauss equation (8.3) and the Gauss-Bonnet formula we get

$$\begin{aligned}
\int_M |A|^2 \Omega &= \int_M |H|^2 \Omega + 2\int_M \left(\sigma_M u_1^2 + \sigma_N u_2^2\right)\Omega - 2\int_M \sigma_{g(t)}\Omega \\
&\leq 2\int_M \sigma_M u_1^2 \Omega - 2\int_M \sigma_{g(t)}\Omega + \int_M |H|^2 \Omega \\
&\leq 2\int_M \sigma_M u_1 \Omega - 2\int_M \sigma_{g(t)}\Omega + \int_M |H|^2 \Omega \\
&\leq 2\int_M \sigma_M \Omega_M - 2\int_M \sigma_{g(t)}\Omega + \int_M |H|^2 \Omega = \int_M |H|^2 \Omega \\
&\leq Ct^{-1},
\end{aligned}$$

where C is a non-negative constant. Passing to the limit, we deduce that

$$\int_M |A_\infty|^2 \Omega_\infty = 0.$$

Thus, $M_\infty = F_\infty(M)$ must be a totally geodesic graphical surface.

This completes the proof. □

References

1. R. Aiyama, Totally real surfaces in the complex 2-space, in *Steps in Differential Geometry (Debrecen, 2000)*. Institute of Mathematical Information, Debrecen (2001), pp. 15–22
2. R. Aiyama, Lagrangian surfaces with circle symmetry in the complex two-space. Michigan Math. J. **52**, 491–506 (2004)
3. F. Almgren, Some interior regularity theorems for minimal surfaces and an extension of Bernstein's theorem. Ann. Math. **84**, 277–292 (1966)
4. B. Andrews, C. Baker, Mean curvature flow of pinched submanifolds to spheres. J. Differ. Geom. **85**, 357–395 (2010)
5. B. Andrews, C. Hopper, *The Ricci flow in Riemannian geometry, A complete proof of the differentiable 1/4-pinching sphere theorem*. Lecture Notes in Mathematics, vol. 2011 (Springer, Heidelberg, 2011)
6. R. Assimos, J. Jost, The geometry of maximum principles and a Bernstein theorem in codimension 2. arXiv:1811.09869 (2018)
7. C. Baker, *The mean curvature flow of submanifolds of high codimension*, PhD Thesis (Australian National University, Mathematical Sciences Institute, Australia, 2010). https://arxiv.org/abs/1104.4409
8. S. Bernstein, Über ein geometrisches Theorem und seine Anwendung auf die partiellen Differentialgleichungen yom elliptischen Typus. Math. Z. **26**, 551–558 (1927). (translation of the original version in Comm. Soc. Math. Kharkov 2-'eme sér. **15**, 38–45 (1915-1917))
9. C. Böhm, B. Wilking, Nonnegatively curved manifolds with finite fundamental groups admit metrics with positive Ricci curvature. Geom. Funct. Anal. **17**, 665–681 (2007)
10. E. Bombieri, E. De Giorgi, E. Giusti, Minimal cones and the Bernstein conjecture. Invent. Math. **7**, 243–268 (1969)
11. P. Breuning, Immersions with bounded second fundamental form. J. Geom. Anal. **25**, 1344–1386 (2015)
12. A. Chau, J. Chen, W. He, Lagrangian mean curvature flow for entire Lipschitz graphs. Calc. Var. Partial Differ. Equations **44**, 199–220 (2012)
13. A. Chau, J. Chen, Y. Yuan, Lagrangian mean curvature flow for entire Lipschitz graphs II. Math. Ann., **357**, 165–183 (2013)
14. J. Cheeger, M. Gromov, M. Taylor, Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds. J. Differ. Geom. **17**, 15–53 (1982)
15. J. Chen, W. He, A note on singular time of mean curvature flow. Math. Z., **266**, 921–931 (2010)
16. J. Chen, J. Li, Mean curvature flow of surface in 4-manifolds. Adv. Math. **163**, 287–309 (2001)
17. J. Chen, J. Li, Singularity of mean curvature flow of Lagrangian submanifolds. Invent. Math. **156**, 25–51 (2004)
18. J. Chen, G. Tian, Moving symplectic curves in Kähler-Einstein surfaces. Acta Math. Sin. (Engl. Ser.) **16**, 541–548 (2000)
19. B.-Y. Chen, J.-M. Morvan, Géométrie des surfaces lagrangiennes de $\mathbf{C}^2$. J. Math. Pures Appl. **66**, 321–325 (1987)
20. S.-S. Chern, Minimal surfaces in an euclidean space of N dimensions, in *Differential and Combinatorial Topology* (Princeton University, Princeton, 1965), pp. 187–198

21. S.-S. Chern, Minimal submanifolds in a Riemannian manifold, in *University of Kansas, Department of Mathematics Technical Report*, vol. 19 (New Series) (University of Kansas Lawrence, Lawrence, 1968)
22. S.S. Chern, R. Osserman, Complete minimal surfaces in Euclidean n-space. J. Anal. Math. **19**, 15–34 (1967)
23. S.-S. Chern, J.G. Wolfson, Minimal surfaces by moving frames. Am. J. Math. **105**, 59–83 (1983)
24. B. Chow, P. Lu, L. Ni, *Hamilton's Ricci Flow* (American Mathematical Society/Science Press Beijing, Providence/New York, 2006)
25. B. Chow, S.-C. Chu, D. Glickenstein, C. Guenther, J. Isenberg, T. Ivey, D. Knopf, P. Lu, F. Luo, L. Ni, The Ricci flow: techniques and applications. Geometric aspects. Part I, in *Mathematical Surveys and Monographs*, vol. 135 (American Mathematical Society, Providence, RI, 2007)
26. B. Chow, S.-C. Chu, D. Glickenstein, C. Guenther, J. Isenberg, T. Ivey, D. Knopf, P. Lu, F. Luo, L. Ni, The Ricci flow: techniques and applications. Part II, in *Mathematical Surveys and Monographs*, vol. 144 (American Mathematical Society, Providence RI, 2008)
27. A. Cooper, *Mean Curvature Flow in Higher Codimension*, Thesis (Ph.D.) (Michigan State University, Michigan, 2011)
28. D. Crowley, T. Schick, The Gromoll filtration, KO-characteristic classes and metrics of positive scalar curvature. Geom. Topol. **17**, 1773–1789 (2013)
29. M. Dajczer, R. Tojeiro, *Submanifold Theory Beyond an Introduction* (Springer, New York, 2019)
30. E. De Giorgi, Una extensione del teorema di Bernstein. Ann. Sci. Norm. Super. Pisa **19**, 79–85 (1965)
31. D. DeTurck, Deforming metrics in the direction of their Ricci tensors. J. Differ. Geom. **18**, 157–162 (1983)
32. K. Ecker, *Regularity Theory for Mean Curvature Flow* (Birkhäuser Boston Inc., Boston, 2004)
33. K. Ecker, G. Huisken, Mean curvature evolution of entire graphs. Ann. Math. **130**, 453–471 (1989)
34. J. Eells, J. Sampson, Harmonic mappings of Riemannian manifolds. Am. J. Math. **86**, 109–160 (1964)
35. L.C. Evans, Partial differential equations, in *Graduate Studies in Mathematics*, vol. 19 (2010)
36. L.C. Evans, A strong maximum principle for parabolic systems in a convex set with arbitrary boundary. Proc. Am. Math. Soc. **138**, 3179–3185 (2010)
37. D. Fischer-Colbrie, Some rigidity theorems for minimal submanifolds of the sphere. Acta Math. **145**, 29–46 (1980)
38. M. Gromov, Homotopical effects of dilatation. J. Differ. Geom. **13**, 303–310 (1978)
39. M. Gromov, Pseudoholomorphic curves in symplectic manifolds. Invent. Math. **82**, 307–347 (1985)
40. M. Gromov, Carnot-Carathéodory spaces seen from within, in *Sub-Riemannian Geometry*. Program of Mathematical, vol. 144 (Birkhäuser, Basel, 1996), pp. 79–323
41. M. Gromov, *Quantitative homotopy theory*, in *Prospects in Mathematics* (American Mathematical Society, Providence, 1999), pp. 45–49
42. L. Guth, Contraction of areas vs. topology of mappings. Geom. Funct. Anal. **23**, 1804–1902 (2013)
43. R. Hamilton, The inverse function theorem of Nash and Moser. Bull. Am. Math. Soc. (N.S.) **7**, 65–222 (1982)
44. R. Hamilton, Three-manifolds with positive Ricci curvature. J. Differ. Geom. **17**, 255–306 (1982)
45. R. Hamilton, Four-manifolds with positive curvature operator. J. Differ. Geom. **24**, 153–179 (1986)
46. R. Hamilton, A compactness property for solutions of the Ricci flow. Am. J. Math. **117**, 545–572 (1995)

47. T. Hasanis, A. Savas-Halilaj, T. Vlachos, Minimal graphs in $\mathbb{R}^4$ with bounded Jacobians. Proc. Am. Math. Soc. **137**, 3463–3471 (2009)
48. T. Hasanis, A. Savas-Halilaj, T. Vlachos, On the Jacobian of minimal graphs in $\mathbb{R}^4$. Bull. Lond. Math. Soc. **43**, 321–327 (2011)
49. A. Hatcher, A proof of the Smale conjecture, $\mathrm{Diff}(\mathbb{S}^3) \simeq \mathbb{O}(4)$. Ann. Math. **117**, 553–607 (1983)
50. S. Hildebrandt, J. Jost, K.O. Widman, Harmonic mappings and minimal submanifolds. Invent. Math. **62**, 269–298 (1980)
51. E. Hopf, On S. Bernstein's theorem on surfaces $z(x, y)$ of nonpositive curvature. Proc. Am. Math. Soc. **1**, 80–85 (1950)
52. G. Huisken, Flow by mean curvature of convex surfaces into spheres. J. Differ. Geom. **20**, 237–266 (1984)
53. G. Huisken, Asymptotic behavior for singularities of the mean curvature flow. J. Differ. Geom. **31**, 285–299 (1990)
54. J. Jost, *Compact Riemann Surfaces: An Introduction to Contemporary Mathematics* (Springer, Berlin, 2006)
55. J. Jost, *Riemannian Geometry and Geometric Analysis* Universitext edn., vol. 7 (Springer, Cham, 2017)
56. J. Jost, Y.-L. Xin, L. Yang, The Geometry of Grassmannian manifolds and Bernstein type theorems for higher codimension. Ann. Sci. Norm. Super. Pisa Cl. Sci. **16**, 1–39 (2016)
57. J. Kazdan, *Applications of Partial Differential Equations to Problems in Differential Geometry* (1983). http://www.math.upenn.edu/~kazdan/japan/japan.pdf
58. H.-B. Lawson, R. Osserman, Non-existence, non-uniqueness and irregularity of solutions to the minimal surface system. Acta Math. **139**, 1–17 (1977)
59. J.M. Lee, Riemannian manifolds, in *Graduate Texts in Mathematical*, vol. 176 (Springer, Berlin, 1997)
60. K.-W. Lee, Y.-I. Lee, Mean curvature flow of the graphs of maps between compact manifolds. Trans. Am. Math. Soc. **363**, 5745–5759 (2011)
61. J. Li, L. Yang, Symplectic mean curvature flows in Kähler surfaces with positive holomorphic sectional curvatures. Geom. Dedicata **170**, 63–69 (2014)
62. F. Lubbe, Curvature estimates for graphical mean curvature flow in higher codimension, in *Technische Informationsbibliothek und Universitätsbibliothek Hannover (TIB), Hannover* (2015)
63. C. Mantegazza, *Lecture Notes on Mean Curvature Flow* (Birkhäuser/Springer, Basel, 2011)
64. I. Medoš, M.-T. Wang, Deforming symplectomorphisms of complex projective spaces by the mean curvature flow. J. Differ. Geom. **87**, 309–341 (2011)
65. E.J. Mickle, A remark on a theorem of Serge Bernstein. Proc. Am. Math. Soc. **1**, 86–89 (1950)
66. J. Morgan, G. Tian, Ricci flow and the Poincaré conjecture, in *Clay Mathematics Monographs*, vol. 3 (American Mathematical Society/Clay Mathematics Institute, Providence/Cambridge, 2007)
67. J. Moser, On Harnack's theorem for elliptic differential equations. Comm. Pure Appl. Math. **14**, 577–591 (1961)
68. L. Ni, A Bernstein type theorem for minimal volume preserving maps. Proc. Am. Math. Soc. **130**, 1207–1210 (2002)
69. R. Osserman, Global properties of minimal surfaces in E^3 and E^n. Ann. Math. (2) **80**, 340–364 (1964)
70. R. Osserman, *A Survey of Minimal Surfaces* (Van Nostrand-Reinhold, New York, 1969)
71. R. Osserman, The minimal surface equation, in *Seminar on Nonlinear Partial Differential Equations, Berkeley, California* (1983), pp. 237–259
72. M.H. Protter, H.F. Weinberger, *Maximum Principles in Differential Equations* (Springer, New York, 1984)
73. K. Reidemeister, Uber die singulären Randpunkte eines konvexen Körpers. Math. Ann. **83**, 116–118 (1921)

74. E.A. Ruh, J. Vilms, The tension field of the Gauss map. Trans. Am. Math. Soc. **149**, 569–573 (1970)
75. A. Savas-Halilaj, K. Smoczyk, Homotopy of area decreasing maps by mean curvature flow. Adv. Math. **255**, 455–473 (2014)
76. A. Savas-Halilaj, K. Smoczyk, Bernstein theorems for length and area decreasing minimal maps. Calc. Var. Partial Differ. Equations **50**, 549–577 (2014)
77. A. Savas-Halilaj, K. Smoczyk, Evolution of contractions by mean curvature flow. Math. Ann. **361**, 725–740 (2015)
78. A. Savas-Halilaj, K. Smoczyk, Mean curvature flow of area decreasing maps between Riemann surfaces. Ann. Global Anal. Geom. **53**, 11–37 (2018)
79. R. Schoen, *The role of harmonic mappings in rigidity and deformation problems*. Lecture Notes in Pure and Application Mathematical, vol. 143 (Dekker, New York, 1993), pp. 179–200
80. J. Simons, Minimal varieties in Riemannian manifolds. Ann. Math. (2) **88**, 62–105 (1968)
81. L. Simon, A Hölder estimate for quasiconformal maps between surfaces in euclidean space. Acta Math. **139**, 19–51 (1977)
82. L. Simon, Asymptotics for a class of nonlinear evolution equations, with applications to geometric problems. Ann. Math. (2) **118**, 525–571 (1983)
83. D. Singley, Smoothness theorems for the principal curvatures and principal vectors of a hypersurface. Rocky Mountain J. Math. **5**, 135–144 (1975)
84. S. Smale, Diffeomorphisms of the 2-sphere. Proc. Am. Math. Soc. **10**, 621–626 (1959)
85. K. Smoczyk, Angle theorems for the Lagrangian mean curvature flow. Math. Z. **240**, 849–883 (2002)
86. K. Smoczyk, Mean curvature flow in higher codimension: introduction and survey, in *Global Differential Geometry, Springer Proceedings of the Mathematical*, vol. 17 (Springer, Heidelberg, 2012), pp. 231–274
87. K. Smoczyk, M.-T. Wang, Mean curvature flows of Lagrangians submanifolds with convex potentials. J. Differ. Geom. **62**, 243–257 (2002)
88. K. Smoczyk, M.-T. Wang, Generalized Lagrangian mean curvature flows in symplectic manifolds. Asian J. Math. **15**, 129–140 (2011)
89. K. Smoczyk, M.-P. Tsui, M.-T. Wang, Curvature decay estimates for mean curvature flow in higher codimensions. Trans. Am. Math. Soc. **368**, 7763–7775 (2016)
90. K. Smoczyk, M.-P. Tsui, M.-T. Wang, Generalized Lagrangian mean curvature flows: the cotangent bundle case. J. Reine Angew. Math. **750**, 97–121 (2019)
91. M.-P. Tsui, M.-T. Wang, Mean curvature flows and isotopy of maps between spheres. Comm. Pure Appl. Math. **57**, 1110–1126 (2004)
92. A. Vogiatzi, *Mean Curvature Flow and Isotopy of Problems*, M.S. Thesis (University of Ioannina, Ioannina, 2020), pp. 1–91
93. X. Wang, A remark on strong maximum principle for parabolic and elliptic systems. Proc. Am. Math. Soc. **109**, 343–348 (1990)
94. M.-T. Wang, Mean curvature flow of surfaces in Einstein four-manifolds. J. Differ. Geom. **57**, 301–338 (2001)
95. M.-T. Wang, Deforming area preserving diffeomorphism of surfaces by mean curvature flow. Math. Res. Lett. **8**, 651–661 (2001)
96. M.-T. Wang, Long-time existence and convergence of graphic mean curvature flow in arbitrary codimension. Invent. Math. **148**, 525–543 (2002)
97. M.-T. Wang, On graphic Bernstein type results in higher codimension. Trans. Am. Math. Soc. **355**, 265–271 (2003)
98. M.-T. Wang, Subsets of Grassmannians preserved by mean curvature flow. Comm. Anal. Geom. **13**, 981–998 (2005)
99. M.-T. Wang, *Lectures on mean curvature flows in higher codimensions*. Handbook of Geometric Analysis, vol. 1 (2008), pp. 525–543
100. M.-T. Wang, Mean curvature flows and isotopy problems, in *Surveys in Differential Geometry. Geometry and Topology* (International Press, Somerville, 2013), pp. 227–235

101. H.F. Weinberger, Invariant sets for weakly coupled parabolic and elliptic systems. Rend. Mat. (6) **8**, 295–310 (1975)
102. Y.-L. Xin, Geometry of harmonic maps, in *Progress in Nonlinear Differential Equations and their Applications*, vol. 23 (Birkhäuser Boston, Inc., Boston, 1996)

Critical Point Theory in Infinite Dimensional Spaces Using the Leray–Schauder Index

Martin Schechter

Abstract Many problems arising in science and engineering call for the solving of the Euler–Lagrange equations of functionals. Thus, solving the Euler–Lagrange equations is tantamount to finding critical points of the corresponding functional. An idea that has been very successful is to find appropriate sets that sandwich the functional. This means that the functional is bounded from above on one of the sets and bounded from below on the other. Two sets of the space are said to form a sandwich if they produce a critical sequence whenever they sandwich a functional. If the critical sequence has a convergent subsequence, then that produces a critical point. Finding sets that sandwich a functional is quite easy, but determining whether or not the sets form a sandwich is quite another story. It appears that the only way we can check to see if two sets form a sandwich is to require that one of them be contained in a finite-dimensional subspace. The reason is that in order to verify the definition, we need to invoke the Brouwer fixed point theorem. Our aim is to find a counterpart that holds true when both sets are infinite dimensional. We adjust our definitions to accommodate infinite dimensions. These definitions reduce to the usual when one set is finite dimensional. In order to prove the corresponding theorems, we make adjustments to the topology of the space and introduce infinite dimensional splitting. This allows us to use a form of compactness on infinite dimensional subspaces that does not exist under the usual topology. We lose the Brouwer index, but we are able to replace it with the Leray–Schauder index. We carry out the details in Sections 5, 7, and 8. In Section 6 we solve a system of equations which require infinite dimensional splitting.

AMS Subject Classification Primary 35J35, 47J30, 49J35, 49J40, 58K05

M. Schechter (✉)
Department of Mathematics, University of California, Irvine, CA, USA
e-mail: mschecht@math.uci.edu

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_21

1 Introduction

Many problems arising in science and engineering call for the solving of the Euler-Lagrange equations of functionals, i.e., equations equivalent to

$$G'(u) = 0, \tag{1}$$

where $G(u)$ is a C^1-functional (usually representing the energy) arising from the given data. By this we mean that functions are solutions of the Euler–Lagrange equations of G iff they satisfy (1). Solutions of (1) are called **critical points** of G. Thus, solving the Euler–Lagrange equations is tantamount to finding critical points of the corresponding functional.

The variational approach to solving differential equations and systems has its roots in the calculus of variations. The original problem was to minimize or maximize a given functional. The approach was to obtain the Euler–Lagrange equations of the functional, solve them, and show that the solutions provided the required minimum or maximum. This worked well for one-dimensional problems. However, when it came to higher dimensions, it was recognized quite early that it was more difficult to solve the Euler–Lagrange equations than it was to find minima or maxima of the corresponding functional. Consequently, the approach was abandoned for many years.

Eventually, when nonlinear partial differential equations and systems arose in applications and people were searching for solutions, they began to check if the equations and systems were the Euler-Lagrange equations of functionals. If so, a natural approach would be to find critical points of the corresponding functionals. The problem is that there is no uniform way of finding them.

The initial approach to finding critical points is to look for maxima or minima. Global extrema are the easiest to find, but they can exist only if the functional is semi bounded. For instance, if the continuously differential functional G is bounded from below, then we can find a minimizing sequence $\{u_k\}$ such that

$$G(u_k) \to a = \inf G > -\infty. \tag{2}$$

If this series converges or has a convergent subsequence, we have a minimum.

However, if the functional G is bounded from below, it can be shown that there is a sequence satisfying

$$G(u_k) \to a, \quad G'(u_k) \to 0. \tag{3}$$

If the sequence has a convergent subsequence, this will produce a minimum. The gain is that a sequence satisfying (3) has a better chance of having a convergent subsequence than a sequence satisfying only (2).

When the functional is not semibounded, there is no clear way of obtaining critical points. In general, one would like to determine when a functional has a

critical sequence, i.e., a sequence satisfying

$$G(u_k) \to a, \quad G'(u_k) \to 0. \tag{4}$$

This would give one the same advantages that one has in the case of semi-bounded functionals.

2 Sandwich Sets

An idea that has been very successful is to find appropriate sets that **sandwich the functional.** By this we mean the following: Two sets A, B **sandwich** the functional $G(u)$ if $G(u)$ is bounded from below on one of them and bounded from above on the other, e.g., if

$$a_0 := \sup_A G < \infty, \quad b_0 := \inf_B G > -\infty. \tag{5}$$

We would like to find sets A and B such that (5) will imply

$$\exists u : \ G(u) \geq b_0, \ G'(u) = 0. \tag{6}$$

This is too much to expect, since even semi-boundedness alone does not imply the existence of a critical point. Consequently, we weaken our requirements and look for sets A, B such that (5) implies the existence of a critical sequence (4) with $a \geq b_0$. This leads to

Definition 1 We shall say that the set A **forms a sandwich** with the set B if (5) implies (4) with $a \geq b_0$ for every C^1-functional $G(u)$.

Of course, (4) is a far cry from (6), but if, e.g., the sequence (4) has a convergent subsequence, then (4) implies (6). Whether or not this is true depends on the functional $G(u)$.

3 The Finite Dimensional Case

It appears that the only way we can check to see if two sets form a sandwich, is to require that one of them is contained in a finite-dimensional subspace. The reason is that in order to verify the definition, we need to invoke the Brouwer fixed point theorem.

The following three results hold when a subspace N is finite dimensional.

Theorem 2 *Let N be a finite dimensional subspace of a Banach space E, and for each $R > R_0$, let $\Omega_R(p)$ be an open bounded set in N containing a point p such*

that $d(\partial\Omega_R(p), p) \to \infty$ as $R \to \infty$. Let F be a continuous map of E onto N such that $F = I$ on N. Assume also that

$$d(A_R, F^{-1}(p)) \to \infty, \quad R \to \infty,$$

where $A_R = \partial\Omega_R(p)$. Let G be a C^1-functional on E such that

$$-\infty < b_0 = \inf_B G, \quad \sup_{A^R} G \le a_0 < \infty, \tag{7}$$

for $R > R_0$, where $A^R = N\backslash\Omega_R(p)$ and $B = F^{-1}(p)$. Then for each $\rho \in \mathcal{Q}$ there is a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c, \ b_0 \le c \le a_0, \quad \rho(d(u_k, B))\|G'(u_k)\| \to 0. \tag{8}$$

Corollary 3 *Let N be a finite dimensional subspace of a Hilbert space E and let $M = N^\perp$. For $G \in C^1(E, \mathbb{R})$, assume*

$$a_0 = \sup_N G < \infty, \quad b_0 = \inf_M G > -\infty. \tag{9}$$

Then for each $\rho \in \mathcal{Q}$ there is a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c, \ b_0 \le c \le a_0, \quad \rho(d(u_k, M))\|G'(u_k)\| \to 0. \tag{10}$$

Theorem 4 *Let N be a finite dimensional subspace of a Hilbert space E with complement $M \oplus \{v_0\}$, where v_0 is an element in E having unit norm, and let δ be any positive number. Let $\varphi(t) \in C^1(\mathbb{R})$ be such that*

$$0 \le \varphi(t) \le 1, \ \varphi(0) = 1,$$

and

$$\varphi(t) = 0, \quad |t| \ge 1.$$

Let

$$F(v+w+sv_0) = v+[s+\delta-\delta\varphi(\|w\|^2/\delta^2)]v_0, \quad w \in M, \ v \in N, \ s \in \mathbb{R}. \tag{11}$$

Let G be a C^1-functional on E such that (5) holds with $A = [N \oplus \{v_0\}]\backslash\mathcal{B}_{R_0}$ and $B = F^{-1}(\delta v_0) = \{w + rv_0 : w \in M, \ r = \delta\varphi(\|w\|^2/\delta^2)\}$. Then for each $\rho \in \mathcal{Q}$ there is a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c, \ b_0 \le c \le a_0, \quad \rho(d(u_k, B))\|G'(u_k)\| \to 0. \tag{12}$$

However, there are many applications for which we would like to obtain critical points in infinite-dimensional situations. It is not obvious how to proceed. It is not clear that we can obtain similar results for such situations. We now describe one method that works in the infinite-dimensional case. It involves adjusting the topology of the underlying space. In order to deal with the infinite dimensional situation, we are forced to make several adjustments. We must restrict the functional G and the mapping F and use a more general index for verification. The restrictions that we have chosen do not apply when one of the subspaces is finite dimensional. While we could not use the Brouwer index, we were able to use the Leray–Schauder index in its place. Thus the new theorems that we prove reverts to the old theorems when one of the subsets is finite dimensional. For this purpose we use flows described in the next section.

4 Flows

Let $\mathcal{Q}$ be a set of positive functions $\rho(t)$ on $[0, \infty)$, which are

(a) locally Lipschitz continuous,
(b) nondecreasing
(c) satisfy

$$\int_0^\infty \frac{dt}{\rho(t)} = \infty. \tag{13}$$

Moreover, $\mathcal{Q}$ is to satisfy

$$\rho_1, \rho_2 \in \mathcal{Q} \Longrightarrow \max(\rho_1, \rho_2) \in \mathcal{Q},$$

and contain functions of the form

$$(1 + |t|)^\beta, \quad \beta \in \mathbb{R}.$$

Let $Q \neq \phi$ be a subset of a Banach space E, and let Σ_Q be the set of all continuous maps $\sigma = \sigma(t)$ from $E \times [0, 1]$ to E such that

1. $\sigma(0)$ is the identity map,
2. for each $t \in [0, 1]$, $\sigma(t)$ is a homeomorphism of E onto E,
3. $\sigma'(t)$ is piecewise continuous and satisfies

$$\|\sigma'(t)u\| \leq C\rho(d(\sigma(t)u, Q)), \quad u \in E, \tag{14}$$

for some $\rho \in \mathcal{Q}$. If $Q = \{0\}$, we write $\Sigma = \Sigma_Q$. The mappings in Σ_Q are called *flows*. We note the following.

Remark 5 If $\sigma_1,\ \sigma_2$ are in Σ_Q, define $\sigma_3 = \sigma_1 \circ \sigma_2$ by

$$\sigma_3(s) = \begin{cases} \sigma_1(2s), & 0 \le s \le \frac{1}{2}, \\ \sigma_2(2s-1)\sigma_1(1), & \frac{1}{2} < s \le 1. \end{cases}$$

Then $\sigma_3 \in \Sigma_Q$, and $\sigma_3(1) = \sigma_2(1)\sigma_1(1)$.

Proof The first two properties are obvious. To check the third, note that

$$\sigma_3'(s) = \begin{cases} 2\sigma_1'(2s), & 0 \le s \le (\frac{1}{2})_-, \\ 2\sigma_2'(2s-1)\sigma_1(1), & (\frac{1}{2})_+ \le s \le 1. \end{cases}$$

Thus, if

$$\|\sigma_i'(t)u\| \le C_i \rho_i(d(\sigma_i(t)u, Q)), \quad u \in E,\ i = 1, 2, \tag{15}$$

then

$$\|\sigma_3'(s)u\| \le \begin{cases} 2\|\sigma_1'(2s)u\|, & 0 \le s \le (\frac{1}{2})_-, \\ 2\|\sigma_2'(2s-1)\sigma_1(1)u\|, & (\frac{1}{2})_+ \le s \le 1, \end{cases}$$

or

$$\|\sigma_3'(s)u\| \le \begin{cases} 2C_1\rho(d(\sigma_3(s)u, Q)), & 0 \le s \le (\frac{1}{2})_-, \\ 2C_2\rho(d(\sigma_3(s)u, Q)), & (\frac{1}{2})_+ \le s \le 1, \end{cases}$$

where $\rho = \max(\rho_1, \rho_2)$. We can now take $C_3 = 2\max(C_1, C_2)$. □

The following theorem can be found in [21, 22].

Theorem 6 *Let $g(t, x)$ be a continuous map from $\mathbb{R} \times X$ to X, where X is a Banach space. Assume that for each point $(\hat{t}, \hat{x}) \in \mathbb{R} \times X$, there are constants $K, b > 0$ such that*

$$\|g(t, x) - g(t, y)\| \le K\|x - y\|, \quad |t - \hat{t}| < b,\ \|x - \hat{x}\| < b,\ \|y - \hat{x}\| < b, \tag{16}$$

and

$$\|g(t, x)\| \le \rho(d(x, Q)), \quad x \in X,\ t \in [t_0, \infty), \tag{17}$$

where Q is a subset of X, with ρ nondecreasing or bounded. Assume

$$\int_0^{u_0} \frac{d\tau}{\rho(\tau)} = \int_{u_0}^{\infty} \frac{d\tau}{\rho(\tau)} = \infty \tag{18}$$

where $u_0 > 0$. Then for each $x_0 \in X$ and $t_0 \in \mathbb{R}$ there is a unique solution $x(t)$ of the equation

$$\frac{dx(t)}{dt} = g(t, x(t)), \quad t \in [t_0, \infty), \quad x(t_0) = x_0. \tag{19}$$

Moreover, $x(t)$ depends continuously on x_0 and satisfies

$$u_1(t) \leq d(x(t), Q) \leq u_2(t), \quad t \in [t_0, \infty), \tag{20}$$

where $u_1(t)$ is the solution of

$$u'(t) = -\rho(u(t)) \tag{21}$$

in $[t_0, \infty)$ satisfying $u(t_0) = u_0 = d(x_0, Q)$, and $u_2(t)$ is the solution of

$$u'(t) = \rho(u(t)) \tag{22}$$

in $[t_0, \infty)$ satisfying $u(t_0) = u_0 = d(x_0, Q)$.

5 The Infinite Dimensional Case

We now describe a sandwich theory that works in the infinite-dimensional case.

Let N be a closed, separable subspace of a Hilbert space E. We can define a new norm $|v|_w$ satisfying $|v|_w \leq \|v\|$ $\forall v \in E$ and such that the topology induced by this norm is equivalent to the weak topology of N on bounded subsets of N. We construct the norm so that $v_j \to v$ weakly in N implies $|v_j - v|_w \to 0$. Conversely, if $\|v_j\|, \|v\| \leq C$ for all $j > 0$ and $|v_j - v|_w \to 0$, then $v_j \to v$ weakly in N.

We adjust our assumptions on G and F for the infinite dimensional case of $\dim N = \infty$, but they reduce to the same assumptions that are made in the finite dimensional case $\dim N < \infty$.

Our requirements on G are given by

Definition 7 Let N be a closed separable subspace of a Hilbert space E. A functional G(u) on E will be called an N-weak-to-weak continuously differentiable functional on E if

$$|v_n - v|_w \to 0 \tag{23}$$

implies that there is a renamed subsequence satisfying

$$G(v_n) \to G(v), \ |G'(v_n) - G'(v)|_w \to 0. \tag{24}$$

This means that G is a continuous functional on E_w, continuously differentiable on E and such that

$$v_n = Pu_n \to v \text{ weakly in } E, \quad w_n = (I - P)u_n \to w \text{ strongly in } E \tag{25}$$

implies that there is a renamed subsequence satisfying

$$G'(v_n + w_n) \to G'(v + w) \text{weakly in } E, \tag{26}$$

where P is the projection of E onto N.

Note that every C' functional is N-weak-to-weak continuously differentiable when $\dim N < \infty$.

Concerning the mapping F we define

Definition 8 Let N be a closed separable subspace of a Hilbert space E. We shall call a map F of E onto N an N-weakly continuous mapping if F is a $|\cdot|_w$-continuous map from E onto N satisfying

- $F_N = I$; F maps any finite dimensional subspace of N containing p into itself; it maps bounded sets into bounded sets;
- There exists a fixed finite-dimensional subspace E_0 of E such that $F(u - v) - (F(u) - F(v)) \in E_0, \ \forall\, u, \ v \in E$;
- F maps finite-dimensional subspaces of E to finite-dimensional subspaces of E;

Note that every continuous map F of E onto N satisfying $F_N = I$ is N-weakly continuous when N is finite dimensional.

We have

Theorem 9 *Let N be a closed separable subspace of a Hilbert space E. For each $R > R_0$, let $\Omega_R(p)$ be an open, convex, bounded set in N containing a point p such that $d(\partial\Omega_R(p), p) \to \infty$ as $R \to \infty$. Let F be a N-weakly continuous mapping of E onto N such that F = I on N. Assume also that*

$$d(A_R, F^{-1}(p)) \to \infty, \quad R \to \infty,$$

where $A_R = \partial\Omega_R(p)$. Let G be a N-weak-to-weak continuously differentiable functional on E such that

$$-\infty < b_0 = \inf_B G, \quad \sup_{A^R} G \le a_0 < \infty, \tag{27}$$

for $R > R_0$, where $A^R = N \backslash \Omega_R(p)$ and $B = F^{-1}(p)$. Then for each $\rho \in \mathcal{Q}$ there is a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c, \ b_0 \le c \le a_0, \quad \rho(d_w(u_k, B))\|G'(u_k)\| \to 0. \tag{28}$$

Note that when dim $N < \infty$, Theorem 9 reduces to Theorem 2.

Corollary 10 *Let N be a closed separable subspace of a Hilbert space E, and let $M = N^{\perp}$. For G a N-weak-to-weak continuously differentiable functional on E assume*

$$a_0 = \sup_N G < \infty, \quad b_0 = \inf_M G > -\infty. \tag{29}$$

Then for each $\rho \in \mathcal{Q}$ there is a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c, \ b_0 \le c \le a_0, \quad \rho(d_w(u_k, M))\|G'(u_k)\| \to 0. \tag{30}$$

Note that when dim $N < \infty$, Corollary 10 reduces to Corollary 3.

Corollary 11 *Let N be a closed, separable subspace of a Banach space E, and for each $R > R_0$ let $\Omega_R(p)$ be an open, convex, bounded set in N containing a point p. Let G be a an N-weak-to-weak continuously differentiable functional on E, and let F be an N-weakly continuous mapping. Assume*

$$d_R = d_w(A^R, F^{-1}(p)) \to \infty, \quad R \to \infty,$$

where $A^R = N \backslash \Omega_R(p)$. Assume

$$-\infty < b_0 = \inf_B G, \quad \sup_{A^R} G \le a_0 < \infty, \tag{31}$$

for $R > R_0$. Then for each sequence $\nu_k \ge 2d_k - \varepsilon$ there is a $\beta > 0$ and a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c, \ b_0 \le c \le b_1, \quad (\nu_k + |u_k|_w)\|G'(u_k)\| \le \beta. \tag{32}$$

Theorem 12 *Let N be a closed separable subspace of a Hilbert space E with complement $M \oplus \{v_0\}$, where v_0 is an element in E having unit norm, and let δ be any positive number. Let $\varphi(t) \in C^1(\mathbb{R})$ be such that*

$$0 \le \varphi(t) \le 1, \ \varphi(0) = 1,$$

and

$$\varphi(t) = 0, \quad |t| \ge 1.$$

Let

$$F(v+w+sv_0) = v + [s + \delta - \delta\varphi(\|w\|^2/\delta^2)]v_0, \quad w \in M, \ v \in N, \ s \in \mathbb{R}. \tag{33}$$

Let G be a N-weak-to-weak continuously differentiable functional on E such that (5) holds with $A = [N \oplus \{v_0\}]\backslash\mathcal{B}_{R_0}$ and $B = F^{-1}(\delta v_0) = \{w + rv_0 : w \in M,\ r = \delta\varphi(\|w\|^2/\delta^2)\}$. Then for each $\rho \in \mathcal{Q}$ there is a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c,\ b_0 \le c \le a_0, \quad \rho(d_w(u_k, B))\|G'(u_k)\| \to 0. \tag{34}$$

Note that when $\dim N < \infty$, Theorem 12 reduces to Theorem 4.

Corollary 13 *Let N be a closed, separable subspace of a Hilbert space E with orthogonal complement $M \oplus \{v_0\}$, where v_0 is an element in E having unit norm and orthogonal to both M and N, and let $\delta < R$ be positive numbers. Let $\varphi(t) \in C^1(\mathbb{R})$ be such that*

$$0 \le \varphi(t) \le 1,\ \varphi(0) = 1,$$

and

$$\varphi(t) = 0, \quad |t| \ge 1.$$

Let

$$F(v+w+sv_0) = v+[s+\delta-\delta\varphi(\|w\|^2/\delta^2)]v_0, \quad w \in M,\ v \in N,\ s \in \mathbb{R}. \tag{35}$$

Let G be a an N-weak-to-weak continuously differentiable functional on E. Assume

$$-\infty < b_0 = \inf_B G, \quad \sup_A G = b_1 < \infty, \tag{36}$$

holds with $A = [N \oplus \{v_0\}]$, $A_R = A \cap \partial\mathcal{B}_R$ and $B = F^{-1}(\delta v_0) = \{w + rv_0 : w \in M,\ r = \delta\varphi(\|w\|^2/\delta^2)\}$. Then for each sequence $v_k \to \infty$ there is a $\beta > 0$ and a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c,\ b_0 \le c \le b_1, \quad (v_k + |u_k|_w)\|G'(u_k)\| \le \beta. \tag{37}$$

6 Applications

Let $\mathcal{A}, \mathcal{B}$ be positive, self-adjoint operators on $L^2(\Omega)$ with compact resolvents, where $\Omega \subset \mathbb{R}^n$ (Ω may be unbounded). Let $F(x, v, w)$ be a Caratheódory function on $\Omega \times \mathbb{R}^2$ such that

$$f(x, v, w) = \partial F/\partial v, \quad g(x, v, w) = \partial F/\partial w \tag{38}$$

are also Caratheódory functions satisfying

$$|f(x, v, w)| + |g(x, v, w)| \leq C_0(|v| + |w| + 1), \quad v, w \in \mathbb{R}. \tag{39}$$

We wish to solve the system

$$\mathcal{A}v = -f(x, v, w) \tag{40}$$

$$\mathcal{B}w = g(x, v, w). \tag{41}$$

The reason for the minus sign is because it leads to a variational problem involving infinite dimensional subspaces. The plus sign leads to a minimization problem.

Let $\lambda_0(\mu_0)$ be the lowest eigenvalue of $\mathcal{A}(\mathcal{B})$.

Theorem 14 *Assume*

$$2F(x, s, 0) \geq -\lambda_0 s^2 - W_1(x), \quad x \in \Omega, s \in \mathbb{R}, \tag{42}$$

where $W_1(x) \in L^1(\Omega)$. *In addition, assume that the eigenfunctions of* λ_0 *and* μ_0 *are bounded and* $\neq 0$ *a.e. in* Ω, *and there is a* $q > 2$ *such that*

$$\|w\|_q^2 \leq Cb(w), \quad w \in M. \tag{43}$$

Assume

$$2F(x, 0, t) \leq \mu(x)t^2, \quad x \in \Omega, \quad t \in \mathbb{R} \tag{44}$$

where

$$\mu(x) \leq\not\equiv \mu_0, \quad x \in \Omega, \tag{45}$$

and for some $\delta > 0$,

$$2F(x, s, t) \leq \mu_0 t^2 - \lambda_0 s^2, \quad |t| + |s| \leq \delta. \tag{46}$$

Also

$$H(x, s, t) \leq W(x) \tag{47}$$

and

$$H(x, s, t) \to -\infty \; as \; |s| + |t| \to \infty, \tag{48}$$

where $W(x) \in L^2(\mathbb{R}^n)$ *and*

$$H(x, s, t) = f(x, s, t)s + g(x, s, t)t - 2F(x, s, t). \tag{49}$$

Then the system (40) (41) has a nontrivial solution.

Proof Let $D = D(\mathcal{A}^{1/2}) \times D(\mathcal{B}^{1/2})$. Then D becomes a Hilbert space with norm given by

$$\|u\|_D^2 = (\mathcal{A}v, v) + (\mathcal{B}w, w), \quad u = (v, w) \in D. \tag{50}$$

We define

$$G(u) = b(w) - a(v) - 2\int_\Omega F(x, v, w)dx, \quad u \in D \tag{51}$$

where

$$a(v) = (\mathcal{A}v, v), \quad b(w) = (\mathcal{B}w, w). \tag{52}$$

Then $G \in C^1(D, \mathbb{R})$ and

$$(G'(u), h)/2 = b(w, h_2) - a(v, h_1) - (f(u), h_1) - (g(u), h_2), \tag{53}$$

where we write $f(u), g(u)$ in place of $f(x, v, w), g(x, v, w)$, respectively. It is readily seen that the system (40), (41) is equivalent to

$$G'(u) = 0. \tag{54}$$

We let N be the set of those $(v, 0) \in D$ and M the set of those $(0, w) \in D$. Then M, N are orthogonal closed subspaces such that

$$D = M \oplus N. \tag{55}$$

If we define

$$Lu = 2(-v, w), \quad u = (v, w) \in D \tag{56}$$

then L is a selfadjoint bounded operator on D. Also

$$G'(u) = Lu + c_0(u) \tag{57}$$

where

$$c_0(u) = -(\mathcal{A}^{-1/2} f(u), \mathcal{B}^{-1/2} g(u)) \tag{58}$$

is compact on D. This follows from (39) and the fact that $\mathcal{A}$ and $\mathcal{B}$ have compact resolvents. It also follows that G' has $D - weak - to - weak$ continuity. For if $u_k \to u$ weakly, then $Lu_k \to Lu$ weakly and $c_0(u_k)$ has a convergent subsequence.

Let N' be the orthogonal complement of $N_0 = \{\varphi_0\}$ in N, where φ_0 is the eigenfunction of $\mathcal{A}$ corresponding to λ_0. Then $N = N' \oplus N_0$. Let M_0 be the subspace of M spanned by the eigenfunctions of $\mathcal{B}$ corresponding to μ_0, and let M' be its orthogonal complement in M. Since N_0 and M_0 are contained in $L^\infty(\Omega)$, there is a positive constant ρ such that

$$a(y) \le \rho^2 \Rightarrow \|y\|_\infty \le \delta/4, \quad y \in N_0 \tag{59}$$

$$b(h) \le \rho^2 \Rightarrow \|h\|_\infty \le \delta/4, \quad h \in M_0 \tag{60}$$

where δ is the number given in (46). If

$$a(y) \le \rho^2, b(w) \le \rho^2, |y(x)| + |w(x)| \ge \delta \tag{61}$$

we write $w = h + w', h \in M_0, w' \in M'$ and

$$\delta \le |y(x)| + |w(x)| \le |y(x)| + |h(x)| + |w'(x)| \le (\delta/2) + |w'(x)|. \tag{62}$$

Thus

$$|y(x)| + |h(x)| \le \delta/2 \le |w'(x)| \tag{63}$$

and

$$|y(x)| + |w(x)| \le 2|w'(x)|. \tag{64}$$

Now by (46) and (64)

$$\begin{aligned} G(y, w) &= b(w) - a(y) - 2\int_\Omega F(x, y, w)\, dx \qquad (65)\\ &\ge b(w) - a(y) - \int\limits_{|y|+|w|<\delta} \{\mu_0 w^2 - \lambda_0 y^2\} dx \\ &\quad - c_0 \int\limits_{|y|+|w|>\delta} (|y| + |w| + 1)^2\, dx \\ &\ge b(w) - a(y) - \mu_0\|w\|^2 + \lambda_0\|y\|^2 - c_1 \int\limits_{2|w'|>\delta} |w'|^q dx \\ &\ge b(w') - \mu_0\|w'\|^2 - c_2 b(w')^{q/2} \\ &\ge \left(1 - \frac{\mu_0}{\mu_1} - c_2 b(w')^{(q/2)-1}\right) b(w'), \quad a(y) \le \rho^2,\ b(w) \le \rho^2 \end{aligned}$$

where μ_1 is the next eigenvalue of $\mathcal{B}$ after μ_0. If we reduce ρ accordingly, we can find a positive constant ν such that

$$G(y, w) \geq \nu b(w'), \quad a(y) \leq \rho^2, \ b(w) \leq \rho^2. \tag{66}$$

I claim that either (40) (41) has a nontrivial solution or there is an $\epsilon > 0$ such that

$$G(y, w) \geq \epsilon, \quad a(y) + b(w) = \rho^2. \tag{67}$$

For suppose (67) did not hold. Then there would be a sequence $\{y_k, w_k\}$ such that $a(y_k)+b(w_k) = \rho^2$ and $G(y_k, w_k) \to 0$. If we write $w_k = w_k' + h_k$, $w_k' \in M'$, $h_k \in M_0$, then (66) tells us that $b(w_k') \to 0$. Thus $a(y_k) + b(h_k) \to \rho^2$. Since N_0, M_0 are finite dimensional, there is a renamed subsequence such that $y_k \to y$ in N_0 and $h_k \to h$ in M_0. Thus $G(y, h) = 0$. By (59) and (60), $\|y\|_\infty \leq \delta/4$ and $\|h\|_\infty \leq \delta/4$. Consequently (46) implies

$$2F(x, y, h) \leq \mu_0 h^2 - \lambda_0 y^2. \tag{68}$$

Since

$$G(y, h) = b(h) - a(y) - 2\int_\Omega F(x, y, h)dx = 0, \tag{69}$$

we have

$$\int_\Omega \{2F(x, y, h) + \lambda_0 y^2 - \mu_0 h^2\}dx = 0. \tag{70}$$

In view of (68), this implies

$$2F(x, y, h) \equiv \mu_0 h^2 - \lambda_0 y^2. \tag{71}$$

For $\zeta \in C_0^\infty(\Omega)$ and $t > 0$ small we have

$$2[F(x, y + t\zeta, h) - F(x, y, h)]/t \leq -\lambda_0[(y + t\zeta)^2 - y^2]/t. \tag{72}$$

Taking $t \to 0$, we have

$$f(x, y, h)\zeta \leq -\lambda_0 y\zeta. \tag{73}$$

Since this is true for all $\zeta \in C_0^\infty(\Omega)$, we have

$$f(x, y, h) = -\lambda_0 y = -\mathcal{A}y. \tag{74}$$

Similarly,

$$2[F(x, y, h + t\zeta) - F(x, y, h)]/t \leq \mu_0[(h + t\zeta)^2 - h^2]/t \tag{75}$$

and consequently

$$g(x, y, h)\zeta \leq \mu_0 h\zeta \tag{76}$$

and

$$g(x, y, h) = \mu_0 h = \mathcal{B}h \tag{77}$$

We see from (74) and (77) that (40) (41) has a nontrivial solution. Thus, we may assume that (67) holds.

Next, we note that there is an $\varepsilon > 0$ depending on ρ such that

$$G(0, w) \geq \varepsilon, \quad b(w) \geq \rho > 0.$$

To see this, suppose that $\{w_k\} \subset M$ is a sequence such that

$$G(0, w_k) \to 0, \quad b(w_k) \geq \rho.$$

If

$$b_k = b(w_k) \leq C,$$

this implies

$$b(w_k) - \mu_0 \|w_k\|^2 \to 0$$

and

$$\int [\mu_0 - \mu(x)]w_k^2 dx \to 0,$$

since

$$G(0, w) \geq b(w) - \mu_0 \|w\|^2 + \int [\mu_0 - \mu(x)]w^2 dx, \quad w \in M.$$

If we write $w_k = w_k' + h_k, w_k' \in M', h_k \in M_0$ as before, then this tells us that $b(w_k') \to 0$. Since M_0 is finite dimensional, there is a renamed subsequence such that $h_k \to h$. But the two conclusions above tell us that $h = 0$. Since $b(h) \geq \rho$, we see that $\varepsilon > 0$ exists for any constant C. If the sequence $\{b_k\}$ is not bounded, we take $\tilde{w}_k = w_k / b_k^{1/2}$. Then

$$G(0, w_k)/b_k \geq b(\tilde{w}_k) - \mu_0 \|\tilde{w}_k\|^2 + \int [\mu_0 - \mu(x)](\tilde{w}_k)^2 dx.$$

Next we note that there is a $\nu > 0$ such that

$$G(0, w) \geq \nu b(w), \quad w \in M. \tag{78}$$

Assuming this for the moment, we see that

$$b_0 := \inf_B G \geq \varepsilon_1 > 0 \tag{79}$$

where

$$B = \{w \in M : b(w) \geq \rho^2\} \cup \{u = (s\varphi_0, w) : s \geq 0, w \in M, \|u\|_D = \rho\}, \tag{80}$$

and $\varepsilon_1 = \min\{\varepsilon, \nu\rho^2\}$. By (14) there is an $R > \rho$ such that

$$\sup_{A^R} G \leq a_0 < \infty, \tag{81}$$

where $A^R = N \backslash \mathcal{B}_R$. By Theorem 9 there is a sequence $\{u_k\} \subset D$ such that (34) holds with $c \geq \varepsilon_1$. To complete the proof, we show that the sequence u_k is bounded in D and has a convergent subsequence. To see this, assume that $r_k = \|u_k\|_D \to \infty$, and let $\tilde{u}_k = u_k/r_k$. Then $\|\tilde{u}_k\|_D = 1$, and there is a renamed subsequence such that $\tilde{u}_k \to \tilde{u}$, weakly in D, strongly in $L^2(\Omega)$, and a.e. in Ω. Since,

$$b(w_k) + a(v_k) = (g(u_k), w_k) - (f(u_k), v_k),$$

we have

$$b(\tilde{w}_k) + a(\tilde{v}_k) \leq C\|\tilde{u}_k\|^2.$$

Hence,

$$1 = \|\tilde{u}_k\|_D^2 \leq C\|\tilde{u}_k\|^2 \to C\|\tilde{u}\|^2.$$

Consequently, $\tilde{u} \not\equiv 0$.

Let Ω_0 be the subset of Ω on which $\tilde{u} \neq 0$. Then

$$|u_k(x)| = r_k|\tilde{u}_k(x)| \to \infty, \quad x \in \Omega_0. \tag{82}$$

If $\Omega_1 = \Omega \setminus \Omega_0$, then we have

$$\int_\Omega H(x,u_k)\,dx = \int_{\Omega_0} + \int_{\Omega_1} \le \int_{\Omega_0} H(x,u_k)\,dx + \int_{\Omega_1} W_1(x)dx \to -\infty. \tag{83}$$

But

$$\int_\Omega H(x,u_k)\,dx = G(u_k) - (G'(u_k),u_k)$$

and

$$|(G'(u_k),u_k)| \le (\nu_k + |u_k|_w)\|G'(u_k)\| \le \beta.$$

Thus

$$|\int_\Omega H(x,u_k)\,dx| \le K.$$

This contradicts (83), and we see that $r_k = \|u_k\|_D$ is bounded. Hence, there is a renamed subsequence converging weakly to a function $u \in D$. Since

$$(G'(u_k),h)/2 = b(w_k,h_2) - a(v_k,h_1) - (f(u_k),h_1) - (g(u_k),h_2) \to 0, \quad h \in D, \tag{84}$$

we have

$$(G'(u),h)/2 = b(w,h_2) - a(v,h_1) - (f(u),h_1) - (g(u),h_2) = 0, \quad h \in D. \tag{85}$$

This shows that (40) and (41) hold. Since $c \neq 0$ and $G(0) = 0$, we see that $u \neq 0$, and we have a nontrivial solution of the system (40) (41).

It therefore remains only to prove (78). Clearly $\nu \ge 0$. If $\nu = 0$, then there is a sequence $\{w_k\} \subset M$ such that

$$G(0,w_k) \to 0, \quad b(w_k) = 1. \tag{86}$$

Thus there is a renamed subsequence such that $w_k \to w$ weakly in M, strongly in $L^2(\Omega)$ and a.e. in Ω. Consequently

$$\int_\Omega [\mu_0 - \mu(x)]w_k^2\,dx \le 1 - \int_\Omega \mu(x)w_k^2\,dx \le G(0,w_k) \to 0 \tag{87}$$

and

$$1 = \int_\Omega \mu(x)w^2 dx \le \mu_0\|w\|^2 \le b(w) \le 1 \tag{88}$$

which means that we have equality throughout. It follows that we must have $w \in E(\mu_0)$, the eigenspace of μ_0. Since $w \not\equiv 0$, we have $w \neq 0$ a.e. But

$$\int_\Omega [\mu_0 - \mu(x)]w^2 dx = 0 \tag{89}$$

implies that the integrand vanishes identically on Ω, and consequently $\mu(x) \equiv \mu_0$, violating (45). This establishes (78) and completes the proof of the theorem. □

7 Infinite Dimensional Splitting

Let N be a closed, separable subspace of a Hilbert space E. We can define a new norm $|v|_w$ satisfying $|v|_w \le \|v\|$ $\forall v \in N$ and such that the topology induced by this norm is equivalent to the weak topology of N on bounded subsets of N. This can be done as follows: Let $\{e_k\}$ be an orthonormal basis for N. Define

$$(u, v)_w = \sum_{k=1}^{\infty} \frac{(u, e_k)(v, e_k)}{2^k}, \quad u, v \in N.$$

This is a scalar product. The corresponding norm squared is

$$|v|_w^2 = \sum_{k=1}^{\infty} \frac{|(v, e_k)|^2}{2^k}, \quad v \in N.$$

Then $|v|_w$ satisfies $|v|_w \le \|v\|$, $v \in N$. If $v_j \to v$ weakly in N, then there is a $C > 0$ such that

$$\|v_j\|, \|v\| \le C, \quad \forall j > 0.$$

For any $\varepsilon > 0$, there exist $K > 0, M > 0$, such that $1/2^K < \varepsilon^2/(8C^2)$ and $|(v_j - v, e_k)| < \varepsilon/2$ for $1 \le k \le K$, $j > M$. Therefore,

$$\begin{aligned} |v_j - v|_w^2 &= \sum_{k=1}^{\infty} \frac{|(v_j - v, e_k)|^2}{2^k} \\ &\le \sum_{k=1}^{K} \frac{\varepsilon^2/4}{2^k} + \sum_{k=K+1}^{\infty} \frac{4C^2}{2^k} \\ &\le \frac{\varepsilon^2}{4} \sum_{k=1}^{\infty} \frac{1}{2^k} + \frac{4C^2}{2^K} \sum_{k=1}^{\infty} \frac{1}{2^k} \\ &\le \frac{\varepsilon^2}{2} + \frac{\varepsilon^2}{2}. \end{aligned}$$

Therefore, $v_j \to v$ weakly in N implies $|v_j - v|_w \to 0$.

Conversely, let $\|v_j\|, \|v\| \le C$ for all $j > 0$ and $|v_j - v|_w \to 0$. Let $\varepsilon > 0$ be given. If $h = \sum_{k=1}^{\infty} \alpha_k e_k \in N$, take K so large that $\|h_K\| < \varepsilon/(4C)$, where $h_K = \sum_{k=K+1}^{\infty} \alpha_k e_k$. Take M so large that $|v_j - v|_w^2 < \varepsilon^2/(4\sum_{k=1}^{K} 2^k|\alpha_k|^2)$ for all $j > M$. Then

$$
\begin{aligned}
|(v_j - v, h - h_K)|^2 &= |\sum_{k=1}^{K} \alpha_k (v_j - v, e_k)|^2 \\
&\le \sum_{k=1}^{K} 2^k |\alpha_k|^2 \sum_{k=1}^{\infty} \frac{|(v_j - v, e_k)|^2}{2^k} \\
&< \varepsilon^2/4
\end{aligned}
$$

for $j > M$. Also, $|(v_j - v, h_K)| \le 2C\|h_K\| < \varepsilon/2$. Therefore,

$$
|(v_j - v, h)| < \varepsilon, \quad \forall j > M,
$$

that is, $v_j \to v$ weakly in N.

For $u = v + h,\ u_1 = v_1 + h_1 \in E = N \oplus N^{\perp}$ with $v, v_1 \in N, h, h_1 \in N^{\perp}$, we define the scalar product $(u, u_1)_w = (v, v_1)_w + (h, h_1)$. Thus, the corresponding norm satisfies $|u|_w \le \|u\|\ \forall u \in E$. Clearly, when $\dim N < \infty$, the norms $\|\cdot\|$ and $|\cdot|_w$ are equivalent.

We denote E equipped with this scalar product and norm by E_w. It is a scalar product space with the same elements as E. In particular, if $(u_n = v_n + w_n)$ is $\|\cdot\|$-bounded and $u_n \overset{|\cdot|_w}{\to} u$, then $v_n \rightharpoonup v$ weakly in N, $w_n \to w$ strongly in $N^{\perp}$, $u_n \rightharpoonup v + w$ weakly in E.

For $u \in E$ and $Q \subset E$, we define

$$
d_w(u, Q) = \inf_{v \in Q} |u - v|_w.
$$

Let L be a bounded, convex, closed subset of N. Then L is $|\cdot|_w$-compact. In fact, since L is bounded with respect to both norms $|\cdot|_w$ and $\|\cdot\|$, for any $v_n \in L$, there is a renamed subsequence such that $v_n \rightharpoonup v_0$ weakly in E. Then $v_0 \in L$ since L is convex, and on the bounded set L the $|\cdot|_w$-topology is equivalent to the weak topology. Thus, $v_n \overset{|\cdot|_w}{\to} v_0$ and L is $|\cdot|_w$-compact.

Let L be a compact subset of E_w. We define $\Sigma_w(L)$ to be the set of all $\sigma(t) \in \Sigma : [0, 1] \times E \mapsto E$ such that

1. $\sigma(t)$ is $|\cdot|_w$-continuous.
2. $\sigma(0)u = u,\ u \in E.$
3. There is a finite dimensional subspace E_f of E such that $\dim E_f > 0$ and $\sigma(t)u - u \in E_f,\ (t,u) \in I \times L.$

Here we use E_f to denote various finite-dimensional subspaces of E when exact dimensions are irrelevant. Note that $\Sigma_w(L)$ is not empty since $\sigma(t) \equiv 1$ is a member.

We let Σ_{wQ} denote the set of those $\sigma \in \Sigma_w$ which satisfy

$$|\sigma'(t)u|_w \le C\rho(d_w(\sigma(t)u, Q)), \quad u \in E, \tag{90}$$

where $Q \subset E$.

We have

Lemma 15 *If L is compact in E_w and $\sigma \in \Sigma_w(L)$, then*

$$\tilde{L} = \{\sigma(t)L : t \in I\}$$

is compact in E_w.

Proof Suppose $\{t_k\} \subset I,\ \{u_k\} \subset L$ are sequences. Then there are renamed subsequences such that

$$t_j \to t_0, \quad |u_k - u_0|_w \to 0.$$

Thus $I \times L$ is a compact subset of $I \times E_w$. By definition, there is a finite dimensional subspace E_f containing the set $\{\sigma(t)u - u,\ t \in I,\ u \in L\}$. Since this set is bounded, every sequence has a convergent subsequence. Since every sequence in L has a convergent subsequence, the same must be true of $\tilde{L}$. □

Lemma 16 *If $\sigma_1,\ \sigma_2 \in \Sigma_w(L)$, then $\sigma_3 = \sigma_1 \circ \sigma_2 \in \Sigma_w(L)$.*

Proof By the definition of $\Sigma_w(L)$, for any $(s_0, u_0) \in I \times L$, there is a $|\cdot|_w$-neighborhood $U_{(s_0,u_0)}$ such that $\{u - \sigma_1(t)u : (t,u) \in U_{(s_0,u_0)} \cap L\} \subset E_f$. Note that, $L \subset \bigcup_{(s,u)\in L} U_{(s,u)}$. Since L is $|\cdot|_w$-compact, $L \subset \bigcup_{i=1}^{j_0} U_{(s_i,u_i)}$ where $(s_i, u_i) \in L$. Consequently, $\{u - \sigma_1(t)u : (t,u) \in L\} \subset E_f$. The same is true of σ_2. Since

$$\sigma_3(s) = \begin{cases} \sigma_1(2s), & 0 \le s \le \frac{1}{2}, \\ \sigma_2(2s-1)\sigma_1(1), & \frac{1}{2} < s \le 1, \end{cases}$$

$u - \sigma_3(t)u \in E_f$ as well. □

8 Some Lemmas

Before giving the proof of Theorem 9, we prove a few lemmas.

Lemma 17 *Let N be a closed, separable subspace of a Banach space E, and let Ω be a bounded, convex, open subset of N containing a point p. Let F be an N-weakly continuous mapping. Assume*

$$\sigma(t)\partial\Omega \cap F^{-1}(p) = \phi, \quad 0 \le t \le 1,$$

for some $\sigma \in \Sigma_w(\overline{\Omega})$. Then

$$\sigma(t)\Omega \cap F^{-1}(p) \neq \phi, \quad 0 \le t \le 1.$$

Proof Assume that there is a $\sigma \in \Sigma_w(\overline{\Omega})$ such that

$$\sigma(t)\partial\Omega \cap F^{-1}(p) = \phi, \quad 0 \le t \le 1, \tag{91}$$

and

$$\sigma(t)\Omega \cap F^{-1}(p) = \phi, \quad 0 \le t \le 1,$$

or, equivalently,

$$F(\sigma(t)\Omega) \cap \{p\} = \phi, \quad 0 < t \le 1. \tag{92}$$

Let

$$\gamma(t)x = F\sigma(t)x, \quad (t, x) \in I \times \overline{\Omega}.$$

Then $\gamma(t) \in C(I \times \overline{\Omega}, E_w \cap N)$ and

$$\gamma(t)x \neq p, \quad x \in \partial\Omega,\ t \in [0, 1]. \tag{93}$$

Also

$$\gamma(0)x = Fx = x, \quad x \in \overline{\Omega}. \tag{94}$$

By hypothesis, there exists a fixed finite-dimensional subspace E_0 of E such that $F(u - v) - (F(u) - F(v)) \in E_0,\ \forall\ u,\ v \in E$. Take $u = \sigma(t)x,\ \ v = x$. Since $\overline{\Omega}$ is compact in E_w and $\sigma \in \Sigma_w(\overline{\Omega})$, there is a finite dimensional subspace E_1 of E such that $\dim E_1 > 0$ and $\sigma(t)u - u \in E_1,\ (t, u) \in I \times \overline{\Omega}$. Hence

$$\gamma(t)x = P_0(F\sigma(t)x - Fx - F[\sigma(t)x - x])$$

$$+ FP_1[\sigma(t)x - x] + x$$
$$= x - \varphi(t)x, \quad (t, x) \in I \times \overline{\Omega},$$

where $\varphi(t)x = -P_0(F\sigma(t)x - Fx - F[\sigma(t)x - x]) - FP_1[\sigma(t)x - x]$, and the P_0, P_1 are projections onto the finite dimensional subspaces E_0, E_1. Thus, $\varphi(t)$ is a compact map from $I \times \overline{\Omega}$ to $I \times E_f$. In view of (91), the Leray–Schauder degree i satisfies

$$i(\gamma(t), \Omega, p) = i(\gamma(0), \Omega, p) = 1$$

for all $t \in [0, 1]$. But this contradicts (92). Hence

$$\sigma(t)\Omega \cap F^{-1}(p) \neq \phi, \quad 0 \leq t \leq 1.$$

□

Lemma 18 *Let N be a closed separable subspace of a Hilbert space E, and let Ω be a bounded, convex, open subset of N containing a point p. Let G be a an N-weak-to-weak continuously differentiable functional on E, and let F be an N-weakly continuous mapping. Assume that $d = d(A, B) > 0$, and*

$$-\infty < b_0 := \inf_B G, \quad b_1 := \sup_{\overline{\Omega}} G < \infty,$$

where $A = \partial\Omega$ and $B = F^{-1}(p)$. If $\overline{\Omega} \subset \mathcal{B}_R$, let $\tilde{B} = B \cap \mathcal{B}_\nu$, where

$$\beta \int_R^\nu \frac{dt}{\rho(t)} > b_1 - b_0, \quad \beta \int_0^d \frac{dt}{\rho(t)} > b_1 - b_0$$

for some $\rho \in \mathcal{Q}$ and $\beta > 0$. Then there is a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c, \ b_0 \leq c \leq b_1, \quad \rho(d_w(u_k, \tilde{B}))\|G'(u_k)\| \leq \beta. \tag{95}$$

Proof If the lemma were false, then there would be a $\delta > 0$ such that

$$\rho(d_w(u, \tilde{B}))\|G'(u)\| > \beta \tag{96}$$

when

$$u \in U = \{u \in E : b_0 - 3\delta \leq G(u) \leq b_1 + 3\delta\}. \tag{97}$$

For $u \in \hat{E} = \{u \in E : G'(u) \neq 0\}$, let $h(u) = G'(u)/\|G'(u)\|$. Then by (96)

$$(G'(u), h(u)) > \beta/\rho(d_w(u, \tilde{B})), \quad u \in U. \tag{98}$$

For each $u \in U$ there is an $\tilde{E}$ neighborhood $W(u)$ of u such that

$$(G'(v), h(u)) > \beta/\rho(d_w(v, \tilde{B})), \quad v \in W(u) \cap U. \tag{99}$$

For otherwise there would be a sequence $\{v_k\} \subset U$ such that

$$|v_k - u|_w \to 0 \; and \; (G'(v_k), h(u)) \leq \beta/\rho(d_w(v_k, \tilde{B})). \tag{100}$$

Since G is an N-weak-to-weak continuously differentiable functional on E, we would have

$$(G'(v_k), h(u)) \to (G'(u), h(u)) \leq \beta/\rho(d_w(u, \tilde{B})), \tag{101}$$

by (24) in view of (100). This contradicts (98). Thus (99) holds.

Let $\tilde{U}$ be the set U with the inherited topology of $\tilde{E}$. It is a metric space, and $W(u) \cap \tilde{U}$ is an open set in this space. Thus, $\{W(u) \cap \tilde{U}\}, u \in \tilde{U}$, is an open covering of the paracompact space $\tilde{U}$ (cf., e.g., [7]). Consequently, there is a locally finite refinement $\{W_\tau\}$ of this cover. For each τ there is an element u_τ such that $W_\tau \subset W(u_\tau)$. Let $\{\psi_\tau\}$ be a partition of unity subordinate to this covering. Each ψ_τ is locally Lipschitz continuous with respect to the norm $|u|_w$ and consequently with respect to the norm of E. Let

$$Y(u) = \sum \psi_\tau(u) h(u_\tau), \quad u \in \tilde{U}. \tag{102}$$

Then $Y(u)$ is locally Lipschitz continuous with respect to both norms. Moreover,

$$\|Y(u)\| \leq \sum \psi_\tau(u) \|h(u_\tau)\| \leq 1 \tag{103}$$

and

$$(G'(u), Y(u)) = \sum \psi_\tau(u)(G'(u), h(u_\tau)) \geq \beta/\rho(d_w(u, \tilde{B})), \quad u \in \tilde{U}. \tag{104}$$

Reduce δ to satisfy

$$\beta \int_\delta^d \frac{dt}{\rho(t)} \geq b_1 - b_0 + \delta.$$

Let

$$Q_0 = \{u \in E : b_0 - 2\delta \leq G(u) \leq b_1 + 2\delta\},$$
$$Q_1 = \{u \in E : b_0 - \delta \leq G(u) \leq b_1 + \delta\},$$
$$Q_2 = E \setminus Q_0,$$

$$\eta(u) = d_w(u, Q_2)/[d_w(u, Q_1) + d_w(u, Q_2)].$$

It is easily checked that $\eta(u)$ is locally Lipschitz continuous (with respect to the E_w norm) on E and satisfies

$$\begin{cases} \eta(u) = 1, & u \in Q_1, \\ \eta(u) = 0, & u \in \bar{Q}_2, \\ \eta(u) \in (0, 1), & \text{otherwise.} \end{cases} \tag{105}$$

Let

$$\tilde{W}(u) = -\eta(u)Y(u)\rho(d_w(u, \tilde{B})).$$

Then

$$\|\tilde{W}(u)\| \le \rho(d_w(u, \tilde{B})) \le \rho(d(u, \tilde{B})), \quad u \in \tilde{U}.$$

By Theorem 6, for each $v \in U$ there is a unique solution $\sigma(t)v$ of

$$\sigma'(t) = \tilde{W}(\sigma(t)),\ t \in \mathbb{R}^+, \quad \sigma(0) = v. \tag{106}$$

Take

$$T = \int_\delta^d \frac{dt}{\rho(t)} \ge (b_1 - b_0 + \delta)/\beta. \tag{107}$$

Let

$$K = \{(u, t) : u = \sigma(t)v,\ v \in \bar{\Omega},\ t \in [0, T]\}.$$

Then K is a compact subset of $\tilde{E} \times \mathbb{R}$. To see this, let (u_k, t_k) be any sequence in K. Then $u_k = \sigma(t_k)v_k$, where $v_k \in \bar{\Omega}$. Since Ω is bounded, there is a subsequence such that $v_k \to v_0$ weakly in E and $t_k \to t_0$ in $[0, T]$. Since $\bar{\Omega}$ is convex and bounded, v_0 is in $\bar{\Omega}$ and $|v_k - v_0|_w \to 0$. Since $\sigma(t)$ is continuous in $\tilde{E} \times \mathbb{R}$, we have

$$u_k = \sigma(t_k)v_k \rightharpoonup \sigma(t_0)v_0 \in K.$$

Each $u_0 \in U$ has a neighborhood $W(u_0)$ in $\tilde{E}$ and a finite dimensional subspace $S(u_0)$ such that $Y(u) \subset S(u_0)$ for $u \in W(u_0) \cap U$. Since $\sigma(t)u$ is continuous in $\tilde{E} \times \mathbb{R}$, for each $(u_0, t_0) \in K$ there is a neighborhood $W(u_0, t_0) \subset \tilde{E} \times \mathbb{R}$ and a finite dimensional subspace $S(u_0, t_0) \subset E$ such that $z_t(u) \subset S(u_0, t_0)$ for $(u, t) \in W(u_0, t_0)$, where

$$z_t(u) := u - \sigma(t)u = \begin{cases} \int_0^t Y(\sigma(s)u)\rho(d_w(\sigma(s)u, \tilde{B}))ds, & u \in U, \\ 0, & u \notin U. \end{cases} \tag{108}$$

Since K is compact, there is a finite number of points $(u_j, t_j) \subset K$ such that $K \subset W = \cup W(u_j, t_j)$. Let S be a finite dimensional subspace of E containing p and all the $S(u_j, t_j)$ and such that $FS \neq \{0\}$. Then for $v \in \bar{\Omega}$ and $t \in [0, T]$ we have $z_t(v) \in S$. Thus $\sigma \in \Sigma_w(\overline{\Omega})$.

We also have

$$\begin{aligned} dG(\sigma(t)v)/dt &= -\eta(\sigma(t)v)(G'(\sigma(t)v), Y(\sigma(t)v))\rho(d_w(\sigma(t)v, \tilde{B})) \\ &\leq -\beta\eta(\sigma). \end{aligned} \tag{109}$$

Let $v \in \overline{\Omega}$. If there is a $t_1 \leq T$ such that $\sigma(t_1)v \notin Q_1$, then

$$G(\sigma(T)v) \leq G(\sigma(t_1)v) \leq b_0 - \delta. \tag{110}$$

On the other hand, if $\sigma(t)v \in Q_1$ for all $t \in [0, T]$, then we have by (109)

$$G(\sigma(T)v) \leq b_1 - \beta T \leq b_0 - \delta.$$

Hence

$$G(\sigma(T)v) \leq b_0 - \delta, \quad v \in \overline{\Omega}. \tag{111}$$

Let $u_1(t)$ be the solution of

$$u'(t) = -\rho(u(t)), \ t \in [0, T], \quad u(0) = d = d(A, \tilde{B}).$$

By Theorem 6,

$$d(\sigma(t)v, \tilde{B}) \geq u_1(t), \ t \in [0, T], \ v \in A.$$

But

$$\int_{u_1(t)}^{d} \frac{d\tau}{\rho(\tau)} = t, \quad t \in [0, T].$$

Consequently,

$$u_1(t) \geq u_1(T) \geq \delta, \quad t \in [0, T],$$

since

$$T = \int_{\delta}^{d} \frac{dt}{\rho(t)} \geq (b_1 - b_0 + \delta)/\beta.$$

Thus,

$$d(\sigma(t)v, \tilde{B}) \geq \delta, \quad t \in [0, T], \ v \in A.$$

Consequently, $\sigma(t)v \cap \tilde{B} = \phi, \ t \in (0, T]$. This means that

$$\sigma(t)v \cap \tilde{B} = \phi, \quad v \in A, \ t \in (0, T].$$

Hence,

$$\sigma(t)A \cap \tilde{B} = \phi, \quad t \in (0, T], \tag{112}$$

and

$$\sup_{\sigma(T)A} G \leq b_0 - \delta.$$

Let $u_2(t)$ be the solution of

$$u'(t) = \rho(u(t)), \ t \in [0, T], \quad u(0) = R.$$

By Theorem 6,

$$\|\sigma(t)v\| \leq u_2(t), \ t \in [0, T], \ v \in A.$$

Now it follows from the choice of ν and the fact that $A \subset \mathcal{B}_R$, that

$$\int_{R}^{u_2(T)} \frac{dt}{\rho(t)} = T \leq \int_{R}^{\nu} \frac{dt}{\rho(t)}.$$

Thus,

$$\|\sigma(t)v\| \leq \nu, \quad v \in A, \ t \in [0, T].$$

Hence,

$$\sigma(t)A \cap [B \backslash \tilde{B}] = \phi, \quad t \in [0, T]. \tag{113}$$

If we combine this with (112), we obtain

$$\sigma(t)A \cap B = \phi, \quad t \in [0, T]. \tag{114}$$

But $\sigma \in \Sigma_w(\overline{\Omega})$. By Lemma 17, this implies

$$\sigma(t)\Omega \cap B \neq \phi, \quad 0 < t \leq T.$$

Thus, there is a $u \in \Omega$ such that $\sigma(T)u \in B$. But that would mean that $G(\sigma(T)u) \geq b_0$, contradicting (111). This completes the proof. □

Lemma 19 *Let N be a closed separable subspace of a Hilbert space E, and let Ω be a bounded, convex, open subset of N containing a point p. Let G be a an N-weak-to-weak continuously differentiable functional on E. Let F be an N-weakly continuous mapping. Assume $d = d(A, B) > 0$, and*

$$-\infty < b_0 := \inf_B G, \quad b_1 := \sup_{\overline{\Omega}} G < \infty,$$

where $A = \partial\Omega$ and $B = F^{-1}(p)$. If $\overline{\Omega} \subset \mathcal{B}_R$, let $\nu > 0$, $\beta > 0$ be such that

$$\beta \ln \frac{2\nu + d}{2\nu} > b_1 - b_0, \quad \beta \ln \frac{3\nu}{2\nu + R} > b_1 - b_0. \tag{115}$$

Then there is a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c, \ b_0 \leq c \leq b_1, \quad (\nu + |u_k|_w)\|G'(u_k)\| \leq \beta. \tag{116}$$

Proof We take $\rho(s) = 2\nu + s$. Then Lemma 18 requires (115) to hold. Since

$$|u|_w \leq \nu + d_w(u, \tilde{B}), \quad u \in E,$$

(116) follows from (95). □

Lemma 20 *Let N be a closed separable subspace of a Hilbert space E, and let Ω_n be a sequence of bounded, convex, open subsets of N containing a point p. Let G be a an N-weak-to-weak continuously differentiable functional on E. Let F be an N-weakly continuous mapping. Assume $d_n = d(A_n, B) \to \infty$, and*

$$-\infty < b_0 := \inf_B G, \quad b_1 := \sup_n \sup_{\overline{\Omega}_n} G < \infty,$$

where $A_n = \partial\Omega_n$ and $B = F^{-1}(p)$. Assume that $\overline{\Omega}_n \subset \mathcal{B}_{R_n}$, and there are $\nu_n > 0, \beta > 0$ be such that

$$\beta \ln \frac{2\nu_n + d_n}{2\nu_n} > b_1 - b_0, \quad \beta \ln \frac{3\nu_n}{2\nu_n + R_n} > b_1 - b_0. \tag{117}$$

Then there is a sequence $\{u_n\} \subset E$ such that

$$G(u_n) \to c,\ b_0 \le c \le b_1, \quad (\nu_n + |u_n|_w)\|G'(u_n)\| \le \beta. \tag{118}$$

Proof By Lemma 19, for each ν_n there is a sequence satisfying

$$G(u_k) \to c,\ b_0 \le c \le b_1, \quad (\nu_n + |u_k|_w)\|G'(u_k)\| \le \beta. \tag{119}$$

Pick one member. □

Lemma 21 *Let N be a closed, separable subspace of a Banach space E, and for each $R > R_0$ let $\Omega_R(p)$ be an open, convex, bounded set in N containing a point p. Let G be a an N-weak-to-weak continuously differentiable functional on E, and let F be an N-weakly continuous mapping. Assume*

$$d_R = d_w(A^R, F^{-1}(p)) \to \infty, \quad R \to \infty,$$

where $A^R = N \backslash \Omega_R(p)$. Assume

$$-\infty < b_0 = \inf_B G, \quad \sup_{A^R} G \le a_0 < \infty, \tag{120}$$

for $R > R_0$. Then for each sequence $\nu_k \ge 2d_k - \varepsilon$ there is a $\beta > 0$ and a sequence $\{u_k\} \subset E$ such that

$$G(u_k) \to c,\ b_0 \le c \le b_1, \quad (\nu_k + |u_k|_w)\|G'(u_k)\| \le \beta. \tag{121}$$

Proof Take $\nu_n = 2R_n = 2(d_n + \varepsilon)$. Apply Lemma 20. □

Lemma 22 *Let N be a closed, separable subspace of a Hilbert space E with orthogonal complement $M \oplus \{v_0\}$, where v_0 is an element in E having unit norm and orthogonal to both M and N, and let $\delta < R$ be positive numbers. Let $\varphi(t) \in C^1(\mathbb{R})$ be such that*

$$0 \le \varphi(t) \le 1,\ \varphi(0) = 1,$$

and

$$\varphi(t) = 0, \quad |t| \ge 1.$$

Let

$$F(v+w+sv_0) = v + [s + \delta - \delta\varphi(\|w\|^2/\delta^2)]v_0, \quad w \in M,\ v \in N,\ s \in \mathbb{R}. \tag{122}$$

Let G be a an N-weak-to-weak continuously differentiable functional on E. Assume

$$-\infty < b_0 = \inf_B G, \quad \sup_A G = b_1 < \infty, \tag{123}$$

holds with $A = [N \oplus \{v_0\}]$, $A_R = A \cap \partial\mathcal{B}_R$ *and* $B = F^{-1}(\delta v_0) = \{w + rv_0 : w \in M,\ r = \delta\varphi(\|w\|^2/\delta^2)\}$. *Then for each sequence* $v_k \to \infty$ *there is a* $\beta > 0$ *and a sequence* $\{u_k\} \subset E$ *such that*

$$G(u_k) \to c,\ b_0 \le c \le b_1, \quad (v_k + |u_k|_w)\|G'(u_k)\| \le \beta. \tag{124}$$

Proof Note that F is an N-weakly continuous mapping. Let

$$\Omega_R = \{v + sv_0 : |v|_w^2 + s^2 < R^2,\ v \in N\},$$

$$A_R = \{v + sv_0 : |v|_w^2 + s^2 = R^2,\ v \in N\}$$

and

$$B = F^{-1}(\delta v_0) = \{w + rv_0 : w \in M,\ r = \delta\varphi(\|w\|^2/\delta^2)\}.$$

Then,

$$\begin{aligned} d_w(A_R, B)^2 &= \inf |v + sv_0 - w - rv_0|_w^2 \\ &= |v|_w^2 + \|w\|^2 + (s - r)^2 \\ &= R^2 - s^2 + \|w\|^2 + s^2 - 2s\delta\varphi(\|w\|^2/\delta^2) + \delta^2\varphi(\|w\|^2/\delta^2)^2 \\ &\ge R^2 - 2R\delta\varphi(\|w\|^2/\delta^2) + \delta^2\varphi(\|w\|^2/\delta^2)^2 \\ &= [R - \delta\varphi(\|w\|^2/\delta^2)]^2 \\ &\ge (R - \delta)^2 \to \infty, \quad R \to \infty. \end{aligned}$$

The hypotheses of Lemma 21 are satisfied. □

Proof of Theorem 9 Apply Lemma 18. □

Proof of Theorem 11 Apply Lemma 21. □

Proof of Corollary 13 Apply Lemma 22. □

Proof of Theorem 2 Apply Theorem 9 □

Proof of Theorem 4 Apply Theorem 12. □

References

1. D.G. Costa, On a class of elliptic systems in $\mathbb{R}^N$. Electron. J. Differ. Equ. 7 (1994), 14 pp. (electronic)

2. D.G. Costa, C.A. Magalhes, A variational approach to subquadratic perturbations of elliptic systems. J. Differ. Equ. **111**(1), 103–122 (1994)
3. D.G. de Figueiredo, P.L. Felmer, On superquadratic elliptic systems. Trans. Am. Math. Soc. **343**(1), 99–116 (1994)
4. M.F. Furtado and E.A.B. Silva, Double resonant problems which are locally non-qmuadratic at infinity, in *Proceedings of the USA–Chile Workshop on Nonlinear Analysis* (Vi-a del Mar-Valparaiso, 2000), 155–171 (electronic), Electron. J. Differ. Equ. Conf., vol. 6, Southwest Texas State Univ., San Marcos, TX, 2001
5. M.F. Furtado, L.A. Maia, E.A.B. Silva, On a double resonant problem in $\mathbb{R}^N$. Differ. Integr. Equ. **15**(11), 1335–1344 (2002)
6. M.F. Furtado, L.A. Maia, E.A.B. Silva, Solutions for a resonant elliptic system with coupling in $\mathbb{R}^N$. Commun. Partial Differ. Equ. **27**(7–8), 1515–1536 (2002)
7. J.L. Kelley, *General Topology* (Van Nostrand Reinhold, 1955)
8. W. Kryszewski, A. Szulkin, Generalized linking theorems with an application to semilinear Schrodinger equation. Adv. Differ. Equ. **3**, 441–472 (1998)
9. G. Li, J. Yang, Asymptotically linear elliptic systems (English summary). Commun. Partial Differ. Equ. **29**(5–6), 925–954 (2004)
10. J. Mawhin, Nonlinear functional analysis and periodic solution of semilinear wave equation, in *Nonlinear Phenomena in Mathematical Sciences*, ed. by Lakshmikantham (Academic Press, London, 1982), pp. 671–681
11. M. Schechter, New saddle point theorems, Generalized functions and their applications (Varanasi, 1991) (Plenum, New York, 1993), pp. 213–219
12. M. Schechter, A generalization of the saddle point method with applications. Ann. Polon. Math. **57**(3), 269–281 (1992)
13. M. Schechter, New linking theorems. Rend. Sem. Mat. Univ. Padova **99**, 255–269 (1998)
14. M. Schechter, Infinite-dimensional linking. Duke Math. J. **94**(3), 573–595 (1998)
15. M. Schechter, Critical point theory with weak-to-weak linking. Commun. Pure Appl. Math. **51**(11–12), 1247–1254 (1998)
16. M. Schechter, Rotationally invariant periodic solutions of semilinear wave equations. Abstr. Appl. Anal. **3**(1–2), 171–180 (1998)
17. M. Schechter, *Linking Methods in Critical Point Theory* (Birkhauser Boston, 1999)
18. M. Schechter, Periodic solutions of semilinear higher dimensional wave equations. Chaos Solitons Fractals **12**(6), 1029–1034 (2001)
19. M. Schechter, Sandwich pairs in critical point theory. Trans. Am. Math. Soc. **360**(6), 2811–2823 (2008)
20. M. Schechter, Strong sandwich pairs. Indiana Univ. Math. J. **57**(3), 1105–1131 (2008)
21. M. Schechter, *Minimax Systems and Critical Point Theory* (Birkhauser Boston, 2009)
22. M. Schechter, Critical point theory: sandwich and linking systems (2020)
23. M. Schechter, W. Zou, Weak linking. Nonlinear Anal. **55**(6), 695–706 (2003)
24. E.A. de B.e. Silva, Linking theorems and applications to semilinear elliptic problems at resonance. Nonlinear Anal. TMA **16**, 455–477 (1991)
25. E.A. B. Silva, Nontrivial solutions for noncooperative elliptic systems at resonance, in *Proceedings of the USA-Chile Workshop on Nonlinear Analysis* (Vi-a del Mar-Valparaiso, 2000), 267–283 (electronic), Electron. J. Differ. Equ. Conf., vol. 6, Southwest Texas State Univ., San Marcos, TX, 2001
26. K. Tintarev, Solutions to elliptic systems of Hamiltonian type in $\mathbb{R}^N$. Electron. J. Differ. Equ. 29 (1999), 11 pp.

Canonical Systems of Partial Differential Equations

Martin Schechter

Abstract We use critical point theory to find solutions of the nonlinear steady state Schrödinger equations arising in the study of photonic lattices.

1 Introduction

Systems of partial differential equations arise in many investigations in the physical sciences. Depending on the application and on the questions asked, different types of systems emerge. Usually, if one is interested in finding steady states solutions, the resulting system is elliptic in nature. Such systems may display severe difficulties when one tries to solve them. Most of the time they admit a trivial solution, where all of the unknown functions are identically zero. However, the physical application requires a solution which is not identically zero. In such cases, the methods of solution may be very difficult. In particular, one has to show that the solution obtained is not trivial. The system that we study is not only deceptive, but it is almost impossible to tell if one has solved the whole system or only parts of the system. I call it "canonical." I shall elaborate on this later.

Many general systems are the form

$$\mathscr{A}v = f(x, v, w), \quad x \in Q \subset \mathbb{R}^n, \tag{1}$$

$$\mathscr{B}w = g(x, v, w), \quad x \in Q \subset \mathbb{R}^n, \tag{2}$$

where $\mathscr{A}$, $\mathscr{B}$ are linear partial differential operators. I call this system "deceptive" if $(v, 0)$ is a solution of (1) and (2) whenever v satisfies

$$\mathscr{A}v = f(x, v, 0), \quad x \in Q \subset \mathbb{R}^n, \tag{3}$$

M. Schechter (✉)
Department of Mathematics, University of California, Irvine, CA, USA
e-mail: mschecht@math.uci.edu

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_22

or $(0, w)$ is a solution whenever w satisfies

$$\mathscr{B}w = g(x, 0, w), \quad x \in Q \subset \mathbb{R}^n. \tag{4}$$

In this case it is very difficult to determine if both components of a solution are nontrivial.

The particular system I have chosen consist of nonlinear Schrödinger equations arising in optics (cf. [16]) describing the propagation of a light wave in induced photonic lattices. They can be written in the form

$$iV_t + \Delta V = \frac{PV}{1 + |V|^2 + |W|^2}$$

$$iW_t + \Delta W = \frac{PW}{1 + |V|^2 + |W|^2}$$

for the periodic wave functions $V(x, t)$, $W(x, t)$ over a periodic bounded spacial domain $\Omega \subset \mathbb{R}^2$, where P, Q are parameters (cf. [2, 21]). To find a steady state solution, we look for solutions of the form

$$V(x, t) = e^{i\lambda t}v(x), \quad W(x, t) = e^{i\lambda t}w(x),$$

where λ is a real constant. This leads to the following system of equations over a periodic domain $\Omega \subset \mathbb{R}^2$:

$$\Delta v = \frac{Pv}{1 + v^2 + w^2} + \lambda v, \tag{5}$$

$$\Delta w = \frac{Qw}{1 + v^2 + w^2} + \lambda w, \tag{6}$$

where P, Q, λ are parameters. The solutions v, w are to be periodic in Ω with the same periods. One wishes to obtain intervals of the parameter λ for which there are nontrivial solutions. This will provide continuous energy spectrum that allows the existence of steady state solutions. This system was studied in [2], where it was shown that

1. If P, Q, λ are all positive, then the only solution is trivial.
2. If $P < 0$ and $0 < \lambda < -P$, then the system (5) and (6) has a nontrivial solution.
3. If P, $Q > 0$, there is a constant $\delta > 0$ such that the system (5) and (6) has a nontrivial solution provided $0 < -\lambda < \delta$.
4. All of these statements are true if we replace P by Q.

Wave propagation in nonlinear periodic lattices has been studied by many reseachers (cf., e.g., [1–11, 17, 20–23] and their bibliographies.)

In the present paper, we wish to cover some remaining situations not mentioned in [2] as well as extending their results to higher dimensions. We shall show that there are many intervals of the parameters in which nontrivial solutions exist. Our results are true in any dimension.

In stating our results, we shall make use of the following considerations. Let Ω be a bounded periodic domain in $\mathbb{R}^n$, $n \geq 1$. Consider the operator $-\Delta$ on functions in $L^2(\Omega)$ having the same periods as Ω. The spectrum of $-\Delta$ consists of isolated eigenvalues of finite multiplicity:

$$0 = \lambda_0 < \lambda_1 < \cdots < \lambda_\ell < \cdots,$$

with eigenfunctions in $L^\infty(\Omega)$. Let λ_ℓ, $\ell \geq 0$, be one of these eigenvalues, and define

$$N = \bigoplus_{\lambda \leq \lambda_\ell} E(\lambda), \quad M = N^\perp.$$

As noted in [2], to prove the existence of a nontrivial solution of system (5) and (6), it suffices to obtain a nontrivial solution of either

$$\Delta v = \frac{Pv}{1+v^2} + \lambda v, \tag{7}$$

or

$$\Delta w = \frac{Qw}{1+w^2} + \lambda w. \tag{8}$$

This stems from the fact that $(v, 0)$ is a solution of (5) and (6) if v is a solution of (7) and $(0, w)$ is a solution of (5) and (6) if w is a solution of (8). The author is unaware if such solutions are desirable from the physical point of view. However, we have been able to find values of P, Q, λ for which the system (5) and (6) has a solution (v, w) where $v \neq 0, w \neq 0$.

We shall prove

Theorem 1 *If $0 < \lambda < -P$ or $0 < \lambda < -Q$, then (5) and (6) has a nontrivial solution.*

Theorem 2 *If $0 < -\lambda < P$ or $0 < -\lambda < Q$, then (5) and (6) has a nontrivial solution.*

Theorem 3 *If $P > 0$, $Q > 0$, $\sigma = -\lambda > 0$, and either $0 \leq \sigma - P < \lambda_1 < \sigma$ or $0 \leq \sigma - Q < \lambda_1 < \sigma$, then (5) and (6) has a nontrivial solution.*

Theorem 4 *If $P > 0$, $Q > 0$, $\sigma = -\lambda > 0$, and either $\lambda_\ell \leq \sigma - P < \lambda_{\ell+1} < \sigma$ or $\lambda_\ell \leq \sigma - Q < \lambda_{\ell+1} < \sigma$ then (5) and (6) has a nontrivial solution.*

Theorem 5 *If $P > 0$, $Q = 0$, $\lambda = -\lambda_\ell < 0$, $-\Delta w = \lambda_\ell w$ and*

$$\lambda_\ell < P \int_\Omega \frac{1}{1+w^2}/|\Omega|,$$

then (5) and (6) has a nontrivial solution. If $w \neq 0$, *then the solution has both components nonzero.*

Theorem 6 *If* $Q > 0,\ P = 0,\ \lambda = -\lambda_\ell < 0,\ -\Delta v = \lambda_\ell v$ *and*

$$\lambda_\ell < Q \int_\Omega \frac{1}{1+v^2}/|\Omega|,$$

then (5) and (6) has a nontrivial solution. If $v \neq 0$, *then the solution has both components nonzero.*

2 Some Lemmas

In proving our results we shall make use of the following lemmas (cf., e.g., [12, 14, 15, 18]). For the definition of linking, cf. [12].

Lemma 1 *Let* M, N *be closed subspaces of a Hilbert space* E *such that one of them is finite dimensional and* $E = M \oplus N$. *Take* $B = \partial\mathscr{B}_\delta \cap M$, *and let* w_0 *be any element in* $\partial\mathscr{B}_1 \cap M$. *Take* A *to be the set of all* u *of the form*

$$u = v + sw_0, \quad v \in N,\ s \in \mathbb{R},$$

satisfying the following

(a) $\|v\|_E \leq R,\ s = 0$
(b) $\|v\|_E \leq R,\ s = 2R_0$
(c) $\|v\|_E = R,\ 0 \leq s \leq 2R_0,$

where $0 < \delta < \min(R, R_0)$. *Then* A *and* B *link each other.*

Lemma 2 *The sets* $\|u\|_E = R > 0$ *and* $\{e_1, e_2\}$ *link each other provided* $\|e_1\|_E < R$ *and* $\|e_2\|_E > R$.

Lemma 3 *If* $G(u) \in C^1(E, \mathbb{R})$ *satisfies*

$$\alpha = \inf_E G > -\infty, \tag{9}$$

then there is a sequence $\{u_k\}$ *such that*

$$G(u_k) \to \alpha, \quad (1 + \|u_k\|_E)\|G'(u_k)\| \to 0. \tag{10}$$

Lemma 4 *If* A *links* B, *and* $G(u) \in C^1(E, \mathbb{R})$ *satisfies*

$$a_0 = \sup_A G \le b_0 = \inf_B G, \tag{11}$$

then there is a sequence $\{u_k\}$ *such that*

$$G(u_k) \to c \ge b_0, \quad (1 + \|u_k\|_E)\|G'(u_k)\| \to 0. \tag{12}$$

We let E be the subspace of $H^{1,2}(\Omega)$ consisting of those functions having the same periodicity as Ω with norm given by

$$\|w\|_E^2 = \|\nabla w\|^2 + \|w\|^2.$$

Assume $P \neq 0$, $Q \neq 0$, $\lambda \neq 0$. Let

$$a(u) = \frac{1}{P}[\ \|\nabla v\|^2 + \lambda\ \|v\|^2] + \frac{1}{Q}[\ \|\nabla w\|^2 + \lambda\ \|w\|^2], \quad v, w \in E \tag{13}$$

and

$$G(u) = a(u) + \int_\Omega \ln(1 + u^2)\, dx. \tag{14}$$

We have

Lemma 5 *If G(u) is given by (14), then every sequence satisfying (10) has a subsequence converging in E. Consequently, there is a* $u \in E$ *such that G(u)=c and* $G'(u) = 0$.

Proof The sequence satisfies

$$\begin{aligned} G(u_k) = {} & \frac{1}{P}\|\nabla v_k\|^2 + \frac{\lambda}{P}\|v_k\|^2 + \frac{1}{Q}\|\nabla w_k\|^2 + \frac{\lambda}{Q}\|w_k\|^2 \\ & + \int_\Omega \ln\{1 + |u_k|^2\}\, dx \to c, \end{aligned} \tag{15}$$

$$\begin{aligned} (G'(u_k), q)/2 = {} & \frac{1}{P}(\nabla v_k, \nabla g) + \frac{\lambda}{P}(v_k, g) \\ & + \frac{1}{Q}(\nabla w_k, \nabla h) + \frac{\lambda}{Q}(w_k, h) \\ & + \int_\Omega \frac{u_k q}{1 + u_k^2}\, dx \to 0, \quad q = (g, h), \end{aligned} \tag{16}$$

$$(G'(u_k), v_k)/2 = \frac{1}{P}(\nabla v_k, \nabla v_k) + \frac{\lambda}{P}(v_k, v_k) \tag{17}$$

$$+\int_\Omega \frac{u_k v_k}{1+u_k^2}\,dx \to 0.$$

and

$$(G'(u_k), w_k)/2 = \frac{1}{Q}(\nabla w_k, \nabla w_k) + \frac{\lambda}{Q}(w_k, w_k) + \int_\Omega \frac{u_k w_k}{1+u_k^2}\,dx \to 0. \tag{18}$$

Thus,

$$\int_\Omega H(x, u_k)\,dx \to c, \tag{19}$$

where

$$H(x,t) = \ln(1+t^2) - \frac{t^2}{1+t^2}. \tag{20}$$

Let $\rho_k = \|u_k\|_H$, where

$$\|u\|_H^2 = \frac{1}{|P|}[\|\nabla v\|^2 + |\lambda|\ \|v\|^2] + \frac{1}{|Q|}[\|\nabla w\|^2 + |\lambda|\ \|w\|^2], \quad u = (v, w) \in E. \tag{21}$$

Assume first that $\rho_k \to \infty$. Let $\tilde{u}_k = u_k/\rho_k$. Then $\|\tilde{u}_k\|_H = 1$. Hence, there is a renamed subsequence such that $\tilde{u}_k \rightharpoonup \tilde{u}$ in E, and $\tilde{u}_k \to \tilde{u}$ in $L^2(\Omega)$ and a.e. Now

$$\|u_k\|_H^2 = \frac{1}{|P|}[\|\nabla v_k\|^2 + |\lambda|\ \|v_k\|^2] + \frac{1}{|Q|}[\|\nabla w_k\|^2 + |\lambda|\ \|w_k\|^2]. \tag{22}$$

By (17) and (18),

$$\|u_k\|_H^2 \le |(G'(u_k), v_k)|/2 + |(G'(u_k), w_k)|/2 + \frac{|\lambda| - \lambda}{|P|}\|v_k\|^2 + \frac{|\lambda| - \lambda}{|Q|}\|w_k\|^2 + \int_\Omega \frac{u_k^2}{1+u_k^2}\,dx.$$

Hence,

$$1 = \|\tilde{u}_k\|_H^2 \le [|(G'(u_k), v_k)|/2 + |(G'(u_k), w_k)|/2]/\rho_k^2 + C\|\tilde{u}_k\|^2. \tag{23}$$

In the limit we have,

$$1 \le C\|\tilde{u}\|^2.$$

This shows that $\tilde{u} \not\equiv 0$. Let Ω_0 be the subset of Ω where $\tilde{u}(x) \neq 0$. Then $|\Omega_0| \neq 0$. Thus

$$\begin{aligned}\int_\Omega H(x, u_k)\, dx &= \int_{\Omega_0} H(x, u_k)\, dx + \int_{\Omega\setminus\Omega_0} H(x, u_k)\, dx \\ &\ge \int_{\Omega_0} H(x, u_k)\, dx \to \infty.\end{aligned}$$

This contradicts (19). Thus, the sequence satisfying (10) is bounded in E. Hence, there is a renamed subsequence such that $u_k \rightharpoonup u_0$ in E, and $u_k \to u_0$ in $L^2(\Omega)$ and a.e. Taking the limit in (17), we obtain

$$\begin{aligned}(G'(u_0), q)/2 =& \frac{1}{P}(\nabla v_0, \nabla g) + \frac{\lambda}{P}(v_0, g) \\ &+ \frac{1}{Q}(\nabla w_0, \nabla h) + \frac{\lambda}{Q}(w_0, h) \\ &+ \int_\Omega \frac{u_0 q}{1+u_0^2}\, dx = 0, \quad q = (g, h),\end{aligned} \tag{24}$$

Thus, u_0 satisfies $G'(u_0) = 0$. Since $u_0 \in E$, it satisfies

$$\begin{aligned}(G'(u_0), u_0)/2 =& \frac{1}{P}(\nabla v_0, \nabla v_0) + \frac{\lambda}{P}(v_0, v_0) \\ &+ \frac{1}{Q}(\nabla w_0, \nabla w_0) + \frac{\lambda}{Q}(w_0, w_0) \\ &+ \int_\Omega \frac{u_0^2}{1+u_0^2}\, dx = 0\end{aligned} \tag{25}$$

Also, from the limit in (17), we have

$$\begin{aligned}\lim \frac{1}{P}\|\nabla v_k\|^2 =& \lim(G'(u_k), v_k)/2 \\ &- \lim[\frac{\lambda}{P}\|v_k\|^2 + \int_\Omega \frac{v_k^2}{1+u_k^2}\, dx]\end{aligned}$$

$$= -\left[\frac{\lambda}{P}\|v\|^2 + \int_\Omega \frac{v^2}{1+u^2}\,dx\right]$$
$$= \frac{1}{P}\|\nabla v\|^2,$$

with a similar statement for $\|\nabla w\|^2$. Consequently, $\nabla u_k \to \nabla u$ in $L^2(\Omega)$. This shows that $G(u_k) \to G(u_0)$. Hence, $G(u_0) = c$.

Lemma 6 *If $G'(u) = 0$, then (v,w) is a solution of (5) and (6).*

Proof From (24) we see that

$$|(\nabla u, \nabla q)| \le C\|q\|, \quad q \in E.$$

From the fact that the functions and Ω are periodic with the same period, it follows that $u \in H^{2,2}(\Omega)$ and satisfies (5) and (6) (cf., e.g., [13]).

Lemma 7

$$\int_\Omega \ln(1+u^2)dx/\|u\|_H^2 \to 0, \quad \|u\|_H \to \infty. \tag{26}$$

Proof Suppose $u_k \in H$ is a sequence such that $\rho_k = \|u_k\|_H \to \infty$. Let $\tilde{u}_k = u_k/\rho_k$. Then $\|\tilde{u}_k\|_H = 1$. Hence, there is a renamed subsequence such that $\tilde{u}_k \rightharpoonup \tilde{u}$ in H, and $\tilde{u}_k \to \tilde{u}$ in $L^2(\Omega)$ and a.e. Now

$$\frac{\ln(1+u_k^2)}{\rho_k^2} = \frac{\ln(1+u_k^2)}{u_k^2}\tilde{u}_k^2 \to 0 \ a.e.$$

and it is dominated a.e. by $\tilde{u}_k^2 \to \tilde{u}^2$ in $L^1(\Omega)$. Thus

$$\int_\Omega \frac{\ln(1+u_k^2)}{\rho_k^2}dx \to 0.$$

Since this is true for any sequence satisfying $\|u_k\|_H \to \infty$, we see that (26) holds.

Corollary 1 *If*

$$I(u) = \|u\|_H^2 - \int_\Omega \ln(1+u^2)\,dx,$$

then

$$I(v) \to \infty \ as \ \|v\|_H \to \infty. \tag{27}$$

Proof We have

$$I(u)/\|u\|_H^2 = 1 - \int_\Omega \ln(1+u^2)dx/\|u\|_H^2 \to 1, \quad \|u\|_H \to \infty$$

by Lemma 7. This gives (27).

Lemma 8

$$\int_\Omega [u^2 - \ln(1+u^2)]dx/\|u\|_H^2 \to 0, \quad \|u\|_H \to 0. \tag{28}$$

Proof Suppose $u_k \in H$ is a sequence such that $\rho_k = \|u_k\|_H \to 0$. In particular, there is a renamed subsequence such that $u_k \to 0$ a.e. Let $\tilde{u}_k = u_k/\rho_k$. Then $\|\tilde{u}_k\|_H = 1$. Hence, there is a renamed subsequence such that $\tilde{u}_k \rightharpoonup \tilde{u} \in H$, and $\tilde{u}_k \to \tilde{u}$ in $L^2(\Omega)$ and a.e. Now

$$\frac{u_k^2 - \ln(1+u_k^2)}{\rho_k^2} \le \frac{u_k^2}{1+u_k^2}\tilde{u}_k^2 \to 0 \ a.e.$$

and it is dominated a.e. by $\tilde{u}_k^2 \to \tilde{u}^2$ in $L^1(\Omega)$. Thus

$$\int_\Omega \frac{u_k^2 - \ln(1+u_k^2)}{\rho_k^2}dx \to 0.$$

Since this is true for any sequence satisfying $\|u_k\|_H \to 0$, we see that (28) holds.

3 Proofs of the Theorems

Proof of Theorem 1 We let E be the subspace of $H^{1,2}(\Omega)$ consisting of those functions having the same periodicity as Ω with norm given by

$$\|w\|_E^2 = \|\nabla w\|^2 + \|w\|^2.$$

Let $u = (v, w)$, where $v, w \in E$ and $u^2 = v^2 + w^2$. If $q = (g, h)$, we write $uq = vg + wh$. Define

$$\begin{aligned}\|u\|_H^2 =& \frac{1}{|P|}[\|\nabla v\|^2 + |\lambda| \ \|v\|^2] \\ &+ \frac{1}{|Q|}[\|\nabla w\|^2 + |\lambda| \ \|w\|^2], \quad v, w \in E.\end{aligned} \tag{29}$$

Assume that P, Q, λ do not vanish. Then $\|u\|_H^2$ is a norm on $H = E \times E$ having a scalar product $(u, h)_H$.

Let

$$I(u) = \|u\|_H^2 - \int_\Omega \ln(1+u^2)\,dx. \tag{30}$$

Then,

$$(I'(u), q)/2 = (u, q)_H - \int_\Omega \frac{uq}{1+u^2}\,dx, \quad q \in H. \tag{31}$$

If $I'(u) = 0$, then

$$\Delta v = \frac{-|P|v}{1+|v|^2+|w|^2} + |\lambda|v, \tag{32}$$

$$\Delta w = \frac{-|Q|w}{1+|v|^2+|w|^2} + |\lambda|w. \tag{33}$$

This is equivalent to (5) and (6) if $P < 0$, $Q < 0$, $\lambda > 0$. To prove the theorem, we must show that there is a nontrivial solution of $I'(u) = 0$ when either $0 < \lambda < -P$ or $0 < \lambda < -Q$.

Assume $0 < \lambda < -P$. We show that $I(u)$ has a minimum $u \neq 0$.

Let the sequence $u_k \in H$ satisfy

$$I(u_k) \searrow \alpha = \inf_H I$$

(which may be $-\infty$). By (27), $\rho_k = \|u_k\|_H$ is bounded. Hence, there is a renamed subsequence such that $u_k \rightharpoonup u_0$ in H, and $u_k \to u_0$ in $L^2(\Omega)$ and a.e. Since

$$\|u_k\|_H^2 - 2([u_k - u_0], u_0)_H = \|u_0\|_H^2 + \|u_k - u_0\|_H^2,$$

we have

$$\begin{aligned} I(u_0) &\le \|u_k\|_H^2 - 2([u_k - u_0], u_0)_H \\ &\quad - \int_Q \ln(1+u_0^2)dx \\ &= I(u_k) - 2([u_k - u_0], u_0)_H \\ &\quad - \int_Q [\ln(1+u_0^2) - \ln(1+u_k^2)]dx \\ &\to \alpha. \end{aligned}$$

Thus,

$$\alpha \le I(u_0) \le \alpha,$$

showing that α is finite and that u_0 is a minimum. Thus, $I'(u_0) = 0$ and u_0 is a solution of

$$\Delta v = \frac{-|P|v}{1+|v|^2+|w|^2} + \lambda v, \tag{34}$$

$$\Delta w = \frac{-|Q|w}{1+|v|^2+|w|^2} + \lambda w. \tag{35}$$

Next, we show that $u_0 \neq 0$. We do this by showing that $\alpha < 0$. Consider a constant function $u = (s, 0)$. Then,

$$I(u) = [\frac{\lambda}{|P|}s^2 - \ln(1+s^2)]|\Omega|, \quad s \in \mathbb{R}.$$

This has a negative minimum if $\lambda < |P|$. Thus $I(u_0) = \alpha < 0$. Since $I(0, 0) = 0$, we see that $u_0 \neq 0$. However, u_0 satisfies (34) and (35), not (5) and (6). To rectify the situation, we merely note that the same method produces a negative minimum v_0 for $I(v, 0)$, and $(v_0, 0)$ is a nontrivial solution of (5) and (6). This completes the proof for the case $0 < \lambda < -P$. The case $0 < \lambda < -Q$ is treated similarly.

Proof of Theorem 2 Assume $0 < \sigma < P$, $0 < \sigma < Q$, and let $a(u)$ and $G(u)$ be given by (13) and (14), respectively. Then $G'(u) = 0$ iff $u = (v, w)$ is a solution of (5) and (6). We search for a nontrivial solution.

Let $\rho_k = \|u_k\|_H$, where

$$\|u\|_H^2 = \frac{1}{|P|}[\|\nabla v\|^2 + |\lambda| \ \|v\|^2] \tag{36}$$
$$+ \frac{1}{|Q|}[\|\nabla w\|^2 + |\lambda| \ \|w\|^2], \quad u = (v, w) \in E.$$

Assume that $\rho_k \to 0$. Let $\tilde{u}_k = u_k/\rho_k$. Then $\|\tilde{u}_k\|_H = 1$. Hence, there is a renamed subsequence such that $\tilde{u}_k \rightharpoonup \tilde{u}$ in E, and $\tilde{u}_k \to \tilde{u}$ in $L^2(\Omega)$ and a.e. We have

$$G(u_k)/\rho_k^2 = \frac{1}{P}\|\nabla \tilde{v}_k\|^2 + \frac{\lambda+P}{P}\|\tilde{v}_k\|^2$$
$$+ \frac{1}{Q}\|\nabla \tilde{w}_k\|^2 + \frac{\lambda+Q}{Q}\|\tilde{w}_k\|^2$$
$$+ \int_\Omega [\ln\{1+|u_k|^2\} - u_k^2]\,dx/\rho_k^2.$$

Since $P > \sigma$ and $Q > \sigma$, we see in view of Lemma 7 that there are positive constants ε, η such that

$$G(u)/\|u\|_H^2 \geq \varepsilon, \quad \|u\|_H \leq \eta.$$

Let A be the set of those $u \in H$ such that $\|u\|_H = \eta$. Consider a constant function $u = (s, 0)$. Then,

$$G(u)/s^2 = [\frac{\lambda}{P} + s^{-2}\ln(1+s^2)]|\Omega| \to \frac{\lambda}{P}|\Omega| < 0, \quad s \to \infty.$$

Hence, there is a $u \in H$ such that $\|u\|_H > \eta$ and $G(u) < \varepsilon\eta^2$. Since $G(0,0) = 0$, there is a $u \in H$ such that $\|u\|_H < \eta$ and $G(u) < \varepsilon\eta^2$. The theorem now follows from Lemmas 2, 4, and 5.

Proof of Theorem 3 Assume $P > 0$, $Q > 0$, $\lambda < 0$ and $\sigma = -\lambda > \max[P, \lambda_1]$. Let

$$a(u) = \frac{1}{P}[\,\|\nabla v\|^2 - \sigma\,\|v\|^2] + \frac{1}{Q}[\,\|\nabla w\|^2 - \sigma\,\|w\|^2], \quad v, w \in E \tag{37}$$

and

$$G(u) = a(u) + \int_\Omega \ln(1+u^2)\,dx. \tag{38}$$

Then $G'(u) = 0$ iff u satisfies (5) and (6).

First, we note that

$$G(u) \le 0, \quad u \in N,$$

if $\sigma \ge P,\ \sigma \ge Q$. To see this, let $u = (c, d) \in N$. Then

$$a(c,d) = -\frac{\sigma}{P}c^2|\Omega| - \frac{\sigma}{Q}d^2|\Omega|$$

and

$$\int_\Omega \ln(1+c^2+d^2)dx \le (c^2+d^2)|\Omega|.$$

Thus,

$$G(u) \le [1 - \frac{\sigma}{P}]c^2|\Omega| + [1 - \frac{\sigma}{Q}]d^2|\Omega|.$$

This means that

$$G(u) \le 0, \quad u \in N, \tag{39}$$

provided $\sigma \ge P,\ \sigma \ge Q$.

Next, let ψ be an eigenfunction of $-\Delta$ corresponding to the eigenvalue λ_1. If we take $u = (\psi + c, \psi + d)$, we have

$$a(u) = \frac{1}{P}[\ \|\nabla(\psi + c)\|^2 - \sigma\ \|(\psi + c)\|^2] + \frac{1}{Q}[\ \|\nabla(\psi + d)\|^2 - \sigma\ \|(\psi + d)\|^2],$$

and this gives

$$a(u) = \frac{1}{P}[(\lambda_1 - \sigma)\ \|\psi\|^2 - \sigma c^2] + \frac{1}{Q}[(\lambda_1 - \sigma)\ \|\psi\|^2 - \sigma d^2],$$

which will be negative if $\sigma > \lambda_1$. Moreover,

$$\int_\Omega \ln(1 + 2\psi^2 + c^2 + d^2)dx/\|u\|_H^2 \to 0, \quad \|u\|_H \to \infty.$$

This follows from the fact that

$$\int_\Omega \ln(1 + u^2)dx/\|u\|_H^2 \to 0, \quad \|u\|_H \to \infty \tag{40}$$

(Lemma 7). Consequently,

$$\limsup_{\|(\psi+c,\psi+d)\|_H \to \infty} G(\psi + c, \psi + d) < 0 \tag{41}$$

provided $\sigma > \lambda_1$.

Next, let $u = (v, w)$ be any function in M. Then $\|\nabla u\|^2 = \|\nabla v\|^2 + \|\nabla w\|^2 \geq \lambda_1\|v\|^2 + \lambda_1\|w\|^2 = \lambda_1\|u\|^2$. Then

$$a(u) + \|u\|^2 \geq \frac{1}{P}[1 - \frac{\sigma - P}{\lambda_1}]\ \|\nabla v\|^2 + \frac{1}{Q}[1 - \frac{\sigma - Q}{\lambda_1}]\ \|\nabla w\|^2.$$

Thus, there is an $\varepsilon > 0$ such that

$$a(u) + \|u\|^2 \geq 2\varepsilon\|\nabla u\|^2, \quad u \in M, \tag{42}$$

when $\sigma - \lambda_1 < \min[P, Q]$.

Now

$$\int_\Omega [u^2 - \ln(1 + u^2)]dx/\|u\|_H^2 \to 0, \quad \|u\|_H \to 0 \tag{43}$$

by Lemma 8. If we combine (42) and (43), we see that there is an $\varepsilon > 0$ such that

$$G(u) \geq \varepsilon \|\nabla u\|^2, \quad u \in M, \tag{44}$$

when $\|\nabla u\|^2$ is small and $\sigma - \lambda_1 < \min[P, Q]$.

Take $A = \partial(N \oplus \{\psi\})$, $B = \partial \mathscr{B}_\delta \cap M$. By (39), (41), and (44) one can apply Lemma 1 to obtain (10) and then Lemma 5 to conclude that (5) and (6) has a nontrivial solution. To see this, note that $a_0 = 0 < \varepsilon \leq b_0$, showing that the solution u_0 satisfies $G(u_0) \geq \varepsilon > 0$. Since $G(0) = 0$, we see that $u_0 \neq 0$. If $\max[P, \lambda_1] < -\lambda < P + \lambda_1$ is true, but $\max[Q, \lambda_1] < -\lambda < Q + \lambda_1$, is not, we can apply the argument used in the proof of Theorem 1. The same is true in the other direction. This completes the proof.

Proof of Theorem 4 First, we note that

$$G(u) \leq 0, \quad u \in N, \tag{45}$$

if $\sigma \geq \lambda_\ell + \max[P, Q]$. To see this, let $u = (v, w) \in N$. Then $\|\nabla u\|^2 = \|\nabla v\|^2 + \|\nabla w\|^2 \leq \lambda_\ell \|v\|^2 + \lambda_\ell \|w\|^2 = \lambda_\ell \|u\|^2$. Then

$$G(u) \leq \frac{1}{P}[\lambda_\ell - \sigma + P] \, \|v\|^2 + \frac{1}{Q}[\lambda_\ell - \sigma + Q] \, \|w\|^2 \leq 0.$$

Next, let g be an eigenfunction of $-\Delta$ corresponding to the eigenvalue $\lambda_{\ell+1}$. If we take $u = (g + v, g + w)$, we have

$$\begin{aligned} a(u) =& \frac{1}{P}[\, \|\nabla(g + v)\|^2 - \sigma \, \|(g + v)\|^2] \\ &+ \frac{1}{Q}[\, \|\nabla(g + w)\|^2 - \sigma \, \|(g + w)\|^2], \end{aligned}$$

and this gives

$$\begin{aligned} a(u) =& \frac{1}{P}[(\lambda_{\ell+1} - \sigma) \, \|g\|^2 + (\lambda_\ell - \sigma)\|v\|^2] \\ &+ \frac{1}{Q}[(\lambda_{\ell+1} - \sigma) \, \|g\|^2 + (\lambda_\ell - \sigma^2)\|w\|^2], \end{aligned}$$

which will be negative if $\sigma > \lambda_{\ell+1}$. Moreover, by Lemma 7,

$$\int_\Omega \ln(1 + 2g^2 + v^2 + w^2) dx / \|u\|_H^2 \to 0, \quad \|u\|_H \to \infty. \tag{46}$$

Consequently, (46) holds provided $\sigma > \lambda_{\ell+1}$.

Next, let $u = (v, w)$ be any function in M. Then $\|\nabla u\|^2 = \|\nabla v\|^2 + \|\nabla w\|^2 \geq \lambda_{\ell+1} \|v\|^2 + \lambda_{\ell+1} \|w\|^2 = \lambda_{\ell+1} \|u\|^2$. Then

$$a(u) + \|u\|^2 \geq \frac{1}{P}[1 - \frac{\sigma - P}{\lambda_{\ell+1}}] \|\nabla v\|^2 + \frac{1}{Q}[1 - \frac{\sigma - Q}{\lambda_{\ell+1}}] \|\nabla w\|^2.$$

Thus, there is an $\varepsilon > 0$ such that

$$a(u) + \|u\|^2 \geq 2\varepsilon \|\nabla u\|^2, \quad u \in M, \tag{47}$$

when $\sigma - \lambda_{\ell+1} < \min[P, Q]$.

Now by Lemma 8,

$$\int_\Omega [u^2 - \ln(1 + u^2)]dx / \|u\|_H^2 \to 0, \quad \|u\|_H \to 0. \tag{48}$$

If we combine (47) and (48), we see that there is an $\varepsilon > 0$ such that

$$G(u) \geq \varepsilon \|\nabla u\|^2, \quad u \in M, \tag{49}$$

when $\|\nabla u\|^2$ is small and $\sigma - \lambda_{\ell+1} < \min[P, Q]$.

By (45), (46), and (49) one can apply Lemma 1 to obtain (10) and then Lemma 5 to conclude that (5) and (6) has a nontrivial solution u_0 taking $A = \partial(N \oplus \{g\})$, $B = \partial \mathscr{B}_\delta \cap M$. Then $a_0 = 0 < \varepsilon \leq b_0$, showing that $G(u_0) \geq \varepsilon > 0$. Since $G(0, 0) = 0$, we see that $u_0 \neq 0$. If $\lambda_\ell < \sigma - P < \lambda_{\ell+1} < \sigma$ is true, but $\lambda_\ell < \sigma - Q < \lambda_{\ell+1} < \sigma$ is not, we can apply the argument used in the proof of Theorem 1. The same is true in the other direction. This completes the proof.

Proof of Theorem 5 If $w = 0$, this follows from Theorem 1 since $0 < \lambda_\ell < -P$. Otherwise, let

$$I_w(v) = \frac{1}{P}\|\nabla v\|^2 - \frac{\lambda_\ell}{P}\|v\|^2 + \int_\Omega \ln\{1 + v^2 + w^2\}\, dx, \quad v \in H. \tag{50}$$

Then,

$$(I_w'(v), g)/2 = \frac{1}{P}(\nabla v, \nabla g) - \frac{\lambda_\ell}{P}(v, g) + \int_\Omega \frac{vg}{1 + v^2 + w^2}\, dx. \tag{51}$$

If $I_w'(v) = 0$, then $u = (v, w)$ satisfies

$$\Delta v = \frac{Pv}{1 + v^2 + w^2} - \lambda_\ell v, \tag{52}$$

$$\Delta w = -\lambda_\ell w, \tag{53}$$

which is (5) and (6) for the case $Q = 0$, $\lambda = -\lambda_\ell$. If we can find a solution $v \neq 0$ of $I_w'(v) = 0$, then we shall have a solution $u = (v, w)$ of (5) and (6) with $v \neq 0$, $w \neq 0$. This was done in Theorem 3 of [19].

The proof of Theorem 6 is similar to that of Theorem 5 and is omitted.

References

1. G. Bartal, O. Manela, O. Cohen, J.W. Fleischer, M. Segev, Observation of second-band vortex solitons in 2D photonic lattices. Phys. Rev. Lett. **95**, 053904 (2005)
2. S. Chen, Y. Lei, Existence of steady-state solutions in a nonlinear photonic lattice model. J. Math. Phys. **52**(6), 063508 (2011)
3. W. Chen, D.L. Mills, Gap solitons and the nonlinear optical response of superlattices. Phys. Rev. Lett. **62**, 1746–1749 (1989)
4. N.K. Efremidis, S. Sears, D.N. Christodoulides, Discrete solitons in photorefractive optically-induced photonic lattices. Phys.Rev.Lett. **85**, 1863–1866 (2000)
5. W.J.W. Fleischer, M. Segev, N.K. Efremidis, D.N. Christodolides, Observation of two-dimensional discrete solitons in optically induced nonlinear photonic lattices. Nature **422**(6928), 147–149 (2003)
6. J.W. Fleischer, G. Bartal, O. Cohen, O. Manela, M. Segev, J. Hudock, D.N. Christodoulides, Observation of vortex-ring discrete solitons in photonic lattices. Phys. Rev. Lett. **92**, 123904 (2004)
7. P. Kuchment, The mathematics of photonic crystals, in *Mathematical Modeling in Optical Science*. Frontiers Application of Mathematical, vol. 22 (SIAM, Philadelphia, 2001), pp. 207–272
8. C. Liu, Q. Ren, On the steady-state solutions of a nonlinear photonic lattice model. J. Math. Phys. **56**, 031501, 1–12 (2015). https://doi.org/10.1063/1.4914333
9. H. Martin, E.D. Eugenieva, Z. Chen, Discrete solitons and soliton-induced dislocations in partially coherent photonic lattices. Martin et al. Phys. Rev. Lett. **92**, 123902 (2004)
10. D.N. Neshev, T.J. Alexander, E.A. Ostrovskaya, Y.S. Kivshar, H. Martin, I. Makasyuk, Z. Chen, Observation of discrete vortex solitons in optically induced photonic lattices. Phys. Rev. Lett. **92**, 123903 (2004)
11. A. Pankov, Periodic nonlinear Schrödinger equation with application to photonic crystals. Milan J. Math. **73**, 259–287 (2005)
12. M. Schechter, *Linking Methods in Critical Point Theory* (Birkhauser, Boston, 1999)
13. M. Schechter, An introduction to nonlinear analysis, in *Cambridge Studies in Advanced Mathematics*, vol. 95 (Cambridge University, Cambridge, 2004)
14. M. Schechter, The use of Cerami sequences in critical point theory theory. Abstr. Appl. Anal. **2007**, 28 (2007). Art. ID 58948
15. M. Schechter, *Minimax Systems and Critical Point Theory* (Birkhauser, Boston, 2009)
16. M. Schechter, Steady state solutions for Schr/"odinger equations governing nonlinear optics. J. Math. Phys. **53**, 043504, 8 pp. (2012)
17. M. Schechter, Photonic lattices. J. Math. Phys. **54**, 061502, 7 pp. (2013)
18. M. Schechter, *Critical Point Theory, Sandwich and Linking Systems* (Birkhauser, Boston, 2020)
19. M. Schechter, Schrodinger equations in nonlinear optics, in *Nonlinear Analysis and Global Optimization*, ed. by Th. M. Rassias, P.M. Pardalos (Spriger, 2021), pp. 449–459
20. Y. Yang, *Solition in Field Theory and Nonlinear Analysis* (Springer, New York, 2001)
21. Y. Yang, R. Zhang, Steady state solutions for nonlinear Schrödinger equation arising in optics. J. Math. Phys. **50**, 053501–053509 (2009)
22. J. Yang, A. Bezryadina, Z. Chen, I. Makasyuk, Observation of two-dimensional lattice vector solitons. Opt. Lett. **29**, 1656 (2004)
23. J. Yang, I. Makasyuk, A. Bezryadina, Z. Chen, Dipole and quadrupole solitons in optically induced two-dimensional photonic lattices: theory and experiment. Studies Appl. Math. **113**, 389–412 (2004)

The Semi-discrete Method for the Approximation of the Solution of Stochastic Differential Equations

Ioannis S. Stamatiou

Abstract We study the numerical approximation of the solution of stochastic differential equations (SDEs) that do not follow the standard smoothness assumptions. In particular, we focus on SDEs that admit solutions which take values in a certain domain; examples of these equations appear in various fields of application such as mathematical finance and natural sciences among others, where the quantity of interest may be the interest rate, which takes non-negative values, or the population dynamics which takes values between zero and one. We review the Semi-Discrete method (SD), a numerical method that has the qualitative feature of domain preservation among other desirable properties.

1 Introduction

We are interested in the numerical approximation of stochastic differential equations (SDEs) that admit solutions in a certain domain and do not satisfy the usual assumptions. Such equations appear in mathematical finance, e.g. interest rate models, but also in other fields of applications such as natural and social sciences. Generally speaking, explicit solutions of these SDEs are unknown, so numerical methods have to be used to simulate them. While numerical methods exist that converge strongly to the true solution of SDEs with non-standard coefficients, few of them are able to maintain the solution process domain. Implicit methods can in some cases succeed in that direction, but they are usually more time-consuming. Let us state the problem in mathematical terms.

Throughout, let $T > 0$ and $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{0\le t\le T}, \mathbb{P})$ be a complete probability space, meaning that the filtration $\{\mathscr{F}_t\}_{0\le t\le T}$ is right continuous and $\mathscr{F}_0$ includes all $\mathbb{P}$-null sets. We are interested in the following SDE in integral form

I. S. Stamatiou (✉)
Department of Biomedical Sciences, University of West Attica, Athens, Greece
e-mail: istamatiou@uniwa.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_23

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s, \quad t \in [0, T], \tag{1}$$

where $W_{t,\omega} : [0, T] \times \Omega \to \mathbb{R}^m$ is an m-dimensional Wiener process adapted to the filtration $\{\mathcal{F}_t\}_{0 \le t \le T}$, the drift coefficient $a : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and the diffusion coefficient $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times m}$ are measurable functions such that (1) has a unique strong solution and X_0 is independent of all $\{W_t\}_{0 \le t \le T}$. SDE (1) has non-autonomous coefficients, i.e. $a(t, x), b(t, x)$ depend explicitly on t. More precisely, we assume the existence of a predictable stochastic process $X : [0, T] \times \Omega \to \mathbb{R}^d$ such that, c.f. [22, Def. 5.2.1], [24, Def. 2.1],

$$\{a(t, X_t)\} \in \mathscr{L}^1([0, T]; \mathbb{R}^d), \quad \{b(t, X_t)\} \in \mathscr{L}^2([0, T]; \mathbb{R}^{d \times m})$$

and

$$\mathbb{P}\left[X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s\right] = 1, \quad \text{for every } t \in [0, T].$$

The drift coefficient a is the infinitesimal mean of the process (X_t) and the diffusion coefficient b is the infinitesimal standard deviation of the process (X_t). SDEs like (1) have rarely explicit solutions so numerical approximations are required for path simulations of the solution process $X_t(\omega)$.

We are interested in strong approximations (mean-square) of (1), in the case of nonlinear drift and diffusion coefficients. Strongly converging numerical schemes have applications in many areas, such as simulating scenarios, filtering or visualizing stochastic dynamics (c.f [20, Sec. 4] and references therein), they are of theoretical interest (they provide basic insight into weak-sense schemes) and usually do not require simulations over long-time periods or of a significant number of trajectories. In the same time we aim for numerical methods that preserve the domain of the original process, or as we say possess an *eternal life time*.

Definition 1 (Eternal Life Time of Numerical Solution) Let $D \subseteq \mathbb{R}^d$ and consider a process (X_t) well defined on the domain $\overline{D}$, with initial condition $X_0 \in \overline{D}$ and such that

$$\mathbb{P}(\{\omega \in \Omega : X(t, \omega) \in \overline{D}\}) = 1,$$

for all $t > 0$. A numerical solution $(Y_{t_n})_{n \in \mathbb{N}}$ has an *eternal life time* if

$$\mathbb{P}(Y_{t_{n+1}} \in \overline{D} \,\big|\, Y_{t_n} \in \overline{D}) = 1.$$

Let us consider the following nonlinear model both in the drift and diffusion coefficient:

$$x_t = x_0 + \int_0^t (\alpha x_s - \beta x_s^2)ds + \int_0^t \sigma x_s^{3/2} dW_s, \quad t \in [0, T], \tag{2}$$

where x_0 is independent of $\{W_t\}_{0\le t\le T}$, $x_0 > 0$ a.s. and $\sigma \in \mathbb{R}$. SDE (2) is referred to as the 3/2-model [18] or the inverse square root process [1] and is used for modeling stochastic volatility. The conditions $\alpha > 0$ and $\beta > 0$ are necessary and sufficient for the stationarity of the process (x_t) and such that neither zero nor infinity is attainable in finite time [1, App. A].

A "good" numerical scheme for the approximation of the solution of an SDE that takes positive values, as (2), should preserve positivity, c.f. [2, 21]. The explicit Euler scheme does not have that property, since its increments are conditionally Gaussian and therefore there is a positive probability of producing negative values. We refer, among other papers, to [23] that considers Euler type schemes, modifications of them to overcome the above drawback, and the importance of positivity.

SDE (2) is a special case of super-linear models of the form (1) where one of the coefficients $a(\cdot), b(\cdot)$ is super-linear, i.e. when we have that

$$a(x) \ge \frac{|x|^\beta}{C},\ b(x) \le C|x|^\alpha,\ \text{for every} \quad |x| \ge C, \tag{3}$$

or

$$b(x) \ge \frac{|x|^\beta}{C},\ a(x) \le C|x|^\alpha,\ \text{for every} \quad |x| \ge C, \tag{4}$$

where $\beta > 1, \beta > \alpha \ge 0, C > 0$.

Another issue that arises at the numerical approximation of super-linear problems like (3) or (4), is that the moments of the scheme may explode, see [19, Th. 1]. A method that overcomes this drawback is the tamed Euler method, which reads in a general from

$$Y^N_{n+1}(\omega) := Y^N_n(\omega) + a_\Delta(Y^N_n(\omega)) \cdot \Delta + b_\Delta(Y^N_n(\omega))\Delta W_n(\omega), \tag{5}$$

for every $n \in \{0, 1, \dots, N-1\}, N \in \mathbb{N}$ and all $\omega \in \Omega$ where $\Delta W_n(\omega) := W_{\frac{(n+1)T}{N}}(\omega) - W_{\frac{nT}{N}}(\omega)$ are the increments of the Wiener process, $Y^N_0(\omega) := x_0(\omega)$ and the control functions are such that $a_\Delta \to a$ and $b_\Delta \to b$ as $\Delta \to 0$, c.f [20, (4)], [31, Rel. (3.1)], [27], for various choices of a_Δ and b_Δ. These balanced type schemes are explicit, do not explode in finite time and converge strongly to the exact solution. Nevertheless, in general they do not preserve positivity. We should also mention here other interesting implicit methods, c.f. [26] and [25], which are unfortunately time-consuming.

We study SDEs of the general type (1) with solutions in a certain domain and our aim is to construct explicit numerical schemes which on the one hand, converge strongly to the solution process and on the other, preserve the domain of the original SDE.

The semi-discrete (SD) method, originally proposed in [7], has all the above properties and more, that is:

- it is explicit in general and therefore does not require a lot of computational time,
- it does not explode in non-linear problems, see [8, Sec. 3], [15, Sec. 4], [11]

- it strongly converges to the exact solution of the original SDE, [7, Sec. 3], [10–15, 28, 29]
- has the qualitative property of domain preservation, [7, Sec. 3.2], [10, 12–14], [15, Sec. 4], [11, 28, 29]
- preserves monotonicity, [7, Sec. 3.1]
- preserves the a.s. asymptotic stability of the underlying SDE, [16].

2 The Semi-discrete Method: Setting and General Results

We address first the scalar differential equation (1), that is the one-dimensional case ($d = 1$), which we rewrite here

$$x_t = x_0 + \int_0^t a(s, x_s)ds + \int_0^t b(s, x_s)dW_s, \quad t \in [0, T]. \tag{6}$$

Consider the equidistant partition $0 = t_0 < t_1 < \ldots < t_N = T$ with step-size $\Delta = T/N$. We assume that there is a unique strong solution a.s. to the following SDE

$$y_t = y_{t_n} + \int_{t_n}^t f(t_n, s, y_{t_n}, y_s)ds + \int_{t_n}^t g(t_n, s, y_{t_n}, y_s)dW_s, \quad t \in (t_n, t_{n+1}], \tag{7}$$

for every $n \in \mathbb{N}, n \leq N - 1$, with $y_0 = x_0$. Here, the auxiliary functions f and g satisfy the following assumption.

Assumption 2.1 *Let $f(s, r, x, y), g(s, r, x, y) : [0, T]^2 \times \mathbb{R}^2 \to \mathbb{R}$ be such that $f(s, s, x, x) = a(s, x), g(s, s, x, x) = b(s, x)$, where f, g satisfy the following conditions:*

$$|f(s_1, r_1, x_1, y_1) - f(s_2, r_2, x_2, y_2)| \leq C_R\,(|s_1 - s_2| + |r_1 - r_2| + |x_1 - x_2| + |y_1 - y_2|)$$
$$|g(s_1, r_1, x_1, y_1) - g(s_2, r_2, x_2, y_2)| \leq C_R\,(|s_1 - s_2| + |r_1 - r_2| + |x_1 - x_2| + |y_1 - y_2| + \sqrt{|x_1 - x_2|}\Big),$$

for any $R > 0$ such that $|x_1| \vee |x_2| \vee |y_1| \vee |y_2| \leq R$, where the constant C_R depends on R and $x \vee y$ denotes the maximum of x, y.

We consider the following interpolation process of the semi-discrete approximation, in a compact form,

$$y_t = y_0 + \int_0^t f(\hat{s}, s, y_{\hat{s}}, y_s)ds + \int_0^t g(\hat{s}, s, y_{\hat{s}}, y_s)dW_s, \tag{8}$$

where $\hat{s} = t_n$ when $s \in [t_n, t_{n+1})$. In that way we may compare with the exact solution x_t, which is a continuous time process. The first and third variable in f, g denote the discretized part of the original SDE. We observe from (8) that in order to solve for y_t, we have to solve, in general, an SDE and not an algebraic equation. We can reproduce the Euler scheme if we choose $f(s, r, x, y) = a(s, x)$ and $g(s, r, x, y) = b(s, x)$. The semi-discrete method (8) can be appropriately modified to produce an implicit scheme that is explicitly and easily solved if necessary (see [11, 14, 29]).

In the case of superlinear coefficients the numerical scheme (8) converges to the true solution x_t of SDE (6) and this is stated in the following, see [15, Th. 2.1].

Theorem 1 (Strong Convergence) *Suppose Assumption 2.1 holds and (7) has a unique strong solution for every $n \leq N - 1$, where $x_0 \in \mathscr{L}^p(\Omega, \mathbb{R})$. Let also*

$$\mathbb{E}(\sup_{0\leq t\leq T} |x_t|^p) \vee \mathbb{E}(\sup_{0\leq t\leq T} |y_t|^p) < A,$$

for some $p > 2$ and $A > 0$. Then the semi-discrete numerical scheme (8) converges to the true solution of (6) in the $\mathscr{L}^2$-sense, that is

$$\lim_{\Delta\to 0} \mathbb{E} \sup_{0\leq t\leq T} |y_t - x_t|^2 = 0. \tag{9}$$

Theorem 1 is an extension of [8, Th. 1] to time-dependent coefficients which covers super-linear diffusion coefficients, like for example of the form $b(t, x) = \beta(t) \cdot x^{3/2}$. In all other cases we may assume the usual local Lipschitz assumption for both f and g.

We understand by the general form of decomposition (7) that we may produce many different semi-discrete numerical schemes. In a sense the method is problem dependent, since the form of the drift and diffusion coefficients, a and b, of the original SDE suggest the way of discretization. We will see in the following Sections 3 and 4 applications of the semi-discrete method which all have in common the qualitative property of domain preservation.

Relation (9) does not reveal the order of convergence. In order to show the order of convergence, we work with a truncated version of the SD method, see [30].

We choose a strictly increasing function $\mu : \mathbb{R}_+ \to \mathbb{R}_+$ such that for every $s, r \leq T$

$$\sup_{|x|\leq u} (|f(s, r, x, y)| \vee |g(s, r, x, y)|) \leq \mu(u)(1 + |y|), \qquad u \geq 1. \tag{10}$$

The inverse function of μ, denoted by μ^{-1}, maps $[\mu(1), \infty)$ to $\mathbb{R}_+$. Moreover, we choose a strictly decreasing function $h : (0, 1] \to [\mu(1), \infty)$ and a constant $\hat{h} \geq 1 \vee \mu(1)$ such that

$$\lim_{\Delta\to 0} h(\Delta) = \infty \quad \text{and} \quad \Delta^{1/6} h(\Delta) \leq \hat{h} \quad \text{for every} \quad \Delta \in (0, 1]. \tag{11}$$

Let $\Delta \in (0, 1]$ and f_Δ, g_Δ defined by

$$\phi_\Delta(s, r, x, y) := \phi\left(s, r, (|x| \wedge \mu^{-1}(h(\Delta)))\frac{x}{|x|}, y\right), \tag{12}$$

for $x, y \in \mathbb{R}$ where we set $x/|x| = 0$ when $x = 0$. Using the truncated auxiliary functions f_Δ and g_Δ we may redefine SDEs (7) and (8), which now read

$$y_t^\Delta = y_{t_n}^\Delta + \int_{t_n}^t f_\Delta(t_n, s, y_{t_n}^\Delta, y_s^\Delta)ds + \int_{t_n}^t g_\Delta(t_n, s, y_{t_n}^\Delta, y_s^\Delta)dW_s, \quad t \in (t_n, t_{n+1}], \tag{13}$$

and

$$y_t^\Delta = y_0 + \int_0^t f_\Delta(\hat{s}, s, y_{\hat{s}}^\Delta, y_s^\Delta)ds + \int_0^t g_\Delta(\hat{s}, s, y_{\hat{s}}^\Delta, y_s^\Delta)dW_s. \tag{14}$$

respectively, with $y_0 = x_0$ a.s.

Assumption 2.2 *Let the truncated versions $f_\Delta(s, r, x, y)$, $g_\Delta(s, r, x, y)$ of f, g satisfy the following condition ($\phi_\Delta \equiv f_\Delta, g_\Delta$)*

$$|\phi_\Delta(s_1, r_1, x_1, y_1) - \phi_\Delta(s_2, r_2, x_2, y_2)| \leq h(\Delta)\Big(|s_1 - s_2| + |r_1 - r_2| + |x_1 - x_2| + |y_1 - y_2|\Big)$$

for all $0 < \Delta \leq 1$ and $x_1, x_2, y_1, y_2 \in \mathbb{R}$, where $h(\Delta)$ is as in (11).

Let us also assume that the coefficients $a(t, x)$, $b(t, x)$ of the original SDE satisfy the Khasminskii-type condition.

Assumption 2.3 *We assume the existence of constants $p \geq 2$ and $C_K > 0$ such that $x_0 \in \mathscr{L}^p(\Omega, \mathbb{R})$ and*

$$xa(t, x) + \frac{p-1}{2}b(t, x)^2 \leq C_K(1 + |x|^2)$$

for all $(t, x) \in [0, T] \times \mathbb{R}$.

Under the local Lipschitz and the Khasminskii-type condition SDE (6) has a unique solution and finite moment bounds of order p, c.f. [24], i.e. for all $T > 0$, there exists a constant $A > 0$ such that $\sup_{0 \leq t \leq T} \mathbb{E}|x_t|^p < A$. We rewrite the main result [30, Th. 3.1].

Theorem 2 (Order of Strong Convergence) *Suppose Assumption 2.2 and Assumption 2.3 hold and (13) has a unique strong solution for every $n \leq N - 1$, where $x_0 \in \mathscr{L}^p(\Omega, \mathbb{R})$ for some $p \geq 14 + 2\gamma$. Let $\epsilon \in (0, 1/3)$ and define for $\gamma > 0$*

$$\mu(u) = \overline{C}u^{1+\gamma}, \quad u \geq 0 \quad \text{and} \quad h(\Delta) = \overline{C} + \sqrt{\ln \Delta^{-\epsilon}}, \quad \Delta \in (0, 1],$$

where $\Delta \leq 1$ *and* $\hat{h}$ *are such that (11) holds. Then the semi-discrete numerical scheme (14) converges to the true solution of (6) in the* $\mathscr{L}^2$*-sense with order arbitrarily close to* $1/2$*, that is*

$$\mathbb{E} \sup_{0\leq t\leq T} |y_t^{\Delta} - x_t|^2 \leq C\Delta^{1-\epsilon}. \tag{15}$$

3 Applications of the Semi-discrete Method: Mathematical Finance

3.1 3/2-*Model*

Let us first consider the more general 3/2-model (2) with super-linear drift and diffusion coefficients, see [15, Sec. 4.1],

$$x_t = x_0 + \int_0^t (k_1(s)x_s - k_2(s)x_s^2)ds + \int_0^t k_3(s)x_s^{3/2}\phi(x_s)dW_s, \quad t \in [0, T], \tag{16}$$

where $\phi(\cdot)$ is a locally Lipschitz and bounded function with locally Lipschitz constant C_R^{ϕ}, bounding constant K_{ϕ}, x_0 is independent of all $\{W_t\}_{0\leq t\leq T}$, $x_0 \in \mathscr{L}^{4p}(\Omega, \mathbb{R})$ for some $2 < p$ and $x_0 > 0$ a.s., $\mathbb{E}(x_0)^{-2} < A$, $k_1(\cdot), k_2(\cdot), k_3(\cdot)$ are positive and bounded functions with $k_{2,\min} > \frac{7}{2}(K_{\phi}k_{3,\max})^2$. It holds that $x_t > 0$ a.s. The following semi-discrete numerical scheme,

$$y_t = y_0 + \int_0^t (k_1(s) - k_2(s)y_{\hat{s}})y_s ds + \int_0^t k_3(s)\sqrt{y_{\hat{s}}}\phi(y_{\hat{s}})y_s dW_s, \tag{17}$$

where $\hat{s} = t_n$, when $s \in [t_n, t_{n+1})$, produces a linear SDE with solution

$$y_t = x_0 \exp\left\{\int_0^t \left(k_1(s) - k_2(s)y_{\hat{s}} - k_3^2(s)\frac{y_{\hat{s}}\phi^2(y_{\hat{s}})}{2}\right)ds + \int_0^t k_3(s)\sqrt{y_{\hat{s}}}\phi(y_{\hat{s}})dW_s\right\}, \tag{18}$$

where $y_t = y_t(t_0, x_0)$. We call (18) an exponential semi-discrete approximation of (16). The exponential semi-discrete numerical scheme (18) converges to the true solution of (16) in the mean square sense, is positive and has finite moments $\mathbb{E}(\sup_{0\leq t\leq T}(y_t)^p)$ for appropriate p, see [15, Sec. 4.1]. See also the very recent work [17], a combination of the Lamperti transformation with the SD method, named LSD method.

3.2 CEV Process

The following SDE

$$x_t = x_0 + \int_0^t (k_1 - k_2 x_s)ds + \int_0^t k_3 (x_s)^q dW_s, \quad t \in [0, T], \tag{19}$$

where k_1, k_2, k_3 are positive and $1/2 < q < 1$ is known as a mean-reverting CEV process. Equation (19) may represent the instantaneous volatility or the instantaneous variance of the underlying financially observable. Here the diffusion coefficient is sub-linear. Feller's test implies that there is a unique non-explosive strong solution such that $x_t > 0$ a.s. when $x_0 > 0$ a.s. c.f. [22, Prop. 5.22]. The steady-state level of x_t is k_1/k_2 and the rate of mean-reversion is k_2.

Here we examine two versions of an implicit SD scheme that are solved explicitly. In [14], we propose

$$y_t = x_0 + \int_0^t (k_1 - k_2(1-\theta)y_{\hat{s}} - k_2\theta y_{\tilde{s}})\, ds + k_3 \int_0^t (y_{\hat{s}})^{q-\frac{1}{2}} \sqrt{y_s} dW_s$$
$$+ \int_t^{t_{n+1}} \left(k_1 - k_2(1-\theta)y_{t_n} - \frac{(k_3)^2}{4(1+k_2\theta\Delta)} (y_{t_n})^{2q-1} - k_2\theta y_t \right) ds, \tag{20}$$

for $t \in (t_n, t_{n+1}]$ where

$$\hat{s}{=}t_j, s \in (t_j, t_{j+1}],\ j{=}0, \ldots, n, \quad \tilde{s}{=} \begin{cases} t_{j+1}, & \text{for } s \in [t_j, t_{j+1}], \\ t, & \text{for } s \in [t_n, t], \end{cases} j{=}0, \ldots, n{-}1$$

and $\theta \in [0, 1]$ represents the level of implicitness. After rearranging

$$y_t(q){=}y_n{+}\int_{t_n}^t \frac{(k_3)^2}{4(1+k_2\theta\Delta)^2} (y_{t_n})^{2q-1} ds{+}\frac{k_3}{1+k_2\theta\Delta} (y_{t_n})^{q-\frac{1}{2}} \int_{t_n}^t \mathrm{sgn}(z_s)\sqrt{y_s} dW_s, \tag{21}$$

with solution

$$y_t(q) = (z_t)^2, \quad z_t := \sqrt{y_n} + \frac{k_3}{2(1+k_2\theta\Delta)} (y_{t_n})^{q-\frac{1}{2}} (W_t - W_{t_n}), \tag{22}$$

where y_n is

$$y_n := y_{t_n} \left(1 - \frac{k_2\Delta}{1+k_2\theta\Delta} \right) + \frac{k_1\Delta}{1+k_2\theta\Delta} - \frac{(k_3)^2}{4(1+k_2\theta\Delta)^2} (y_{t_n})^{2q-1} \Delta.$$

The SD method (22) is positive by construction and under some conditions on the coefficients k_i, the level of implicitness θ and the step-size Δ, it strongly converges to the solution of (19) with a logarithmic rate if also $\mathbb{E}(x_0)^p < A$ for some $p \geq 4$

and with a polynomial rate of convergence of magnitude $\frac{1}{2}(q-\frac{1}{2})$ if $x_0 \in \mathbb{R}$, see [14, Th.1 and Th.2]. The other version of the implicit SD scheme, see [12], is written in each sub-interval,

$$\widetilde{y}_t(q) = \widetilde{y}_n + \int_{t_n}^{t} \frac{q(k_3)^2}{2}(\widetilde{y}_s)^{2q-1} ds + k_3 \int_{t_n}^{t} \operatorname{sgn}(\widetilde{z}_s)(\widetilde{y}_s)^q dW_s \tag{23}$$

with solution

$$\widetilde{y}_t(q) = |\widetilde{z}_t|^{1/(1-q)}, \quad \widetilde{z}_t := (\widetilde{y}_n)^{1-q} + k_3(1-q)(W_t - W_{t_n}), \tag{24}$$

where

$$\widetilde{y}_n := \widetilde{y}_{t_n}(1 - k_2\Delta) + k_1\Delta - \frac{q(k_3)^2\Delta}{2}(\widetilde{y}_{t_n})^{2q-1}.$$

The SD method (24) is again positive by construction and under some conditions on the coefficients k_i, the level of implicitness θ and the step-size Δ, it strongly converges to the solution of (19) with a polynomial rate of convergence of magnitude $q(q-\frac{1}{2})$ if $x_0 \in \mathbb{R}$. See also how LSD performs [17].

3.3 CIR/CEV Delay Models with Jump

Here we study a general model of type (19) including delay and jump terms. In particular we consider the following stochastic delay differential equation (SDDE) with jump,

$$x_t = \begin{cases} \xi_0 + \int_0^t (k_1 - k_2 x_{s^-})ds + \int_0^t k_3 b(x_{s-\tau}) x_{s^-}^{\alpha} dW_s + \int_0^t g(x_{s^-}) d\widetilde{N}_s, & t \in [0, T], \\ \xi(t), & t \in [-\tau, 0], \end{cases} \tag{25}$$

where $x_{s^-} = \lim_{r \uparrow s} x_r$, the coefficient $b \in \mathscr{C}(\mathbb{R}_+, \mathbb{R}_+)$[1] and is assumed to be γ-Hölder continuous with $\gamma > 0$, the jump coefficient $g : \mathbb{R} \mapsto \mathbb{R}$ is assumed deterministic for simplicity, the function $\xi \in \mathscr{C}([-\tau, 0], (0, \infty))$ and $\tau > 0$ is a positive constant which represents the delay. Process $\widetilde{N}(t) = N(t) - \lambda t$ a compensated Poisson process with intensity $\lambda > 0$ independent of W_t. (25) has a unique and nonnegative solution and under some conditions on $\|\xi\|$ and the step-size Δ the following scheme strongly converges to the solution of (25) with polynomial or logarithmic rate, see [29],

[1] $\mathscr{C}(A, B)$ the space of continuous functions $\phi : A \mapsto B$ with norm $\|\phi\| = \sup_{u \in A} \phi(u)$.

$$\begin{cases} y_{t_{k+1-}} = (z_{t_{k+1}})^2, \\ y_{t_{k+1}} = y_{t_{k+1-}} + g(y_{t_{k+1-}})\Delta\widetilde{N}_k, \end{cases} \tag{26}$$

where

$$z_t = \sqrt{y_{t_k}\left(1 - \frac{k_2\Delta_k}{1+k_2\theta\Delta_k}\right) + \frac{k_1\Delta_k}{1+k_2\theta\Delta_k} - \frac{(k_3)^2}{4(1+k_2\theta\Delta_k)^2}\frac{b^2(y_{t_{k-\tau}})}{(1+b(y_{t_{k-\tau}})\Delta_k^m)^2}(y_{t_k})^{2\alpha-1}\Delta_k}$$
$$+\frac{k_3}{2(1+k_2\theta\Delta_k)}\frac{b(y_{t_{k-\tau}})}{1+b(y_{t_{k-\tau}})\Delta_k^m}(y_{t_k})^{\alpha-\frac{1}{2}}(W_t - W_{t_k})$$

$y_t = \xi(t)$ when $t \in [-\tau, 0]$ and for $k = 0, 1, \ldots, n_T - 1$, and $\Delta_k = t_{k+1} - t_k$, $\Delta\widetilde{N}_k := \widetilde{N}(t_{k+1}) - \widetilde{N}(t_k) = \Delta N_k - \lambda\Delta_k$ and $\theta \in [0, 1]$ represents the level of implicitness, with $m = 1/4$. The SD scheme (26) combines the semi-discrete idea with a taming procedure. For the case $\alpha = 1/2$, known as the CIR model, where no delay and jump terms, see also [7, 11] and the application of the LSD method [17]. For extensions of the SD method to the two-factor CIR, see [10].

3.4 Aït-Sahalia Model

Let

$$x_t = x_0 + \int_0^t (\frac{a_1}{x_s} - a_2 + a_3 x_s - a_4 x_s^r)ds + \sigma\int_0^t x_s^\rho dW_s, \tag{27}$$

where $x_0 > 0$, the coefficients a_i are nonnegative and $r > 1, \rho > 1$. SDE (27), known as the *Aït-Sahalia* model, is used as an interest rate model and satisfies $x_t > 0$ a.s. The approximation of (27), by a combination of the splitting step method and the semi-discrete method, is proposed in [13]. In fact the SD approximation for the transformed process $z_t = x_t^2$ takes place first with dynamics given by

$$z_t = z_0 + \int_0^t (2a_1 z_s - 2a_2\sqrt{z_s} + 2a_3 z_s - 2a_4 z_s^{(r+1)/2} + \sigma^2 z_s^\rho)ds + 2\sigma\int_0^t z_s^{(\rho+1)/2}dW_s. \tag{28}$$

Splitting (28) in each subinterval with $t \in [t_n, t_{n+1}]$ as

$$z_1(t) = z_2(t_n) + \int_{t_n}^t (\ln(4/3)z_1(s) - 2a_2\sqrt{z_1(s)})ds \tag{29}$$

$$z_2(t) = z_1(t_{n+1}) + \int_{t_n}^t (2a_1 + (2a_3 - \ln(4/3))z_2(s) - 2a_4 z_2^{(r+1)/2}(s) + \sigma^2 z_2^\rho(s))ds$$
$$+2\sigma\int_{t_n}^t z_2^{(\rho+1)/2}(s)dW_s, \tag{30}$$

where $z_2(0) = x_0$ suggests that we may take the solution of (29)

$$z_1(t) = \left(\frac{2a_2}{\ln(4/3)} + \left(\sqrt{z_2(t_n)} - \frac{2a_2}{\ln(4/3)} \right) \left(\frac{4}{3} \right)^{(t-t_n)/2} \right)^2 \tag{31}$$

and approximate (30) with

$$\begin{aligned} \widetilde{z}_2(t) &= z_1(t_{n+1}) + 2a_1\Delta \\ &\quad + \int_{t_n}^{t} \left(2a_3 - \ln(4/3) - 2a_4 \widetilde{z}_2^{(r-1)/2}(\hat{s}) + \sigma^2 \widetilde{z}_2^{(\rho-1)/2}(\hat{s}) \right) \widetilde{z}_2(s) ds \\ &\quad + 2\sigma \int_{t_n}^{t} \widetilde{z}_2^{(\rho-1)/2}(\hat{s}) \widetilde{z}_2(s) dW_s. \end{aligned} \tag{32}$$

We end up with the following SD numerical scheme for the transformed process z_t

$$\begin{aligned} \widetilde{z}_{n+1} &= \left(2a_1\Delta + \left(\frac{2a_2}{\ln(4/3)} + \left(\sqrt{\widetilde{z}_n} - \frac{2a_2}{\ln(4/3)} \right) \left(\frac{4}{3} \right)^{\Delta/2} \right)^2 \right) \\ &\quad \times \exp\{(2a_3 - \ln(4/3) - 2a_4 \widetilde{z}_n^{(r-1)/2} - \sigma^2 \widetilde{z}_n^{\rho-2})\Delta + 2\sigma \widetilde{z}_n^{(\rho-1)/2} \Delta W_n\} \end{aligned} \tag{33}$$

and then take $y_n = \sqrt{\widetilde{z}_n}$ for the approximation of the original Aït-Sahalia model, which is positive, strongly convergent with finite moment bounds, when $r+1 > 2\rho$, with $\rho \geq 2$, see [13]. See also the performance of LSD [17].

4 Applications of the Semi-discrete Method: Population Dynamics and Biology

4.1 Wright-Fisher Model

The next class of SDEs appears in population dynamics to describe fluctuations in gene frequency of reproducing individuals among finite populations [5] and ion channel dynamics within cardiac and neuronal cells, (cf. [3, 4, 6] and references therein),

$$x_t = x_0 + \int_0^t (k_1 - k_2 x_s) ds + k_3 \int_0^t \sqrt{x_s(1-x_s)} dW_s, \tag{34}$$

where $k_i > 0, i = 1, 2, 3$. If $x_0 \in (0, 1)$ and $(k_1 \wedge (k_2 - k_1)) \geq (k_3)^2/2$, then $0 < x_t < 1$ a.s. The process

$$y_t = y_{t_n} + \int_{t_n}^{t_{n+1}} \left(k_1 - \frac{(k_3)^2}{4} + y_{t_n}\left(\frac{(k_3)^2}{2} - k_2\right)\right) ds + \int_{t_n}^{t} \frac{(k_3)^2}{4}(1 - 2y_s)ds$$
$$+k_3 \int_{t_n}^{t} \sqrt{y_s(1 - y_s)}\, \mathrm{sgn}(z_s) dW_s, \tag{35}$$

for $t \in (t_n, t_{n+1}]$, with $y_0 = x_0$ a.s. and $z_t = \sin\left(k_3 \Delta W_n^t + 2\arcsin(\sqrt{y_n})\right)$, where $y_n := y_{t_n} + \left(k_1 - \frac{(k_3)^2}{4} + y_{t_n}\left(\frac{(k_3)^2}{2} - k_2\right)\right) \cdot \Delta$ has the following solution

$$y_t = \sin^2\left(\frac{k_3}{2}\Delta W_n^t + \arcsin(\sqrt{y_n})\right), \tag{36}$$

which has the pleasant feature that $y_t \in (0, 1)$ when $y_0 \in (0, 1)$. Process (36) is well defined when $0 < y_n < 1$, which is achieved for appropriate Δ. To simplify conditions on the parameters and the step size d we may adopt the strategy presented in [28] considering a perturbation of order Δ in the initial condition. Here we used an additive discretization of the drift coefficient and the eternal life time SD scheme (36) strongly converges to the solution of (34), see [28]. Moreover, in [28], an application of the SD method to an extension of the Wright-Fisher model to the multidimensional case is treated, producing a strongly converging and boundary preserving scheme.

4.2 Predator-Prey Model

The following system of SDEs, c.f. [20],

$$X_t^{(1)} = X_0^{(1)} + \int_0^t (aX_s^{(1)} - bX_s^{(1)}X_s^{(2)})ds + \int_0^t k_1 X_s^{(1)} dW_s^{(1)},$$
$$X_t^{(2)} = X_0^{(2)} + \int_0^t (cX_s^{(1)}X_s^{(2)} - dX_s^{(2)})ds + \int_0^t k_2 X_s^{(2)} dW_s^{(2)},$$

where $a, b, c, d > 0$ and $k_1, k_2 \in \mathbb{R}$ with independent Brownian motions $W_t^{(1)}$, $W_t^{(2)}$ was studied in [9]. Under some moment bound conditions for $(X_t^{(i)})$, $i = 1, 2$ and when $X_0^{(1)} > 0$ and $X_0^{(2)} > 0$ then $X_t^{(1)} > 0$ and $X_t^{(2)} > 0$ a.s. Transforming the second equation $Z_t^{(2)} = \ln(X_t^{(2)})$ produces the following system

$$X_t^{(1)} = X_0^{(1)} + \int_0^t (a - be^{Z_s^{(2)}})X_s^{(1)} ds + \int_0^t k_1 X_s^{(1)} dW_s^{(1)},$$
$$Z_t^{(2)} = Z_0^{(2)} + \int_0^t (cX_s^{(1)} - d - (k_2)^2)ds + k_2 W_t^{(2)},$$

which is approximated by the following SD scheme

$$Y_t^{(1)} = X_0^{(1)} + \int_0^t (a - be^{Y_{\hat{s}}^{(2)}}) Y_s^{(1)} ds + \int_0^t k_1 Y_s^{(1)} dW_s^{(1)},$$
$$Y_t^{(2)} = Y_0^{(2)} + \int_0^t (cY_{\hat{s}}^{(1)} - d - (k_2)^2) ds + k_2 W_t^{(2)},$$

which reads

$$Y_{t_{n+1}}^{(1)} = Y_{t_n}^{(1)} \exp\{(a - be^{Y_{t_n}^{(2)}} - \frac{(k_1)^2}{2})\Delta + k_1 \Delta W_n^{(1)}\}$$
$$Y_{t_{n+1}}^{(2)} = Y_{t_n}^{(2)} + (cY_{t_n}^{(1)} - d - (k_2)^2)\Delta + k_2 \Delta W_n^{(2)}.$$

References

1. D-H. Ahn, B. Gao, A parametric nonlinear model of term structure dynamics. Rev. Financ. Stud. **12**(4), 721–762 (1999)
2. J.A.D. Appleby, M. Guzowska, K. Cónall, A. Rodkina, Preserving positivity in solutions of discretised stochastic differential equations. Appl. Math. Comput. **217**(2), 763–774 (2010)
3. C.E. Dangerfield, D. Kay, S. MacNamara, K. Burrage, A boundary preserving numerical algorithm for the Wright-Fisher model with mutation. BIT Numer. Math. **52**(2), 283–304 (2012)
4. C.E. Dangerfield, D. Kay, K. Burrage, Modeling ion channel dynamics through reflected stochastic differential equations. Phys. Rev. E **85**(5), 051907 (2012)
5. W.J. Ewens, *Mathematical Population Genetics 1: Theoretical Introduction*, vol. 27 (Springer Science & Business Media, New York, 2012)
6. J.H. Goldwyn, N.S. Imennov, M. Famulare, E. Shea-Brown, Stochastic differential equation models for ion channel noise in Hodgkin-Huxley neurons. Phys. Rev. E **83**(4), 041908 (2011)
7. N. Halidias, Semi-discrete approximations for stochastic differential equations and applications. Int. J. Comput. Math. **89**(6), 780–794 (2012)
8. N. Halidias, A novel approach to construct numerical methods for stochastic differential equations. Numer. Algorithms **66**(1), 79–87 (2014)
9. N. Halidias, Construction of positivity preserving numerical schemes for some multidimensional stochastic differential equations. Discrete Contin. Dynam. Syst. B **20**(1), 153–160 (2015)
10. N. Halidias, Constructing positivity preserving numerical schemes for the two-factor CIR model. Monte Carlo Methods Appl. **21**(4), 313–323 (2015)
11. N. Halidias, A new numerical scheme for the CIR process. Monte Carlo Methods Appl. **21**(3), 245–253 (2015)
12. N. Halidias, An explicit and positivity preserving numerical scheme for the mean reverting CEV model. Japan J. Ind. Appl. Math. **32**(2), 545–552 (2015)
13. N. Halidias, *On the construction of boundary preserving numerical schemes*. Monte Carlo Methods Appl. **22**, 277–289 (2016)
14. N. Halidias, I.S. Stamatiou, Approximating explicitly the mean-reverting CEV process. J. Probab. Stat. (2015), Article ID 513137, 20 pp.
15. N. Halidias, I.S. Stamatiou, On the numerical solution of some non-linear stochastic differential equations using the semi-discrete method. Comput. Methods Appl. Math. **16**(1), 105–132 (2016)

16. N. Halidias, I.S. Stamatiou, A note on the asymptotic stability of the semi-discrete method for stochastic differential equations (2020). https://arxiv.org/abs/2008.03148
17. N. Halidias, I.S. Stamatiou, Lamperti Semi-Dicrete method (2020). http://arxiv.org/abs/2104.06149
18. S.L. Heston, A simple new formula for options with stochastic volatility, OLIN-97-23 (1997). http://ssrn.com/abstract=86074
19. M. Hutzenthaler, A. Jentzen, P.E. Kloeden, Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. Proc. R. Soc. Lond. A Math. Phys. Eng. Sci. **467**(2130), 1563–1576 (2011)
20. M. Hutzenthaler, A. Jentzen, Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. Memoirs Am. Math. Soc. **236**(1112), 1611–1641 (2015)
21. C. Kahl, M. Günther, T. Rossberg, Structure preserving stochastic integration schemes in interest rate derivative modeling. Appl. Numer. Math. **58**(3), 284–295 (2008)
22. I. Karatzas, S.E. Shreve, *Brownian Motion and Stochastic Calculus* (Springer, New York, 1988)
23. P. Kloeden, A. Neuenkirch, Convergence of numerical methods for stochastic differential equations in mathematical finance, in *Recent Developments in Computational Finance: Foundations, Algorithms and Applications* (2013), pp. 49–80
24. X. Mao, *Stochastic Differential Equations and Applications*, 2nd ed. (Horwood Publishing, Chichester, 2007)
25. X. Mao, L. Szpruch, Strong convergence and stability of implicit numerical methods for stochastic differential equations with non-globally Lipschitz continuous coefficients. J. Comput. Appl. Math. **238**, 14–28 (2013)
26. X. Mao, L. Szpruch, Strong convergence rates for backward Euler–Maruyama method for nonlinear dissipative-type stochastic differential equations with super-linear diffusion coefficients. Stochastics **85**(1), 144–171 (2013)
27. S. Sabanis, Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients. Ann. Appl. Probab. **26**(4), 2083 (2016)
28. I.S. Stamatiou, A boundary preserving numerical scheme for the Wright–Fisher model. J. Comput. Appl. Math. **328**, 132–150 (2018)
29. I.S. Stamatiou, An explicit positivity preserving numerical scheme for CIR/CEV type delay models with jump. J. Comput. Appl. Math. **360**, 78–98 (2019)
30. I.S. Stamatiou, N. Halidias, Convergence rates of the semi-discrete method for stochastic differential equations. Theory Stoch. Process. **24**(40), 89–100 (2019)
31. M.V. Tretyakov, Z. Zhang, A fundamental mean-square convergence theorem for SDEs with locally Lipschitz coefficients and its applications. SIAM J. Numer. Anal. **51**(6), 3135–3162 (2013)

Homotopic Metric-Interval L-Contractions in Gauge Spaces

Mihai Turinici

Abstract A functional version—under the lines in Leader [Math. Japonica, 24 (1979), 17–24]—is given for the 1967 contraction mapping principle in Gheorghiu [Stud. Cerc. Mat., 19 (1967), 119–122]. As a by-product of this, an appropriate functional extension is proposed for the homotopic fixed point result in gauge spaces due to Frigon [L. Notes Nonlin. Anal., 16 (2017), 9–91].

1 Introduction

Let X be a nonempty set. Call the subset Y of X, *almost-singleton* (in short: *asingleton*), provided $[y_1, y_2 \in Y$ implies $y_1 = y_2]$; and *singleton* if, in addition, Y is nonempty; note that in this case $Y = \{y\}$, for some $y \in X$.

Further, let $d : X \times X \to R_+ := [0, \infty[$ be a *metric* over X; and take some $T \in \mathscr{F}(X)$. [Here, given the nonempty sets A and B, $\mathscr{F}(A, B)$ stands for the class of all functions from A to B; when $A = B$, we write $\mathscr{F}(A, A)$ as $\mathscr{F}(A)$]. Denote $\mathrm{Fix}(T) := \{z \in X; z = Tz\}$; any point of it will be called *fixed* under T. These points are to be determined in the context below (cf. Rus [50, Ch 2, Sect 2.2]):

(pic-1) We say that T is *fix-asingleton*, if $\mathrm{Fix}(T)$ is an asingleton; and *fix-singleton*, if $\mathrm{Fix}(T)$ is a singleton

(pic-2) We say that $x \in X$ is a *Picard point* (modulo (d, T)) when $(T^n x; n \geq 0)$ is d-Cauchy. If this property holds for all $x \in X$, we say that T is a *Picard operator* (modulo d)

AMS Subject Classification: 47H10 (Primary), 54H25 (Secondary).

M. Turinici (✉)
A. Myller Mathematical Seminar, A. I. Cuza University, Iaşi, Romania
e-mail: mturi@uaic.ro

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_24

(pic-3) We say that $x \in X$ is a *strongly Picard point* (modulo (d, T)) when $(T^n x; n \geq 0)$ is d-convergent with $\lim_n (T^n x) \in \mathrm{Fix}(T)$. If this property holds for all $x \in X$, we say that T is a *strongly Picard operator* (modulo d).

In this perspective, a basic answer to the posed question [referred to as *Banach contraction principle*; in short: (B-cp)], may be stated as follows. Given $k \geq 0$, let us say that T is *(Banach)* $(d; k)$*-contractive*, provided

$$\text{(B-contr)} \quad d(Tx, Ty) \leq k d(x, y), \quad \text{for all } x, y \in X.$$

Theorem 1 *Suppose that T is $(d; k)$-contractive, for some $k \in [0, 1[$. In addition, let X be d-complete. Then,*

(11-a) T is fix-singleton: $\mathrm{Fix}(T) = \{z\}$*, for some $z \in X$*

(11-b) T is a strongly Picard operator (modulo d): $T^n x \xrightarrow{d} z$, $\forall x \in X$.

This result, obtained in 1922 by Banach [7], found a multitude of applications in operator equations theory; so, it was the subject of many extensions. Essentially, there are two main directions of doing this.

(I) The initial metric d remains as it is, but the contractive condition is taken in the implicit way

$$\text{(i-contr)} \quad F(d(Tx, Ty), d(x, y), d(x, Tx), d(y, Ty), d(x, Ty), d(Tx, y)) \leq 0,$$
$$\text{for all } x, y \in X, x \mathscr{R} y$$

where $F : R_+^6 \to R$ is a function and $\mathscr{R} \subseteq X \times X$ is a *relation* over X. When the function F appearing here admits the explicit form

$$F(t_1, t_2, t_3, t_4, t_5, t_6) = t_1 - G(t_2, t_3, t_4, t_5, t_6), (t_1, t_2, t_3, t_4, t_5, t_6) \in R_+^6$$

(where $G : R_+^5 \to R_+$ is a function), one gets the explicit functional version of this (functional) contraction

$$\text{(e-contr)} \quad d(Tx, Ty) \leq G(d(x, y), d(x, Tx), d(y, Ty), d(x, Ty), d(Tx, y)),$$
$$\text{for all } x, y \in X, x \mathscr{R} y.$$

In particular, when $\mathscr{R} = X \times X$ (the *trivial relation* over X), some outstanding explicit results have been established in Boyd and Wong [11], Reich [48], and Matkowski [40]; see also the survey paper by Rhoades [49]. And, for the implicit setting above, certain technical aspects have been considered by Leader [39] and Turinici [54]. On the other hand, when $\mathscr{R}$ is a *(partial) order* on X, some appropriate extensions of Matkowski fixed point principle we just quoted were obtained in the 1986 papers by Turinici [60, 61]; two decades later, these results have been re-discovered—at the level of (Banach) contractive maps—by

Ran and Reurings [47]; see also Nieto and Rodriguez-Lopez [45]. Further, an extension—to the same framework—of Leader's contribution was performed in Agarwal et al. [2] and Turinici [65]. Finally, when $\mathscr{R}$ is a (general) relation, some results in this direction were obtained by Jachymski [34], within a graph setting, and by Samet and Turinici [51] under a general perspective.

(II) The initial metric d is substituted by a separated family $D = (d_\lambda; \lambda \in \Lambda)$ of semimetrics on X; and the contractive condition is to written as

$$\text{(G-contr)}\ \ d_\lambda(Tx, Ty) \leq k_\lambda d_{\varphi(\lambda)}(x, y),\ \text{ for all } x, y \in X,\ \text{ and all } \lambda \in \Lambda;$$

where $(k_\lambda; \lambda \in \Lambda)$ is a family of positive numbers and $\varphi : \Lambda \to \Lambda$ is a mapping. The first fixed point principle in this direction was obtained in 1967 by Gheorghiu [25]. Further refinement of it were obtained in Gheorghiu and Rotaru [27]; see also Gheorghiu [26]. Note that, by the very intricate form of contractive condition, genuine functional extensions of this principle were not yet obtained.

Recently, an interesting application of Gheorghiu contraction principle to the homotopic fixed point theory was obtained in Frigon [23]. It is our aim in the following to state a functional refinement of this result. We use here a metrical maximality principle as well as a metric interval, in contrast with—respectively—the Zorn-Bourbaki maximal one [9] and the standard (real) interval considered in that paper. Further aspects will be delineated elsewhere.

2 Conv-Cauchy Structures

Throughout this exposition, the axiomatic system in use is Zermelo-Fraenkel's (abbreviated: ZF); cf. Cohen [16, Ch 2]. The notations and basic facts to be considered are standard. Some important ones are described below.

(A) Let X be a nonempty set. By a *relation* over X, we mean any (nonempty) part $\mathscr{R} \subseteq X \times X$; then, $(X, \mathscr{R})$ will be referred to as a *relational structure*. For simplicity, we sometimes write $(x, y) \in \mathscr{R}$ as $x\mathscr{R}y$. Note that $\mathscr{R}$ may be regarded as a mapping between X and $\exp[X]$ (=the class of all subsets in X). In fact, denote for $x \in X$:

$$X(x, \mathscr{R}) = \{y \in X; x\mathscr{R}y\} \text{ (the } section \text{ of } \mathscr{R} \text{ through } x);$$

then, the desired mapping representation is $[\mathscr{R}(x) = X(x, \mathscr{R}), x \in X]$.

A basic example of relational structure is to be constructed as below. Let $N = \{0, 1, \ldots\}$ be the set of *natural* numbers, endowed with the usual addition and (partial) order; note that

$$(N, \leq) \text{ is well ordered: any (nonempty) subset of } N \text{ has a first element.}$$

Further, denote for $p, q \in N$, $p \leq q$,

$$N[p,q] = \{n \in N;\ p \leq n \leq q\},\ N]p,q[= \{n \in N;\ p < n < q\},$$
$$N[p,q[= \{n \in N;\ p \leq n < q\},\ N]p,q] = \{n \in N;\ p < n \leq q\};$$

as well as, for $r \in N$,

$$N[r,\infty[= \{n \in N;\ r \leq n\},\ N]r,\infty[= \{n \in N;\ r < n\}.$$

By definition, $N[0,r[= N(r,>)$ is referred to as the *initial interval* (in N) induced by r. Any set P with $P \sim N$ (in the sense: there exists a bijection from P to N) will be referred to as *effectively denumerable*. In addition, given some natural number $n \geq 1$, any set Q with $Q \sim N(n,>)$ will be said to be *n-finite*; when n is generic here, we say that Q is *finite*. Finally, the (nonempty) set Y is called (at most) *denumerable* iff it is either effectively denumerable or finite.

Let X be a nonempty set. By a *sequence* in X, we mean any mapping $x : N \to X$, where $N = \{0, 1, \ldots\}$ is the set of *natural* numbers. For simplicity reasons, it will be useful to denote it as $(x(n); n \geq 0)$, or $(x_n; n \geq 0)$; moreover, when no confusion can arise, we further simplify this notation as $(x(n))$ or (x_n), respectively. Also, any sequence $(y_n := x_{i(n)}; n \geq 0)$ with

$$(i(n); n \geq 0) \text{ is } divergent\ (i(n) \to \infty \text{ as } n \to \infty)$$

will be referred to as a *subsequence* of $(x_n; n \geq 0)$. Note that, under such a convention, the relation "subsequence of" is transitive; i.e.:

(z_n)=subsequence of (y_n) and (y_n)=subsequence of (x_n)

imply (z_n)=subsequence of (x_n).

(B) Remember that, an outstanding part of (ZF) is the *Axiom of Choice* (abbreviated: AC); which, in a convenient manner, may be written as

(AC) For each couple (J, X) of nonempty sets and each $F : J \to \exp(X)$,

there exists a (selective) function $f : J \to X\ [f(\nu) \in F(\nu), \forall \nu \in J]$.

Here, $\exp(X)$ stands for the class of all nonempty elements in $\exp[X]$. Sometimes, when the index set J is denumerable, the existence of such a selective function may be determined by using a weaker form of (AC), called: *Dependent Choice* principle (in short: DC). Call the relation $\mathscr{R}$ over X, *proper* when

$$(X(x, \mathscr{R}) =)\mathscr{R}(x) \text{ is nonempty, for each } x \in X.$$

Then, $\mathscr{R}$ is to be viewed as a mapping between X and $\exp(X)$; and the couple $(X, \mathscr{R})$ will be referred to as a *proper relational structure*. Further, given $a \in X$, let us say that the sequence $(x_n; n \geq 0)$ in X is $(a; \mathscr{R})$-*iterative*, provided

$$x_0 = a, \text{ and } (x_n; n \geq 0) \text{ is } \mathscr{R}\text{-increasing } [x_n \mathscr{R} x_{n+1} \text{ (i.e.: } x_{n+1} \in \mathscr{R}(x_n)), \forall n].$$

Proposition 1 *Let the relational structure $(X, \mathscr{R})$ be proper. Then, for each $a \in X$ there is at least an $(a, \mathscr{R})$-iterative sequence in X.*

This principle—proposed, independently, by Bernays [8] and Tarski [53]—is deductible from (AC), but not conversely; cf. Wolk [69]. Moreover, by the developments in Moskhovakis [44, Ch 8], and Schechter [52, Ch 6], the reduced system (ZF-AC+DC) it comprehensive enough so as to cover the "usual" mathematics; see also Moore [43, Appendix 2].

A basic consequence of (DC) is the so-called *Denumerable Axiom of Choice* [in short: AC(N)].

Proposition 2 *Let $F : N \to \exp(X)$ be a function. Then, for each $a \in F(0)$ there exists a function $f : N \to X$ with $f(0) = a$ and $(f(n) \in F(n), \forall n)$.*

Proof Denote $Q = N \times X$; and let us introduce the (proper) relation $\mathscr{R}$ over it, according to:

$$\mathscr{R}(n, x) = \{n + 1\} \times F(n + 1), \quad n \geq 0, x \in X.$$

By an application of (DC) to the proper relational structure $(Q, \mathscr{R})$ the conclusion follows; we do not give details.

As a consequence of the above facts,

$$\text{(DC)} \implies \text{(AC(N)) in (ZF-AC); or, equivalently:}$$

$$\text{(AC(N)) is deductible in the system (ZF-AC+DC).}$$

The reciprocal of the written inclusion is not true; see Moskhovakis [44, Ch 8, Sect 8.25] for details.

(C) In the following, the concept of conv-Cauchy structure over a metric space is introduced.

Let X be a nonempty set. Further, let $d : X \times X \to R_+$ be a mapping with

(m-1) d is *triangular*: $d(x, z) \leq d(x, y) + d(y, z), \forall x, y, z \in X$
(m-2) d is *reflexive-sufficient*: $d(x, y) = 0$ iff $x = y$
(m-3) d is *symmetric*: $d(x, y) = d(y, x)$, for all $x, y \in X$.

We then say that $d(., .)$ is a *metric* on X; and the couple (X, d) will be then referred to as a *metric space*.

Given the sequence (x_n) in X and the point $x \in X$, we say that (x_n), *d-converges* to x (written as: $x_n \stackrel{d}{\longrightarrow} x$), provided $d(x_n, x) \to 0$ as $n \to \infty$; i.e.,

$$\forall \varepsilon > 0, \exists i = i(\varepsilon): \ i \leq n \Longrightarrow d(x_n, x) < \varepsilon; \text{ or, equivalently:}$$
$$\forall \varepsilon > 0, \exists i = i(\varepsilon): \ i \leq n \Longrightarrow d(x_n, x) \leq \varepsilon.$$

The set of all such points x will be denoted $\lim_n(x_n)$; when it is nonempty, then (x_n) is called *d-convergent*. By this very definition, we have the properties:

(conv-1) $((\stackrel{d}{\longrightarrow})$ is *hereditary*)
$x_n \stackrel{d}{\longrightarrow} x$ implies $y_n \stackrel{d}{\longrightarrow} x$, for each subsequence (y_n) of (x_n)

(conv-2) $((\stackrel{d}{\longrightarrow})$ is *reflexive*)
$(\forall u \in X)$: the constant sequence $(x_n = u; n \geq 0)$ fulfills $x_n \stackrel{d}{\longrightarrow} u$.

As a consequence, $(\stackrel{d}{\longrightarrow})$ has all properties required in Kasahara [36]; in addition—as d is triangular symmetric—the following extra property is holding here

(conv-3) $(\stackrel{d}{\longrightarrow})$ is *separated* (referred to as d is *separated*):
$\lim_n(x_n)$ is an asingleton, for each sequence (x_n) in X.

The introduced concepts allow us to give a useful property.

Proposition 3 *The mapping $(x, y) \mapsto d(x, y)$ is d-Lipschitz, in the sense*

(23-1) $|d(x, y) - d(u, v)| \leq d(x, u) + d(y, v)$, $\forall (x, y), (u, v) \in X \times X$.

As a consequence, this map is d-continuous; i.e.,

(23-2) $x_n \stackrel{d}{\longrightarrow} x$, $y_n \stackrel{d}{\longrightarrow} y$ *imply* $d(x_n, y_n) \to d(x, y)$.

Further, call (x_n), *d-Cauchy* when $d(x_m, x_n) \to 0$ as $m, n \to \infty$, $m < n$; that is,

$$\forall \varepsilon > 0, \exists j = j(\varepsilon): \ j \leq m < n \Longrightarrow d(x_m, x_n) < \varepsilon; \text{ or, equivalently:}$$
$$\forall \varepsilon > 0, \exists j = j(\varepsilon): \ j \leq m < n \Longrightarrow d(x_m, x_n) \leq \varepsilon.$$

The class of all such sequences will be denoted as $Cauchy(d)$. As before, from this very definition one has the properties

(Cauchy-1) ($Cauchy(d)$ is *hereditary*)
(x_n) is d-Cauchy implies (y_n) is d-Cauchy, for each subsequence (y_n) of (x_n)

(Cauchy-2) ($Cauchy(d)$ is *reflexive*)
$(\forall u \in X)$: the constant sequence $(x_n = u; n \geq 0)$ is d-Cauchy.

Hence, $Cauchy(d)$ is a Cauchy structure, under the lines in Turinici [64].

Now—according to the quoted work—term the couple $((\xrightarrow{d}), Cauchy(d))$, a *conv-Cauchy structure* induced by d. The following regularity conditions about this structure are to be (optionally) considered

(CC-1) d is *regular*: each d-convergent sequence in X is d-Cauchy
(CC-2) d is *complete*: each d-Cauchy sequence in X is d-convergent.

Clearly, the former of these is always obtainable, via d=triangular symmetric; but the latter one is not in general valid.

(D) In the following, some d-Cauchy criteria will be stated.
Let us say that the sequence $(x_n; n \geq 0)$ is *d-asymptotic*, provided

$$r_n := d(x_n, x_{n+1}) \to 0 \text{ as } n \to \infty.$$

Clearly, each d-Cauchy sequence is d-asymptotic too; the reciprocal of this is not in general true. This tells us that the d-Cauchy criteria we are looking for are to be sought in the class of d-asymptotic sequences. To get concrete examples of such properties, we need some conventions and auxiliary facts.

Given the d-asymptotic sequence $(x_n; n \geq 0)$ and the number $\varepsilon > 0$, let us say that $i \in N$ is *ε-regular*, provided

$$i \leq n \text{ implies } d(x_n, x_{n+1}) < \varepsilon.$$

The class $\mathscr{Z}(\varepsilon)$ of all these ranks is nonempty; so that

$$(\forall \varepsilon > 0): \ Z(\varepsilon) = \min \mathscr{Z}(\varepsilon) \text{ is well defined, as an element of } N;$$

with, in addition: $d(x_n, x_{n+1}) < \varepsilon$, for all $n \geq Z(\varepsilon)$.

Define the subsets of $N \times N$

$$(\leq; N) = \{(m, n) \in N \times N; m \leq n\}, \ (<; N) = \{(m, n) \in N \times N; m < n\};$$

these are just the graph over N of the relations $(\leq)$ and $(<)$, respectively.

Further, let us say that the subset Θ of $R^0_+ :=]0, \infty[$ is $(>)$-*cofinal* in R^0_+, when

$$\text{for each } \varepsilon > 0, \text{ there exists } \theta \in \Theta \text{ with } \varepsilon > \theta.$$

Given this subset, define (as before)

$$(\geq; \Theta) = \{(\beta, \gamma) \in \Theta \times \Theta; \beta \geq \gamma\}, \ (>; \Theta) = \{(\beta, \gamma) \in \Theta \times \Theta; \beta > \gamma\};$$

these are just the graph over Θ of the relations $(\geq)$ and $(>)$, respectively. In addition, for each $(\beta, \gamma) \in (>; \Theta)$, denote

$$B(\beta, \gamma) = \{n \in N(2, \leq); 3^{-n} < \min\{\beta - \gamma, \gamma\};$$

clearly, $B(\beta, \gamma)$ is nonempty and hereditary:

$$i \in B(\beta, \gamma) \text{ and } i \leq j \text{ imply } j \in B(\beta, \gamma).$$

In this case, it makes sense denoting

$$b(\beta, \gamma) = \min B(\beta, \gamma), (\beta, \gamma) \in (>; \Theta).$$

If no confusion arise, we write $b(\beta, \gamma)$ as b, for simplicity; hence, by definition

$$(b \geq 2 \text{ and}) \, k \in N(b, \leq) \text{ implies } 3^{-k} < \min\{\beta - \gamma, \gamma\}.$$

From the perspective of our initial problem, the points $\beta \in \Theta$ appearing here must be chosen according to some admissible properties. Some conventions are in order. Let (x_n) be a sequence in X. We say that $\beta \in \Theta$ is (x_n)-*admissible*, when

(rela-be) $E(j; \beta) := \{(m, n) \in (<; N); \, j \leq m < n, d(x_m, x_n) > \beta\} \neq \emptyset, \forall j.$

The class of all these points will be denoted as $\text{adm}(\Theta; (x_n))$.

Proposition 4 *Let (x_n) be a sequence in X that is not d-Cauchy; and Θ be a $(>)$-cofinal part of R_+^0. Then, necessarily,*

(24-1) $\text{adm}(\Theta; (x_n))$ *is (nonempty and) $(>)$-cofinal in R_+^0*
(24-2) for each $\beta \in \text{adm}(\Theta; (x_n))$ and each $\gamma \in \Theta$, we have $(\beta, \gamma) \in (>; \Theta)$ iff $(\beta, \gamma) \in (>; \text{adm}(\Theta; (x_n)))$.

Proof

(i): By definition, the d-Cauchy property of our sequence writes:

$$\forall \varepsilon \in R_+^0, \exists a \in N, \forall (m, n) \in (<; N) : \; a \leq m < n \Longrightarrow d(x_m, x_n) \leq \varepsilon.$$

As Θ is a $(>)$-cofinal part in R_+^0, this property may be also written as

$$\forall \theta \in \Theta, \exists \alpha \in N, \forall (m, n) \in (<; N) : \; \alpha \leq m < n \Longrightarrow d(x_m, x_n) \leq \theta.$$

The negation of this property means: there exists $\beta \in \Theta$ such that

$$E(j; \beta) := \{(m, n) \in (<; N); \, j \leq m < n, d(x_m, x_n) > \beta\} \neq \emptyset, \forall j;$$

which shows that $\beta \in \text{adm}(\Theta; (x_n))$. Moreover, as

$$(\forall j) : \; \beta_1 \geq \beta_2 \text{ implies } E(j; \beta_1) \subseteq E(j; \beta_2)$$

it is clear that (under our standard notations)

$$\beta \in \mathrm{adm}(\Theta; (x_n)) \text{ implies } \Theta(\beta, >) \subseteq \Theta(\beta, \geq) \subseteq \mathrm{adm}(\Theta; (x_n));$$

and proves (by means of $(>)$-cofinal property of Θ) the first conclusion.
(ii): Evident, in view of the preceding stage.

Finally, the following conventions are needed. For each sequence of ranks $(\lambda(k); k \geq 0)$ in $N(1, \leq)$, define the property

$$(\lambda(k); k \geq 0) \text{ is } \textit{strictly ascending}: \ i < j \text{ implies } \lambda(i) < \lambda(j);$$

$$\text{hence, } (\lambda(k)) \text{ is divergent } (\lim_k \lambda(k) = \infty).$$

On the other hand, for each sequence (r_n) in R and each point $r \in R$, let us write

$$r_n \to r+ \text{ (also written as: } \lim_n (r_n) = r+) \text{ if } r_n \to r \text{ and } (r_n > r, \forall n)$$

$$r_n \to\to r+ \text{ (also written as: } \lim_n (r_n) = r++) \text{ if } r_n \to r \text{ and } (r_n > r, \forall\forall n).$$

Here, given a property $\pi(k)$ depending on $k \in N$, let us say that it holds for *nearly all* k [written: $(\pi(k), \forall\forall k)$] provided

$$\text{there exists } c = c(\pi) \in N \text{ such that } (\pi(k) \text{ is true, for all } k \geq c).$$

In particular, given the sequence $(w_k; k \geq 0)$ in X, the subset Y of X, and the property $[\pi(k)$ holds iff $w_k \in Y]$, we introduce the convention

$$(\pi(k) \text{ holds for nearly all } k) \text{ is referred to as } ((w_k) \text{ is } \textit{nearly} \text{ in } Y).$$

The following result, referred to as *Boyd-Wong non-Cauchy Criterion* (in short: (BW-n-CC)) is now available.

Theorem 2 *Let the sequence* $(x_n; n \geq 0)$ *in* X *be such that*

(21-i) $(x_n; n \geq 0)$ *is d-asymptotic* $(r_n := d(x_n, x_{n+1}) \to 0$ *as* $n \to \infty)$
(21-ii) $(x_n; n \geq 0)$ *is not d-Cauchy.*

Further, let the subset Θ *of* R_+^0 *be* $(>)$*-cofinal in* R_+^0*; hence (see above)*

$$\mathrm{admc}(\Theta; (x_n)) \text{ (=the class of } (x_n)\text{-admissible couples) is nonempty.}$$

Then, for each couple $(\beta, \gamma) \in (>; \mathrm{adm}(\Theta; (x_n)))$ *(with the associated rank* $b = b(\beta, \gamma)$*), and each strictly ascending rank sequence* $(\lambda(k); k \geq 0)$ *in* $N(1, \leq)$*, there exists a rank sequence* $(J(k); k \geq 0)$ *in* $N(1, \leq)$ *and a couple of rank sequences* $(m(k); k \geq 0)$ *and* $(n(k); k \geq 0)$ *in* $N(1, \leq)$*, so that*

(21-a) $k + 1 \leq J(k) \leq m(k) < m(k) + 3\lambda(k) < n(k), \forall k$

(21-b) $J(k) \leq m(k) < m(k) + 2\lambda(k) < n(k) - 1 < n(k)$, *and* $d(x_{m(k)}, x_{n(k)}) > \gamma$, $d(x_{m(k)}, x_{n(k)-1}) \leq 3^{-b-k} + \gamma$, $\forall k$,

(21-c) $U_k := d(x_{m(k)}, x_{n(k)}) \to \gamma+$ *as* $k \to \infty$

(21-d) for each couple of rank sequences $(\mu(k))$ *and* $(\nu(k))$ *in* N *with* $(\mu(k), \nu(k) \leq 3\lambda(k), \forall k)$, $V_k := d(x_{m(k)+\mu(k)}, x_{n(k)+\nu(k)}) \to \gamma+$ *as* $k \to \infty$

(21-e) for each couple of ranks (i, j) *in* $N \times N$ *with* $(i, j \leq 3\lambda(0))$, $S_k := d(x_{m(k)+i}, x_{n(k)+j}) \to \gamma+$ *as* $k \to \infty$

(21-f) for each couple of ranks (i, j) *in* $N \times N$, *one has* $T_k := d(x_{m(k)+i}, x_{n(k)+j}); k \geq 0) \to\to \gamma+$ *as* $k \to \infty$.

Proof Let the couple $(\beta, \gamma) \in (>; \text{adm}(\Theta; (x_n)))$ be fixed in the sequel; so, $[\beta \in \text{adm}(\Theta; (x_n))$ and $\beta, \gamma) \in (>; \Theta)]$; remember that the first property means

(rela-1) $E(j) := \{(m, n) \in (<; N); j \leq m < n, d(x_m, x_n) > \beta\} \neq \emptyset, \forall j$.

Further, take a strictly ascending rank sequence $(\lambda(k); k \geq 0)$ in $N(1, \leq)$. With the aid of these data, define the rank sequence $(J(k); k \geq 0)$ in $N(1, \leq)$, according to

$$(J(k) = 1 + k + Z(3^{-2b-k}/\lambda(k)); k \geq 0);$$

where $b = b(\beta, \gamma)$ (see above) and the mapping $\varepsilon \mapsto Z(\varepsilon)$ is introduced by the d-asymptotic property of (x_n). Then, denote

$$(A(k) = E(J(k)); k \geq 0); \text{ hence, by definition,}$$

$$A(k) := \{(m, n) \in (<; N); J(k) \leq m < n, d(x_m, x_n) > \beta\}, k \geq 0;$$

with, in addition: $A(k)$ is a nonempty relation over N, for each $k \geq 0$.

By the triangle inequality (and the choice of b)

$$(\forall k): (m, n) \in A(k) \text{ implies}$$

$$d(x_{m+s}, x_{n+t}) \geq d(x_m, x_n) - d(x_m, x_{m+s}) - d(x_n, x_{n+t}) >$$

$$\beta - 3^{-2b-k+1} - 3^{-2b-k+1} > \beta - 3^{-b-k} > \gamma, \forall s, t \in N[0, 3\lambda(k)];$$

which tells us that

$$(\forall k): B(k) := \{(m, n) \in (<; N); J(k) \leq m < n,$$

$$d(x_{m+s}, x_{n+t}) > \gamma, \forall s, t \in N[0, 3\lambda(k)]\} \text{ is a nonempty relation over } N.$$

Having this precise, denote for each $k \geq 0$

$$m(k) = \min \text{Dom}(B(k)), \ n(k) = \min B(k)(m(k)).$$

By this very definition, we get

(pro-1) $(\forall k)$: $k+1 \leq J(k) \leq m(k) < n(k)$,
(pro-2) $(\forall k)$: $d(x_{m(k)+s}, x_{n(k)+t}) > \gamma, \forall s, t \in N[0, 3\lambda(k)]$.

We claim that the couple $(\beta, \gamma) \in (>; \mathrm{adm}(\Theta; (x_n)))$, the rank sequence $(J(k))$ in $N(1, \leq)$, and the couple of rank-sequences $[(m(k)), (n(k))]$ fulfill all desired conclusions.

(i): By (pro-1), it is clear that the first, second, and third relation of (21-a) holds.

(ii): Suppose by contradiction that

$$(m(k) <) n(k) \leq m(k) + 3\lambda(k), \text{ for some } k \geq 0.$$

By $m(k) \geq J(k) \geq Z(3^{-2b-k}/\lambda(k))$, the triangle inequality, and the choice of b,

$$d(x_{m(k)}, x_{n(k)}) \leq 3^{-2b-k+1} < 3^{-b-k} < \gamma;$$

in contradiction with (pro-2); whence, the fourth relation in (21-a) holds too.

(iii): The first and second part of (21-b) are directly obtainable from the preceding stage and (pro-2), respectively. Concerning the third part of (21-b), let $k \geq 0$ be arbitrary fixed. By definition, $n(k)$ is the minimum of all ranks $p \in N$ with

$(m(k), p) \in B(k)$; that is:

$J(k) \leq m(k) < p$ and $d(x_{m(k)+s}, x_{p+t}) > \gamma, \forall s, t \in N[0, 3\lambda(k)]$.

As $m(k) < m(k) + 2\lambda(k) < n(k) - 1$, we must have (by this minimal property)

(pro-3) $d(x_{m(k)+s}, x_{n(k)-1+t}) \leq \gamma$, for some $s, t \in N[0, 3\lambda(k)]$.

But, in view of (pro-2) once again,

(pro-4) $d(x_{m(k)+u}, x_{n(k)-1+v}) > \gamma, \forall u \in N[0, 3\lambda(k)], \forall v \in N[1, 3\lambda(k)]$.

This, combined with (pro-3), tells us that, necessarily,

(pro-5) $d(x_{m(k)+s}, x_{n(k)-1}) \leq \gamma$, for some $s \in N[0, 3\lambda(k)]$.

By $m(k) \geq Z(3^{-2b-k}/\lambda(k))$, the triangle inequality, and $b \geq 2$, we then have

$$d(x_{m(k)}, x_{n(k)-1}) \leq$$

$$d(x_{m(k)}, x_{m(k)+s}) + d(x_{m(k)+s}, x_{n(k)-1}) \leq 3^{-2b-k+1} + \gamma \leq 3^{-b-k} + \gamma$$

and the last conclusion of (21-b) follows.

(iv): From these facts and triangular inequality,

$$\gamma < d(x_{m(k)}, x_{n(k)}) \leq d(x_{m(k)}, x_{n(k)-1}) + r_{n(k)-1} \leq 3^{-b-k} + \gamma + r_{n(k)-1}, \ \forall k.$$

Passing to limit in this double inequality gives (21-c).

(v): By the very definition of $[(\mu(k)), (\nu(k))]$, and (pro-2),

$$V_k > \gamma, \text{ for all } k \geq 0; \text{ so, the first half of (21-d) holds.}$$

Moreover, from a metrical property of d, the very definition of $(m(k); k \geq 0)$ and $(n(k)); k \geq 0)$, and the choice of b

$$|d(x_{m(k)}, x_{n(k)}) - d(x_{m(k)+\mu(k)}, x_{n(k)+\nu(k)})| \leq$$
$$d(x_{m(k)}, x_{m(k)+\mu(k)}) + d(x_{n(k)}, x_{n(k)+\nu(k)}) \leq$$
$$3^{-2b-k+1} + 3^{-2b-k+1} < 3^{-2b-k+2} \leq 3^{-b-k}, \text{ for all } k \geq 0.$$

Passing to limit in the relation between the first and the last member of this relation gives the second half of (21-d).

(vi): By the strict ascending property of $(\lambda(k); k \geq 0)$, the sequences $(\mu(k) = i; k \geq 0)$ and $(\nu(k) = j; k \geq 0)$, fulfill $(\mu(k), \nu(k) \leq 3\lambda(k), \forall k)$; and this, along with the preceding stage, yields the desired conclusion.

(vii): Let $(i, j) \in N \times N$ be given. By the strict ascending property of $(\lambda(k); k \geq 0)$, there exists an index $L = L(i, j) \in N$ with

$$i, j \leq 3\lambda(k), \text{ for all } k \geq L.$$

Then, define the couple of sequences $(\mu(k); k \geq 0)$ and $(\nu(k); k \geq 0)$ as

$$(\mu(k) = \nu(k) = 0;\ k \leq L);\ (\mu(k) = i, \nu(k) = j;\ k > L).$$

Clearly, $(\mu(k), \nu(k) \leq 3\lambda(k), \forall k)$; and this, along with the preceding stage, yields

$$T_k^* = d(x_{m(k)+\mu(k)}, x_{n(k)+\nu(k)}) \to \gamma + \text{ as } n \to \infty.$$

It will now suffice observing that

$$T_k = T_k^*, \text{ for all } k > L \text{ (hence, for all } k \geq L + 1)$$

to get the written conclusion. The proof is complete.

In particular, when $\Theta = R_+^0$ and $(\lambda(k) = 1 + k; k \geq 0)$, the obtained statement covers the 1969 one in Boyd and Wong [11]; so, it is natural that this result be referred to in the proposed way. Further aspects may be found in Reich [48]; see also Khan et al. [37].

3 Admissible Real Functions

In the following, some classes of admissible real functions are introduced. Their usefulness will become clear from our next developments.

(A) Denote for simplicity

- $\mathscr{F}_0(R_+)$=the subclass of all $\varphi \in \mathscr{F}(R_+)$, with $\varphi(0) = 0$
- $\mathscr{F}_0(re)(R_+)$=the subclass of all $\varphi \in \mathscr{F}_0(R_+)$, with the *regressive* property: $\varphi(t) < t, \forall t \in R_+^0$
- $\mathscr{F}_0(in)(R_+)$=the subclass of all increasing functions $\varphi \in \mathscr{F}_0(R_+)$
- $\mathscr{F}_0(re, in)(R_+) = \mathscr{F}_0(re)(R_+) \cap \mathscr{F}_0(in)(R_+)$.

For each $\varphi \in \mathscr{F}_0(re)(R_+)$, let us introduce the sequential properties

(M-a) φ is *Matkowski admissible*:
for each (t_n) in R_+^0 with $(t_{n+1} \leq \varphi(t_n), \forall n)$ we have $\lim_n t_n = 0$

(n-d-a) φ is *non-diagonally admissible*:
there are no sequences $(t_n; n \geq 0)$ in R_+^0
and no elements $\varepsilon \in R_+^0$ with $t_n \to \varepsilon+$, $\varphi(t_n) \to \varepsilon+$

(MK-a) φ is *Meir-Keeler admissible*:
$\forall \varepsilon > 0, \exists \delta > 0$, such that $\varepsilon < s < \varepsilon + \delta \Longrightarrow \varphi(s) \leq \varepsilon$.

The relationships between these properties are discussed in the statement below

Theorem 3 *For each $\varphi \in \mathscr{F}_0(re)(R_+)$, we have in (ZF-AC+DC)*

$$(M\text{-}a) \Longrightarrow (n\text{-}d\text{-}a) \Longrightarrow (MK\text{-}a) \Longrightarrow (M\text{-}a).$$

Hence, for each $\varphi \in \mathscr{F}_0(re)(R_+)$, the properties (M-a), (n-d-a), and (MK-a) are equivalent to each other.

Proof There are three cases to be discussed.

(i) Suppose that φ is Matkowski admissible; we assert that φ is non-diagonally admissible. For, if φ is not endowed with such a property, there must be a sequence $(t_n; n \geq 0)$ in R_+^0 and a number $\varepsilon > 0$, such that

$$t_n \to \varepsilon + \text{ and } \varphi(t_n) \to \varepsilon +, \text{ as } n \to \infty.$$

Put $i(0) = 0$. As $\varepsilon < \varphi(t_{i(0)})$ and $t_n \to \varepsilon+$, we have that

$A(i(0)) := \{n > i(0); t_n < \varphi(t_{i(0)})\}$ is not empty;

hence, $i(1) := \min(A(i(0)))$ is an element of it, and $t_{i(1)} < \varphi(t_{i(0)})$.

Likewise, as $\varepsilon < \varphi(t_{i(1)})$ and $t_n \to \varepsilon+$, we have that

$A(i(1)) := \{n > i(1); t_n < \varphi(t_{i(1)})\}$ is not empty;

hence, $i(2) := \min(A(i(1)))$ is an element of it, and $t_{i(2)} < \varphi(t_{i(1)})$.

This procedure may continue indefinitely; and yields (without any choice technique) a strictly ascending rank sequence $(i(n); n \geq 0)$ (hence, $i(n) \to \infty$ as $n \to \infty$) for which the attached subsequence $(s_n := t_{i(n)}; n \geq 0)$ of (t_n) fulfills

$$s_{n+1} < \varphi(s_n) (< s_n), \text{ for all } n.$$

On the other hand, by this very subsequence property,

$$(s_n > \varepsilon, \forall n) \text{ and } \lim_n s_n = \lim_n t_n = \varepsilon.$$

The obtained relations are in contradiction with the Matkowski property of φ; hence, the working condition cannot be true; and we are done.

(ii) Suppose that φ is non-diagonally admissible; we show that, necessarily, φ is Meir-Keeler admissible. For, if φ is not endowed with such a property, we must have (for some $\varepsilon > 0$)

$$H(\delta) := \{t \in R_+^0; \varepsilon < t < \varepsilon + \delta, \varphi(t) > \varepsilon\} \text{ is not empty, for each } \delta > 0.$$

Taking a strictly descending sequence $(\delta_n; n \geq 0)$ in R_+^0 with $\delta_n \to 0$, we get by the Denumerable Axiom of Choice (AC(N)) [deductible, as precise, in (ZF-AC+DC)], a sequence $(t_n; n \geq 0)$ in R_+^0, so as

$$(\forall n) : t_n \text{ is an element of } H(\delta_n);$$

or, equivalently (by the very definition above and φ=regressive)

$$(\forall n) : \varepsilon < \varphi(t_n) < t_n < \varepsilon + \delta_n;$$

hence, in particular: $\varphi(t_n) \to \varepsilon +$ and $t_n \to \varepsilon +$.

But, these relations are in contradiction with the non-diagonal admissible property of our function; hence, the assertion follows.

(iii) Suppose that φ is Meir-Keeler admissible; we have to establish that φ is Matkowski admissible. Let $(s_n; n \geq 0)$ be a sequence in R_+^0 with the property $(s_{n+1} \leq \varphi(s_n); n \geq 0)$. Clearly, (s_n) is strictly descending in R_+^0; hence, $\sigma := \lim_n s_n$ exists in R_+. Suppose by contradiction that $\sigma > 0$; and let $\rho > 0$ be given by the Meir-Keeler admissible property of φ. By the above convergence relations, there exists some rank $n(\rho)$, such that

$$n \geq n(\rho) \text{ implies } \sigma < s_n < \sigma + \rho.$$

But then, under the notation ($t_n := \varphi(s_n)$; $n \geq 0$), we get (for the same ranks)

$$\sigma < s_{n+1} \leq t_n < s_n < \sigma + \rho;$$

in contradiction with the Meir-Keeler admissible property. Hence, necessarily, $\sigma = 0$; and conclusion follows. The proof is complete.

In the following, some important examples of such objects will be given. For any $\varphi \in \mathscr{F}_0(re)(R_+)$ and any $s \in R_+^0$, put

$$\Lambda^+\varphi(s) = \inf_{\varepsilon>0} \Phi(s+)(\varepsilon); \text{ where } \Phi(s+)(\varepsilon) = \sup \varphi(]s, s+\varepsilon[), \varepsilon > 0.$$

From the regressive property of φ, these quantities are finite; precisely,

$$0 \leq \Lambda^+\varphi(s) \leq s, \ \forall s \in R_+^0.$$

The following completion of this will be useful.

Proposition 5 *Let $\varphi \in \mathscr{F}_0(re)(R_+)$ and $s \in R_+^0$ be arbitrary fixed. Then,*

(31-1) $\limsup_n(\varphi(t_n)) \leq \Lambda^+\varphi(s)$, *for each sequence (t_n) in R_+^0 with $t_n \to s+$*
(31-2) there exists a sequence (r_n) in R_+^0 with $r_n \to s+$ and $\varphi(r_n) \to \Lambda^+\varphi(s)$.

Proof Denote, for simplicity,

$$\alpha = \Lambda^+\varphi(s); \text{ hence}, \alpha = \inf_{\varepsilon>0} \Phi(s+)(\varepsilon), \text{ and } 0 \leq \alpha \leq s.$$

(i): Given $\varepsilon > 0$, there exists a rank $p(\varepsilon) \geq 0$ such that $s < t_n < s + \varepsilon$, for all $n \geq p(\varepsilon)$; hence

$$\limsup_n(\varphi(t_n)) \leq \sup\{\varphi(t_n); n \geq p(\varepsilon)\} \leq \Phi(s+)(\varepsilon).$$

Passing to infimum over $\varepsilon > 0$, yields (see above)

$$\limsup_n(\varphi(t_n)) \leq \inf_{\varepsilon>0} \Phi(s+)(\varepsilon) = \alpha; \text{ and the claim follows.}$$

(ii): Define ($\beta_n := \Phi(s+)(2^{-n})$; $n \geq 0$); this is a descending sequence in R_+, with

$$(\beta_n \geq \alpha, \ \forall n) \text{ and } \inf_n \beta_n = \alpha; \text{ hence } \lim_n \beta_n = \alpha.$$

By these properties, there may be constructed a sequence $(\gamma_n; n \geq 0)$ in R, with

$$\gamma_n < \beta_n, \ \forall n; \lim_n \gamma_n = \lim_n \beta_n = \alpha.$$

(For example, we may take $(\gamma_n = \beta_n - 3^{-n}; n \geq 0)$; we do not give details). Let $n \geq 0$ be arbitrary fixed. By the supremum definition, there exists $r_n \in]s, s + 2^{-n}[$ such that $\varphi(r_n) > \gamma_n, \forall n$; moreover (again by definition), $\varphi(r_n) \leq \beta_n$. The obtained sequence $(r_n; n \geq 0)$ fulfills $r_n \to s+$ and $\varphi(r_n) \to \alpha$; wherefrom, all is clear.

We may now pass to the announced examples of such functions.

Example 1 Call $\varphi \in \mathscr{F}_0(re)(R_+)$, *Boyd-Wong admissible* [11], if

$$\Lambda^+\varphi(s) < s, \text{ for all } s > 0.$$

In particular, this holds provided

$$\varphi \text{ is upper semicontinuous at the right on } R_+^0 : \ \Lambda^+\varphi(s) \leq \varphi(s), \ \forall s \in R_+^0.$$

Moreover, the written property is fulfilled when

$$\varphi \text{ is continuous at the right on } R_+^0;$$

for, in such a case, $[\Lambda^+\varphi(s) = \varphi(s), \forall s \in R_+^0]$.

Under these conditions, we have (cf. Meir and Keeler [42]):

$$(\forall \varphi \in \mathscr{F}_0(re)(R_+)) : \ \varphi \text{ is Boyd-Wong admissible implies}$$

$$\varphi \text{ is Meir-Keeler admissible [or, equivalently: Matkowski admissible].}$$

In fact, suppose that $\varphi \in \mathscr{F}_0(re)(R_+)$ is Boyd-Wong admissible; and fix some $\gamma > 0$. As $\Lambda^+\varphi(\gamma) < \gamma$, there exists $\beta = \beta(\gamma) > 0$ such that

$$\Phi(\gamma+)(\beta) < \gamma; \text{ wherefrom, } \gamma < t < \gamma + \beta \text{ implies } \varphi(t) < \gamma;$$

and this gives the desired property.

Concerning the reverse inclusion, let us consider the function $\varphi \in \mathscr{F}_0(re)(R_+)$, according to (for some $r > 0$):

$$(\varphi(t) = t/2, \text{ if } 0 \leq t \leq r), \ (\varphi(t) = r, \text{ if } t > r).$$

Clearly, φ is Matkowski admissible, as it can be directly seen. On the other hand,

$$\Lambda^+\varphi(r) = r; \text{ whence, } \varphi \text{ is not Boyd-Wong admissible;}$$

proving that the reverse inclusion is not in general valid. For an extended example of this type, see Turinici [55] and the references therein.

Example 2 According to its definition, the Matkowski admissible property of some $\varphi \in \mathscr{F}_0(re, in)(R_+)$ writes (cf. Matkowski [41])

$$\varphi^n(t) \to 0 \text{ as } n \to \infty, \text{ for all } t > 0.$$

[Here, for each $n \in N$, φ^n stands for the n-th iterate of φ]. This, by a previous equivalence result, tells us that φ is Meir-Keeler admissible. A different way of proving this may be described as follows (cf. Jachymski [33]):

Assume that $\varphi \in \mathscr{F}_0(re, in)(R_+)$ is Matkowski admissible. If the underlying property fails, then (for some $\gamma > 0$):

$$\forall \beta > 0, \exists t \in]\gamma, \gamma + \beta[, \text{ such that } \varphi(t) > \gamma.$$

As φ is increasing, this yields

$$\varphi(t) > \gamma, \forall t > \gamma; \text{ hence, by induction: } \varphi^n(t) > \gamma, \forall n \in N, \forall t > \gamma.$$

Taking some $t > \gamma$ and passing to limit as $n \to \infty$, one gets $0 \geq \gamma$; contradiction.

A sufficient condition for this property is to be obtained from the result above, by simply noting that

$$(\forall \varphi \in \mathscr{F}_0(re, in)(R_+)) : \ \Lambda^+\varphi(s) = \varphi(s + 0), \ \forall s \in R_+^0.$$

Precisely, we have the practical characterization

$$(\forall \varphi \in \mathscr{F}_0(re, in)(R_+)) : \ \varphi \text{ is Boyd-Wong admissible iff}$$
$$\varphi \text{ is } \textit{strongly regressive } [\varphi(s + 0) < s, \text{ for all } s > 0].$$

Example 3 Call $\varphi \in \mathscr{F}_0(re)(R_+)$, *Geraghty admissible* [24] provided

$$\text{each sequence } (t_n) \text{ in } R_+^0 \text{ with } \varphi(t_n)/t_n \to 1 \text{ fulfills } t_n \to 0.$$

The connection between this property and the preceding ones is described as:

$$(\text{for each } \varphi \in \mathscr{F}_0(re)(R_+)) :$$
$$\varphi \text{ is Geraghty admissible implies } \varphi \text{ is Boyd-Wong admissible.}$$

In fact, suppose that $\varphi \in \mathscr{F}_0(re)(R_+)$ is not Boyd-Wong admissible. From a previous relation, there exists some $s \in R_+^0$ with $\Lambda^+\varphi(s) = s$. Combining with a preceding auxiliary fact, there exists a sequence $(r_n; n \geq 0)$ in R_+^0 with

$$r_n \to s + \text{ and } \varphi(r_n) \to s; \text{ whence } \varphi(r_n)/r_n \to 1;$$

i.e.: φ is not Geraghty admissible. The obtained contradiction proves our claim.

Concerning the reverse inclusion, note that, for the (continuous) Boyd-Wong admissible function $[\varphi(t) = t(1 - e^{-t}), t \geq 0]$ in $\mathscr{F}_0(re, in)(R_+)$, and the sequence $(t_n = n + 1; n \geq 0)$ in R_+^0, we have

$$\varphi(t_n)/t_n \to 1; \text{ but, evidently, } t_n \to \infty.$$

Hence, φ is not Geraghty admissible; so that the reciprocal is not in general true.

(B) A basic application involving these concepts may be described as below. Fix some $\omega \in \mathscr{F}_0(re, in)(R_+)$ and let $\omega^* \in \mathscr{F}(R_+)$ be defined as

$(\omega^*(t) = t - \omega(t); t \in R_+)$ [in short: $\omega^* = I - \omega$]; called: the *complement* of ω.

We say that ω is *complementary coercive*, in case

$$\omega^* \text{ is coercive: } \omega^*(t) \to \infty \text{ as } t \to \infty; \text{ or, equivalently:}$$
$$\{t \in R_+; \omega^*(t) \le a\} \text{ is bounded, for each } a \ge 0.$$

Note that, under such a condition, the function $\gamma \in \mathscr{F}(R_+)$ given as

$$\gamma(s) = \sup A(s) \text{ where } A(s) = \{t \in R_+;\ t \le \omega(s + t)\}, s \in R_+$$

[referred to as the *right complementary inverse* (in short: *rc-inverse*) associated to ω] is well defined. In fact, it will suffice noting that, for each $s \in R_+$,

$$A(s) = \{t \in R_+; \omega^*(s + t) \le s\};\ \text{whence, } A(s) \text{ is bounded in } R_+.$$

Clearly, $\gamma(0) = 0$. In addition, by the properties of ω, we have

(p-1) (Increasing): $s_1 \le s_2$ implies $A(s_1) \subseteq A(s_2)$; whence γ is increasing
(p-2) (Representation formula): $\gamma(s) \le \omega(s + \gamma(s))$, for each $s \in R_+$.

In fact, let $t \in A(s)$ be arbitrary fixed; hence, $t \le \omega(s + t)$. By the definition of supremum, $(t \le \omega(s + \gamma(s)), \forall t \in A(s))$; wherefrom (again by the underlying definition) the conclusion follows.

Further properties of $\gamma(.)$ are available under extra properties of $\omega(.)$.

Proposition 6 *Suppose that, in addition,*

$$\omega \text{ is strongly regressive: } \omega(t + 0) < t, \forall t > 0.$$

Then (in addition to the above)

$$(\lim_{t \to 0+} \gamma(t) =) \gamma(0 + 0) = 0;\ \textit{that is: } s_n \to 0 \textit{ implies } \gamma(s_n) \to 0.$$

Proof As γ is increasing, $\delta := \gamma(0+0)$ exists. Assume by contradiction that $\delta > 0$. Let (s_n) be a strictly descending sequence in R_+^0 with $s_n \to 0$; whence, $\gamma(s_n) \to \delta$. From the preceding representation formula

$$\gamma(s_n) \le \omega(s_n + \gamma(s_n)),\ \text{for all } n.$$

As $n \to \infty$, we have $s_n + \gamma(s_n) \to \delta+$; so that, passing to limit in this relation yields $\delta \le \omega(\delta + 0) < \delta$; contradiction. Hence, $\gamma(0 + 0) = 0$, as claimed.

Sometimes, we may ask for the associated rc-inverse $\gamma(.)$ a property like

$$\gamma \text{ is strongly regressive: } \gamma(t+0) < t, \text{ for all } t \in R_+^0;$$

the initial function ω will be referred to as *rc-regressive* in such a case. A useful answer in this direction is obtained by imposing regularity conditions upon

$$(\eta(t) = \omega(2t);\ t \geq 0) \text{ [the } double\ function \text{ attached to } \omega].$$

Note that, by the increasing property of ω (and η)

$$(\forall t > 0):\ \eta(t+0) = \inf_{s>0} \eta(t+s) = \inf_{s>0} \omega(2t+2s) = \omega(2t+0);$$

this will be useful in the sequel.

Theorem 4 *Let the function $\omega \in \mathscr{F}_0(re, in)(R_+)$ be such that*

(32-i) the double function η is strongly regressive ($\eta(t+0) < t, \forall t > 0$); expressed as: ω is double strongly regressive

(32-ii) the double function η is complementary coercive ($\eta^* := I - \eta$ is coercive); expressed as: ω is double complementary coercive.

Then,

(32-a) its associated complement $\omega^* := I - \omega$ is coercive; so, the right complementary inverse

$$\gamma(s) = \sup A(s) \text{ where } A(s) = \{t \in R_+;\ t \leq \omega(s+t)\},\ s \in R_+$$

is well defined as an element of $\mathscr{F}_0(R_+)$

(32-b) The right complementary inverse $\gamma(.)$ fulfills

(p-1) γ is increasing and $\gamma(s) \leq \omega(s+\gamma(s))$, for all $s \in R_+$
(p-2) γ is zero-continuous: $\gamma(t) \to 0 = \gamma(0+0)$ as $t \to 0$

(32-c) Finally, we have the properties

(p-3) γ is strongly regressive; hence, in particular, regressive
(p-4) γ is complementary coercive: $\gamma^* := I - \gamma$ is coercive.

Proof By a preceding relation,

$$(\forall t > 0):\ \omega(t+0) < t/2;\ \text{whence, } t - \omega(t) > t - t/2 = t/2;$$

proving that ω is complementary coercive: $\omega^* := I - \omega$ is coercive.

(i): Clearly, the right complementary inverse $\gamma : R_+ \to R_+$ is well defined.

(ii): The properties (p-1) and (p-2) of γ are directly obtainable by the preceding facts, in view of

$$\omega \text{ is double strongly regressive implies } \omega \text{ is strongly regressive.}$$

(iii): Fix $s > 0$; and let (s_n) be a strictly descending sequence with

$$s_n \to s+ \text{ as } n \to \infty; \text{ so that, } \gamma(s_n) \to r := \gamma(s+0) \text{ as } n \to \infty.$$

From the above representation formula,

(rep-seq) $\gamma(s_n) \leq \omega(s_n + \gamma(s_n)), \forall n.$

Denote for simplicity $u = s + r$. By the posed hypotheses,

$$s_n + \gamma(s_n) \to u+; \text{ whence } \omega(s_n + \gamma(s_n)) \to \omega(u+0)$$

Passing to limit in (rep-seq) gives (as ω is double strongly regressive)

$$r \leq \omega(u+0) < (1/2)u = (1/2)(s+r); \text{ wherefrom, } r < s;$$

and the conclusion follows.

(iv): Let $a \geq 0$ be arbitrary fixed. We have to establish that

$$A := \{t \in R_+; t \leq \gamma(t) + a\} \text{ is bounded.}$$

But, according to definition (and the properties of ω and γ)

$$A \subseteq \{t \in R_+; t \leq \omega(t + \gamma(t)) + a\} \subseteq \{t \in R_+; t \leq \omega(2t) + a\} =$$
$$\{t \in R_+; t \leq \eta(t) + a\} = \{t \in R_+; \eta^*(t) \leq a\};$$

and this, combined with the last subset being bounded, ends our argument.

The obtained properties of the right complementary inverse $\gamma \in \mathscr{F}(R_+)$ associated to ω are basic tools for getting a lot of relative type statements involving real sequences, to be needed further. For simplicity reasons, we will express these as generic results (without explicitly mentioning the conditions under which these properties were derived).

Theorem 5 *Let the sequences $(a_n; n \geq 0)$ and $(b_n; n \geq 0)$ in R_+, and the function $\varphi \in \mathscr{F}_0(re, in)(R_+)$ be such that*

(33-i) $a_n \leq b_n + \varphi(a_n)$, for all n.

In addition, suppose that

(33-ii) φ is strongly regressive ($\varphi(t+0) < t$, $\forall t > 0$)
(33-iii) φ is complementary coercive ($\varphi^* := I - \varphi$ is coercive).

Then, necessarily,

(33-a) (b_n) is bounded implies (a_n) is bounded
(33-b) (b_n) is zero-convergent implies (a_n) is zero-convergent.

Proof

(i): By definition, $\beta := \sup\{b_n; n \geq 0\}$ exists in R_+. This, by the imposed hypothesis, yields

$$(\forall n) : \ a_n \leq \beta + \varphi(a_n); \ \text{that is } \varphi^*(a_n) \leq \beta.$$

Combining with the coerciveness of φ^*, yields the desired fact.

(ii): Suppose that (b_n) is zero-convergent. In particular, (b_n) is bounded; hence, by the preceding stage, (a_n) is bounded too. By a standard result, (a_n) has convergent subsequences. Let $(a_{i(n)}; n \geq 0)$ be one of these; hence, $a_{i(n)} \to a$ as $n \to \infty$, for some $a \in R_+$. Suppose by contradiction that $a > 0$. Passing to limit as $n \to \infty$ in

$$a_{i(n)} \leq b_{i(n)} + \varphi(a_{i(n)}), n \geq 0$$

one derives (by the strongly regressive property of φ)

$$a \leq \varphi(a + 0) < a; \ \text{absurd; whence, } a = 0.$$

In other words, all convergent subsequences $(a_{i(n)}; n \geq 0)$ of (a_n) fulfill $a_{i(n)} \to 0$ as $n \to \infty$. This necessarily gives $a_n \to 0$ as $n \to \infty$; and the conclusion follows.

A bi-dimensional version of this result may be stated as follows. By a *pseudometric* over N, we mean any mapping $a : N \times N \to R_+$. If, in addition,

$$a(.,.) \text{ is reflexive: } a(n, n) = 0, \ \forall n$$

we say that $a(.,.)$ is a *r-pseudometric*. Given the r-pseudometric $a(.,.)$, call it *Cauchy*, in case of $[a(n, m) \to 0$ as $n, m \to \infty, n \leq m]$; and *asymptotic*, provided $[a(n, n + 1) \to 0$ as $n \to \infty]$. Clearly,

$$a(.,.) \text{ is Cauchy implies } a(.,.) \text{ is asymptotic;}$$

the reciprocal is not in general true.

Theorem 6 *Let the r-pseudometrics $a(.,.)$ and $b(.,.)$ over N and the function $\varphi \in \mathscr{F}_0(re, in)(R_+)$ be such that*

(34-i) $a(n, m) \leq b(n, m) + \varphi(a(n, m))$, *for all* $n, m \in N$, $n \leq m$.

In addition, suppose that

(34-ii) φ *is strongly regressive* ($\varphi(t + 0) < t, \forall t > 0$)

(34-iii) φ *is complementary coercive* ($\varphi^* := I - \varphi$ *is coercive).*

Then, necessarily,

(34-a) $b(.,.)$ is asymptotic implies $a(.,.)$ is asymptotic
(34-b) $b(.,.)$ is Cauchy implies $a(.,.)$ is Cauchy.

Proof There are two steps to be passed.

Step 1. Suppose that

$$b(.,.) \text{ is asymptotic } [(\beta_n := b(n, n+1); n \geq 0) \text{ fulfills } \beta_n \to 0].$$

We have to establish that

$$a(.,.) \text{ is asymptotic } [(\alpha_n := a(n, n+1); n \geq 0) \text{ fulfills } \alpha_n \to 0].$$

This, however, by the sequential condition

(seq) $\alpha_n \leq \beta_n + \varphi(\alpha_n)$, for all $n \in N$,

is deductible from the preceding statement; so that, our claim follows.

Step 2. Now, let us show that

$$b(.,.) \text{ is Cauchy implies } a(.,.) \text{ is Cauchy.}$$

The last property means (by definition)

$$\forall \varepsilon > 0, \text{ there exists } j(\varepsilon), \text{ such that } j(\varepsilon) \leq n \leq m \text{ implies } a(n,m) \leq \varepsilon;$$

or, equivalently [passing to the successor $J(\varepsilon) := j(\varepsilon) + 1$ and remembering that $a(.,.)$ is r-pseudometric]

$$\forall \varepsilon > 0, \text{ there exists } j(\varepsilon), \text{ such that } j(\varepsilon) < n < m \text{ implies } a(n,m) \leq \varepsilon.$$

The negation of this means: there exists $\varepsilon > 0$, such that

$$C(j) = \{(n,m) \in N \times N;\ j < n < m, a(n,m) > \varepsilon\} \neq \emptyset, \forall j.$$

Denote, for each j

$$n(j) = \min \mathrm{Dom}(C(j)), m(j) = \min C(j)(n(j)).$$

Fix $j(0) \in N$. By this construction, there exists a couple of ranks $(j(1) = n(j(0)),$ $j(2) = m(j(0))$ with $j(0) < j(1) < j(2), a(j(1), j(2)) > \varepsilon$. Further, given this index $j(2)$ there exists a couple of ranks $(j(3) = n(j(2)),\ j(4) = m(j(2))$ with $j(2) < j(3) < j(4), a(j(3), j(4)) > \varepsilon$. The procedure may continue indefinitely (without any choice techniques); and yields a strictly ascending sequence of ranks $(j(n); n \geq 0)$ [hence, $j(n) \to \infty$ as $n \to \infty$], with

$$\alpha_n := a(j(2n+1), j(2n+2)) > \varepsilon, \forall n.$$

On the other hand, by the Cauchy property

$$\beta_n := b(j(2n+1),\, j(2n+2)) \to 0,\ \text{as } n \to \infty.$$

This, in combination with the sequential condition

(seqq) $\alpha_n \leq \beta_n + \varphi(\alpha_n)$, for all $n \in N$

gives a contradiction with respect to the preceding statement. Hence, necessarily, $a(.,.)$ is Cauchy; as claimed.

An interesting question is that of such conclusions being retainable when one of the basic hypotheses about φ is to be dropped. The answer to this is negative, in general; we do not give details.

4 Topological Preliminaries

In the following, some basic concepts and results involving topological structures are given, with a special emphasis on gauge spaces.

Let $N = \{0, 1, \ldots\}$ denote the set of natural numbers. Given $p \geq 1$, each set M equivalent with $N(p, >) := \{0, \ldots, p-1\}$ (in the sense: there exists a bijection between M and $N(p, >)$) is called *p-finite*. For completeness, we accept that the empty set $\emptyset$ is 0-finite. Finally, let us say that the set M is *finite*, provided it is q-finite, for some $q \in N$.

Let E be a nonempty set. Remember that $\exp[E]$ denotes the class of all subsets in E. Given the subset $\mathscr{D} \subseteq \exp[E]$ of (i.e.: a family of subsets in E), remember that its union $\cup\mathscr{D}$ is introduced, axiomatically, as

$$z \in \cup\mathscr{D} \text{ iff } z \in H,\ \text{for some } H \in \mathscr{D};\ \text{hence, in particular, } \cup\emptyset = \emptyset.$$

On the other hand, whenever $\mathscr{D}$ is nonempty, then $\cap\mathscr{D}$ is introduced as

$$z \in \cap\mathscr{D} \text{ iff } z \in H,\ \text{for each } H \in \mathscr{D}.$$

Finally, when $\mathscr{D} = \emptyset$, we put $\cap\mathscr{D} = E$. We must stress that this last definition is entirely *locally*; i.e.: it is restricted to the ambient set E and the class $\exp[E]$.

(A) **[General aspects]**

Let X be a nonempty set. We say that the family $\mathscr{G} \subseteq \exp[X]$ is

$$\textit{semi-normal},\ \text{if } \emptyset \in \mathscr{G};\ \textit{normal},\ \text{if } \{\emptyset, X\} \subseteq \mathscr{G}.$$

By a *topology* on X, we mean any normal family $\mathscr{T} \subseteq \exp[X]$, with the properties (cf. Bourbaki [10, Ch I, Sect 1])

(top-1) the union of any subset in $\mathscr{T}$ belongs to $\mathscr{T}$
(top-2) the intersection of any finite subset of $\mathscr{T}$ is in $\mathscr{T}$.

In this case, the couple $(X, \mathscr{T})$ will be called a *topological space*. Any $D \in \mathscr{T}$ will be called *open*; and its complement $X \setminus D$ is termed *closed* (with respect to $\mathscr{T}$).

An equivalent way of introducing such a concept is the *complementary* one. Let us say that the normal family $\mathscr{S} \subseteq \exp[X]$ is a *cotopology* on X, when

(cotop-1) the intersection of any subset in $\mathscr{S}$ belongs to $\mathscr{S}$
(cotop-2) the union of any finite subset of $\mathscr{S}$ is in $\mathscr{S}$.

In this case, the couple $(X, \mathscr{S})$ will be called a *cotopological space*.

Given the topology $\mathscr{T}$ on X, the (normal) family $\mathscr{S} \subseteq \exp[X]$ introduced as $[E \in \mathscr{S}$ iff $X \setminus E \in \mathscr{T}]$ is a cotopology on X. This will be referred to as the *cotopology induced* by the topology $\mathscr{T}$; denoted as: $\mathscr{S} = X \setminus \mathscr{T}$. Conversely, given the cotopology $\mathscr{S}$ on X, the (normal) family $\mathscr{T} \subseteq \exp[X]$ introduced as $[D \in \mathscr{T}$ iff $X \setminus D \in \mathscr{S}]$ is a topology on X. This will be referred to as the *topology induced* by the cotopology $\mathscr{S}$; denoted as: $\mathscr{T} = X \setminus \mathscr{S}$.

(B) **[Topological constructions]**

Let X be a nonempty set. There are several ways of constructing a topology on X; the basic ones are described below.

(I) We say that the selfmap $A \mapsto \mathrm{Klo}(A)$ of $\exp[X]$ is a *(Kuratowski) closure* over X, provided (cf. Kuratowski [38, Ch I, Sect 4])

(Klo-1) $\emptyset = \mathrm{Klo}(\emptyset)$, $X = \mathrm{Klo}(X)$
(Klo-2) $A \subseteq \mathrm{Klo}(A)$, for each $A \in \exp[X]$
(Klo-3) $\mathrm{Klo}(A \cup B) = \mathrm{Klo}(A) \cup \mathrm{Klo}(B)$, $\forall A, B \in \exp[X]$
(Klo-4) $\mathrm{Klo}(\mathrm{Klo}(A)) = \mathrm{Klo}(A)$, for each $A \in \exp[X]$.

Note that, as a direct consequence of (Klo-3),

(Klo-incr) $A \subseteq B$ implies $\mathrm{Klo}(A) \subseteq \mathrm{Klo}(B)$.

In fact, let $A, B \in \exp[X]$ be such that $A \subseteq B$. As $B = A \cup (B \setminus A)$, we must have $[\mathrm{Klo}(B) = \mathrm{Klo}(A) \cup \mathrm{Klo}(B \setminus A) \supseteq \mathrm{Klo}(A)]$; wherefrom, the assertion follows.

Let $\mathscr{T}$ be a topology over X. Define a selfmap $A \mapsto \mathrm{cl}(A)$ of $\exp[X]$, as:

$$(\forall A \in \exp[X]) : \ \mathrm{cl}(A) = \text{the intersection of all } G \in \exp[X] \text{ with } A \subseteq G = \text{closed}.$$

It is not hard to see that all properties (Klo-1)-(Klo-4) hold; whence, $A \mapsto \mathrm{cl}(A)$ is a Kuratowski closure over X; referred to as the *closure operator* induced by $\mathscr{T}$. Conversely, suppose that $A \mapsto \mathrm{Klo}(A)$ is a Kuratowski closure. Then, the (normal) class $\mathscr{S} = \{E \in \exp[X]; E = \mathrm{Klo}(E)\}$ is a cotopology over X (see above); so that, $\mathscr{T} = X \setminus \mathscr{S}$ is a topology over X in the above discussed sense. Moreover, with respect to this topology, we have

$$\mathrm{cl}(A) = \mathrm{Klo}(A), \ \text{for each } A \in \exp[X].$$

(II) Suppose that, to each $x \in X$ we attached a class $\mathscr{W}(x) \subseteq \exp[X]$ (referred to as: *neighborhoods system* of x) such that the family $(\mathscr{W}(x); x \in X)$ fulfills

(ns-1) for each $x \in X$ and each $W \in \mathscr{W}(x)$, we have $x \in W$
(ns-2) $(\forall x \in X)$: $A \in \mathscr{W}(x)$ and $A \subseteq B$ imply $B \in \mathscr{W}(x)$; hence, in particular, $X \in \mathscr{W}(x)$
(ns-3) $(\forall x \in X)$: $W_1, W_2 \in \mathscr{W}(x)$ imply $W_1 \cap W_2 \in \mathscr{W}(x)$
(ns-4) for each $x \in X$ and each $A \in \mathscr{W}(x)$ there exists $D \in \mathscr{W}(x)$ such that $A \in \mathscr{W}(y)$, for each $y \in D$ (hence, $A \supseteq D$).

We then say that $(\mathscr{W}(x); x \in X)$ is a *neighborhoods system* over X. A concrete example is to be given as follows. Let $\mathscr{T}$ be a topology over X. Given $x \in X$, let us say that $Y \in \exp[X]$ is a $\mathscr{T}$*-neighborhood* of x, provided

$$x \in D \subseteq Y, \text{ for some } D \in \mathscr{T};$$

the class of all sets will be denoted as $\mathscr{V}(x)$. It is not hard to see that all properties (ns-1)-(ns-4) hold; whence, $(\mathscr{V}(x); x \in X)$ is a neighborhoods system over X; referred to as the neighborhoods system generated by $\mathscr{T}$. Concerning the reciprocal question, the following statement holds (cf. Costinescu [17, Ch II, Sect 1]):

Proposition 7 *Suppose that $(\mathscr{W}(x); x \in X)$ is a neighborhoods system over X; and put $\mathscr{T}$ = the set of all $D \in \exp[X]$ with $(D \in \mathscr{W}(x), \forall x \in D)$. Then,*

(41-1) $\mathscr{T}$ is (a normal family and) a topology over X
(41-2) with respect to this topology, $(\mathscr{V}(x) = \mathscr{W}(x)$, for all $x \in X)$.

A variant of this construction is to be described as follows. Suppose that, to each $x \in X$ we attached a class $\mathscr{W}^*(x) \subseteq \exp[X]$ (referred to as: neighborhoods subsystem of x) such that the family $(\mathscr{W}^*(x); x \in X)$ fulfills the properties

(nss-1) for each $x \in X$ and each $W^* \in \mathscr{W}^*(x)$, we have $x \in W^*$
(nss-2) $(\forall x \in X)$: for each $W_1^*, W_2^* \in \mathscr{W}^*(x)$ there exists $W_3^* \in \mathscr{W}^*(x)$, such that $W_1^* \cap W_2^* \supseteq W_3^*$
(nss-3) for each $x \in X$ and each $A^* \in \mathscr{W}^*(x)$ there exists $D^* \in \mathscr{W}^*(x)$, such that for each $y \in D^*$, there exists $B_y^* \in \mathscr{W}^*(y)$ with $A^* \supseteq B_y^*$.

We then say that $(\mathscr{W}^*(x); x \in X)$ is a *neighborhoods subsystem* over X. A concrete example is to be given as follows. Let $\mathscr{T}$ be a topology over X; and, for each $x \in X$, let $\mathscr{V}(x)$ be the neighborhoods system over X generated by $\mathscr{T}$. Then, for each $x \in X$, let the subset $\mathscr{V}^*(x)$ of $\mathscr{V}(x)$ be taken as

$$\text{for each } V \in \mathscr{V}(x) \text{ where exists } V^* \in \mathscr{V}^*(x) \text{ with } V \supseteq V^*.$$

Then, $(\mathscr{V}^*(x); x \in X)$ is a neighborhoods subsystem over X; referred to as: the neighborhoods subsystem over X generated by $\mathscr{T}$. For example, $(\mathscr{V}^*(x); x \in X)$ may be taken as the *open neighborhoods subsystem* over X generated by $\mathscr{T}$:

$$\mathscr{V}^*(x) = \{D \in \mathscr{T}; x \in D\}, \ x \in X.$$

Conversely, let $(\mathscr{W}^*(x); x \in X)$ be a neighborhoods subsystem over X. Then, the class $(\mathscr{W}(x); x \in X)$ introduced as

$$(\forall x \in X): \ W \in \mathscr{W}(x) \text{ iff } W \supseteq W^*, \text{ for some } W^* \in \mathscr{W}^*(x)$$

is a neighborhoods system over X, as it can be directly seen. By a preceding statement, there exists a (uniquely determined) topology $\mathscr{T}$ over X such that $\mathscr{W}(x) = \mathscr{V}(x)$ (the neighborhood system on X generated by $\mathscr{T}$). But then, $(\mathscr{W}^*(x); x \in X)$ is a neighborhoods subsystem on X generated by $\mathscr{T}$.

(III) Let us say that the subset $\mathscr{B} \subseteq \exp[X]$ is a *basis* for $\mathscr{T}$, provided

$$\text{each } D \in \mathscr{T} \text{ is the union of a subset in } \mathscr{B}.$$

In particular, this means that $\mathscr{B} \subseteq \mathscr{T}$; but, in general, we cannot have $\mathscr{B} = \mathscr{T}$.

A useful characterization of this concept is to be obtained by means of neighborhood subsystems generated by $\mathscr{T}$. Let $\mathscr{B} \subseteq \mathscr{T}$ be a family of open sets. Denote

$$\mathscr{B}(x) = \{B \in \mathscr{B}; x \in B\}, \ x \in X.$$

We have (see Costinescu [17, Ch II, Sect 2] for details)

Proposition 8 *Let $\mathscr{B} \subseteq \mathscr{T}$ be a family of open sets. The following are equivalent:*

(42-1) $\mathscr{B}$ is a basis for the topology $\mathscr{T}$
(42-2) $(\mathscr{B}(x); x \in X)$ is a neighborhoods subsystem over X generated by $\mathscr{T}$: for each $x \in X$ and each $V \in \mathscr{V}(x)$ there exists $B \in \mathscr{B}(x)$ with $V \supseteq B$.

Let $\mathscr{B} \subseteq \exp[X]$ be a class of sets. We may ask of to what extent there exists a topology $\mathscr{T}$ on X such that $\mathscr{B}$ is the basis for $\mathscr{T}$. To this end, note that if $\mathscr{B} \subseteq \exp[X]$ is a basis for the topology $\mathscr{T}$, then (by the properties of $(\mathscr{B}(x); x \in X)$)

(s-dir) $\mathscr{B}$ is *strongly directed*: for each $B_1, B_2 \in \mathscr{B}$ and each $x \in B_1 \cap B_2$ there exists $B_3 \in \mathscr{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$
(B-total) $\mathscr{B}$ is *total*: $X = \cup \mathscr{B}$ $(\forall x \in X, \exists B \in \mathscr{B}$ such that $x \in B)$.

Theorem 7 *Let the subset $\mathscr{B} \subseteq \exp[X]$ be strongly directed, total; and put*

$$\mathscr{T} = \text{the class of all } D \in \exp[X] \text{ with } D = \text{union of a subset in } \mathscr{B}.$$

Then,

(41-a) $\mathscr{T}$ is (a normal family and) a topology on X
(41-b) $\mathscr{B}$ is a basis of the topology $\mathscr{T}$.

Proof (cf. Engelking [22, Ch 1, Sect 1.2]). There are two steps to be passed.

Step 1. Clearly, $\mathscr{T}$ is a normal family, by the total property. In addition, each union of elements in $\mathscr{T}$ belongs to $\mathscr{T}$. It remains to establish that the intersection

of any finite subset in $\mathscr{T}$ belongs to $\mathscr{T}$; this, clearly, amounts to

$$D_1, D_2 \in \mathscr{T} \text{ imply } D_1 \cap D_2 \in \mathscr{T}.$$

Let $D_1, D_2 \in \mathscr{T}$ be given. For the arbitrary fixed $x \in D_1 \cap D_2$, there exist $B_1, B_2 \in \mathscr{B}$ such that $x \in B_1 \subseteq D_1$, $x \in B_2 \subseteq D_2$; hence, $x \in B_1 \cap B_2 \subseteq D_1 \cap D_2$. From the strongly directed property, there exists $B_3 \in \mathscr{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2 \subseteq D_1 \cap D_2$. This, along with the arbitrariness of x in $D_1 \cap D_2$, tells us that

$$D_1 \cap D_2 = \text{union of a subset in } \mathscr{B};\ \text{whence, } D_1 \cap D_2 \in \mathscr{T};$$

so that, $\mathscr{T}$ is a topology on X.

Step 2. By the preceding step, $(B = \cup\{B\} \in \mathscr{T}, \forall B \in \mathscr{B})$; and this, along with definition of $\mathscr{T}$, gives the desired conclusion.

(IV) Given the topology $\mathscr{T}$ over X, let us say that the family $\mathscr{A} \subseteq \exp[X]$ is a *subbasis* of it, when

$$\text{the class } \mathscr{B} \text{ of intersections of finite subsets in } \mathscr{A} \text{ is a basis for } \mathscr{T}.$$

Clearly, this in particular means that $\mathscr{A} \subseteq \mathscr{B}$. Moreover, as $\cup\mathscr{A} = \cup\mathscr{B}$ and $\cup\mathscr{B} = X$, we have $\cup\mathscr{A} = X$; i.e.: $\mathscr{A}$ is *total*.

Conversely, let $\mathscr{A} \subseteq \exp[X]$ be a class of sets. As before, we may ask whether there exists a topology $\mathscr{T}$ on X such that $\mathscr{A}$ is a subbasis for $\mathscr{T}$. By the above developments, this happens when

$$\mathscr{B} = \text{the class of intersections of finite subsets in } \mathscr{A}$$

is endowed with the properties

(s-dir) $\mathscr{B}$ is *strongly directed*: for each $B_1, B_2 \in \mathscr{B}$ and each $x \in B_1 \cap B_2$ there exists $B_3 \in \mathscr{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$

(B-total) $\mathscr{B}$ is *total*: $X = \cup\mathscr{B}$ $(\forall x \in X, \exists B \in \mathscr{B}$ such that $x \in B)$.

The former of these is evident (in our case), via

$$B_1, B_2 \in \mathscr{B} \text{ implies } B_1 \cap B_2 \in \mathscr{B}.$$

So, it remains to verify the latter; which, under $\cup\mathscr{A} = \cup\mathscr{B}$, means

(A-total) $\mathscr{A}$ is *total*: $X = \cup\mathscr{A}$ $(\forall x \in X, \exists A \in \mathscr{A}$ such that $x \in A)$.

Summing up, the following answer to the posed question is available.

Theorem 8 *Let the family $\mathscr{A} \subseteq \exp[X]$ be total; and put*

$$\mathscr{T} = \{D \in \exp[X];\ D = \textit{union of intersections of finite subsets in } \mathscr{A}\}.$$

Then,

(42-a) $\mathscr{T}$ *is (a normal family and) a topology on* X
(42-b) $\mathscr{A}$ *is a subbasis of the topology* $\mathscr{T}$.

(C) **[Comparison criteria]**
In the following, some elementary facts involving comparison of topologies will be discussed.

Given the couple of topological spaces $(X, \mathscr{T})$ and $(Y, \mathscr{S})$, let us say that $f : X \to Y$ is $(\mathscr{T}, \mathscr{S})$-continuous, when

$$\text{for each } G \in \mathscr{S} \text{ we have } f^{-1}(G) \in \mathscr{T}.$$

Note that, an equivalent characterization of this property is by means of cotopologies; precisely, $f : X \to Y$ is $(\mathscr{T}, \mathscr{S})$-continuous, iff

$$\text{for each } H \in Y \setminus \mathscr{S} \text{ we have } f^{-1}(H) \in X \setminus \mathscr{T}.$$

This concept may be used towards the comparison of topologies. Some preliminaries are needed. Let X be a nonempty set. Given the topologies $\mathscr{T}, \mathscr{S}$ over X, let us introduce the relation

$$\mathscr{T} \subseteq \mathscr{S} :\ \text{each } \mathscr{T}\text{-open set is } \mathscr{S}\text{-open}.$$

This will be referred to as: $\mathscr{T}$ is *coarser* than $\mathscr{S}$; or: $\mathscr{S}$ is *finer* than $\mathscr{T}$.

Let $i_X : X \to X$ stand for the *identity selfmap* $(i_X(x) = x; x \in X)$. The following result is now immediate; so, we do not give details.

Proposition 9 *For the topologies* $\mathscr{T}, \mathscr{S}$ *over* X*, the following are equivalent:*

(43-1) $\mathscr{T}$ *is coarser than* $\mathscr{S}$ *(i.e.:* $\mathscr{S}$ *is finer than* $\mathscr{T}$*)*
(43-2) the identical application $i_X : X \to X$ *is* $(\mathscr{S}, \mathscr{T})$*-continuous.*

This characterization is a handy tool for constructing topologies on a (nonempty) set X. Given a nonempty index set Λ, let $((Y_\lambda, \mathscr{S}_\lambda); \lambda \in \Lambda)$ be a family of topological spaces, and $(f_\lambda : X \to Y_\lambda; \lambda \in \Lambda)$ be a family of maps. We are interested to determine the (minimal) topology $\mathscr{T}$ over X with respect to which

$$f_\lambda \text{ is } (\mathscr{T}, \mathscr{S}_\lambda)\text{-continuous, for each } \lambda \in \Lambda.$$

As we shall see, this topology is to be introduced by the subbase

$$\mathscr{A} = \cup\{f_\lambda^{-1}(\mathscr{S}_\lambda); \lambda \in \Lambda\};\ \text{where, by definition,}$$
$$f_\lambda^{-1}(\mathscr{S}_\lambda) = \{f_\lambda^{-1}(E_\lambda);\ E_\lambda \in \mathscr{S}_\lambda\}, \lambda \in \Lambda.$$

In fact, $\mathscr{A}$ is total ($X = \cup\mathscr{A}$); because [$\emptyset$, $X \in f_\lambda^{-1}(\mathscr{S}_\lambda)$, for each $\lambda \in \Lambda$]. By the preceding statement, $\mathscr{A}$ is a subbase of a topology $\mathscr{T}$ over X, represented as

$$\mathscr{T} = \{D \in \exp[X];\ D = \text{union of intersections of finite subsets in } \mathscr{A}\}.$$

The obtained topology is the desired one. Precisely, we have

Theorem 9 *Under the above conventions,*

(43-a) f_λ *is* $(\mathscr{T}, \mathscr{S}_\lambda)$*-continuous, for each* $\lambda \in \Lambda$

(43-b) If the topology $\mathscr{Z}$ *over* X *fulfills [*f_λ *is* $(\mathscr{Z}, \mathscr{S}_\lambda)$*-continuous,* $\forall\lambda \in \Lambda$*] then, necessarily,* $\mathscr{T} \subseteq \mathscr{Z}$ *(*$\mathscr{T}$ *is coarser than* $\mathscr{Z}$*).*

Remark 1 The following alternate representation of this topology is useful in many concrete cases. For each $\lambda \in \Lambda$, let $\mathscr{S}_\lambda^*$ be a subbasis of $\mathscr{S}_\lambda$; note that, by definition, $\mathscr{S}_\lambda^*$ is total. Then, let us denote

$$\mathscr{A}^* = \cup\{f_\lambda^{-1}(\mathscr{S}_\lambda^*);\ \lambda \in \Lambda\};\ \text{where, by definition,}$$
$$f_\lambda^{-1}(\mathscr{S}_\lambda^*) = \{f_\lambda^{-1}(E_\lambda^*);\ E_\lambda^* \in \mathscr{S}_\lambda^*\},\ \lambda \in \Lambda.$$

From the above observation, $\mathscr{A}^*$ is total too; so, the formula

$$\mathscr{T}^* = \{D \in \exp[X];\ D = \text{union of intersections of finite subsets in } \mathscr{A}^*\}$$

defines a topology over X. We claim that, necessarily,

$$\mathscr{T} = \mathscr{T}^*;\ \text{i.e.: } \mathscr{T} \subseteq \mathscr{T}^* \text{ and } \mathscr{T}^* \subseteq \mathscr{T}.$$

The latter inclusion is clear, in view of $\mathscr{A}^* \subseteq \mathscr{A}$; so, it remains to establish the former inclusion. This, in turn, reduces to

(f-int) each intersection of a finite subset in $\mathscr{A}$ belongs to $\mathscr{T}^*$;

for, in such a case,

$$D \in \mathscr{T} \text{ implies } D\text{=union of elements in } \mathscr{T}^*;\ \text{whence } D \in \mathscr{T}^*.$$

Now, (f-int) means, ultimately: $A, B \in \mathscr{A}$ implies $A \cap B \in \mathscr{T}^*$. To verify this, let $A, B \in \mathscr{A}$; hence,

$$A = f_\lambda^{-1}(G_\lambda),\ B = f_\mu^{-1}(H_\mu),\ \text{where } \lambda, \mu \in \Lambda,\ G_\lambda \in \mathscr{S}_\lambda,\ H_\mu \in \mathscr{S}_\mu.$$

Take some $x \in A \cap B$. As $f_\lambda(x) \in G_\lambda$, there exists G_λ^*=intersection of a finite subset in $\mathscr{S}_\lambda^*$, such that $f_\lambda(x) \in G_\lambda^* \subseteq G_\lambda$. Likewise, as $f_\mu(x) \in H_\mu$, there exists H_μ^*=intersection of a finite subset in $\mathscr{S}_\mu^*$, such that $f_\mu(x) \in H_\mu^* \subseteq H_\mu$. Combining these, yields

$$x \in f_\lambda^{-1}(G_\lambda^*) \cap f_\mu^{-1}(H_\mu^*) \subseteq A \cap B;\ \text{with}$$

$$f_\lambda^{-1}(G_\lambda^*) \cap f_\mu^{-1}(H_\mu^*) = \text{intersection of a finite subset in } \mathscr{A}^*;$$

and this, by the arbitrariness of $x \in A \cap B$, gives

$$A \cap B = \text{union of intersections of finite subsets in } \mathscr{A}^*;\ \text{so, } A \cap B \in \mathscr{T}^*.$$

Further aspects may be found in Dugundji [20, Ch IV, Sect 1].

A basic particular case of these developments corresponds to the choice

$$Y_\lambda = X,\ f_\lambda = i_X,\ \text{for all } \lambda \in \Lambda.$$

Precisely, let Λ be a nonempty index set; and $(\mathscr{S}_\lambda;\ \lambda \in \Lambda)$ be a family of topologies over X. We have

Theorem 10 *Under the described setting, there exists a unique topology $\mathscr{T}$ over X having as subbase the class of sets $\mathscr{A} = \cup\{\mathscr{S}_\lambda;\ \lambda \in \Lambda\}$, and representable as*

$$\mathscr{T} = \{D \in \exp[X];\ D = \textit{union of intersections of finite subsets in } \mathscr{A}\}.$$

Moreover, the following properties hold:

(44-a) $\mathscr{S}_\lambda \subseteq \mathscr{T}$, for each $\lambda \in \Lambda$
(44-b) If the topology $\mathscr{Z}$ over X fulfills $(\mathscr{S}_\lambda \subseteq \mathscr{Z},\ \forall \lambda \in \Lambda)$ then, $\mathscr{T} \subseteq \mathscr{Z}$.

In other words:

(44-c) $\mathscr{T} = \sup\{\mathscr{S}_\lambda;\ \lambda \in \Lambda\}$ (in the inclusion sense).

Remark 2 The following alternate representation of this topology is useful in many concrete cases. For each $\lambda \in \Lambda$, let $\mathscr{S}_\lambda^*$ be a subbasis of $\mathscr{S}_\lambda$; note that, by definition, $\mathscr{S}_\lambda^*$ is total. Then, denote $\mathscr{A}^* = \cup\{\mathscr{S}_\lambda^*;\ \lambda \in \Lambda\}$. From the above observation, $\mathscr{A}^*$ is also total; so, the formula

$$\mathscr{T}^* = \{D \in \exp[X];\ D = \text{union of intersections of finite subsets in } \mathscr{A}^*\}$$

defines a topology over X. We claim that, necessarily,

$$\mathscr{T} = \mathscr{T}^*;\ \text{i.e.: } \mathscr{T} \subseteq \mathscr{T}^* \text{ and } \mathscr{T}^* \subseteq \mathscr{T}.$$

The latter inclusion is clear, in view of $\mathscr{A}^* \subseteq \mathscr{A}$; so, it remains to establish the former inclusion. This, in turn, reduces to

(f-int-id) each finite intersection of elements in $\mathscr{A}$ belongs to $\mathscr{T}^*$;

for, in such a case,

$$D \in \mathscr{T} \text{ implies } D = \text{union of elements in } \mathscr{T}^*;\ \text{whence } D \in \mathscr{T}^*.$$

Now, (f-int-id) means, ultimately: $A, B \in \mathscr{A}$ implies $A \cap B \in \mathscr{T}^*$. To verify this, let $A, B \in \mathscr{A}$; hence,

$$A = G_\lambda, B = H_\mu, \text{ where } \lambda, \mu \in \Lambda, G_\lambda \in \mathscr{S}_\lambda, H_\mu \in \mathscr{S}_\mu.$$

Take some $x \in A \cap B$. As $x \in G_\lambda$, there exists G_λ^*=intersection of a finite subset in $\mathscr{S}_\lambda^*$, such that $x \in G_\lambda^* \subseteq G_\lambda$. Likewise, as $x \in H_\mu$, there exists H_μ^*=intersection of a finite subset in $\mathscr{S}_\mu^*$, such that $x \in H_\mu^* \subseteq H_\mu$. Combining these, yields

$$x \in G_\lambda^* \cap H_\mu^* \subseteq A \cap B; \text{ with } G_\lambda^* \cap H_\mu^* \text{=intersection of a finite subset in } \mathscr{A}^*;$$

and this, by the arbitrariness of $x \in A \cap B$, gives

$$A \cap B \text{=union of intersections of finite subsets in } \mathscr{A}^*; \text{ so, } A \cap B \in \mathscr{T}^*.$$

(D) **[Basic concepts]**

In the following, some completions of the above facts will be provided. Let $(X, \mathscr{T})$ be a topological space.

Given $A \in \exp[X]$, let us say that $x \in X$ is *interior* to A, when $A \in \mathscr{V}(x)$; clearly, $x \in A$. The class of all these points will be denoted as $\text{int}(A)$ (the *interior* of A). A global characterization of this operator is given as

$$\text{int}(A) \text{=the union of all } D \in \mathscr{T} \text{ with } D \subseteq A.$$

Given $A \in \exp[X]$, let us say that $x \in X$ is *adherent* to A, when $V \cap A \neq \emptyset$, for all $V \in \mathscr{V}(x)$; clearly, $x \in A$ in not in general true. The class of all these points is just $\text{cl}(A)$ (the *closure* (or, *adherence*) of A); clearly, $\text{int}(A) \subseteq \text{cl}(A)$.

The introduced operators $A \mapsto \text{int}(A)$ and $A \mapsto \text{cl}(A)$ (from $\exp[X]$ to itself) are *duals*, in the sense

$$X \setminus \text{int}(A) = \text{cl}(X \setminus A), \ X \setminus \text{cl}(A) = \text{int}(X \setminus A), \ A \in \exp[X].$$

This means that each property involving int(.) has a dual property involving cl(.), and vice versa; we do not give details.

Given $A \in \exp[X]$, we say that $x \in X$ is *boundary* to A, when $x \in \text{cl}(A) \cap \text{cl}(X \setminus A)$; as before, such a point need not be in A. The class of all these writes

$$\text{bd}(A) = \text{cl}(A) \cap \text{cl}(X \setminus A) = \text{cl}(A) \setminus \text{int}(A) \text{ (the } boundary \text{ of } A);$$

clearly, $\text{bd}(A) \subseteq \text{cl}(A)$. As a consequence of this,

$$\text{cl}(A) = (\text{cl}(A) \setminus \text{int}(A)) \cup \text{int}(A) = \text{bd}(A) \cup \text{int}(A), \ A \in \exp[X].$$

Let again $(X, \mathscr{T})$ be a topological space. Define the properties

(H-sep) $(X, \mathscr{T})$ is *(Hausdorff) separated*: $\forall x, y \in X$ with $x \neq y$, there exist $V \in \mathscr{V}(x)$, $W \in \mathscr{V}(y)$ with $V \cap W = \emptyset$

(q-comp) $(X, \mathscr{T})$ is quasi-compact: $X = \cup\mathscr{G}$ for $\mathscr{G} \subseteq \mathscr{T}$ implies $X = \cup\mathscr{H}$, where $\mathscr{H}$ =finite part of $\mathscr{G}$.

When both these properties hold, we say that $(X, \mathscr{T})$ is *compact*. The basic properties of these concepts are well known; so, further details are not provided.

(E) **[Metrical structures]**

Some basic applications of these facts are to be given in a metrical context.

(I) Let M be a nonempty set. Given the map (=pseudometric) $d : M \times M \to R_+$, consider the properties

(m-1) d is *reflexive*: $d(x, x) = 0$, for all $x \in X$
(m-2) d is *triangular*: $d(x, y) \leq d(x, z) + d(z, y)$, $\forall x, y, z \in X$
(m-3) d is *symmetric*: $d(x, y) = d(y, x)$, for all $x, y \in X$
(m-4) d is *sufficient*: $d(x, y) = 0$ implies $x = y$.

When (m-1)-(m-3) hold, we say that $d(., .)$ is a *semimetric* on M; and (M, d) is called a *semimetric space*. On the other hand, when (m-1)-(m-4) hold, we say that $d(., .)$ is a *metric* on M; and (M, d) is called a *metric space*.

Let (M, d) be a semimetric space. The topology to be considered here is generated by the family of open spheres in M. Precisely, given $a \in M$, $\rho > 0$, denote

$$M(a, \rho)(d) = \{x \in M; d(a, x) < \rho\},\ M[a, \rho](d) = \{x \in M; d(a, x) \leq \rho\};$$

these will be referred to as the *open* (respectively, *closed*) sphere with center $a \in M$ and radius $\rho > 0$. (Clearly, both these spheres are nonempty; because $a \in M$ is an element of them).

Define the family of (nonempty) sets

$$\mathscr{W}^*(a)(d) = \{M(a, \rho)(d); \rho > 0\}, a \in M.$$

It is not hard to see that $(\mathscr{W}^*(a)(d); a \in M)$ is a neighborhoods subsystem over M. Its attached family of (nonempty) sets $(\mathscr{W}(a)(d); a \in M)$ defined as

$$(\forall a \in M) :\ W \in \mathscr{W}(a)(d) \text{ iff } W \supseteq W^*,\ \text{for some } W^* \in \mathscr{W}^*(a)(d)$$

is a neighborhoods system over M. By a preceding statement, there exists a unique topology $\mathscr{T}(d)$ over M such that

$$\mathscr{W}(a)(d) \text{ is just the } \mathscr{T}(d)\text{-neighborhoods system of } a,\ \text{for each } a \in M.$$

According to its definition, we have, for each nonempty $D \in \mathscr{T}(d)$ and each $a \in D$,

$$D \in \mathscr{W}(a)(d);\ \text{so, there exists } \rho > 0 \text{ such that } M(a, \rho)(d) \subseteq D.$$

Moreover, $\mathscr{B}^*(d) = \{\mathscr{W}^*(a)(d); a \in M\}$ is a subbasis of $\mathscr{T}(d)$; so,

$$(\text{for each } D \in \mathscr{T}(d)):\ D=\text{union of intersections of finite subsets in } \mathscr{B}^*(d).$$

In particular, $\mathscr{B}^*(d) \subseteq \mathscr{T}(d)$; whence,

$$M(a, \rho)(d) \text{ is open, for each } a \in M, \rho > 0.$$

On the other hand, by the semimetrical properties of d,

$$(\forall a \in M, \forall \rho > 0):\ M^e(a, \rho)(d) := \{x \in M; d(a, x) > \rho\} \text{ is open;}$$

$$\text{whence, } M[a, \rho](d) = M \setminus M^e(a, \rho)(d) \text{ is closed.}$$

Finally, when no confusion can arise, it would be convenient to drop the index (d) from all notations above.

(II) Let (M, d) be a metric space. The (topological) compactness property of this space was already introduced. A related notion is the following. Let us say that (M, d) is *sequentially compact*, provided

(s-comp) each sequence in M has a convergent subsequence.

Proposition 10 *The following is valid, in (ZF-AC+DC)*

$$(M, d) \textit{ is compact iff } (M, d) \textit{ is sequentially compact.}$$

The verification is very similar with the one in Costinescu [17, Ch VII, Sect 3].

(III) Let (P, d) and (Q, e) be a couple of metric spaces; and $f : P \to Q$ be a mapping. Let us say that f is *sequentially continuous*, when

$$\forall \text{ sequence } (x_n) \text{ in } P \text{ and } \forall \text{ element } x \in P :\ x_n \xrightarrow{d} x \text{ implies } f(x_n) \xrightarrow{e} f(x).$$

Proposition 11 *We have, in (ZF-AC+DC),*

$$f \textit{ is continuous iff } f \textit{ is sequentially continuous.}$$

Proof

(i): Suppose that f is continuous (after the general definition); and let the sequence (x_n) in P and the point $x \in P$ be such that $x_n \xrightarrow{d} x$. Let $\varepsilon > 0$ be arbitrary fixed. By the topological definition of continuity, there exists $\delta > 0$ such

that $f(P(x,\delta)) \subseteq Q(f(x),\varepsilon)$. Further, given this $\delta > 0$, there exists (by hypothesis) an index $n(\delta) \in N$ such that $n \geq n(\delta)$ implies $x_n \in P(x,\delta)$. Combining these gives

$$n \geq n(\delta) \text{ implies } f(x_n) \in Q(f(x),\varepsilon);$$

and this, by the arbitrariness of $\varepsilon > 0$, tells us that $f(x_n) \xrightarrow{e} f(x)$.

(ii): Suppose that f is sequentially continuous; but f is not continuous (after the topological definition): there exists $\varepsilon > 0$ such that

$$(\forall \delta > 0): \ H(\delta) := f(P(x,\delta)) \setminus Q(f(x),\varepsilon) \text{ is nonempty.}$$

Let $(\delta_n; n \geq 0)$ be a strictly descending sequence in R_+^0 with $\delta_n \to 0$. (For, example, one may take $(\delta_n = 2^{-n}; n \geq 0)$; but this is not the only possible choice). By the Denumerable Axiom of Choice (AC(N)) (deductible in (ZF-AC+DC)), there may be determined a sequence (x_n) in P with

$$(\forall n): \ x_n \in H(\delta_n) \text{ (that is; } x_n \in P(x,\delta_n) \text{ and } f(x_n) \notin Q(f(x),\varepsilon)).$$

By the first half of this relation, $x_n \xrightarrow{d} x$; so that, by the sequential continuity,

$f(x_n) \xrightarrow{e} f(x)$; wherefrom:

there exists some index $n(\varepsilon) \geq 0$, with $f(x_n) \in Q(f(x),\varepsilon), \forall n \geq n(\varepsilon)$.

This, however, is in contradiction with the second half of underlying relation; and then, the conclusion follows.

(IV) Let again (P,d) and (Q,e) be a couple of metric spaces; and $f: P \to Q$ be a mapping. Let us say that f is *uniformly continuous*, when

$$\forall \varepsilon > 0, \exists \delta > 0, \text{ such that } d(x,y) \leq \delta \text{ implies } e(f(x),f(y)) \leq \varepsilon.$$

In this case, we may construct a mapping $\psi: R_+ \to R_+ \cup \{\infty\}$, according to

$$\psi(t) = \sup\{e(f(x),f(y)); d(x,y) \leq t\}, t \in R_+.$$

This will be referred to as the *uniform continuity modulus* of f; the class of all these will be denoted as $\mathscr{F}_u(R_+, R_+ \cup \{\infty\})$. Clearly,

(uc-1) ψ is increasing and zero-continuous $[\psi(0) = 0 = \psi(0+0)]$,
(uc-2) $e(f(x),f(y)) \leq \psi(d(x,y)), \forall x,y \in P$;

since the verification is immediate, we do not give details.

A basic example of this type is to be obtained in the compactness context.

Theorem 11 *Suppose that* (P, d) *is sequentially compact. Then, the generic inclusion is valid, in (ZF-AC+DC):*

$$(\forall f \in \mathscr{F}(P, Q)) : \ f \text{ is continuous implies } f \text{ is uniformly continuous}.$$

Proof Suppose that f is continuous, but not uniformly continuous. There exists then $\varepsilon > 0$, such that

$$C(\delta) := \{(x, y) \in P \times P; d(x, y) \leq \delta, e(f(x), f(y)) > \varepsilon\} \neq \emptyset, \text{ for each } \delta > 0.$$

Let $(\delta_n; n \geq 0)$ be a sequence in R_+^0 with $\delta_n \to 0$. (For example, we may take $(\delta_n = 2^{-n}; n \geq 0)$; hence, this is not depending on (AC)). From the Denumerable Axiom of Choice (AC(N)) (deductible, as precise, in (ZF-AC+DC)), we get a sequence $((x_n, y_n); n \geq 0)$ in $P \times P$, such that

$$(x_n, y_n) \in C(\delta_n), \forall n; \text{ that is: } d(x_n, y_n) \leq \delta_n, e(f(x_n), f(y_n)) > \varepsilon, \forall n.$$

From the sequential compactness of (P, d), there exists subsequences (x_n^*), (y_n^*) of (x_n) and (y_n) respectively, and elements $x^*, y^* \in P$, such that (from the first part of this relation)

$$d(x_n^*, y_n^*) \leq \delta_n, \forall n, x_n \xrightarrow{d} x^*, y_n^* \xrightarrow{d} y^*; \text{ whence, } x^* = y^*.$$

Combining with the second part of the same relation, we get (by a limit process)

$$0 = e(f(x^*), f(y^*)) \geq \varepsilon; \text{ contradiction.}$$

Hence, our working assumption is not acceptable; and the conclusion follows.

(F) **[Gauge spaces]**

Let X be a nonempty set; and Λ be some (nonempty) index set. For each $\lambda \in \Lambda$, let $d_\lambda : X \times X \to R_+$ be a semimetric over X; and $\mathscr{T}(d_\lambda)$ stand for the associated topology (see above).

Technically speaking, these data generate two basic structures on X.

(I) The former of these is represented by the supremum topology of the family $\{\mathscr{T}(d_\lambda); \lambda \in \Lambda\}$. Precisely, note that for each $\lambda \in \Lambda$,

$$\mathscr{A}_\lambda = \{X(x, \rho)(d_\lambda); x \in X, \rho > 0\} \text{ is a subbase of } \mathscr{T}(d_\lambda);$$

hence, in particular, $\mathscr{A}_\lambda$ is total. Denote further

$$\mathscr{A} = \cup\{\mathscr{A}_\lambda; \lambda \in \Lambda\}; \text{ clearly, } \mathscr{A} \text{ is total too.}$$

By a preceding result, the formula

$$\mathscr{T} = \{D \in \exp[X];\ D = \text{union of intersections of finite subsets in } \mathscr{A}\}$$

defines a topology over X. This, as precise, is just the supremum topology of the family $\{\mathscr{T}(d_\lambda);\ \lambda \in \Lambda\}$. Let cl(.) stand for the associated closure operator. Note that, according to its definition, we have for each (nonempty) subset Y of X

(g-int) $x \in \text{int}(Y)$ iff there exists an intersection W of a finite subset in $\mathscr{A}$ with $x \in W \subseteq Y$

(g-clo) $x \in \text{cl}(Y)$ iff each intersection W of a finite subset in $\mathscr{A}$ with $x \in W$ fulfills $W \cap Y \neq \emptyset$.

(II) The latter of these is represented by the (sequential) conv-Cauchy structure induced by the family $D = (d_\lambda;\ \lambda \in \Lambda)$ of these semimetrics.

Take an arbitrary sequence $(x_n;\ n \geq 0)$ in X. Given $\lambda \in \Lambda$, the d_λ-*convergence* of this sequence towards an $x \in X$ [depicted as: $x_n \xrightarrow{d_\lambda} x$], means:

$$d_\lambda(x_p, x) \to 0 \text{ as } p \to \infty$$

$$\text{(i.e.: } \forall \varepsilon > 0, \exists i = i(\lambda, \varepsilon), \text{ such that } i \leq p \Longrightarrow d_\lambda(x_p, x) \leq \varepsilon).$$

If this holds for all $\lambda \in \Lambda$, then $(x_n;\ n \geq 0)$ is said to D-*converge* towards x [written as: $x_n \xrightarrow{D} x$]. The set of all such points x will be denoted as $D - \lim_n(x_n)$; when it is nonempty, then $(x_n;\ n \geq 0)$ is called D-*convergent*. On the other hand, given $\lambda \in \Lambda$, the d_λ-Cauchy property of $(x_n;\ n \geq 0)$ means:

$$d_\lambda(x_p, x_q) \to 0 \text{ as } p, q \to \infty,\ p \leq q$$

$$\text{(i.e.: } \forall \varepsilon > 0, \exists j := j(\lambda, \varepsilon), \text{ such that } j \leq p \leq q \Longrightarrow d_\lambda(x_p, x_q) \leq \varepsilon).$$

If this holds for each $\lambda \in \Lambda$, we say that (x_n) is D-*Cauchy*; the class of all such sequences will be denoted as $Cauchy(D)$. By definition, the triple $(X, \Lambda; D)$ endowed with the conv-Cauchy structure $((\xrightarrow{D}), Cauchy(D))$ and the regularity condition

$$D \text{ is } sufficient:\ x, y \in X \text{ and } (d_\lambda(x, y) = 0, \forall \lambda \in \Lambda) \text{ imply } x = y$$

is called a *gauge space*. Note that, in this setting, any D-convergent sequence is D-Cauchy too; the reciprocal is not in general valid.

Using the previous conventions, we may introduce a D-*closure* operator $A \mapsto \text{Dcl}(A)$ from $\exp[X]$ to itself, as: for each $A \in \exp[X]$,

$$y \in \text{Dcl}(A) \text{ iff } y = D - \lim_n(y_n), \text{ for some sequence } (y_n) \text{ in } A.$$

It is not hard to see that $A \mapsto \mathrm{Dcl}(A)$ is a *semi-closure* over X, in the sense

(Dcl-1) (identity) $\emptyset = \mathrm{Dcl}(\emptyset)$, $X = \mathrm{Dcl}(X)$,
(Dcl-2) (progressiveness) $Y \subseteq \mathrm{Dcl}(Y)$, $\forall Y \in \exp[X]$,
(Dcl-3) (additivity) $\mathrm{Dcl}(U \cup V) = \mathrm{Dcl}(U) \cup \mathrm{Dcl}(V)$, $\forall U, V \in \exp[X]$.

Note that, as a direct consequence of these, we also have

(Dcl-4) (monotonicity) $Y_1 \subseteq Y_2$ implies $\mathrm{Dcl}(Y_1) \subseteq \mathrm{Dcl}(Y_2)$.

Further, call $A \in \exp[X]$, *D-closed* provided $A = \mathrm{Dcl}(A)$; note that in this case,

$$\text{the } D\text{-limit of each sequence in } A \text{ belongs to } A.$$

Denote, for simplicity

$$\mathscr{K}[D]=\text{the class of all } D\text{-closed subsets in } X.$$

Clearly, $\mathscr{K}[D] \subseteq \exp[X]$ is a normal family, with the finite union property

(K-D-1) the union of any finite subset of $\mathscr{K}[D]$ is in $\mathscr{K}[D]$.

However, the arbitrary intersection property

(K-D-2) the intersection of any subset in $\mathscr{K}[D]$ belongs to $\mathscr{K}[D]$

is not in general true; so that, $\mathscr{K}[D]$ is not a cotopology on X; we then say that $\mathscr{K}[D]$ is an *almost cotopology* on X. Equivalently, this tells us that

$$\mathscr{T}[D] = X \setminus \mathscr{K}[D] \text{ (in the sense: } Z \in \mathscr{T}[D] \text{ iff } X \setminus Z \in \mathscr{K}[D])$$

is a normal family with the finite intersection property

(T-D-1) the intersection of any finite subset of $\mathscr{T}[D]$ is in $\mathscr{T}[D]$.

However, the arbitrary union property

(T-D-2) the union of any subset in $\mathscr{T}[D]$ belongs to $\mathscr{T}[D]$

is not in general true; so that, $\mathscr{T}[D]$ is not topology on X; we then say that $\mathscr{T}[D]$ is an *almost topology* on X [referred to as the *gauge almost topology*]. An explanation of this bad property is due to the closure operator $A \mapsto \mathrm{Dcl}(A)$ being *not involutive*; i.e., a property like

(Dcl-inv) $\mathrm{Dcl}(\mathrm{Dcl}(A)) = \mathrm{Dcl}(A)$, for each $A \in \exp[X]$

is not in general true; so that, $A \mapsto \mathrm{Dcl}(A)$ is not a closure over X according to Kuratowski [38, Ch I, Sect 4]. Further aspects of this problem may be found in Engelking [22, Ch 1, Sect 1.2].

Finally, let us compare the supremum topology $\mathscr{T} = \sup\{\mathscr{T}(d_\lambda); \lambda \in \lambda\}$ with the gauge almost topology $\mathscr{T}[D]$. In this direction, we have the properties

(p-1) $(\forall Y \in \exp[X])$: $\mathrm{Dcl}(Y) \subseteq \mathrm{cl}(Y)$
(p-2) $\mathscr{T} \subseteq \mathscr{T}[D]$; i.e.: Y=closed (open) implies Y=D-closed (D-open).

In fact, letting $x \in \mathrm{Dcl}(Y)$, there exists a sequence (x_n) in Y with

$$x_n \stackrel{D}{\longrightarrow} x; \text{ hence, } x_n \stackrel{d_\lambda}{\longrightarrow} x, \text{ for each } \lambda \in \Lambda.$$

Let $W = \cap\{X(x, \rho_i)(d_{\lambda(i)}); i \in I\}$ be a finite intersection (of open spheres) that includes x; where, I is a (nonempty) finite index set. By the above convergence property, there must be some index $n(x)$, such that

$$\{x_n; n \geq n(x)\} \subseteq W; \text{ whence, } Y \cap W \neq \emptyset.$$

This yields $x \in \mathrm{cl}(Y)$; and proves the first property. As a consequence,

$$Y\text{=closed implies } \mathrm{cl}(Y) = Y \subseteq \mathrm{Dcl}(Y) \subseteq \mathrm{cl}(Y); \text{ hence, } Y\text{=}D\text{-closed;}$$

wherefrom, the second property holds too.

5 Gheorghiu Functional Contractions

Let X be a nonempty set; and Λ be some nonempty (index) set. For each $\lambda \in \Lambda$, let $d_\lambda : X \times X \to R_+$ be a semimetric over X_λ; and $\mathscr{T}(d_\lambda)$ stand for the associated to d_λ topology over X.

Technically speaking, these data generate two basic structures on X.

(I) The former of these is represented by the supremum topology of the family $\{\mathscr{T}(d_\lambda); \lambda \in \Lambda\}$. For the moment, this topology is not essential for us; however, in the homotopical fixed point theory, a limited use of it will be made in the reasonings to be developed.
(II) The latter of these is represented by the (sequential) conv-Cauchy structure induced by the family $D = (d_\lambda; \lambda \in \Lambda)$ of these semimetrics. This, essentially, may be described in the following way.

Take an arbitrary sequence $(x_n; n \geq 0)$ in X. Given $\lambda \in \Lambda$, the *d_λ-convergence* of this sequence towards an $x \in X$ [depicted as: $x_n \stackrel{d_\lambda}{\longrightarrow} x$], means:

$$d_\lambda(x_p, x) \to 0 \text{ as } p \to \infty$$
$$\text{(i.e.: } \forall \varepsilon > 0, \exists i = i(\lambda, \varepsilon), \text{ such that } i \leq p \Longrightarrow d_\lambda(x_p, x) \leq \varepsilon).$$

If this holds for all $\lambda \in \Lambda$, then $(x_n; n \geq 0)$ is said to *D-converge* towards x [written as: $x_n \stackrel{D}{\longrightarrow} x$]. The set of all such points x will be denoted as $D - \lim_n(x_n)$; when it is nonempty, then $(x_n; n \geq 0)$ is called *D-convergent*. On the other hand, given $\lambda \in \Lambda$, the d_λ-Cauchy property of $(x_n; n \geq 0)$ means:

$$d_\lambda(x_p, x_q) \to 0 \text{ as } p, q \to \infty,\ p \le q$$

$$\text{(i.e.: } \forall \varepsilon > 0, \exists j := j(\lambda, \varepsilon), \text{ such that } j \le p \le q \Longrightarrow d_\lambda(x_p, x_q) \le \varepsilon).$$

If this holds for each $\lambda \in \Lambda$, we say that $(x_n; n \ge 0)$ is *D-Cauchy*. Note that any D-convergent sequence is D-Cauchy too; the reciprocal is not in general valid. By definition, the triple $(X, \Lambda; D)$ endowed with this conv-Cauchy structure and with the regularity condition

$$D \text{ is } sufficient: \ x, y \in X \text{ and } (d_\lambda(x, y) = 0, \forall \lambda \in \Lambda) \text{ imply } x = y$$

is called a *gauge space*.

Let $(X, \Lambda; D)$ be a gauge space. Remember that the subset Y of X, is called *asingleton*, if $[y_1, y_2 \in Y$ imply $y_1 = y_2]$; and *singleton* if, in addition, Y is nonempty; note that, in this case $Y = \{y\}$, for some $y \in X$.

Further, let $T : X \to X$ be a selfmap of X. As usual, $\mathrm{Fix}(T) := \{z \in X; z = Tz\}$ stands for the set of all fixed points of T in X. These are to be determined in the context below, comparable with the one in Rus [50, Ch 2, Sect 2.2]:

(gpic-1) We say that T is *fix-asingleton*, if $\mathrm{Fix}(T)$ is an asingleton; and *fix-singleton*, if $\mathrm{Fix}(T)$ is a singleton

(gpic-2) We say that $x \in X$ is a *Picard point* (modulo (D, T)), when the iterative sequence $(T^n x; n \ge 0)$ is D-Cauchy. If this property holds for all $x \in X$, we say that T is a *Picard operator* (modulo D)

(gpic-3) We say that $x \in X$ is a *strongly Picard point* (modulo (D, T)), when $(T^n x; n \ge 0)$ is D-convergent with $\lim_n(T^n x) \in \mathrm{Fix}(T)$. If this property holds for all $x \in X$, we say that T is a *strongly Picard operator* (modulo D).

The sufficient (regularity) conditions for such properties are being founded on *orbital* concepts (in short: o-concepts). Given $x \in X$, let us say that the sequence $(z_n; n \ge 0)$ in X is *T-orbital* with respect to x, when $(z_n = T^n x; n \ge 0)$; if $x \in X$ is generic, we then say that $(z_n; n \ge 0)$ is a *T-orbital* (or, equivalently: an *o-sequence*).

(greg-1) Call X, *(o;D)-complete*, provided (for each o-sequence) D-Cauchy $\Longrightarrow$ D-convergent

(greg-2) We say that T is *(o;D)-continuous*, if [(z_n)=o-sequence and $z_n \xrightarrow{D} z$] imply $Tz_n \xrightarrow{D} Tz$.

As a completion of these, we must now formulate the metrical contractive conditions upon our data. Let $\Omega = (\omega_\lambda; \lambda \in \Lambda)$ be a family over $\mathscr{F}(R_+)$. Then, let us introduce the mappings: for each $x, y \in X$, and each $\lambda \in \Lambda$

$$M_1(x, y; T; d_\lambda) = d_\lambda(x, y),$$

$$M_2(x, y; T; d_\lambda) = \max\{d_\lambda(x, Tx), d_\lambda(y, Ty)\},$$

$$M_3(x, y; T; d_\lambda) = (1/2)[d_\lambda(x, Ty) + d_\lambda(y, Tx)],$$

$$M(x, y; T; d_\lambda) = \max\{M_1(x, y; T; d_\lambda), M_2(x, y; T; d_\lambda), M_3(x, y; T; d_\lambda)\}.$$

We say that T is (Ω, M)-*contractive*, if

$$d_\lambda(Tx, Ty) \le \omega_\lambda(M(x, y; T; d_\lambda)),\ \forall x, y \in X, \forall \lambda \in \Lambda.$$

The deep part of the main result to be stated is an auxiliary fact involving almost Picard properties over semimetric structures. Some preliminaries are needed.

Let X be a nonempty set; and $d : X \times X \to R_+$ be a semimetric over X; the couple (X, d) will be referred to as a *semimetric space*. Further, take some $T \in \mathscr{F}(X)$. In the following, sufficient conditions are given so that

(AP) T is almost Picard (modulo d); in the sense
(ap-1) $(\forall x \in X)$: the iterative sequence $(T^n x; n \ge 0)$ is d-Cauchy
(ap-2) $(\forall x \in X, \forall z \in X)$: $T^n x \xrightarrow{d} z$ implies $d(z, Tz) = 0$.

The conditions in question are to be written as follows. Let the functional classes

$$\mathscr{F}_0(R_+), \mathscr{F}_0(re)(R_+), \mathscr{F}_0(in)(R_+), \mathscr{F}_0(re, in)(R_+)$$

be introduced as before. Then, let us define the mappings: for each $x, y \in X$,

$$M_1(x, y; T; d) = d(x, y),$$

$$M_2(x, y; T; d) = \max\{d(x, Tx), d(y, Ty)\},$$

$$M_3(x, y; T; d) = (1/2)[d(x, Ty) + d(y, Tx)],$$

$$M(x, y; T; d) = \max\{M_1(x, y; T; d), M_2(x, y; T; d), M_3(x, y; T; d)\}.$$

Given $\omega \in \mathscr{F}(R_+)$, we say that T is (ω, M)-*contractive*, provided

$$d(Tx, Ty) \le \omega(M(x, y; T; d)), \forall x, y \in X.$$

The following almost Picard type statement involving these data is available.

Theorem 12 *Suppose that the selfmap T is (ω, M)-contractive, where the function $\omega \in \mathscr{F}_0(re, in)(R_+)$ fulfills*

ω is Meir-Keeler admissible; or, equivalently: Matkowski admissible.

Then, T is an almost Picard operator (see above).

A direct verification was provided in Leader [39]. For completeness reasons, we will however develop the argument, with certain modifications.

Proof Fix some point $x_0 \in X$, and put $(x_n = T^n x_0; n \geq 0)$.

Step 1. From the contraction hypothesis, we have

$$d(x_{n+1}, x_{n+2}) \leq \omega(\max\{d(x_n, x_{n+1}), d(x_{n+1}, x_{n+2}), (1/2)d(x_n, x_{n+2})\}), \forall n.$$

In view of

$$1/2)d(x_n, x_{n+2}) \leq \max\{d(x_n, x_{n+1}), d(x_{n+1}, x_{n+2})\}, \forall n,$$

the underlying relation becomes

$$d(x_{n+1}, x_{n+2}) \leq \omega(\max\{d(x_n, x_{n+1}), d(x_{n+1}, x_{n+2})\}), \forall n.$$

If, for some n, one has

$$d(x_{n+1}, x_{n+2}) > d(x_n, x_{n+1})(\geq 0),$$

then, by the above relation (and regressiveness of ω)

$$d(x_{n+1}, x_{n+2}) \leq \omega(d(x_{n+1}, x_{n+2})) < d(x_{n+1}, x_{n+2}); \text{ a contradiction.}$$

Hence, necessarily,

$$d(x_{n+1}, x_{n+2}) \leq d(x_n, x_{n+1}); \text{ whence, } d(x_{n+1}, x_{n+2}) \leq \omega(d(x_n, x_{n+1})), \ \forall n;$$

and this (along with ω being Matkowski admissible) yields

$$(x_n) \text{ is } d\text{-asymptotic: } r_n := d(x_n, x_{n+1}) \to 0 \text{ as } n \to \infty.$$

Step 2. Let $\varepsilon > 0$ be arbitrary fixed; and $\delta > 0$ be attached to it, by the Meir-Keeler admissible property of ω:

$$\varepsilon < t < \varepsilon + \delta \text{ implies } \omega(t) \leq \varepsilon; \text{ or, equivalently}$$
$$t < \varepsilon + \delta \text{ implies } \omega(t) \leq \varepsilon \text{ (as } \omega\text{=regressive, } \omega(0) = 0);$$

clearly, without loss, one may assume that $\delta < \varepsilon$. By the d-asymptotic property we just established, there exists a rank $n(\delta)$, such that

(d-asy) $(\forall n \geq n(\delta))$: $d(x_n, x_{n+1}) < \delta/4$; hence, $d(x_n, x_{n+2}) < \delta/2$.

We prove by induction that, for each $j \geq 1$, the following relation holds

(d-C;j) $d(x_n, x_{n+j}) < \varepsilon + \delta/2, \forall n \geq n(\delta)$;

wherefrom, $(x_n; n \geq 0)$ is d-Cauchy. The case $j \in \{1, 2\}$ is evident, by the evaluation (d-asy). Assume that (d-C;j) holds for all $j \in \{1, \ldots, p\}$, where

$p \geq 2$; we must establish that (d-C;p+1) holds too. Let $n \geq n(\delta)$ be arbitrary fixed. From the inductive hypothesis (and our asymptotic relation)

$$d(x_n, x_{n+p}), d(x_{n+1}, x_{n+p}) < \varepsilon + \delta/2$$
$$d(x_n, x_{n+1}), d(x_{n+p}, x_{n+p+1}) < \delta/4 < \varepsilon + \delta/2.$$

This, along with the triangular inequality, gives us

$$d(x_n, x_{n+p+1}) \leq d(x_n, x_{n+p}) + d(x_{n+p}, x_{n+p+1}) < \varepsilon + 3\delta/4;$$
$$\text{so that, by definition, } M(x_n, x_{n+p}; T; d) < \varepsilon + \delta.$$

Combining with contractive hypothesis and Meir-Keeler property of ω, gives

$$d(x_{n+1}, x_{n+p+1}) = d(Tx_n, Tx_{n+p}) \leq \omega(M(x_n, x_{n+p}; T; d)) \leq \varepsilon;$$

wherefrom (again by the triangular inequality)

$$d(x_n, x_{n+p+1}) \leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+p+1}) < \varepsilon + \delta/4 < \varepsilon + \delta/2;$$

and our claim follows.

Step 3. Suppose that $z \in X$ is such that

$$x_n \xrightarrow{d} z \text{ (i.e.: } d(x_n, z) \to 0), \text{ as } n \to \infty.$$

We claim that, necessarily, $d(z, Tz) = 0$. Suppose not: $b := d(z, Tz) > 0$. From the contractive condition,

$$d(x_{n+1}, Tz) \leq \omega(M(x_n, z; T; d)), \forall n.$$

By the convergence and d-asymptotic properties, there exists a rank $n(b)$, with

$$(\forall n \geq n(b)) : \ d(x_n, z), d(x_n, Tx_n), d(Tx_n, z) < b/2;$$
$$\text{whence, } M_1(x_n, z; T; d) < b/2, M_2(x_n, z; T; d) = b.$$

This, along with the triangle inequality, gives

$$(\forall n \geq n(b)) : \ d(x_n, Tz) \leq d(x_n, z) + d(z, Tz) < 3b/2;$$
$$\text{wherefrom } M_3(x_n, z; T; d) < b, M(x_n, z; T; d) = b.$$

Replacing these into the contractive condition, we derive

$$d(x_{n+1}, Tz) \leq \omega(b), \forall n \geq n(b).$$

Passing to limit as $n \to \infty$ gives (by a previous auxiliary fact)

$$b \leq \omega(b) < b;\ \text{contradiction.}$$

Hence, our initial assumption about z cannot be true; so that, $d(z, Tz) = 0$. The proof is thereby complete.

Remark 3 Remember that $\omega \in \mathscr{F}_0(re)(R_+)$ is Matkowski admissible, when

$$\omega \text{ is Boyd-Wong admissible: } \Lambda^+\omega(s) < s, \forall s \in R_+^0.$$

In particular, when $\omega \in \mathscr{F}_0(re, in)(R_+)$, this last condition means

$$\omega \text{ is } strongly\ regressive\ [\omega(s+0) < s,\ \text{for all } s > 0].$$

It is worth noting that, under the Boyd-Wong property, there is a specific way of completing the second part of statement above. For technical reasons, we will describe it in what follows.

Suppose that the d-asymptotic sequence (x_n) is not d-Cauchy; and fix some couple $(\beta, \gamma) \in (>; \text{adm}(R_+^0; (x_n)))$ as well as the rank sequence $(\lambda(k) = 1; k \geq 0)$. By a previous auxiliary statement, there exist $b \in R_+^0$, a rank sequence $(J(k); k \geq 0)$ in $N(1, \leq)$ and a couple of rank-sequences $(m(k); k \geq 0)$, $(n(k); k \geq 0)$, with

(p-1) $k + 1 \leq J(k) \leq m(k) < m(k) + 3\lambda(k) < n(k), \forall k \geq 0$
(p-2) $\forall p, q \in N[0, 3\lambda(0)], u_k(p, q) := d(x_{m(k)+p}, x_{n(k)+q}) \to b+$ as $k \to \infty$.

From the contractive hypothesis,

$$u_k(1, 1) \leq \omega(v_k), \forall k;\ \text{where, by definition}$$

$$v_k = \max\{u_k(0, 0), r_{m(k)}, r_{n(k)}, (1/2)[u_k(0, 1) + u_k(1, 0)]\}, k \geq 0.$$

By the above convergence properties,

$$v_k \to b+ \text{ as } k \to \infty;\ \text{because } u_k(0, 0), u_k(0, 1), u_k(1, 0) \to b+ \text{ as } k \to \infty.$$

Passing to lim sup as $k \to \infty$, in the previous relation, gives (cf. a preceding result)

$$b \leq \Lambda^+\omega(b) < b;\ \text{contradiction.}$$

Hence, our working assumption is not acceptable; and the conclusion follows.

Under these preliminaries, we may now pass to our basic fixed point result in this exposition.

Theorem 13 *Suppose that T is (Ω, M)-contractive, where $\Omega = (\omega_\lambda; \lambda \in \Lambda)$ is a family over $\mathscr{F}_0(re, in)(R_+)$, fulfilling*

$$(\forall \lambda \in \Lambda)\text{: } \omega_\lambda \text{ is Meir-Keeler admissible}$$

(or, equivalently: Matkowski admissible).

In addition, let X be (o;D)-complete. Then,

(52-a) T is fix-asingleton
(52-b) T is strongly Picard (modulo D).

Proof There are several steps to be passed.

Part 1 We prove that T is fix-asingleton. Let $z_1, z_2 \in \mathrm{Fix}(T)$ be arbitrary fixed. By the contractive condition,

$$d_\lambda(z_1, z_2) \le \omega_\lambda(d_\lambda(z_1, z_2)),\ \lambda \in \Lambda.$$

This, by the regressive property of the functions in Ω, gives

$$d_\lambda(z_1, z_2) = 0,\ \forall \lambda \in \Lambda;\ \text{wherefrom (as } D \text{ is sufficient) } z_1 = z_2.$$

Part 2 Take some $x_0 \in X$, and put $(x_n := T^n x_0; n \ge 0)$; clearly, (x_n) is orbital. By the preceding semimetric statement, one derives that

(52-c) (x_n) is D-Cauchy (i.e.: d_λ-Cauchy, for each $\lambda \in \Lambda$).

Part 3 Summing up, $(x_n; n \ge 0)$ is an orbital D-Cauchy sequence. As X is (o;D)-complete, there must be some (uniquely determined) $z \in X$, with

$$x_n \xrightarrow{D} z \text{ (hence, } x_n \xrightarrow{d_\lambda} z, \text{ for each } \lambda \in \Lambda).$$

Again by the preceding semimetric statement, we get

(52-d) $d_\lambda(z, Tz) = 0$, for each $\lambda \in \Lambda$.

This, along with D=sufficient, gives $z = Tz$; i.e.: $z \in \mathrm{Fix}(T)$.

In particular, when when Λ is a singleton, this main result covers the one in Leader [39]. Passing to the general case, note that our main result includes the trivial quasi-order version of the related 2017 one in Turinici [67]; and has large overlaps with the fixed point result in Agarwal et al. [1]. In fact, it is not difficult to get a quasi-order version of our main result so as to include both these statements; we do not give details. An interesting problem to be posed is that of this statement having Maia type extensions like in Gheorghiu [26]; further aspects will be discussed elsewhere. Some applications of such results to integral equations may be found in Gheorghiu and Turinici [28] or Cherichi and Samet [15]; see also the Angelov's monograph [4, Ch II, Sect 2.2].

6 Homotopic Fixed Points

Let X be a nonempty set; and Λ be some nonempty (index) set. For each $\lambda \in \Lambda$, let $d_\lambda : X \times X \to R_+$ be a semimetric over X; and $\mathscr{T}(d_\lambda)$ stand for the associated topology. Remember that there are two basic structures on X to be used further.

(I) The former of these is represented by the supremum topology of the family $\{\mathscr{T}(d_\lambda);\ \lambda \in \Lambda\}$. Precisely, note that for each $\lambda \in \Lambda$,

$$\mathscr{A}_\lambda = \{X(x, \rho)(d_\lambda);\ x \in X,\ \rho > 0\} \text{ is a subbase of } \mathscr{T}(d_\lambda);$$

hence, in particular, $\mathscr{A}_\lambda$ is total. Denote further

$$\mathscr{A} = \cup\{\mathscr{A}_\lambda;\ \lambda \in \Lambda\};\ \text{clearly, } \mathscr{A} \text{ is total too.}$$

By a preceding result, the formula

$$\mathscr{T} = \{D \in \exp[X];\ \ D = \text{union of intersections of finite subsets in } \mathscr{A}\}$$

defines a topology over X. This, as precise, is just the supremum topology of the family $\{\mathscr{T}(d_\lambda);\ \lambda \in \Lambda\}$. Note that, according to its definition, we have for each (nonempty) subset Y of X and each $x \in X$,

(g-int) $x \in \text{int}(Y)$ iff there exists an intersection W of a finite subset in $\mathscr{A}$ with $x \in W \subseteq Y$

(g-clo) $x \in \text{cl}(Y)$ iff each intersection W of a finite subset in $\mathscr{A}$ with $x \in W$ fulfills $W \cap Y \neq \emptyset$.

(II) The latter of these is represented by the (sequential) conv-Cauchy structure induced by the family $D = (d_\lambda;\ \lambda \in \Lambda)$ of these semimetrics. This, essentially, may be described in the following way.

Take an arbitrary sequence $(x_n;\ n \geq 0)$ in X. Given $\lambda \in \Lambda$, the *d_λ-convergence* of this sequence towards an $x \in X$ [depicted as: $x_n \xrightarrow{d_\lambda} x$], means:

$$d_\lambda(x_p, x) \to 0 \text{ as } p \to \infty$$

$$\text{(i.e.: } \forall \varepsilon > 0, \exists i = i(\lambda, \varepsilon), \text{ such that } i \leq p \Longrightarrow d_\lambda(x_p, x) \leq \varepsilon).$$

If this holds for all $\lambda \in \Lambda$, then $(x_n;\ n \geq 0)$ is said to *D-converge* towards x [written as: $x_n \xrightarrow{D} x$]. The set of all such points x will be denoted as $D - \lim_n(x_n)$; when it is nonempty, then $(x_n;\ n \geq 0)$ is called *D-convergent*. On the other hand, given $\lambda \in \Lambda$, the d_λ-Cauchy property of $(x_n;\ n \geq 0)$ means:

$$d_\lambda(x_p, x_q) \to 0 \text{ as } p, q \to \infty,\ p \leq q$$

$$\text{(i.e.: } \forall \varepsilon > 0, \exists j := j(\lambda, \varepsilon), \text{ such that } j \leq p \leq q \Longrightarrow d_\lambda(x_p, x_q) \leq \varepsilon).$$

If this holds for each $\lambda \in \Lambda$, we say that $(x_n;\ n \geq 0)$ is *D-Cauchy*. Note that any D-convergent sequence is D-Cauchy too; the reciprocal is not in general valid. By definition, the triple $(X, \Lambda; D)$ endowed with this conv-Cauchy structure and the regularity condition

$$D \text{ is } sufficient: \ x, y \in X \text{ and } (d_\lambda(x, y) = 0, \forall \lambda \in \Lambda) \text{ imply } x = y$$

is called a *gauge space*.

Suppose that we fixed such a structure. Define a D-closure operator $A \mapsto \mathrm{Dcl}(A)$ from $\exp[X]$ to itself, as: for each $A \in \exp[X]$,

$$y \in \mathrm{Dcl}(A) \text{ iff } y = D - \lim_n (y_n), \text{ for some sequence } (y_n) \text{ in } A.$$

It is not hard to see that $A \mapsto \mathrm{Dcl}(A)$ is a *semi-closure* over X, in the sense

(Dcl-1) (identity) $\emptyset = \mathrm{Dcl}(\emptyset)$, $X = \mathrm{Dcl}(X)$,
(Dcl-2) (progressiveness) $Y \subseteq \mathrm{Dcl}(Y)$, $\forall Y \in \exp[X]$,
(Dcl-3) (additivity) $\mathrm{Dcl}(U \cup V) = \mathrm{Dcl}(U) \cup \mathrm{Dcl}(V)$, $\forall U, V \in \exp[X]$.

Note that, as a direct consequence of these, we also have

(Dcl-4) (monotonicity) $Y_1 \subseteq Y_2$ implies $\mathrm{Dcl}(Y_1) \subseteq \mathrm{Dcl}(Y_2)$.

Unfortunately, $A \mapsto \mathrm{Dcl}(A)$ is not involutive; i.e.,

(Dcl-inv) $\mathrm{Dcl}(\mathrm{Dcl}(A)) = \mathrm{Dcl}(A)$, for each $A \in \exp[Y]$

is not in general true; so that, $A \mapsto \mathrm{Dcl}(A)$ is not a closure over X according to Kuratowski [38, Ch I, Sect 4].

Further, let us say that $A \in \exp[X]$ is D-*closed*, provided $A = \mathrm{Dcl}(A)$; note that

$$\text{the } D\text{-limit of each sequence in } A \text{ belongs to } A.$$

Having these precise, let (J, g) be a metric space; and $(\leq)$ be a (partial) order over J; the triple $(J, g, \leq)$ will be then referred to as an *ordered metric space*. Let $(<)$ stand for the *attached strict order*:

$$t < s \text{ iff } t \leq s \text{ and } t \neq s \text{ [clearly, } (<) \text{ is irreflexive and transitive].}$$

Call $r \in J$, $(\leq)$-*maximal* provided

$$J(r, \leq) = \{r\}; \text{ or, equivalently, } J(r, <) = \emptyset;$$

the class of all these will be denoted as $\max(J, \leq)$. The negation of this property $(J(r, <) \neq \emptyset)$ will be referred to as r is $(\leq)$-*nonmaximal*; denote the class of all such points as $\mathrm{nmax}(J, \leq)$. The following properties upon $(J, g, \leq)$ will be admitted:

(Jgle-1) $(J, g, \leq)$ is *conv-bd-regular*: each ascending sequence (t_n) in J is g-convergent (hence, g-Cauchy) in J, with $t_n \leq \lim_n (t_n)$, $\forall n$
(Jgle-2) $(J, g, \leq)$ is *nonmaximal accessible*: each $r \in \mathrm{nmax}(J, \leq)$ is the g-limit of some sequence (p_n) in $J(r, <)$.

Further, take a mapping $(t, x) \mapsto H(t, x) = H(t)x$ from $J \times X$ into X; hence,

(s-map) $H(t)$ is a selfmap of X, for each $t \in J$.

Given the subset $U \in \exp(X)$, the family $\Psi = (\psi_\lambda; \lambda \in \Lambda)$ over $\mathscr{F}(R_+)$, and the family $\Omega = (\omega_\lambda; \lambda \in \Lambda)$ over $\mathscr{F}(R_+)$, let us say that H is a $(\Psi, \Omega; M; U)$-*homotopic mapping*, when the following conditions hold

(hom-1) H is first variable Ψ-contractive:
$d_\lambda(H(t, x), H(s, x)) \leq \psi_\lambda(g(t, s))$, for all $t, s \in J$ and all $x \in U$

(hom-2) H is second variable Ω-contractive:
$d_\lambda(H(t, x), H(t, y)) \leq \omega_\lambda(M(x, y; H(t); d_\lambda)), \forall t \in J, \forall x, y \in U$.

Here, letting d be any semimetric over X and $T : X \to X$ be any selfmap of X, we introduced the mappings (M_1, M_2, M_3, M) as: for each $x, y \in X$,

$$M_1(x, y; T; d) = d(x, y),$$

$$M_2(x, y; T; d) = \max\{d(x, Tx), d(y, Ty)\},$$

$$M_3(x, y; T; d) = (1/2)[d(x, Ty) + d(y, Tx)],$$

$$M(x, y; T; d) = \max\{M_1(x, y; T; d), M_2(x, y; T; d), M_3(x, y; T; d)\}.$$

Finally, let the functional classes

$$\mathscr{F}_0(R_+), \mathscr{F}_0(re)(R_+), \mathscr{F}_0(in)(R_+), \mathscr{F}_0(re, in)(R_+)$$

be introduced as before. Given $\omega \in \mathscr{F}_0(re, in)(R_+)$, let $(\eta(t) = \omega(2t); t \geq 0)$ stand for the *double function* attached to it; clearly, $\eta \in \mathscr{F}_0(in)(R_+)$. The conditions to be considered are essentially related to this last function; precisely

(s-reg) η is strongly regressive $(\eta(t + 0) < t, \forall t > 0)$;
referred to as: ω is *double strongly regressive*

(c-co) η is complementary coercive $[\eta^* = I - \eta$ is coercive];
referred to as: ω is *double complementary coercive*.

Denote for simplicity

$$J(H; U) = \{t \in J; \mathrm{Fix}(H(t)) \cap U \neq \emptyset\}.$$

We say that the homotopic map H is *transversal*, provided

$$J(H; U) \neq \emptyset \text{ implies } J(H; U) \cap \max(J, \leq) \neq \emptyset;$$

or, in other words: the generic property (involving points of J)

$$P(t) : \ \mathrm{Fix}(H(t)) \cap U \text{ is nonempty}$$

may jump from the points of J to the points of $\max(J, \leq)$.

The following homotopic fixed point theorem is available.

Theorem 14 *Let the ordered metric space $(J, g, \leq)$ with*

(61-i) $(J, g, \leq)$ *is conv-bd-regular and nonmaximal accessible*

and the subset $U \in \exp(X)$ be such that the map $(t, x) \mapsto H(t, x) = H(t)x$ in $\mathscr{F}(J \times X, X)$ is $(\Psi, \Omega; M; U)$-homotopic, where the family $\Psi = (\psi_\lambda; \lambda \in \Lambda)$ over $\mathscr{F}(R_+)$, and the family $\Omega = (\omega_\lambda; \lambda \in \Lambda)$ over $\mathscr{F}_0(re, in)(R_+)$, fulfill

(61-ii) $(\forall \lambda \in \Lambda)$*: ψ_λ is zero-continuous ($\psi_\lambda(t) \to 0 = \psi_\lambda(0)$ as $t \to 0$)*

(61-iii) $(\forall \lambda \in \Lambda)$*: the double function $(\eta_\lambda(t) = \omega_\lambda(2t); t \geq 0)$ is strongly regressive and complementary coercive.*

In addition, suppose that

(61-iv) X is D-complete and U is D-closed

(61-v) $J(H; U \setminus \mathrm{int}(U)) = \emptyset$*; i.e.:* $\mathrm{Fix}(H(t)) \cap (U \setminus \mathrm{int}(U)) = \emptyset$*, for all $t \in J$.*

Then, the homotopic map H is transversal in (ZF-AC+DC); i.e.,

(61-a) $J(H; U) \neq \emptyset$ *implies* $J(H; U) \cap \max(J, \leq) \neq \emptyset$.

Proof There are several steps to be passed.

Step 1. By the Zorn-Bourbaki metric maximality principle (ZB-m-mp) (see below), $\max(J, \leq)$ is nonempty. Moreover, by the posed hypothesis, $J(H; U)$ is nonempty. We claim that

(61-b) the ordered metric subspace $(J(H; U), g, \leq)$ is conv-bd-regular.

To this end, let (r_n) be an ascending sequence in $J(H; U)$; hence

$$((r_n)\text{=ascending and) } \mathrm{Fix}(H(r_n)) \cap U \text{ is nonempty, for each } n.$$

By the Denumerable Axiom of Choice (deductible, as above said, in (ZF-AC+DC)), there exists a sequence (x_n) in U, with

$$(\forall n)\text{: } x_n \in \mathrm{Fix}(H(r_n)); \text{ that is: } x_n = H(r_n)x_n.$$

As $(J, g, \leq)$ is conv-bd-regular, there exists $r_\infty := \lim_n(r_n)$ in J with $(r_n \leq r_\infty, \forall n)$. We now claim that $r_\infty \in J(H; U)$; and, from this, all is clear. Remember that, by a previous auxiliary fact, we have that, for each $\lambda \in \Lambda$,

(p-1) right complementary inverse $(\gamma_\lambda(s) = \sup\{t \in R_+; t \leq \omega_\lambda(s + t)\}, s \in R_+)$ belongs to $\mathscr{F}_0(re, in)(R_+)$

(p-2) moreover, γ_λ is strongly regressive $(\gamma_\lambda(s + 0) < s, \forall s > 0)$ as well as complementary coercive ($\gamma_\lambda^* := I - \gamma_\lambda$ is coercive).

For the arbitrary fixed $\lambda \in \Lambda$ and each couple of ranks (n, m) with $n \leq m$, we have (by the contractive conditions)

(61-c) $d_\lambda(x_n, x_m) = d_\lambda(H(r_n)x_n, H(r_m)x_m) \leq$
$d_\lambda(H(r_n)x_n, H(r_m)x_n) + d_\lambda(H(r_m)x_n, H(r_m)x_m) \leq$
$\psi_\lambda(g(r_n, r_m)) + d_\lambda(H(r_m)x_n, H(r_m)x_m)$.

Let us now evaluate the last expression. By definition (and preceding facts)

$$M_1(x_n, x_m; H(r_m); d_\lambda) = d_\lambda(x_n, x_m),$$
$$M_2(x_n, x_m; H(r_m); d_\lambda) = \max\{d_\lambda(x_n, H(r_m)x_n), d(x_m, H(r_m)x_m)\} =$$
$$d_\lambda(x_n, H(r_m)x_n) = d_\lambda(H(r_n)x_n, H(r_m)x_n) \le \psi_\lambda(g(r_n, r_m)),$$
$$M_3(x_n, x_m; H(r_m); d_\lambda) = (1/2)[d_\lambda(x_n, H(r_m)x_m) + d_\lambda(x_m, H(r_m)x_n)] =$$
$$(1/2)[d_\lambda(x_n, x_m) + d_\lambda(H(r_m)x_m, H(r_m)x_n)].$$

Denote for each couple (n, m),

$$a_\lambda(n, m) = d_\lambda(x_n, x_m), b_\lambda(n, m) = \psi_\lambda(g(r_n, r_m)),$$
$$C_\lambda(n, m) = d_\lambda(H(r_m)x_n, H(r_m)x_m)$$

By the above evaluations,

$$M(x_n, x_m; H(r_m); d_\lambda) \le a_\lambda(n, m) + b_\lambda(n, m) + C_\lambda(n, m)$$

wherefrom (as ω_λ is increasing)

$$C_\lambda(n, m) \le \omega_\lambda(M(x_n, x_m; H(r_m); d_\lambda)) \le$$
$$\omega_\lambda(a_\lambda(n, m) + b_\lambda(n, m) + C_\lambda(n, m)), \text{ for } n \le m.$$

Combining with the complementary coerciveness of ω_λ, gives

$$C_\lambda(n, m) \le \gamma_\lambda(a_\lambda(n, m) + b_\lambda(n, m));$$

and this, along with (61-c), implies

$$a_\lambda(n, m) \le b_\lambda(n, m) + \gamma_\lambda(a_\lambda(n, m) + b_\lambda(n, m));$$

wherefrom (by a simple addition)

$$a_\lambda(n, m) + b_\lambda(n, m) \le 2b_\lambda(n, m) + \gamma_\lambda(a_\lambda(n, m) + b_\lambda(n, m)).$$

As (r_n) is g-Cauchy and ψ_λ is zero-continuous, it is clear that $b_\lambda(.,.)$ (hence, $2b_\lambda(.,.)$ as well) appears as Cauchy. This, along with γ_λ being strongly regressive and complementary coercive tells us (by a preceding auxiliary fact) that $a_\lambda(.,.) + b_\lambda(.,.)$ is Cauchy too; wherefrom

$$(\forall \lambda \in \Lambda)\colon d_\lambda(x_n, x_m) = a_\lambda(n, m) \to 0 \text{ as } n, m \to \infty, n \le m;$$

proving that (x_n) is D-Cauchy; so that (by the properties of (X, U))

$$x_n \xrightarrow{D} x_\infty \text{ as } n \to \infty, \text{ for some } x_\infty \in U.$$

Step 2. For each $\lambda \in \Lambda$ and each $n \in N$, we have (by the contractive conditions)

(61-d) $d_\lambda(x_\infty, H(r_\infty)x_\infty) - d_\lambda(x_\infty, x_n) \le$
$d_\lambda(x_n, H(r_\infty)x_\infty) = d_\lambda(H(r_n)x_n, H(r_\infty)x_\infty) \le$
$d_\lambda(H(r_n)x_n, H(r_\infty)x_n) + d_\lambda(H(r_\infty)x_n, H(r_\infty)x_\infty) \le$
$\psi_\lambda(g(r_n, r_\infty)) + d_\lambda(H(r_\infty)x_n, H(r_\infty)x_\infty)$.

Let us now evaluate the last expression. By definition (and preceding facts)

$$M_1(x_n, x_\infty; H(r_\infty); d_\lambda) = d_\lambda(x_n, x_\infty),$$

$$M_2(x_n, x_\infty; H(r_\infty); d_\lambda) = \max\{d_\lambda(x_n, H(r_\infty)x_n), d_\lambda(x_\infty, H(r_\infty)x_\infty)\} =$$

$$\max\{d_\lambda(H(r_n)x_n, H(r_\infty)x_n), d_\lambda(x_\infty, H(r_\infty)x_\infty)\} \le$$

$$\max\{\psi_\lambda(g(r_n, r_\infty)), d_\lambda(x_\infty, H(r_\infty)x_\infty)\},$$

$$M_3(x_n, x_\infty; H(r_\infty); d_\lambda) = (1/2)[d_\lambda(x_n, H(r_\infty)x_\infty) + d_\lambda(x_\infty, H(r_\infty)x_n)] \le$$

$$d_\lambda(x_n, x_\infty) + (1/2[d_\lambda(x_\infty, H(r_\infty)x_\infty) + d_\lambda(H(r_n)x_n, H(r_\infty)x_n)] \le$$

$$d_\lambda(x_n, x_\infty) + (1/2)[d_\lambda(x_\infty, H(r_\infty)x_\infty) + \psi_\lambda(g(r_n, r_\infty))] \le$$

$$d_\lambda(x_n, x_\infty) + \max\{\psi_\lambda(g(r_n, r_\infty)), d_\lambda(x_\infty, H(r_\infty)x_\infty)\}.$$

Denote, for $n \in N$

$$E_\lambda = d_\lambda(x_\infty, H(r_\infty)x_\infty), a_\lambda(n) = \psi_\lambda(g(r_n, r_\infty)),$$
$$b_\lambda(n) = d_\lambda(x_n, x_\infty), c_\lambda(n) = a_\lambda(n) + b_\lambda(n) + E_\lambda.$$

By the above evaluations,

$$M(x_n, x_\infty; H(r_\infty); d_\lambda) \le b_\lambda(n) + \max\{a_\lambda(n), E_\lambda\};$$
$$\text{whence, } M(x_n, x_\infty; H(r_\infty); d_\lambda) \le c_\lambda(n);$$

and this, along with ω_λ=increasing, gives

$$d_\lambda(H(r_\infty)x_n, H(r_\infty)x_\infty) \le \omega_\lambda(M(x_n, x_\infty; H(r_\infty); d_\lambda)) \le \omega_\lambda(c_\lambda(n)).$$

Combining with (61-d), one gets

(61-e) $(\forall n)$: $E_\lambda \le a_\lambda(n) + b_\lambda(n) + \omega_\lambda(c_\lambda(n))$; wherefrom
$c_\lambda(n) \le 2(a_\lambda(n) + b_\lambda(n)) + \omega_\lambda(c_\lambda(n))$.

As $2(a_\lambda(n) + b_\lambda(n)) \to 0$ if $n \to \infty$ and ω_λ is strongly regressive and complementary coercive, one gets (by a previous auxiliary fact)

$$(\forall \lambda \in \Lambda) : c_\lambda(n) \to 0 \text{ as } n \to \infty; \text{ so that, } E_\lambda = 0.$$

This, by definition, yields

$$d_\lambda(x_\infty, H(r_\infty)x_\infty) = 0, \forall \lambda \in \Lambda; \text{ whence, } x_\infty = H(r_\infty)x_\infty;$$

telling us that $r_\infty \in J(H; U)$.

Step 3. Summing up, the Zorn-Bourbaki metric maximality principle (ZB-m-mp) applies to $(J(H; U), g, \leq)$. Hence, for the fixed $q_0 \in J(H; U)$ (assured by hypothesis), there exists an ascending sequence (q_n) in $J(H; U)$, with

$$q^* := \lim_n(q_n) \text{ exists in } J(H; U) \text{ and } q^* \in \max(J(H; U), \leq);$$

$$\text{that is: } q^* \leq r \in J(H; U) \text{ implies } q^* = r.$$

Note that, as $q^* \in J(H; U)$, one has

$$y^* = H(q^*)y^*, \text{ for some } y^* \in U.$$

We show that, necessarily, $q^* \in \max(J, \leq)$. Assume not: $J(q^*, <)$ is not empty. By the imposed hypotheses

$$\text{there exists a sequence } (p_m; m \geq 0) \text{ in } J(q^*, <) \text{ with } \lim_m(p_m) = q^*.$$

On the other hand (again by hypothesis)

$$y^* \in \text{Fix}(H(q^*)) \cap U \text{ implies } y^* \notin U \setminus \text{int}(U); \text{ so that, } y^* \in \text{int}(U);$$

where int(.) means: the interior with respect to the supremum topology $\mathscr{T}$ (see above). From the definition of underlying topology, there exists a finite part Λ^* of Λ and a number $s = s(\Lambda^*)$ in R^0_+, such that

$$V := \cap\{X[y^*, s](d_\lambda); \lambda \in \Lambda^*\} \subseteq U.$$

We now claim that there exists some index $i = i(V^*; s)$, such that

$$y_m := H(p_m)y \in V, \text{ for all } m \geq i, y \in V.$$

In fact, for each $\lambda \in \Lambda^*$, $m \in N$, $y \in V$ we have (in view of ω_λ=increasing)

(61-f) $d_\lambda(y_m, y^*) = d_\lambda(H(p_m)y, H(q^*)y^*) \leq$
$d_\lambda(H(p_m)y, H(q^*)y) + d_\lambda(H(q^*)y, H(q^*)y^*) \leq$
$\psi_\lambda(g(p_m, q^*)) + d_\lambda(H(q^*)y, H(q^*)y^*).$

Let us now evaluate the last expression. By definition (and preceding facts)

$$M_1(y, y^*; H(q^*); d_\lambda) = d_\lambda(y, y^*) \le s$$
$$M_2(y, y^*; H(q^*); d_\lambda) = \max\{d_\lambda(y, H(q^*)y), d_\lambda(y^*, H(q^*)y^*)\} =$$
$$d_\lambda(y, H(q^*)y) \le d_\lambda(y, y^*) + d_\lambda(H(q^*)y, H(q^*)y^*) \le$$
$$s + d_\lambda(H(q^*)y, H(q^*)y^*),$$
$$M_3(y, y^*; H(q^*); d_\lambda) = (1/2)[d_\lambda(y, H(q^*)y^*) + d_\lambda(y^*, H(q^*)y)] =$$
$$(1/2)[d_\lambda(y, y^*) + d_\lambda(H(q^*)y, H(q^*)y^*)] \le (1/2)[s + d_\lambda(H(q^*)y, H(q^*)y^*)].$$

This yields directly

$$M(y, y^*; H(q^*); d_\lambda)) \le s + d_\lambda(H(q^*)y, H(q^*)y^*);$$

wherefrom (as ω_λ is increasing)

$$d_\lambda(H(q^*)y, H(q^*)y^*) \le \omega_\lambda(M(y, y^*; H(q^*); d_\lambda)) \le$$
$$\omega_\lambda(s + d_\lambda(H(q^*)y, H(q^*)y^*)).$$

As ω_λ is strongly regressive and complementary coercive, one gets (see above)

$$d_\lambda(H(q^*)y, H(q^*)y^*) \le \gamma_\lambda(s);$$

and this, along with (61-f), yields

$$d_\lambda(y_m, y^*) \le \psi_\lambda(g(p_m, q^*)) + \gamma_\lambda(s).$$

In view of

$$\gamma_\lambda(s) < s,\ \lim_m \psi_\lambda(g(p_m, q^*)) = 0, \forall \lambda \in \Lambda^*,$$

there must be some index $i = i(V^*; s)$, such that

$$\psi_\lambda(g(p_m, q^*)) + \gamma_\lambda(s) < s, \forall m \ge i, \forall \lambda \in \Lambda^*.$$

But then, from the preceding evaluation,

$$y_m := H(p_m)y \in V, \text{ for all } m \ge i,\ y \in V,$$

and our assertion follows.

By the obtained facts, $H(p_i)$ is a Ω-contractive selfmap of the D-closed subset V of U, with (cf. the hypotheses)

$$(\forall \lambda \in \Lambda):\ \omega_\lambda \text{ is strongly regressive; hence, Boyd-Wong admissible.}$$

Combining with our basic fixed point result it follows that, for the starting point $y^* \in V$, there exists another point $z^* \in V$ with

$$z^* = \lim_n H(p_i)^n y^*, \text{ and } z^* = H(p_i)z^* \text{ (whence, } p_i \in J(H; U)).$$

This, in view of $q^* < p_i$, contradicts the maximality of q^* in $J(H; U)$. Hence, necessarily, $q^* \in \max(J, \leq)$; and the conclusion follows.

A basic particular case of these developments corresponds to

$$(\forall \lambda \in \Lambda) : \ \omega_\lambda(t) = k_\lambda t, t \in R_+, \text{ for some } k_\lambda \in R_+.$$

Precisely, let $(J, g, \leq)$ be an ordered metric space; and take a mapping $(t, x) \mapsto H(t, x) = H(t)x$ from $J \times X$ into X. Given the (nonempty) subset $U \in \exp(X)$, the family $\Psi = (\psi_\lambda; \lambda \in \Lambda)$ over $\mathscr{F}(R_+)$, and the family $k = (k_\lambda; \lambda \in \Lambda)$ over R_+ suppose that the following conditions hold

(homm-1) H is first variable Ψ-contractive:
$d_\lambda(H(t, x), H(s, x)) \leq \psi_\lambda(g(t, s))$, for all $t, s \in J$ and all $x \in U$

(homm-2) H is second variable k-contractive:
$d_\lambda(H(t, x), H(t, y)) \leq k_\lambda M(x, y; H(t); d_\lambda)$, for all $t \in J$ and all $x, y \in U$.

We then say that H is a $(\Psi, k; M; U)$-*homotopic mapping*.

The following (linear) homotopic fixed point theorem is available.

Theorem 15 *Let the ordered metric space $(J, g, \leq)$ with*

(62-i) $(J, g, \leq)$ is conv-bd-regular and nonmaximal accessible

and the subset $U \in \exp(X)$ be such that the map $(t, x) \mapsto H(t, x) = H(t)x$ from $J \times X$ into X be $(\Psi, k; M; U)$-homotopic, where the family $\Psi = (\psi_\lambda; \lambda \in \Lambda)$ over $\mathscr{F}(R_+)$ and the family $k = (k_\lambda; \lambda \in \Lambda)$ over R_+ are such that

(62-ii) $(\forall \lambda \in \Lambda)$: ψ_λ is zero-continuous ($\psi_\lambda(t) \to 0 = \psi(0)$ as $t \to 0$)
(62-iii) $(\forall \lambda \in \Lambda)$: k_λ is double subunitary ($0 \leq k_\lambda < 1/2$).

In addition, suppose that

(62-iv) X is D-complete and U is D-closed
(62-v) $J(H; U \setminus \text{int}(U)) = \emptyset$; i.e.: $\text{Fix}(H(t)) \cap (U \setminus \text{int}(U)) = \emptyset$, for all $t \in J$.

Then, the homotopic map H is transversal in (ZF-AC+DC); i.e.,

$$J(H; U) \neq \emptyset \text{ implies } J(H; U) \cap \max(J, \leq) \neq \emptyset.$$

In particular, when the data (X, U) and $(J, g, \leq)$ are taken as

(p-1) X is D-complete and U is $\mathscr{T}$-closed
(p-2) $J = [0, 1]$, $(g, \leq)$=the usual metric and ordering in R

the obtained result is comparable with a similar one due to Ariza-Ruiz and Jimenez-Melado [5, 6]; which, in turn, extends a related one in Frigon [23]. Concerning this aspect, two remarks are in order.

(I) The first variable contraction upon H considered in the quoted papers is

$$(\forall \lambda \in \Lambda)\colon d_\lambda(H(t,x), H(s,x)) \le |\varphi_\lambda(t) - \varphi_\lambda(s)|,$$

for all $t, s \in J$ and all $x \in U$, where $\varphi_\lambda : J \to R$ is continuous.

But, in view of any continuous function on the compact $J = [0,1]$ being uniformly continuous, we must have, $\forall \lambda \in \Lambda$,

$$|\varphi_\lambda(t) - \varphi_\lambda(s)| \le \psi_\lambda(|t-s|), t, s \in J,$$

where $\psi_\lambda \in \mathscr{F}(R_+)$ is zero-continuous;

and, from this, we arrive at the posed first variable condition above.

(II) On the other hand, the linear homotopic result above holds over the reduced system (ZF-AC+DC) by working with metrical type intervals; while, the quoted result is holding over the complete system (ZF) and having as essential tool a connectedness characterization of (nonempty) real intervals.

As a consequence of this, the homotopic results above include the precise one. Note that, the metrical and order character of the interval space $(J, g, \le)$ may be put in an extended context, so as to include a related statement in O'Regan and Precup [46]. Finally, one may ask whether such results are applicable to practical Ulam-Hyers-Rassias stability for nonlinear equations over Banach spaces, under the lines in Wang and Fečkan [68]; we conjecture that a positive answer to this is available.

7 Kang-Park Principles

Let X be a nonempty set. By a *pseudometric* over X we mean any map $d : X \times X \to R_+$. Fix such a map, endowed with

(ref) d is *reflexive*: $d(x,x) = 0, \forall x \in X$;

we then say that it is a *r-pseudometric* on X. Further, let $(\preceq)$ be a *quasi-order* (i.e.: reflexive and transitive relation) over X; the triple $(X, d, \preceq)$ will be referred to as a *quasi-ordered r-pseudometric space*. Given $M \in \exp(X)$, call $z \in X$,

(max-1) $(d, \preceq)$-*maximal* over M, if $(u, v \in M, z \preceq u \preceq v) \Longrightarrow d(u,v) = 0$
(max-2) $(\preceq)$-*maximal* over M, if $(u \in M, z \preceq u) \Longrightarrow d(z,u) = 0$.

It is our aim in the following to give sufficient conditions for such properties. Then, an application is given to Ekeland variational principles.

(A) Call the sequence (x_n) in X, *d-Cauchy* when $d(x_m, x_n) \to 0$ as $m, n \to \infty$, $m \leq n$; that is

$$\forall \varepsilon > 0, \exists n(\varepsilon), \text{ such that } n(\varepsilon) \leq p \leq q \Longrightarrow d(x_p, x_q) \leq \varepsilon;$$

or, equivalently (by the reflexive property)

$$\forall \varepsilon > 0, \exists n(\varepsilon), \text{ such that } n(\varepsilon) < p < q \Longrightarrow d(x_p, x_q) \leq \varepsilon;$$

and *d-asymptotic*, if $\lim_n d(x_n, x_{n+1}) = 0$; that is

$$\forall \varepsilon > 0, \exists n(\varepsilon), \text{ such that } n(\varepsilon) \leq p \Longrightarrow d(x_p, x_{p+1}) \leq \varepsilon.$$

Then, let us consider the global conditions

(C-reg) $(M, d, \preceq)$ is *Cauchy regular*:
each $(\preceq)$-ascending sequence in M is d-Cauchy
(A-reg) $(M, d, \preceq)$ is *asymptotic regular*:
each $(\preceq)$-ascending sequence in M is d-asymptotic.

As each d-Cauchy sequence is d-asymptotic too, it follows that (C-reg) $\Longrightarrow$ (A-reg). The reverse implication also holds, in the sense

Proposition 12 *We have, in (ZF-AC),*

$$\textit{(A-reg)} \Longrightarrow \textit{(C-reg); whence, (A-reg)} \Longleftrightarrow \textit{(C-reg).}$$

Proof Suppose that (A-reg) holds; but some $(\preceq)$-ascending (x_n) is not entitled with the d-Cauchy property; i.e. (for some $\varepsilon > 0$)

$$C(n) = \{(p, q) \in N \times N; n < p < q, d(x_p, x_q) > \varepsilon\} \neq \emptyset, \forall n.$$

Denote, for simplicity

$$p(n) = \min \mathrm{Dom}(C(n)), q(n) = \max(C(n)(p(n)), n \in N;$$

clearly, no choice techniques are used in this construction. Fix some rank $i(0)$. By this assumption, there exist $i(1) = p(i(0)), i(2) = q(i(0))$ with $i(0) < i(1) < i(2)$, $d(x_{i(1)}, x_{i(2)}) > \varepsilon$. Further, given the rank $i(2)$, there exist $i(3) = p(i(2)), i(4) = q(i(2))$ with $i(2) < i(3) < i(4), d(x_{i(3)}, x_{i(4)}) > \varepsilon$, and so on. By induction, we get a $(\preceq)$-ascending subsequence $(y_n = x_{i(n)})$ of (x_n) with $d(y_{2n+1}, y_{2n+2}) > \varepsilon$, for all n. This contradicts (A-reg); hence the claim.

By definition, either of these conditions will be referred to as: $(M, d, \preceq)$ is *regular*. A basic consequence of this property is the following.

Proposition 13 *We have, in (ZF-AC+DC)*

$(M, d, \preceq)$ *is regular implies*

$(M, d, \preceq)$ *is weakly regular:* $\forall x \in M, \forall \varepsilon > 0,\ \exists y = y(x, \varepsilon) \in M(x, \preceq)$,

such that: $(u, v \in M,\ y \preceq u \preceq v) \Longrightarrow d(u, v) \leq \varepsilon$.

Proof Assume this would be false; that is (for some $x \in M$, $\varepsilon > 0$)

$$\text{for each } y \in M(x, \preceq) \text{ there exist } u, v \in M \text{ with } y \preceq u \preceq v, d(u, v) > \varepsilon.$$

This, by definition, yields (for the same (x, ε)):

$$\forall y \in M(x, \preceq), \exists\, (u, v) \in (\preceq) :\ \ y \preceq u, d(u, v) > \varepsilon;$$

where $(\preceq) := \{(a, b) \in M \times M; a \preceq b\}$. Put $Q := \{(a, b) \in (\preceq); x \preceq a\}$; and fix a couple $(y_0, y_1) \in Q$; for example, $y_0 = y_1 = x$. Define a relation $\mathscr{R} = \mathscr{R}(\varepsilon)$ on Q as

$$(a_1, b_1)\mathscr{R}(a_2, b_2) \text{ if and only if } b_1 \preceq a_2, d(a_2, b_2) > \varepsilon.$$

From the imposed condition, $Q((a, b), \mathscr{R}) \neq \emptyset$, $\forall (a, b) \in Q$. So, by (DC), it follows that, for the starting point $w_0 = (y_0, y_1)$ in Q there exists a sequence $(w_n := (y_{2n}, y_{2n+1}); n \geq 0)$ in Q with

$$w_n \mathscr{R} w_{n+1}, \text{ for all } n; \text{ hence, by definition,}$$

$$y_{2n+1} \preceq y_{2n+2},\ \ d(y_{2n+2}, y_{2n+3}) > \varepsilon,\ \text{ for all } n.$$

As a consequence, $(y_n; n \geq 0)$ is $(\preceq)$-ascending and not d-asymptotic; in contradiction with the regularity of $(M, d, \preceq)$; hence the claim.

Starting from the quasi-order $(\preceq)$, let us introduce a *quasi-order convergence* $\mathscr{B} = \mathscr{B}(\preceq)$ over X as follows: given the sequence (x_n) in X and the point $x \in X$,

$$x_n \xrightarrow{\mathscr{B}} x \text{ if } \exists m(x) \in N, \text{ such that } n \geq m(x) \text{ implies } x_n \preceq x;$$

also referred to as: x is a $\mathscr{B}$*-limit* of (x_n). The class of all these will be denoted as $\mathscr{B} - \lim_n(x_n)$; when it is nonempty, we say that (x_n) is $\mathscr{B}$*-convergent*. Given $Y \in \exp[X]$, let us say that $w \in X$ is a $(\mathscr{B}, \preceq)$-*adherence* point of it when

$$w \in \mathscr{B} - \lim_n(z_n), \text{ for some } (\preceq)\text{-ascending sequence } (z_n) \text{ of } Y;$$

the class of all these will be denoted as $\mathrm{ocl}(Y)$. It is not hard to see that $Y \mapsto \mathrm{ocl}(Y)$ is a semi-closure over X, in the sense

(ocl-1) $\mathrm{ocl}(\emptyset) = \emptyset$, $\mathrm{ocl}(X) = X$,
(ocl-2) $\mathrm{ocl}(U \cup V) = \mathrm{ocl}(U) \cup \mathrm{ocl}(V)$, $\forall U, V \in \exp[X]$
(ocl-3) $Y \subseteq \mathrm{ocl}(Y)$, $\forall Y \in \exp[X]$.

Unfortunately, $Y \mapsto \mathrm{ocl}(Y)$ is not involutive; i.e.,

(ocl-inv) $\mathrm{ocl}(\mathrm{ocl}(Y)) = \mathrm{ocl}(Y)$, for each $Y \in \exp[X]$

is not in general true; so that, $Y \mapsto \mathrm{ocl}(Y)$ is not a closure over X according to Kuratowski [38, Ch I, Sect 4]. In this context, let us say that $Y \in \exp[X]$ is *o-closed*, provided $Y = \mathrm{ocl}(Y)$; note that, in this case,

$$\text{the } \mathscr{B}\text{-limit of each } (\preceq)\text{-ascending sequence in } Y \text{ is included in } Y.$$

Putting these together, the following maximal principle (referred to as: Kang-Park pseudometric maximal principle; in short: (KP-p-mp)) is available.

Theorem 16 *Let the quasi-ordered r-pseudometric space $(X, d, \preceq)$ and the subset $M \in \exp(X)$ be such that*

(71-i) $(M, d, \preceq)$ *is regular: each $(\preceq)$-ascending sequence in M is d-Cauchy*
(71-ii) $(M, d, \preceq)$ *is $\mathscr{B}$-complete:*
each ascending d-Cauchy sequence in M is $\mathscr{B}$-convergent (in X).

Then, the conclusion below holds in (ZF-AC+DC)

$$\forall u \in M, \exists v \in \mathrm{ocl}(M), \text{ with } u \preceq v \text{ and } v \text{ is } (d, \preceq)\text{-maximal over } M.$$

Proof By an auxiliary fact above, $(M, d, \preceq)$ is weakly regular. Hence, given $u \in M$, one may construct a $(\preceq)$-ascending sequence (u_n) in M with

$$u \preceq u_0, \text{ and } [(\forall n), (\forall y, z \in M)]\colon u_n \preceq y \preceq z \implies d(y, z) \leq 2^{-n}.$$

In particular, this tells us that (u_n) is d-Cauchy (in M). Combining with the completeness hypothesis, there exists $v \in \mathrm{ocl}(M)$ such that

$$u_n \xrightarrow{\mathscr{B}} v; \text{ whence, } u_n \preceq v, \forall n.$$

This, via inclusion above, tells us that v is $(d, \preceq)$-maximal over M.

Formally, this result is comparable with the 1990 one in Kang and Park [35]. However, its basic lines were already set up in the papers by Turinici [58, 62]. Note that, both these results extend the 1976 Brezis-Browder ordering principle [12]. Further aspects may be found in Altman [3].

Technically speaking, the completeness hypothesis above is not very appropriate for the local version of Kang-Park pseudometric maximal principle (KP-p-mp); because, for many subsets $M \in \exp(X)$, the points in $\mathrm{ocl}(M)$ may be pretty far from the ones in M. To remove this inconvenient, we need some preliminaries.

Define a d-convergence structure on X under the precise way. Given $Y \in \exp[X]$, let us say that $w \in X$ is a $(d, \preceq)$-*adherence* point of it when

$$w \in d - \lim_n (z_n), \text{ for some } (\preceq)\text{-ascending sequence } (z_n) \text{ of } Y;$$

the class of all these will be denoted as $\mathrm{docl}(Y)$. It is not hard to see that $Y \mapsto \mathrm{docl}(Y)$ is a semi-closure over X, in the sense

(docl-1) $\mathrm{docl}(\emptyset) = \emptyset$, $\mathrm{docl}(X) = X$,
(docl-2) $\mathrm{docl}(U \cup V) = \mathrm{docl}(U) \cup \mathrm{docl}(V)$, $U, V \in \exp[X]$
(docl-3) $Y \subseteq \mathrm{docl}(Y)$, $\forall Y \in \exp[X]$.

Unfortunately, $Y \mapsto \mathrm{docl}(Y)$ is not involutive; i.e.,

(docl-inv) $\mathrm{docl}(\mathrm{docl}(Y)) = \mathrm{docl}(Y)$, for each $Y \in \exp[X]$

is not in general true; so that, $Y \mapsto \mathrm{docl}(Y)$ is not a closure over X according to Kuratowski [38, Ch I, Sect 4]. Further, let us say that $Y \in \exp[X]$ is *do-closed*, provided $Y = \mathrm{docl}(Y)$; note that, in this case,

$$\text{the } d\text{-limit of each } (\preceq)\text{-ascending sequence in } Y \text{ is included in } Y.$$

The following version of the statement above (referred to as: Kang-Park strong pseudometric maximal principle; in short: (KP-sp-mp)) is now available.

Theorem 17 *Let the quasi-ordered r-pseudometric space $(X, d, \preceq)$ and the subset $M \in \exp(X)$ be such that*

(72-i) $(M, d, \preceq)$ *is regular: each $(\preceq)$-ascending sequence in M is d-Cauchy*
(72-ii) $(M, d, \preceq)$ *is $(d, \mathscr{B})$-complete: each ascending d-Cauchy sequence (x_n) in M is d-convergent and $\mathscr{B}$-convergent (in X) with (in addition) $(d - \lim_n (x_n)) \cap (\mathscr{B} - \lim_n (x_n)) \neq \emptyset$.*

Then, the conclusion below holds in (ZF-AC+DC)

$$\forall u \in M, \exists v \in \mathrm{docl}(M), \text{ with } u \preceq v \text{ and } v \text{ is } (d, \preceq)\text{-maximal over } M.$$

A basic particular case of these corresponds to d=metric on X.

(I) As a first consequence of this choice, we have

$$(\forall M \in \exp(X), \forall z \in M) : \ z \text{ is } (d, \preceq)\text{-maximal over } M \text{ iff}$$

$$z \text{ is } (\preceq)\text{-maximal over } M \ [z \preceq w \in M \text{ implies } z = w].$$

In fact, the left to right inclusion follows via $z \preceq z \preceq w$ and $z, w \in M$ imply $z = w$. On the other hand, the right to left inclusion is immediate, by the very definition of $(\preceq)$-maximal element, combined with

$$z \preceq u \preceq v \Longrightarrow (z \preceq u \text{ and } z \preceq v).$$

(II) A second consequence of the same choice is contained in

Proposition 14 *Suppose that* $(X, d, \preceq)$ *is regular. Then, necessarily,*

$$(\preceq) \text{ is antisymmetric; hence, an ordering.}$$

Proof ***(cf. Hamel [30, Ch 4, Sect 4.1])*** Let $u, v \in X$ be such that $u \preceq v$ and $v \preceq u$. The sequence $(y_{2n} = u,\ y_{2n+1} = v;\ n \geq 0)$ is ascending; hence, d-asymptotic by hypothesis. This yields $d(u, v) = 0$; whence (as d=metric), $u = v$.

(III) For the last consequence of our choice let us introduce the usual d-closure operator $Y \mapsto \mathrm{dcl}(Y)$, as:

$$z \in \mathrm{dcl}(Y) \text{ iff } z = \lim_n(y_n), \text{ for some sequence } (y_n) \text{ in } Y.$$

Note that, an equivalent way of describing this operator is

$$z \in \mathrm{dcl}(Y) \text{ iff } \forall \varepsilon > 0, \exists y = y(\varepsilon) \in Y\colon d(z, y) < \varepsilon.$$

Starting from this, it is not hard to see that $Y \mapsto \mathrm{dcl}(Y)$ has the properties

(dcl-1) $\mathrm{dcl}(\emptyset) = \emptyset$, $\mathrm{dcl}(X) = X$,
(dcl-2) $\mathrm{dcl}(U \cup V) = \mathrm{dcl}(U) \cup \mathrm{dcl}(V)$, $U, V \in \exp[X]$
(dcl-3) $Y \subseteq \mathrm{dcl}(Y)$, $\forall Y \in \exp[X]$
(dcl-4) $\mathrm{dcl}(\mathrm{dcl}(Y)) = \mathrm{dcl}(Y)$, for each $Y \in \exp[X]$;

so, $Y \mapsto \mathrm{dcl}(Y)$ is a closure over X according to Kuratowski [38, Ch I, Sect 4]. In this case, $Y \in \exp[X]$ is called *d-closed*, provided $Y = \mathrm{dcl}(Y)$.

Putting these together, gives the following maximal statement (referred to as: Zorn-Bourbaki metric maximal principle; in short: (ZB-m-mp)).

Theorem 18 *Let the ordered metric space* $(X, d, \preceq)$ *and the subsets* $M, Y \in \exp(X)$ *be such that*

(73-i) $M \subseteq Y$ *and* Y *is* d*-closed* $(Y = \mathrm{dcl}(Y))$
(73-ii) $(Y, d, \preceq)$ *is (conv,*$\mathscr{B}$*)-regular: each* $(\preceq)$*-ascending sequence* (x_n) *in* Y *is* d*-convergent in* X *and* $x_n \preceq \lim_n(x_n)$, $\forall n$.

Then, the conclusion below holds in (ZF-AC+DC)

for each $u \in M$ *there exists some other point* $v \in Y$, *with* $u \preceq v$ *and*

v *is* $(\preceq)$*-maximal over* M*:* $v \preceq y \in M$ *imply* $v = y$.

Proof By the posed hypothesis,

(73-a) $(Y, d, \preceq)$ is regular: each $(\preceq)$-ascending sequence in Y is d-Cauchy.

Moreover, under the same condition, it is clear that

(73-b) $(Y, d, \preceq)$ is $(d, \mathscr{B})$-complete: each $(\preceq)$-ascending d-Cauchy sequence (x_n) in Y is d-convergent in X and $\lim_n(x_n) \in \mathscr{B} - \lim_n(x_n)$.

Putting these together, we have that the Kang-Park strong pseudometric maximal principle (KP-sp-mp) is applicable here. Hence, for the starting $u \in M \subseteq Y$ there exists some other point $v \in \mathrm{docl}(Y)$ with

$$u \preceq v \text{ and } v \text{ is } (d, \preceq)\text{-maximal over } Y.$$

But, evidently, $\mathrm{docl}(Y) \subseteq Y$ (as Y=d-closed); and this yields

$$v \in Y; \text{ hence (see above) } v \text{ is } (\preceq)\text{-maximal over } Y.$$

As a direct consequence of definition, we get

$$v \preceq y \in M \Longrightarrow v \preceq y \in Y; \text{ whence } v = y;$$

and the proof is complete.

This result may be viewed as a metrical variant of the Zorn-Bourbaki maximal principle [9]. An early version of it was formulated in Turinici [57]; note that it includes the one due to Dancs et al. [18]. Further aspects may be found in the related papers by Turinici [56, 59].

(B) A basic application of these facts is to (local) variational principles. Let X be a nonempty set; and $d(.,.)$ be a metric on X; the couple (X, d) will be referred to as a *metric space*. Further, let the function $\varphi : X \to R \cup \{\infty\}$ be such that the following admissible conditions hold:

(adm-1) φ is inf-proper ($\mathrm{Dom}(\varphi) \neq \emptyset$ and $\inf[\varphi(X)] > -\infty$)
(adm-2) φ is d-lsc: $[\varphi \leq t] := \{x \in X; \varphi(x) \leq t\}$ is d-closed, $\forall t \in R$.

The following variational statement (referred to as: Ekeland variational principle on metric spaces; in short: (EVP-m)) is available.

Theorem 19 *Let the metric space (X, d) and the function $\varphi : X \to R \cup \{\infty\}$ be taken according to*

$$(X, d) \textit{ is complete and } \varphi \textit{ is admissible}.$$

Further, take some $u \in \mathrm{Dom}(\varphi)$. Then, in the reduced system (ZF-AC+DC), there exists $v \in \mathrm{Dom}(\varphi)$, with

(74-a) $d(u, v) \leq \varphi(u) - \varphi(v)$ *(hence $\varphi(u) \geq \varphi(v)$)*
(74-b) $d(v, x) > \varphi(v) - \varphi(x)$, *for each $x \in X \setminus \{v\}$.*

Proof Denote $X[u] = \{x \in X; \varphi(u) \geq \varphi(x)\}$; note that, by the posed conditions, $X[u]$ is a nonempty closed subset of X, with $X[u] \subseteq \mathrm{Dom}(\varphi)$. Let $(\preceq)$ be the relation on X defined as

$$x \preceq y \text{ iff } d(x, y) + \varphi(y) \leq \varphi(x).$$

Clearly, $(\preceq)$ appears as a quasi-order on X; moreover, $(\preceq)$ is antisymmetric (hence an order) on $\mathrm{Dom}(\varphi)$; hence, on $X[u]$ as well. Let also $\mathscr{B}$ stand for the order convergence structure (over X). We claim that the Zorn-Bourbaki metric maximal principle (ZB-m-mp) is applicable over the ordered metric space $(X[u], d, \preceq)$, under the choice $M = Y = X[u]$; and this will complete the argument. In fact, let (x_n) be a $(\preceq)$-ascending sequence in $X[u]$:

(74-c) $d(x_n, x_m) \leq \varphi(x_n) - \varphi(x_m)$, if $n \leq m$.

The sequence $(\varphi(x_n))$ is descending bounded from below; hence a Cauchy one. This, along with the preceding relation, shows that (x_n) is a $(\preceq)$-ascending d-Cauchy sequence in $X[u]$. By the completeness hypothesis (and closeness of $X[u]$) it follows that $x := \lim_n(x_n)$ exists in $X[u]$; moreover,

$$\varphi(x_n) \geq \varphi(x), \text{ for all } n \text{ [as } \varphi \text{ is } d\text{-lsc]}.$$

Finally, take some rank n. From (74-c) (and the triangular property)

$$d(x_n, x) \leq d(x_n, x_m) + d(x_m, x) \leq \varphi(x_n) - \varphi(x) + d(x_m, x), \ \forall m \geq n;$$

wherefrom (passing to limit as $m \to \infty$)

$$d(x_n, x) \leq \varphi(x_n) - \varphi(x) \text{ (i.e.: } x_n \preceq x), \forall n;$$

and our claim follows. Summing up, the Zorn-Bourbaki metric maximal principle (ZB-m-mp) is indeed applicable to our data. According to the quoted result, we have that, for the starting $u \in X[u]$, there exists another point $v \in X[u]$, with

(74-d) $u \preceq v$ and v is $(\preceq)$-maximal on $X[u]$: $v \preceq w \in X[u]$ implies $v = w$.

The first half of this gives our first conclusion. Concerning our second conclusion, suppose by absurd that

$$d(v, y) \leq \varphi(v) - \varphi(y) \text{ (i.e.: } v \preceq y), \text{ for some } y \in X, y \neq v.$$

As $u \preceq v$, one gets $u \preceq y$; hence, $y \in X[u]$. This, under the second half of (74-d), yields $v \preceq y \in X[u]$; hence, $v = y$; a contradiction; so that, we are done.

Remark 4 A local version of this result is immediately obtainable from the Zorn-Bourbaki metric maximal principle (ZB-m-mp) by simply taking the couple (M, Y) appearing there as (M_u, Y_u), where

$$M_u := M \cap X[u], Y_u := \mathrm{dcl}(M) \cap X[u];$$

here $M \in \exp(X)$ and $u \in M \cap \mathrm{Dom}(\varphi)$ are fixed elements.

This principle, due to Ekeland [21], found some basic applications to control and optimization, generalized differential calculus, critical point theory and global

analysis; we refer to the quoted paper for a survey of these. So, it cannot be surprising that, soon after its formulation, many extensions of (EVP-m) were proposed. For example, the *dimensional* way of extension refers to the ambient space (R) of $\varphi(X)$ being substituted by a (topological or not) vector space. An account of the results in this area is to be found in Goepfert et al. [29, Ch 3]. The metrical extension of the same consists in conditions imposed upon our metric being relaxed. Some of these extensions were already stated; for the remaining ones, we refer to Hyers, Isac and Rassias [31, Ch 5]; see also Turinici [63].

By the developments above, we therefore have the implications:

$$\text{(DC)} \Longrightarrow \text{(KP-p-mp)} \Longrightarrow \text{(KP-sp-mp)} \Longrightarrow \text{(ZB-m-mp)} \Longrightarrow \text{(EVP-m)}.$$

So, we may ask whether these may be reversed. Clearly, the natural setting for solving this problem is the strongly reduced system (ZF-AC).

Let X be a nonempty set; and $(\leq)$ be a (partial) order on it. We say that $(\leq)$ has the *inf-lattice* property, provided: $x \wedge y := \inf(x, y)$ exists, for all $x, y \in X$. Remember that $z \in X$ is a $(\leq)$-*maximal* element if $X(z, \leq) = \{z\}$; the class of all these points will be denoted as $\max(X, \leq)$. Call $(\leq)$, a *Zorn order* when

$$\max(X, \leq) \text{ is nonempty and } cofinal \text{ in } X$$

$$\text{(for each } u \in X \text{ there exists a } (\leq)\text{-maximal } v \in X \text{ with } u \leq v).$$

Further aspects are to be described in a metric setting. Let $d(.,.)$ be a metric over X; and $\varphi : X \to R_+$ be some function. Then, the natural choice for $(\leq)$ above is

$$x \leq_{(d,\varphi)} y \text{ iff } d(x, y) \leq \varphi(x) - \varphi(y);$$

referred to as the *Brøndsted order* [13] attached to (d, φ). Denote

$$X(x, \rho) = \{u \in X; d(x, u) < \rho\}, x \in X, \rho > 0$$

[the *open* sphere with center x and radius ρ]. Call (X, d), *discrete* when

$$\text{for each } x \in X \text{ there exists } \rho = \rho(x) > 0 \text{ such that } X(x, \rho) = \{x\}.$$

Note that, under such an assumption, any function $\psi : X \to R$ is continuous over X. However, this is not extendable to the *d-Lipschitz property*

$$|\psi(x) - \psi(y)| \leq Ld(x, y), x, y \in X, \text{for some } L > 0;$$

hence, all the more, to the *d-nonexpansive property* ($L = 1$).

Now, the statement below is a particular case of (EVP-m):

Theorem 20 *Let the metric space (X, d) and the function $\varphi : X \to R_+$ satisfy*

(75-i) (X, d) *is discrete bounded and complete*
(75-ii) $(\leq_{(d,\varphi)})$ *has the inf-lattice property*
(75-iii) φ *is d-nonexpansive and $\varphi(X)$ is countable.*

Then, $(\leq_{(d,\varphi)})$ *is a Zorn order.*

We shall refer to it as: the discrete Lipschitz countable version of (EVP-m) (in short: (EVP-m-dLc)). Clearly, (EVP-m) $\Longrightarrow$ (EVP-m-dLc). The remarkable fact is that this last principle yields (DC); and completes the circle between all these.

Proposition 15 *We have the inclusion (EVP-m-dLc) $\Longrightarrow$ (DC) [in the strongly reduced system (ZF-AC)]. So (by the above),*

(74-1) *the maximal/variational principles (KP-p-mp), (KP-sp-mp), (ZB-m-mp) and (EVP-m) are all equivalent with (DC); hence, mutually equivalent*
(74-2) *each intermediary maximal/variational statement (VP) with (DC) $\Longrightarrow$ (VP) $\Longrightarrow$ (EVP-m) is equivalent with both (DC) and (EVP-m).*

For a complete proof, see Turinici [66]. In particular, when the inf-lattice, nonexpansive, and countable properties are ignored in (EVP-m-dLc), the last result above reduces to the one in Brunner [14]. Note that, in the same particular setting, a different proof of the underlying inclusion was provided in Dodu and Morillon [19]; see also Schechter [52, Ch 19, Sect 19.51]. For a number of finite dimensional stability criteria involving the maximal points in question, we refer to the monograph by Hyers et al. [32, Ch 11].

References

1. R.P. Agarwal, D. O'Regan, N. Shahzad, Fixed point theory for generalized contractive maps of Meir-Keeler type. Math. Nachr. **276**, 3–22 (2004)
2. R.P. Agarwal, M.A. El-Gebeily, D. O'Regan, Generalized contractions in partially ordered metric spaces. Appl. Anal. **87**, 109–116 (2008)
3. M. Altman, A generalization of the Brezis-Browder principle on ordered sets. Nonlin. Anal. **6**, 157–165 (1982)
4. V.G. Angelov, *Fixed Points in Uniform Spaces and Applications* (Cluj University Press, Cluj-Napoca, 2009)
5. D. Ariza-Ruiz, A. Jimenez-Melado, A continuation method for weakly contractive mappings under the interior condition. Fixed Point Th. Appl. **2009**, Article ID 809315 (2009)
6. D. Ariza-Ruiz, A. Jimenez-Melado, A continuation method for weakly Kannan maps. Fixed Point Th. Appl. **2010**, Article ID 321594 (2010)
7. S. Banach, Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. Fund. Math. **3**, 133–181 (1922)
8. P. Bernays, A system of axiomatic set theory: Part III. Infinity and enumerability analysis. J. Symbolic Logic **7**, 65–89 (1942)
9. N. Bourbaki, Sur le théorème de Zorn. Archiv Math. **2**, 434–437 (1949/1950)
10. N. Bourbaki, *General Topology (Chs 1–4)* (Springer, Berlin, 1995)
11. D.W. Boyd, J.S.W. Wong, On nonlinear contractions. Proc. Amer. Math. Soc. **20**, 458–464 (1969)

12. H. Brezis, F.E. Browder, A general principle on ordered sets in nonlinear functional analysis. Adv. Math. **21**, 355–364 (1976)
13. A. Brøndsted, Fixed points and partial orders. Proc. Amer. Math. Soc. **60**, 365–366 (1976)
14. N. Brunner, Topologische Maximalprinzipien. Zeitschr. Math. Logik Grundl. Math. **33**, 135–139 (1987)
15. M. Cherichi, B. Samet, Fixed point theorems on ordered gauge spaces with applications to integral equations. Fixed Point Th. Appl. **2012**, 13 (2012)
16. P.J. Cohen, *Set Theory and the Continuum Hypothesis* (Benjamin, New York, 1966)
17. O. Costinescu, *Elements of General Topology* (Romanian), Ed. (Tehnică, Bucureşti, 1970)
18. S. Dancs, M. Hegedus, P. Medvegyev, A general ordering and fixed-point principle in complete metric space. Acta Sci. Math. (Szeged) **46**, 381–388 (1983)
19. J. Dodu, M. Morillon, The Hahn-Banach property and the axiom of choice. Math. Logic Q. **45**, 299–314 (1999)
20. J. Dugundji, *Topology* (Allyn and Bacon, Boston, 1966)
21. I. Ekeland, Nonconvex minimization problems. Bull. Amer. Math. Soc. (New Series) **1**, 443–474 (1979)
22. R. Engelking, *General Topology* (Heldermann Verlag, Berlin, 1989)
23. M. Frigon, Fixed point results in Fréchet and gauge spaces and applications to differential and integral equations. Lect. Notes. Nonlin. Anal. **16**, 9–91 (2017)
24. M.A. Geraghty, On contractive mappings. Proc. Amer. Math. Soc. **40**, 604–608 (1973)
25. N. Gheorghiu, Contraction theorem in uniform spaces (Romanian). Stud. Cerc. Mat. **19**, 119–122 (1967)
26. N. Gheorghiu, Fixed point theorems in uniform spaces. An. Şt. Univ. Al. I. Cuza Iaşi (Sect. I-a, Mat.) **28**, 17–18 (1982)
27. N. Gheorghiu, E. Rotaru, A fixed point theorem in uniform spaces. An. Şt. Univ. Al. I. Cuza Iasi (Sect. I-a, Mat) **18**, 311–314 (1972)
28. N. Gheorghiu, M. Turinici, Equations intégrales dans les espaces localement convexes. Rev. Roum. Math. Pures Appl. **23**, 33–40 (1978)
29. A. Goepfert, H. Riahi, C. Tammer, C. Zălinescu, *Variational Methods in Partially Ordered Spaces*. Canad. Math. Soc. Books Math., vol. 17 (Springer, New York, 2003)
30. A. Hamel, *Variational Principles on Metric and Uniform Spaces*. Habilitation Thesis. Martin-Luther University, Halle-Wittenberg, 2005
31. D.H. Hyers, G. Isac, Th. M. Rassias, *Topics in Nonlinear Analysis and Applications* (World Scientific Publishing, Singapore, 1997)
32. D.H. Hyers, G. Isac, Th.M. Rassias, *Stability of Functional Equations in Several Variables* (Birkhäuser, Boston, 1998)
33. J. Jachymski, Common fixed point theorems for some families of mappings. Indian J. Pure Appl. Math. **25**, 925–937 (1994)
34. J. Jachymski, The contraction principle for mappings on a metric space with a graph. Proc. Amer. Math. Soc. **136**, 1359–1373 (2008)
35. B.G. Kang, S. Park, On generalized ordering principles in nonlinear analysis. Nonlin. Anal. **14**, 159–165 (1990)
36. S. Kasahara, On some generalizations of the Banach contraction theorem. Publ. Res. Inst. Math. Sci. Kyoto Univ. **12**, 427–437 (1976)
37. M.S. Khan, M. Swaleh, S. Sessa, Fixed point theorems by altering distances between the points. Bull. Austral. Math. Soc. **30**, 1–9 (1984)
38. K. Kuratowski, *Topology* (Academic Press, New York, 1966)
39. S. Leader, Fixed points for general contractions in metric spaces. Math. Japonica **24**, 17–24 (1979)
40. J. Matkowski, *Integrable Solutions of Functional Equations*. Dissertationes Math., vol. 127 (Polish Scientific Publishers, Warsaw, 1975)
41. J. Matkowski, Fixed point theorems for mappings with a contractive iterate at a point. Proc. Amer. Math. Soc. **62**, 344–348 (1977)

42. A. Meir, E. Keeler, *A theorem on contraction mappings*. J. Math. Anal. Appl. **28**, 326–329 (1969)
43. G.H. Moore, *Zermelo's Axiom of Choice: its Origin, Development and Influence* (Springer, New York, 1982)
44. Y. Moskhovakis, *Notes on Set Theory* (Springer, New York, 2006)
45. J.J. Nieto, R. Rodriguez-Lopez, Contractive mapping theorems in partially ordered sets and applications to ordinary differential equations. Order **22**, 223–239 (2005)
46. D. O'Regan, R. Precup, Continuation theory for contractions on spaces with two vector-valued metrics. Applicable Anal. **82**, 131–144 (2003)
47. A.C.M. Ran, M.C. Reurings, A fixed point theorem in partially ordered sets and some applications to matrix equations. Proc. Amer. Math. Soc. **132**, 1435–1443 (2004)
48. S. Reich, Fixed points of contractive functions. Boll. Un. Mat. Ital. **5**, 26–42 (1972)
49. B.E. Rhoades, A comparison of various definitions of contractive mappings. Trans. Amer. Math. Soc. **226**, 257–290 (1977)
50. I.A. Rus, *Generalized Contractions and Applications* (Cluj University Press, Cluj-Napoca, 2001)
51. B. Samet, M. Turinici, Fixed point theorems on a metric space endowed with an arbitrary binary relation and applications. Commun. Math. Anal. **13**, 82–97 (2012)
52. E. Schechter, *Handbook of Analysis and its Foundation* (Academic Press, New York, 1997)
53. A. Tarski, Axiomatic and algebraic aspects of two theorems on sums of cardinals. Fund. Math. **35**, 79–104 (1948)
54. M. Turinici, Fixed points of implicit contraction mappings. An. Şt. Univ. Al. I. Cuza Iaşi (Sect. I-a, Mat.) **22**, 177–180 (1976)
55. M. Turinici, Nonlinear contractions and applications to Volterra functional equations. An. Şt. Univ. Al. I. Cuza Iaşi (Sect. I-a, Mat) **23**, 43–50 (1977)
56. M. Turinici, Maximal elements in a class of order complete metric spaces. Math. Japonica **25**, 511–517 (1980)
57. M. Turinici, Maximality principles and mean value theorems. An. Acad. Brasil. Cienc. **53**, 653–655 (1981)
58. M. Turinici, A generalization of Brezis-Browder's ordering principle. An. Şt. Univ. Al. I. Cuza Iaşi (Sect. I-a, Mat) **28**, 11–16 (1982)
59. M. Turinici, A generalization of Altman's ordering principle. Proc. Amer. Math. Soc. **90**, 128–132 (1984)
60. M. Turinici, Fixed points for monotone iteratively local contractions. Dem. Math. **19**, 171–180 (1986)
61. M. Turinici, Abstract comparison principles and multivariable Gronwall-Bellman inequalities. J. Math. Anal. Appl. **117**, 100–127 (1986)
62. M. Turinici, Pseudometric extensions of the Brezis-Browder ordering principle. Math. Nachrichten **130**, 91–103 (1987)
63. M. Turinici, A monotone version of the variational Ekeland's principle. An. Şt. Univ. Al. I. Cuza Iaşi (Sect. I-a, Mat.) **36**, 329–352 (1990)
64. M. Turinici, Function pseudometric VP and applications. Bul. Inst. Polit. Iaşi (S. Mat., Mec. Teor., Fiz.) **53**(57), 393–411 (2007)
65. M. Turinici, Ran-Reurings theorems in ordered metric spaces. J. Indian Math. Soc. **78**, 207–214 (2011)
66. M. Turinici, Sequential maximality principles, in *Mathematics Without Boundaries*, ed. by T.M. Rassias, P.M. Pardalos (Springer, New York, 2014), pp. 515–548
67. M. Turinici, *Fixed Point Results in Ordered Gauge Spaces*. Selected Topics in Metrical Fixed Point Theory, Paper 3–4 (Pim Editorial House, Iaşi, 2017)
68. J.R. Wang, M. Fečkan, Practical Ulam-Hyers-Rassias stability for nonlinear equations. Math. Bohemica **142**, 47–56 (2017)
69. E.S. Wolk, On the principle of dependent choices and some forms of Zorn's lemma. Canad. Math. Bull. **26**, 365–367 (1983)

Analytic Methods in Rhoades Contractions Theory

Mihai Turinici

Abstract A lot of implicit analytic methods is proposed for the study of Rhoades contractions over a class of relational metric spaces, via regulated functions. Technical connections with some particular statements in the area due to Vujaković et al. (Mathematics 767:8 (2020)) are also discussed.

1 Introduction

Let X be a nonempty set. Call the subset Y of X, *almost singleton* (in short: *asingleton*), provided $[y_1, y_2 \in Y$ implies $y_1 = y_2]$; and *singleton* if, in addition, Y is nonempty; note that in this case $Y = \{y\}$, for some $y \in X$. Further, take a metric $d : X \times X \to R_+ := [0, \infty[$ over X; the couple (X, d) will be referred to as a *metric space*. Finally, let $T \in \mathscr{F}(X)$ be a selfmap of X. [Here, for each couple A, B of nonempty sets, $\mathscr{F}(A, B)$ denotes the class of all functions from A to B; when $A = B$, we write $\mathscr{F}(A)$ in place of $\mathscr{F}(A, A)$]. Denote $\mathrm{Fix}(T) = \{x \in X; x = Tx\}$; each point of this set is referred to as *fixed* under T. The determination of such points is carried out in the context below, comparable with the one in Rus [32, Ch 2, Sect 2.2]:

pic-1) We say that T is a *Picard operator* (modulo d) if, for each $x \in X$, the iterative sequence $(T^n x; n \geq 0)$ is d-Cauchy

pic-2) We say that T is a *strong Picard operator* (modulo d) if, for each $x \in X$, $(T^n x; n \geq 0)$ is d-convergent with $\lim_n (T^n x) \in \mathrm{Fix}(T)$.

pic-3) We say that T is *fix-asingleton* if $\mathrm{Fix}(T)$ is asingleton; and *fix-singleton*, provided $\mathrm{Fix}(T)$ is singleton.

AMS Subject Classification: 47H10 (Primary), 54H25 (Secondary)

M. Turinici (✉)
A. Myller Mathematical Seminar, A. I. Cuza University, Iaşi, Romania
e-mail: mturi@uaic.ro

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_25

Concerning the existence and uniqueness of such points, a basic result (referred to as: Banach fixed point theorem; in short: (B-fpt)) may be stated as follows. Call the selfmap T, $(d;\alpha)$-*contractive* (where $\alpha \geq 0$), if

$$\text{(con)}\ \ d(Tx,Ty) \leq \alpha d(x,y),\ \text{ for all } x,y \in X.$$

Theorem 1 *Suppose that T is $(d;\alpha)$-contractive, for some $\alpha \in [0,1[$. In addition, let X be d-complete. Then, T is a strong Picard operator and fix-asingleton (or, equivalently: fix-singleton).*

This result, established in 1922 by Banach [2], found some important applications to the operator equations theory. Consequently, a multitude of extensions for (B-fpt) were proposed. From the perspective of this exposition, the set implicit ones are of interest. Denote, for $x, y \in X$

$$Q_1(x,y) = d(x,Tx),\ Q_2(x,y) = d(x,y),$$
$$Q_3(x,y) = d(x,Ty),\ Q_4(x,y) = d(Tx,y),$$
$$Q_5(x,y) = d(Tx,Ty),\ Q_6(x,y) = d(y,Ty),$$
$$\mathscr{Q}(x,y) = (Q_1(x,y), Q_2(x,y), Q_3(x,y), Q_4(x,y), Q_5(x,y), Q_6(x,y)).$$

Then, the underlying contractions may be written as

$$\text{(i-s-con)}\ \ \mathscr{Q}(x,y) \in \Upsilon,\ \text{ for all } x,y \in X \text{ with } x\mathscr{R}y;$$

where $\Upsilon \subseteq R_+^6$ is a (nonempty) subset and $\mathscr{R}$ is a relation over X. In particular, when Υ is the zero-section of a certain function $F : R_+^6 \to R$; i.e.,

$$\Upsilon = \{(t_1,\ldots,t_6) \in R_+^6;\ F(t_1,\ldots,t_6) \leq 0\},$$

the implicit contractive condition above has the functional form:

$$\text{(i-f-con)}\ \ F(\mathscr{Q}(x,y)) \leq 0,\ \text{ for all } x,y \in X \text{ with } x\mathscr{R}y.$$

The natural case discussed in the immense majority of papers is $\mathscr{R} = X \times X$ (the *trivial* relation over X). Concerning the "genuine" implicit case, some recent contributions in the area may be found in Akkouchi [1], Berinde and Vetro [3], Nashine et al. [22], or Popa and Mocanu [26]. We stress that in almost all papers based on implicit techniques—including the ones we just quoted—it is claimed that the starting point in the area is represented by the contribution due to Popa [25]. Unfortunately, this claim is not true: fixed point results based on implicit techniques were obtained more than two decades ago in two papers by Turinici [36, 37]. But, we must note that some partial aspects of the set-implicit theory have been discussed in the (classical by now) 1969 Meir-Keeler fixed point principle [18].

On the other hand, for the explicit case, some basic contributions were obtained in Boyd and Wong [5], Leader [16], Reich [29], or Matkowski [17]; see also the survey paper by Rhoades [30].

Having these precise, it is our aim in the following to give a lot of fixed point results—involving six-dimensional contractive conditions—for selfmaps acting upon relational metric spaces. Precisely, as a by-product of our developments, two classes of fixed point statements were build up for contractive conditions of the type discussed in the papers by Rhoades [31] and Dutta and Choudhury [12]; namely:

$$\text{(RDC)}\quad \psi(u \circ \mathcal{Q}(x, y)) \leq \psi(v \circ \mathcal{Q}(x, y)) - \varphi(w \circ \mathcal{Q}(x, y)), x, y \in X, x\mathcal{R}y,$$

where (ψ, φ) is a functional couple over $\mathcal{F}(R_+^0, R)$, (u, v, w) is a triple of functions over $\mathcal{F}(R_+^6, R_+)$, and $\mathcal{R}$ is a relation over X. Note that, the fixed point results of the first class involve contractions like before based on ψ being a regulated function—hence, not in general increasing. Then, as another by-product of these, some fixed point results are given for a class of contractive maps introduced by Cosentino and Vetro [9] and refined by Vujaković et al. [44]. Finally, as a special application of the used asymptotic techniques, an almost regulated version is given for the standard result in the area due to Wardowski [46]. Some anticipative type variants of these results, including the 2015 contribution in the area due to Dung and Hang [11] will be discussed elsewhere.

2 Dependent Choice Principle

Throughout this exposition, the axiomatic system in use is Zermelo-Fraenkel's (abbreviated: ZF), as described by Cohen [8, Ch 2]. The notations and basic facts to be considered in this system are more or less standard. Some important ones are described below.

(A) Let X be a nonempty set. By a *relation* over X, we mean any nonempty part $\mathcal{R} \subseteq X \times X$; then, $(X, \mathcal{R})$ will be referred to as a *relational structure*. For simplicity, we sometimes write $(x, y) \in \mathcal{R}$ as $x\mathcal{R}y$. Note that $\mathcal{R}$ may be regarded as a mapping between X and $\exp[X]$ (=the class of all subsets in X). In fact, denote for $x \in X$:

$$X(x, \mathcal{R}) = \{y \in X; x\mathcal{R}y\} \text{ (the } section \text{ of } \mathcal{R} \text{ through } x);$$

then, the desired mapping representation is $[\mathcal{R}(x) = X(x, \mathcal{R}), x \in X]$.

A basic example of relational structure is to be constructed as below. Let $N = \{0, 1, \ldots\}$ be the set of *natural* numbers, endowed with the usual addition and (partial) order; note that

$(N, \leq)$ is well ordered: any (nonempty) subset of N has a first element.

Further, denote for $p, q \in N$, $p \leq q$,

$$N[p,q] = \{n \in N;\ p \leq n \leq q\},\ N]p,q[= \{n \in N;\ p < n < q\},$$

$$N[p,q[= \{n \in N;\ p \leq n < q\},\ N]p,q] = \{n \in N;\ p < n \leq q\};$$

as well as, for $r \in N$,

$$N[r,\infty[= \{n \in N;\ r \leq n\},\ \ N]r,\infty[= \{n \in N;\ r < n\}.$$

By definition, $N[0,r[= N(r,>)$ is referred to as the *initial interval* (in N) induced by r. Any set P with $P \sim N$ (in the sense: there exists a bijection from P to N) will be referred to as *effectively denumerable*. In addition, given some natural number $n \geq 1$, any set Q with $Q \sim N(n,>)$ will be said to be *n-finite*; when n is generic here, we say that Q is *finite*. Finally, the (nonempty) set Y is called (at most) *denumerable* iff it is either effectively denumerable or finite.

Let X be a nonempty set. By a *sequence* in X, we mean any mapping $x : N \to X$, where $N = \{0, 1, \ldots\}$ is the set of *natural* numbers. For simplicity reasons, it will be useful to denote it as $(x(n);\ n \geq 0)$, or $(x_n;\ n \geq 0)$; moreover, when no confusion can arise, we further simplify this notation as $(x(n))$ or (x_n), respectively. Also, any sequence $(y_n := x_{i(n)};\ n \geq 0)$ with

$$(i(n);\ n \geq 0) \text{ is divergent } (i(n) \to \infty \text{ as } n \to \infty)$$

will be referred to as a *subsequence* of $(x_n;\ n \geq 0)$. Note that, under such a convention, the relation "subsequence of" is transitive; i.e.:

$$(z_n) = \text{subsequence of } (y_n) \text{ and } (y_n) = \text{subsequence of } (x_n)$$

$$\text{imply} (z_n) = \text{subsequence of } (x_n).$$

(B) Remember that, an outstanding part of (ZF) is the *Axiom of Choice* (abbreviated: AC); which, in a convenient manner, may be written as

(AC) For each couple (J, X) of nonempty sets and each function
$F : J \to \exp(X)$, there exists a (selective) function
$f : J \to X$, with $f(\nu) \in F(\nu)$, for each $\nu \in J$.

Here, $\exp(X)$ stands for the class of all nonempty elements in $\exp[X]$. Sometimes, when the index set J is denumerable, the existence of such a selective function may be determined by using a weaker form of (AC), called: *Dependent Choice* principle (in short: DC). Call the relation $\mathscr{R}$ over X, *proper* when

$$(X(x,\mathscr{R}) =)\mathscr{R}(x) \text{ is nonempty, for each } x \in X.$$

Then, $\mathscr{R}$ is to be viewed as a mapping between X and $\exp(X)$; and the couple $(X, \mathscr{R})$ will be referred to as a *proper relational structure*. Further, given $a \in X$, let us say that the sequence $(x_n;\ n \geq 0)$ in X is $(a;\mathscr{R})$-*iterative*, provided

$$x_0 = a, \text{ and } (x_n; n \geq 0) \text{ is } \mathscr{R}\text{-increasing :}$$
$$x_n \mathscr{R} x_{n+1} \text{ (i.e.: } x_{n+1} \in \mathscr{R}(x_n)), \text{ for all } n.$$

Proposition 1 *Let the relational structure $(X, \mathscr{R})$ be proper. Then, for each $a \in X$ there is at least an $(a, \mathscr{R})$-iterative sequence in X.*

This principle—proposed, independently, by Bernays [4] and Tarski [35]—is deductible from (AC), but not conversely; cf. Wolk [47]. Moreover, by the developments in Moskhovakis [20, Ch 8], and Schechter [33, Ch 6], the reduced system (ZF-AC+DC) it comprehensive enough so as to cover the "usual" mathematics; see also Moore [19, Appendix 2].

A basic consequence of (DC) is the so-called *Denumerable Axiom of Choice* [in short: AC(N)].

Proposition 2 *Let $F : N \to \exp(X)$ be a function. Then, for each $a \in F(0)$ there exists a function $f : N \to X$ with $f(0) = a$ and $f(n) \in F(n)$, $\forall n \geq 0$.*

Proof Denote $Q = N \times X$; and let us introduce the (proper) relation $\mathscr{R}$ over it, according to:

$$\mathscr{R}(n, x) = \{n + 1\} \times F(n + 1), \quad n \geq 0, x \in X.$$

By an application of (DC) to the proper relational structure $(Q, \mathscr{R})$ the conclusion follows; we do not give details.

As a consequence of the above facts,

$$\text{(DC)} \implies \text{(AC(N)) in (ZF-AC); or, equivalently :}$$
$$\text{(AC(N)) is deductible in the system (ZF-AC + DC).}$$

The reciprocal of the written inclusion is not true; see Moskhovakis [20, Ch 8, Sect 8.25] for details.

3 Conv-Cauchy Structures

Let X be a nonempty set. Further, let $d : X \times X \to R_+$ be a mapping with

(m-1) d is *triangular*: $d(x, z) \leq d(x, y) + d(y, z)$, $\forall x, y, z \in X$
(m-2) d is *reflexive-sufficient*: $d(x, y) = 0$ iff $x = y$
(m-3) d is *symmetric*: $d(x, y) = d(y, x)$, for all $x, y \in X$.

We then say that $d(., .)$ is a *metric* on X; and the couple (X, d) will be referred to as a *metric space*.

(A) We introduce a d-convergence and d-Cauchy structure on X as follows. Given the sequence (x_n) in X and the point $x \in X$, we say that (x_n), *d-converges* to x (written as: $x_n \stackrel{d}{\longrightarrow} x$), provided $d(x_n, x) \to 0$ as $n \to \infty$; i.e.,

$$\forall \varepsilon > 0, \exists i = i(\varepsilon) : \ i \le n \Longrightarrow d(x_n, x) < \varepsilon;$$

or, equivalently:

$$\forall \varepsilon > 0, \exists i = i(\varepsilon) : \ i \le n \Longrightarrow d(x_n, x) \le \varepsilon.$$

The set of all such points x will be denoted $\lim_n(x_n)$; when it is nonempty, then (x_n) is called *d-convergent*. By this very definition, we have the properties:

(conv-1) $((\stackrel{d}{\longrightarrow})$ is $hereditary)$

$x_n \stackrel{d}{\longrightarrow} x$ implies $y_n \stackrel{d}{\longrightarrow} x$, for each subsequence (y_n) of (x_n)

(conv-2) $((\stackrel{d}{\longrightarrow})$ is $reflexive)$

$(\forall u \in X)$: the constant sequence $(x_n = u; n \ge 0)$ fulfills $x_n \stackrel{d}{\longrightarrow} u$.

As a consequence, $(\stackrel{d}{\longrightarrow})$ has all properties required in Kasahara [14]; in addition—as d is triangular symmetric—the following extra property is holding here

(conv-3) $(\stackrel{d}{\longrightarrow})$ is *separated* (referred to as d is *separated*):
$\lim_n(x_n)$ is an asingleton, for each sequence (x_n) in X.

The introduced concepts allow us to give a useful property.

Proposition 3 *The mapping $(x, y) \mapsto d(x, y)$ is d-Lipschitz, in the sense*

$$(31\text{-}1) \ |d(x, y) - d(u, v)| \le d(x, u) + d(y, v), \ \forall (x, y), (u, v) \in X \times X.$$

As a consequence, this map is d-continuous; i.e.,

$$(31\text{-}2) \ x_n \stackrel{d}{\longrightarrow} x, \ y_n \stackrel{d}{\longrightarrow} y \ imply \ d(x_n, y_n) \to d(x, y).$$

Further, call the sequence (x_n), *d-Cauchy* when $d(x_m, x_n) \to 0$ as $m, n \to \infty$, $m < n$; that is,

$$\forall \varepsilon > 0, \exists j = j(\varepsilon) : \ j \le m < n \Longrightarrow d(x_m, x_n) < \varepsilon;$$

or, equivalently:

$$\forall \varepsilon > 0, \exists j = j(\varepsilon) : \ j \le m < n \Longrightarrow d(x_m, x_n) \le \varepsilon.$$

the class of all these will be denoted as $Cauchy(d)$. As before, from this very definition one has the properties

(Cauchy-1) ($Cauchy(d)$ is *hereditary*)
(x_n) is d-Cauchy implies (y_n) is d-Cauchy,
for each subsequence (y_n) of (x_n)
(Cauchy-2) ($Cauchy(d)$ is *reflexive*)
($\forall u \in X$): the constant sequence ($x_n = u; n \geq 0$) is d-Cauchy.

Hence, $Cauchy(d)$ is a Cauchy structure, under the lines in Turinici [38].

Now—according to the quoted work—term the couple $((\xrightarrow{d}), Cauchy(d))$, a *conv-Cauchy structure* induced by d. The following regularity conditions about this structure are to be (optionally) considered

(CC-1) d is *regular*: each d-convergent sequence in X is d-Cauchy
(CC-2) d is *complete*: each d-Cauchy sequence in X is d-convergent.

Clearly, the former of these is always obtainable, via d=triangular symmetric; but the latter one is not in general valid.

(B) In the following, some d-Cauchy criteria will be stated. Some preliminaries are needed.

Let us say that ($x_n; n \geq 0$) is *d-asymptotic*, provided

$$r_n := d(x_n, x_{n+1}) \to 0 \text{ as } n \to \infty.$$

Clearly, each d-Cauchy sequence is d-asymptotic too; the reciprocal of this is not in general true. This tells us that the d-Cauchy criteria we are looking for are to be sought in the class of d-asymptotic sequences. Precisely, suppose that the following inclusion is valid

(for each d-asymptotic sequence (x_n)) :

(x_n) is not d-Cauchy implies (x_n) has the property π.

Then, the d-Cauchy criterion in question writes

(for each d-asymptotic sequence (x_n)) :

(x_n) is not endowed with the property π implies (x_n) is d-Cauchy.

To get concrete examples of such properties, we need some other conventions. Given the d-asymptotic sequence ($x_n; n \geq 0$) and $\varepsilon > 0$, let us say that $i \in N$ is *ε-regular*, provided

$$i \leq n \text{ implies } d(x_n, x_{n+1}) < \varepsilon.$$

The class $\mathscr{Z}(\varepsilon)$ of all these numbers is nonempty; so that

$$(\forall \varepsilon > 0): \ Z(\varepsilon) = \min \mathscr{Z}(\varepsilon) \text{ is well defined, as an element of } N;$$

$$\text{with, in addition: } d(x_n, x_{n+1}) < \varepsilon, \ \text{for all } n \geq Z(\varepsilon).$$

Further, note that by this very definition,

$$(\forall h \geq 1): \Gamma_n(h) := \max\{d(x_n, x_{n+i}); i \in N[0, h]\} < \varepsilon,$$

$$\text{for all } n \geq Z(\varepsilon/h); \ \text{whence, } \Gamma_n(h) \to 0 \text{ as } n \to \infty.$$

Define the subsets of $N \times N$

$$(\leq; N) = \{(m, n) \in N \times N; m \leq n\}, \ (<; N) = \{(m, n) \in N \times N; m < n\};$$

these are just the graphs of the relations $(\leq)$ and $(<)$ introduced over N.

Further, let us say that the subset Θ of $R_+^0 :=]0, \infty[$ is $(>)$-*cofinal* in R_+^0, when for each $\varepsilon > 0$, there exists $\theta \in \Theta$ with $\varepsilon > \theta$.

Finally, given the sequence $(r_n; n \geq 0)$ in R and the point $r \in R$, let us write

$$r_n \to r+ \ (\text{also written as: } \lim_n r_n = r+),$$

$$\text{if } r_n \to r \text{ and } r_n > r, \ \text{for all } n \geq 0.$$

The following result, referred to as *Boyd-Wong Criterion* is now available.

Theorem 2 *Let the sequence $(x_n; n \geq 0)$ in X be such that*

(31-i) $(x_n; n \geq 0)$ *is d-asymptotic ($r_n := d(x_n, x_{n+1}) \to 0$ as $n \to \infty$)*
(31-ii) $(x_n; n \geq 0)$ *is not d-Cauchy.*

Further, let the subset Θ of R_+^0 be $(>)$-cofinal in R_+^0; and $h \geq 1$ be some rank. There exist then a number $\gamma \in \Theta$, a rank $J := J(\gamma, h) \geq 1$, and a couple of rank-sequences $(m(k); k \geq 0)$, $(n(k); k \geq 0)$, with

(31-a) $J + k \leq m(k) < m(k) + 3h < n(k), \forall k$
(31-b) $J + k \leq m(k) < m(k) + 2h < n(k) - 1 < n(k)$, *and* $d(x_{m(k)}, x_{n(k)}) > \gamma$, $d(x_{m(k)}, x_{n(k)-1}) \leq \Gamma_{m(k)}(3h) + \gamma, \forall k$
(31-c) $U_k := d(x_{m(k)}, x_{n(k)}) \to \gamma+$ *as* $k \to \infty$
(31-d) $\forall s, t \in N[0, 3h]$: $V_k(s, t) := d(x_{m(k)+s}, x_{n(k)+t}) \to \gamma+$ *as* $k \to \infty$
(31-e) $\forall p, q \in N$: $V_k(p, q) := d(x_{m(k)+p}, x_{n(k)+q}) \to \gamma$ *as* $k \to \infty$.

Proof By definition, the d-Cauchy property of our sequence writes:

$$\forall \varepsilon \in R_+^0, \exists a \in N, \forall (m, n) \in (<; N): \ a \leq m < n \Longrightarrow d(x_m, x_n) \leq \varepsilon.$$

As Θ is a $(>)$-cofinal part in R_+^0, this property may be also written as

$$\forall \theta \in \Theta, \exists \alpha \in N, \forall (m, n) \in (<; N): \ \alpha \leq m < n \Longrightarrow d(x_m, x_n) \leq \theta.$$

The negation of this property means: there exists $\beta \in \Theta$ such that

$$\text{(rela-1)}\quad E(j) := \{(m,n) \in (<; N);\ j \leq m < n, d(x_m, x_n) > \beta\} \neq \emptyset,\ \forall j.$$

As $(x_n; n \geq 0)$ is d-asymptotic, we must have (under d=metric)

$$d(x_n, x_{n+i}) \to 0 \text{ as } n \to \infty, \text{ for each } i \geq 0.$$

Let the number $\gamma \in \Theta$ be given according to

$$\beta > 3\gamma \text{ (possible, since } \Theta \text{ is } (>)\text{-cofinal in } R_+^0).$$

Further; let us take

$$\text{(rela-2)}\quad J(\gamma, h) = Z(\gamma/3h);\ \text{hence, in particular}$$
$$\Gamma_n(3h) := \max\{d(x_n, x_{n+i}); i \in N[0, 3h]\} < \gamma,\ \forall n \geq J(\gamma, h).$$

Denote for simplicity $J = J(\gamma, h)$ and

$$(A(k) = E(J + k); k \geq 0);\ \text{hence, by definition,}$$
$$A(k) := \{(m,n) \in (<; N); J + k \leq m < n, d(x_m, x_n) > \beta\}, k \geq 0;$$
$$\text{with, in addition: } A(k) \text{ is nonempty, for each } k \geq 0.$$

By the triangle inequality, we have for each $(m,n) \in A(k)$

$$d(x_{m+s}, x_{n+t}) \geq d(x_m, x_n) - d(x_m, x_{m+s}) - d(x_n, x_{n+t}) >$$
$$\beta - \gamma - \gamma = \beta - 2\gamma > \gamma, \forall s, t \in N[0, 3h];$$

which tells us that

$$B(k) := \{(m,n) \in (<; N); J + k \leq m < n, d(x_{m+s}, x_{n+t}) > \gamma, \forall s, t \in N[0, 3h]\}$$

is nonempty, for all $k \geq 0$.

Having this precise, denote for each $k \geq 0$

$$m(k) = \min \text{Dom}(B(k)),\ n(k) = \min B(k)(m(k)).$$

By this very definition, we get

(pro-1) $J + k \leq m(k) < n(k), \forall k \geq 0$
(pro-2) $d(x_{m(k)+s}, x_{n(k)+t}) > \gamma, \forall s, t \in N[0, 3h], \forall k \geq 0.$

We claim that the couple (γ, J) and the couple of rank-sequences $(m(k); k \geq 0)$ and $(n(k); k \geq 0)$ fulfill all conclusions in the statement.

i): By (pro-1), it is clear that the first and the second relation of (31-a) holds. Concerning the third relation of the same, suppose by contradiction that

$$(m(k) <)n(k) \leq m(k) + 3h, \text{ for some } k \geq 0.$$

Then by (rela-2) (and $m(k) \geq J$)

$$d(x_{m(k)}, x_{n(k)}) \leq \Gamma_{m(k)}(3h) < \gamma;$$

in contradiction with (pro-2); whence, the third relation in (31-a) holds too.

ii): The first and second relation of (31-b) are directly obtainable from the preceding stage and (pro-2), respectively. Concerning the third relation of (31-b), let $k \geq 0$ be arbitrary fixed. By definition, $n(k)$ is the minimum of all ranks $p \in N$ with

$$(m(k), p) \in B(k); \text{ that is:}$$
$$J + k \leq m(k) < p \text{ and } d(x_{m(k)+s}, x_{p+t}) > \gamma, \forall s, t \in N[0, 3h].$$

As $m(k) < m(k) + 2h < n(k) - 1$ we must have (by this minimal property)

$$\text{(pro-3) } d(x_{m(k)+s}, x_{n(k)-1+t}) \leq \gamma, \text{ for some } s, t \in N[0, 3h].$$

But, in view of (pro-2) once again,

$$\text{(pro-4) } d(x_{m(k)+u}, x_{n(k)-1+v}) > \gamma, \text{ for all } u \in N[0, 3h], v \in N[1, 3h].$$

This, combined with (pro-3), tells us that, necessarily,

$$\text{(pro-5) } d(x_{m(k)+s}, x_{n(k)-1}) \leq \gamma, \text{ for some } s \in N[0, 3h].$$

By the triangular inequality, we have (under the precise index s)

$$d(x_{m(k)}, x_{n(k)-1}) \leq d(x_{m(k)}, x_{m(k)+s}) + d(x_{m(k)+s}, x_{n(k)-1}) \leq \Gamma_{m(k)}(3h) + \gamma;$$

and the last conclusion of (31-b) follows.

iii): By the very definition of $(B(k); k \geq 0)$,

$$U_k > \gamma, \text{ for all } k \geq 0.$$

Moreover, taking the triangular inequality into account,

$$\gamma < d(x_{m(k)}, x_{n(k)}) \leq d(x_{m(k)}, x_{n(k)-1}) + r_{n(k)-1} \leq \Gamma_{m(k)}(3h) + \gamma + r_{n(k)-1}, \forall k.$$

Passing to limit in this double inequality gives (31-c).

iv): Let $s, t \in N[0, 3h]$ be arbitrary fixed. By the very definition of $(B(k); k \geq 0)$,

$$V_k(s, t) > \gamma, \text{ for all } k \geq 0.$$

Moreover, from a metrical property of d,

$$|d(x_{m(k)}, x_{n(k)}) - d(x_{m(k)+s}, x_{n(k)+t})| \leq d(x_{m(k)}, x_{m(k)+s}) + d(x_{n(k)}, x_{n(k)+t}) \leq$$
$$\Gamma_{m(k)}(3h) + \Gamma_{n(k)}(3h) < 2\gamma, \text{ for all } k \geq 0.$$

Passing to limit in the relation between the first and the third member of this relation gives (31-d).

v): Fix $p, q \in N$. From the metrical property of d we just evoked,

$$|d(x_{m(k)}, x_{n(k)}) - d(x_{m(k)+p}, x_{n(k)+q})| \leq d(x_{m(k)}, x_{m(k)+p}) + d(x_{n(k)}, x_{n(k)+q}) \leq$$
$$\Gamma_{m(k)}(p) + \Gamma_{n(k)}(q), \text{ for all } k \geq 0.$$

Passing to limit in the relation between the first and the third member of this relation and noting that

$$\Gamma_{m(k)}(p) \to 0 \text{ and } \Gamma_{n(k)}(q) \to 0 \text{ as } k \to \infty$$

gives the last conclusion (31-e). The proof is complete.

Remark 1 A natural problem to be posed is that of the last conclusion above being retainable in terms of right convergence; that is,

$$\forall p, q \in N : V_k(p, q) := d(x_{m(k)+p}, x_{n(k)+q}) \to \gamma + \text{ as } k \to \infty.$$

This is not in general true; because, fixing some rank i, and putting $p = n(i) - m(i)$, we have

$$d(x_{m(i)+p}, x_{n(i)}) = 0 < \gamma;$$

so that, a relation like $d(x_{m(k)+p}, x_{n(k)}) \to \gamma +$ as $k \to \infty$ is not possible.

This contradicts an affirmation in Vujaković et al. [44, Lemma 1]; but, fortunately, it has no impact upon the remaining statements in that paper.

By definition, the quadruple $[\gamma; J; (m(k); k \geq 0); (n(k); k \geq 0)]$ given by this result will be referred to as a *Boyd-Wong* (Θ, h)*-system* attached to (x_n). In this case, the result above may be expressed as below.

Theorem 3 *Let the sequence* $(x_n; n \geq 0)$ *in X be d-asymptotic. Then, the following conditions/properties are equivalent*

(32-a) $(x_n; n \geq 0)$ *is not d-Cauchy*

(32-b) for each $(>)$*-cofinal subset* Θ *of* R_+^0 *and each* $h \geq 1$*, there is at least one Boyd-Wong* $[\Theta, h]$*-system attached to* (x_n).

Proof By the preceding result, the first condition includes the second one. Conversely, if the second condition holds then, under the choice $\Theta = R_+^0$ and $h = 1$, there must be one Boyd-Wong $[\Theta, h]$-system attached to (x_n); hence, in particular

$$\lim_k d(x_{m(k)}, x_{n(k)}) = \gamma+, \text{ where } \gamma > 0.$$

This necessarily gives us that $(x_n; n \geq 0)$ is not d-Cauchy; for otherwise—under a d-Cauchy condition upon our starting sequence—it follows that

$$\lim_k d(x_{m(k)}, x_{n(k)}) = 0;$$

in contradiction with the limiting property above.

In particular, when $\Theta = R_+^0$, the obtained statement covers the 1969 one in Boyd and Wong [5]; so, it is natural that this result be referred to in the proposed way. Further aspects may be found in Reich [29]; see also Khan et al. [15].

4 Statement of the Problem

Let (X, d) be a metric space. For each relation $\mathscr{A}$ on X and each (nonempty) subset Z of X, let us introduce the convention

$$(\mathscr{A}; Z) = \mathscr{A} \cap (Z \times Z) \text{ (the restriction of } \mathscr{A} \text{ to } Z).$$

In particular, $\mathscr{A}$ is identical with $(\mathscr{A}; X)$.

Let $T \in \mathscr{F}(X)$ be a selfmap of X. In the following, sufficient conditions are given for the existence and/or uniqueness of elements in $\mathrm{Fix}(T)$.

4-I) The proposed problem will be developed in the setting of

(it-seq) the fixed points of T are ultimately chosen among the limit points (if any) $T^\omega x_0 := \lim_n (T^n x_0)$, where $x_0 \in X$ is arbitrary fixed.

(Here, ω is the first transfinite ordinal). To do this, a lot of technical facts about iterative processes is needed.

Let x_0 be some point in X. By an *orbital* (or: *iterative*) sequence attached to x_0 (and T), we mean any sequence $X_0 := (x_n; n \geq 0)$ (or, simply, $X_0 = (x_n)$) defined as $(x_n = T^n x_0; n \geq 0)$. When x_0 is generic here, the resulting object will be referred to as an *orbital* sequence (in short: *o-sequence*) on X.

Fix in the following such an object $X_0 = (x_n)$. Then, denote

$$[X_0] = \{x_n; n \geq 0\} \text{ (the } trajectory \text{ attached to } X_0 = (x_n))$$

$$[[X_0]] := \mathrm{cl}([X_0]) \text{ (the } complete\ trajectory \text{ attached to } X_0 = (x_n)).$$

Proposition 4 *Under these conventions,*

$$[[X_0]] = [X_0] \cup \{z\}, \ \textit{whenever } z := \lim_n (x_n) \textit{ exists}.$$

Proof Denote for simplicity $U_0 = [X_0]$, $V_0 = [[X_0]]$; hence, $V_0 = \mathrm{cl}(U_0)$. Clearly, $V_0 \supseteq U_0 \cup \{z\}$. Suppose that there exists $v \in V_0$ that is outside $U_0 \cup \{z\}$. By the limit definition, there exists $\sigma > 0$ such that

$$X(v, \sigma) := \{x \in X; d(v, x) < \sigma\} \text{ is disjoint from } U_0 \cup \{z\}.$$

In particular, this tells us that v cannot belong to $\mathrm{cl}(U_0) = V_0$; contradiction. Consequently, $V_0 = U_0 \cup \{z\}$; and we are done.

4-II) Passing to the basic part of our setting, denote for $x, y \in X$

$$Q_1(x, y) = d(x, Tx), \ Q_2(x, y) = d(x, y),$$
$$Q_3(x, y) = d(x, Ty), \ Q_4(x, y) = d(Tx, y),$$
$$Q_5(x, y) = d(Tx, Ty), \ Q_6(x, y) = d(y, Ty),$$
$$\mathcal{Q}(x, y) = (Q_1(x, y), Q_2(x, y), Q_3(x, y), Q_4(x, y), Q_5(x, y), Q_6(x, y)).$$

Further, let us construct the family of functions [for $x, y \in X$]

$$B_0(x, y) = \min\{Q_2, Q_5\}(x, y),$$
$$B_1(x, y) = \min\{Q_1, Q_2, Q_5, Q_6\}(x, y)$$
$$B_2(x, y) = \min\{Q_2, Q_3, Q_4, Q_5\}(x, y),$$
$$B_3(x, y) = \min\{Q_1, Q_2, Q_3, Q_4, Q_5, Q_6\}(x, y),$$

and introduce the relations (over X)

$$\text{(rela-B)} \ \ (B_i > 0) = \{(x, y) \in X \times X; B_i(x, y) > 0\}, i \in \{0, 1, 2, 3\}.$$

Finally, let Υ be a nonempty subset of R_+^6; and let $\mathscr{R} = \mathscr{R}(\Upsilon)$ stand for the relation

$$\mathscr{R} = \mathcal{Q}^{-1}(\Upsilon); \text{ hence, } \mathcal{Q}(\mathscr{R}) \subseteq \Upsilon.$$

Technically, the couple $(\Upsilon, \mathscr{R})$ will suffice for setting up a class of (implicit) fixed point statements over the ambient space based on the contractive property

$$\text{(R-contr)} \ \ T \text{ is } (d, \mathscr{R}; \Upsilon)\textit{-contractive}, \text{ provided}$$
$$\mathcal{Q}(x, y) \in \Upsilon, \forall (x, y) \in \mathscr{R}; \text{ that is: } \mathcal{Q}(\mathscr{R}) \subseteq \Upsilon.$$

The starting point of it is the construction of an orbital sequence $X_0 = (x_n)$ with

$$\text{(R-asc)}\ \ X_0 = (x_n) \text{ is } \mathscr{R}\text{-ascending: } x_n \mathscr{R} x_{n+1},\ \text{for all } n.$$

The general aspects of this method will be discussed elsewhere. Here, we will consider a particular case of it—that, in fact, includes a large number of such contractions—based on the restrictive condition

$$\text{(B1-adm)}\ \ \mathscr{R} \text{ is } B_1\text{-}admissible : (B_1 > 0) \subseteq \mathscr{R}.$$

This tells us that, in place of our general contraction, we study its restrictive part

$$\text{(B1-contr)}\ \ T \text{ is } (d, (B_1 > 0); \Upsilon)\text{-}contractive, \text{ provided}$$
$$\mathscr{Q}(x, y) \in \Upsilon, \forall (x, y) \in (B_1 > 0).$$

Another condition to be used, in certain moments of our exposition, is

$$\text{(B0-adm)}\ \ \mathscr{R} \text{ is } B_0\text{-admissible: } (B_0 > 0) \subseteq \mathscr{R};$$

with the associated (stronger) contractive condition

$$\text{(B0-contr)}\ \ T \text{ is } (d, (B_0 > 0); \Upsilon)\text{-}contractive, \text{ provided}$$
$$\mathscr{Q}(x, y) \in \Upsilon, \forall (x, y) \in (B_0 > 0).$$

This, as we will see, has an impact upon the uniqueness property.

Returning to the B_1-admissible setting we stress that, in this case, our initial objective of constructing a $\mathscr{R}$-ascending orbital sequence is directly attainable, in terms of the coarser relation $(B_1 > 0)$. This may be carried out as follows. Call the orbital sequence $X_0 = (x_n)$, $(B_1 > 0)$-*ascending* provided

$$(x_n) \text{ is } (B_1 > 0)\text{-ascending} : \ x_n (B_1 > 0) x_{n+1},\ \text{for all } n.$$

Note that, by the very choice of this relation, our convention writes

$$B_1(x_n, x_{n+1}) > 0,\ \text{for all } n;\ \text{referred to as: } (x_n) \text{ is } B\text{-}ascending.$$

We then say that the o-sequence $X_0 = (x_n)$ has the *Ba-property*; or, equivalently, that $X_0 = (x_n)$ is (Ba-o). The possibility o reaching such a property is a consequence of the reasoning below. Given the orbital sequence $X_0 = (x_n)$, we have two alternatives.

Alt-1) The orbital sequence $X_0 = (x_n)$ is *telescopic*, in the sense there exists $h \geq 0$, such that $d(x_h, x_{h+1}) = 0$; i.e.: $x_h = x_{h+1}$.

By the iterative definition of our sequence, one derives

$$x_h = x_n,\ \text{for all } n \geq h;\ \text{whence, } z := x_h \text{ is an element of } \mathrm{Fix}(T).$$

Consequently, this case is completely clarified from the fixed point perspective.

Alt-2) The orbital sequence $X_0 = (x_n)$ is *non-telescopic*, in the sense $d(x_n, x_{n+1}) > 0$, for all n.

Note that, in this case,

$$(B_1(x_n, x_{n+1}) > 0, \forall n);\ \text{hence, } X_0 = (x_n) \text{ is (Ba-o)}.$$

Summing up, the orbital sequences to be used in the sequel are endowed with the (Ba-o) property. In this case, the basic directions under which the investigations be conducted are described in the list below.

(pic-1) We say that the (Ba-o) sequence $X_0 = (x_n)$ is *semi-Picard* (modulo $(d, \mathscr{R}; T)$) when (x_n) is d-asymptotic

(pic-2) We say that the (Ba-o) sequence $X_0 = (x_n)$ is *Picard* (modulo $(d, \mathscr{R}; T)$) when (x_n) is d-Cauchy

(pic-3) We say that the (Ba-o) sequence $X_0 = (x_n)$ is *strongly Picard* (modulo $(d, \mathscr{R}; T)$) when $x_\omega := \lim_n (x_n)$ exists with $x_\omega \in \text{Fix}(T)$

(pic-4) Call the subset Y of X, *$\mathscr{R}$-almost-singleton* (in short: *$\mathscr{R}$-asingleton*) provided $y_1, y_2 \in Y$, $y_1 \mathscr{R} y_2 \Longrightarrow y_1 = y_2$; and *$\mathscr{R}$-singleton* when, in addition, Y is nonempty. Then, let us say that

(fix-R-asing) T is *fix-$\mathscr{R}$-asingleton*, if $\text{Fix}(T)$ is $\mathscr{R}$-asingleton
(fix-R-sing) T is *fix-$\mathscr{R}$-singleton*, in case $\text{Fix}(T)$ is $\mathscr{R}$-singleton.

Likewise (cf. a previous convention), we say that

(fix-asing) T is *fix-asingleton*, if $\text{Fix}(T)$ is asingleton
(fix-sing) T is *fix-singleton*, in case $\text{Fix}(T)$ is singleton.

As a completion of these, we list the sufficient conditions to be used for getting such properties.

(reg-1) We say that X is *(Ba-o,d)-complete*, provided: for any (Ba-o) sequence $Y_0 = (y_n; n \geq 0)$, one has: (y_n) is d-Cauchy implies (y_n) is d-convergent

(reg-2) We say that T is *(Ba-o,d)-continuous*, if: for any (Ba-o) sequence $Y_0 = (y_n; n \geq 0)$, one has: $y_n \xrightarrow{d} z$ implies $Ty_n \xrightarrow{d} Tz$.

4-III) To solve our posed problem along the precise directions, a lot of convergence type requirements is needed for the six-dimensional (geometric) contractive conditions we just introduced; these are strongly connected with the related developments in Turinici [40]. Letting $X_0 = (x_n)$ be a (Ba-o) sequence in X, denote

$$[X_0] = \{x_n; n \geq 0\},\ [[X_0]] = \text{cl}([X_0]).$$

(I) The first condition upon our data is of *asymptotic* type. Two variants of it are of interest.

I-a) Call Υ, *asymptotic* on $[X_0]$ when

(asy) for each sequence (r_n) in R_+^0 and each sequence (p_n) in R_+ with $((r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \in \mathscr{Q}(B_1 > 0; [X_0]) \cap \Upsilon$ and $|p_n - r_n| \leq r_{n+1}$, $\forall n)$, we have $r_n \to 0$; hence, $p_n \to 0$ as well.

I-b) Call Υ, *descending asymptotic* on $[X_0]$ when

(desc-asy) for each sequence (r_n) in R_+^0 and each sequence (p_n) in R_+ with $((r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \in \mathscr{Q}(B_1 > 0; [X_0]) \cap \Upsilon$ and $|p_n - r_n| \leq r_{n+1}, \forall n)$, we have that (r_n)=strictly descending and $r_n \to 0$; hence (via $(0 < r_n - r_{n+1} \leq p_n \leq r_n + r_{n+1}, \forall n)$, (p_n) is a sequence in R_+^0 with (in addition) $p_n \to 0$.

Clearly, the inclusion below holds

Υ is descending asymptotic on $[X_0]$ implies Υ is asymptotic on $[X_0]$.

However, as we will see, the immense majority of contractive conditions is based on the descending asymptotic hypothesis; and not on the asymptotic one.

(II) The second condition to be considered is related to *right* properties. Take some point $c = (c_1, \ldots, c_6)$ in R_+^6. We say that the sequence $(t^n := (t_1^n, \ldots, t_6^n); n \geq 0)$ in R_+^6 is *right* at c, if

$$\text{(r-c)} \quad (t_i^n \to c_i, \text{ as } n \to \infty, \forall i) \text{ and } (t_i^n \to c_i+, \text{ as } n \to \infty, \text{ when } c_i > 0).$$

Given $b > 0$, let us say that Υ is *nright* at b on $[X_0]$, if

(nright) for each sequence $(t^n; n \geq 0)$ in $\mathscr{Q}(B_3 > 0; [X_0]) \cap \Upsilon$, the right property at $(0, b, b, b, b, 0)$ is not true.

The class of all these $b > 0$ will be denoted as $\text{nright}(\Upsilon; [X_0])$. In this case, define

(a-n-r) Υ is *almost nright* on $[X_0]$, if $\Theta := \text{nright}(\Upsilon; [X_0])$ is $(>)$-cofinal in R_+^0 (for each $\varepsilon \in R_+^0$ there exists $\theta \in \Theta$ with $\varepsilon > \theta$)

(n-r) Υ is *nright* on $[X_0]$, if $\Theta := \text{nright}(\Upsilon; [X_0])$ is identical with R_+^0.

(III) The third condition involves *point* properties. Take some point $c = (c_1, \ldots, c_6)$ in R_+^6. We say that the sequence $(t^n := (t_1^n, \ldots, t_6^n); n \geq 0)$ in R_+^6 is *point* at c, if

$$\text{(p-c)} \quad [t_i^n \to c_i \text{ as } n \to \infty, \forall i], \text{ and } [t_6^n = c_6, \forall n].$$

Given $b > 0$, let us say that Υ is *npoint* at b on $[[X_0]]$, if

(npoint) for each sequence $(t^n; n \geq 0)$ in $\mathscr{Q}(B_3 > 0; [[X_0]]) \cap \Upsilon$, the point property at $(0, 0, b, 0, b, b)$ is not true.

The class of all these $b > 0$ will be denoted as $\text{npoint}(\Upsilon; [[X_0]])$. Then, define

(a-n-p) Υ is *almost npoint* on $[[X_0]]$, if $\Theta := \text{npoint}(\Upsilon; [[X_0]])$ is $(>)$-cofinal in R_+^0

(n-p) Υ is *npoint* on $[[X_0]]$, if $\Theta := \text{npoint}(\Upsilon; [[X_0]])$ is identical with R_+^0.

(IV) The fourth condition to be posed is *normality*. Given $a > 0$, let us say that

(norm) Υ is *normal* at a, when $(0, a, a, a, a, 0) \in \Upsilon$
(n-norm) Υ is *nnormal* at a, when $(0, a, a, a, a, 0) \notin \Upsilon$.

The class of all nnormal $a > 0$ will be denoted as $\text{nnorm}(\Upsilon)$. Then, define

(a-n-n) Υ is *almost nnormal*, if $\Theta := \text{nnorm}(\Upsilon)$ is $(>)$-cofinal in R_+^0
(n-n) Υ is *nnormal*, if $\Theta := \text{nnorm}(\Upsilon)$ is identical with R_+^0.

Some concrete examples of such objects will be given a bit further.

5 Main Result

Let (X, d) be a metric space. Further, let $T \in \mathscr{F}(X)$ be a selfmap of X. As precise, we are interested to determine sufficient conditions for (uniqueness and) existence of elements in $\text{Fix}(T)$, via contractive type requirements involving iterative processes $X_0 = (x_n)$ starting from an element x_0 of X, and their associated sets

$$[X_0] = \{x_n; n \geq 0\}, [[X_0]] = \text{cl}([X_0]).$$

The basic directions and regularity conditions (relative to X_0) under which the problem of determining fixed points of T is to be solved were already listed; and the contractive type framework (involving X_0) was settled.

We are now in position to state our main result in this exposition.

Theorem 4 *Let the subset $\Upsilon \in \exp(R_+^6)$ and its attached relation $\mathscr{R}$ be such that (T is $(d, \mathscr{R}; \Upsilon)$-contractive, and)*

(51-i) *$\mathscr{R}$ is B_1-admissible: $(B_1 > 0) \subseteq \mathscr{R}$*
(51-ii) *Υ is asymptotic and almost nright on $[Y_0]$, for each (Ba-o) sequence $Y_0 = (y_n)$.*

Further, assume that X is (Ba-o,d)-complete; and let $X_0 = (x_n)$ be a (Ba-o) sequence. Then,

(51-a) *$X_0 = (x_n)$ is Picard (modulo $(d, \mathscr{R}; T)$)*
(51-b) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever T is (Ba-o,d)-continuous*
(51-c) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever Υ is npoint on $[[X_0]]$*
(51-d) *T is fix-$\mathscr{R}$-asingleton (hence, fix-$\mathscr{R}$-singleton, under any of these extra requirements) when, in addition to the conditions above, Υ is nnormal ($(0, a, a, a, a, 0) \notin \Upsilon$, $\forall a > 0$)*
(51-e) *T is fix-asingleton (hence, fix-singleton, under any of these extra requirements) when, in addition to the conditions above, $\mathscr{R}$ is B_0-admissible and Υ is nnormal.*

Proof There are some steps to be passed.

Step 1. Denote for simplicity

$$(r_n = d(x_n, x_{n+1}); n \geq 0), (p_n := d(x_n, x_{n+2}); n \geq 0).$$

From the triangular inequality,

$$\text{(rela-1)} \quad (\forall n): \ |p_n - r_n| \leq r_{n+1}.$$

Moreover, we have by definition

$$(\forall n): \ B_1(x_n, x_{n+1}) = \min\{r_n, r_{n+1}\} > 0; \ \text{whence}, (x_n, x_{n+1}) \in \mathscr{R};$$

if we remember that $\mathscr{R}$ is B_1-admissible. The contractive condition is applicable to the couples (x_n, x_{n+1}), for all n; and gives (under these notations)

$$\text{(rela-2)} \quad (r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \in \mathscr{Q}(B_1 > 0; [X_0]) \subseteq \Upsilon, \ \text{for all } n.$$

By (rela-1)+(rela-2) (and the asymptotic property of Υ), one derives

$r_n \to 0$ as $n \to \infty$; so that,
$X_0 = (x_n)$ is semi-Picard (modulo $(d, \mathscr{R}; T)$).

Step 2. Summing up, $X_0 = (x_n)$ is (Ba-o) and d-asymptotic. On the other hand, as Υ is almost nright, $\Theta := \text{nright}(\Upsilon; [X_0])$ appears as $(>)$-cofinal in R_+^0. We show that, under these conditions,

$X_0 = (x_n)$ is d-Cauchy; hence,
$X_0 = (x_n)$ is Picard (modulo $(d, \mathscr{R}; T)$).

Suppose that this is not true; and fix some index $h \geq 1$. By a preliminary statement, there exist a number $\gamma \in \Theta$, a rank $J := J(\gamma, h) \geq 1$, and a couple of rank-sequences $(m(k); k \geq 0)$, $(n(k); k \geq 0)$, with

(prop-1) $\quad J + k \leq m(k) < m(k) + 3h < n(k), \forall k$
(prop-2) $\quad \forall s, t \in N[0, 3h]$: $V_k(s, t) := d(x_{m(k)+s}, x_{n(k)+t}) \to \gamma+$ as $k \to \infty$.

By the (Ba-o) and d-asymptotic properties of $X_0 = (x_n)$,

(prop-3) $\quad (t_1^k := r_{m(k)}; k \geq 0)$ and $(t_6^k := r_{n(k)}; k \geq 0)$ are sequences in R_+^0 with $t_1^k, t_6^k \to 0$ as $k \to \infty$.

Moreover, taking (prop-2) into account, yields

$$(t_2^k := V_k(0, 0); k \geq 0), (t_3^k := V_k(0, 1); k \geq 0), \ \text{and}$$
$$(t_4^k := V_k(1, 0); k \geq 0), \ (t_5^k := V_k(1, 1); k \geq 0)$$
$$\text{are sequences in } R_+^0 \text{ with } t_2^k, t_3^k, t_4^k, t_5^k \to \gamma + \ \text{as } k \to \infty;$$

hence, putting these together,

$$\text{(t-right) the vectorial sequence } (t^k := (t_1^k, t_2^k, t_3^k, t_4^k, t_5^k, t_6^k); k \geq 0)$$
$$\text{is right at } (0, \gamma, \gamma, \gamma, \gamma, 0).$$

Concerning the contractive property of the same object $(t^k; k \geq 0)$, note that by (prop-2) and (prop-3),

$$(\forall k): \ B_3(x_{m(k)}, x_{n(k)}) = \min\{r_{m(k)}, V_k(0,0), V_k(0,1), V_k(1,0), V_k(1,1), r_{n(k)}\} > 0;$$

whence, $t^k \in \mathscr{Q}(B_3 > 0; [X_0]) \subseteq \Upsilon$.

This, via (t-right), contradicts the choice of γ as element of $\Theta := \text{nright}(\Upsilon; [X_0])$. Hence, our working assumption is not acceptable; and the assertion follows.

Step 3. From these developments, we have, as X is (Ba-o,d)-complete

$$X_0 = (x_n) \text{ is } d\text{-convergent: } x_n \xrightarrow{d} z_0 \text{ as } n \to \infty, \text{ for some } z_0 \in X.$$

We now claim that z_0 is a fixed point of T. Two possible cases—treated in the steps below—are to be discussed.

Step 4. Suppose that T is (Ba-o,d)-continuous. Then,

$$u_n := Tx_n \xrightarrow{d} Tz_0 \text{ as } n \to \infty.$$

On the other hand, $(u_n = x_{n+1}; n \geq 0)$ is a subsequence of $(x_n; n \geq 0)$; whence

$$u_n \xrightarrow{d} z_0 \text{ as } n \to \infty.$$

Combining with d=separated, yields $z_0 = Tz_0$.

Step 5. Suppose that Υ is npoint on $[[X_0]]$. Three alternatives occur.

Alter-1) Suppose that

$$\text{(Tz-rela-1)} \ \ H_1 := \{n \in N; x_n = Tz_0\} \text{ is unbounded (in } N).$$

By a direct procedure (avoiding any use of (AC)) there may be obtained a strictly ascending sequence of ranks $(i(n); n \geq 0)$, such that

$$a_n := x_{i(n)} = Tz_0, \ \text{for all } n.$$

But, $(a_n; n \geq 0)$ is a subsequence of $(x_n; n \geq 0)$; so that $\lim_n(a_n) = z_0$. Passing to limit in the relation above, gives $z_0 = Tz_0$.

Alter-2) Suppose that

(Tz-rela-2) $H_2 := \{n \in N; Tx_n = Tz_0\}$ is unbounded (in N).

By the same procedure (avoiding any use of (AC)) there may be obtained a strictly ascending sequence of ranks $(j(n); n \geq 0)$, such that

$$b_n := Tx_{j(n)} = Tz_0, \text{ for all } n.$$

But, $(b_n = x_{j(n)+1}; n \geq 0)$ is a subsequence of $(x_n; n \geq 0)$; so that $\lim_n(b_n) = z_0$. Passing to limit in the relation above, gives $z_0 = Tz_0$.

Alter-3) Suppose that

both subsets H_1 and H_2 are bounded (in N).

This tells us that

$$\exists i = i(z_0) \in N, \text{ such that:}$$
$$n \geq i \text{ implies } Tx_n \neq Tz_0 \text{ (hence, } x_n \neq z_0) \text{ and } x_n \neq Tz_0.$$

Denote for simplicity $(u_n = x_{n+i}; n \geq 0)$; clearly, by the preceding relation,

(non-id) $(\forall n): Tu_n \neq Tz_0$ (hence, $u_n \neq z_0$) and $u_n \neq Tz_0$.

Again combining with the (Ba-o) property of $X_0 = (x_n)$, one derives

$$\text{(posi-1) } (\forall n): Q_1(u_n, z_0) = d(u_n, Tu_n) > 0,$$
$$Q_2(u_n, z_0) = d(u_n, z_0) > 0, \ Q_3(u_n, z_0) = d(u_n, Tz_0) > 0,$$
$$Q_4(u_n, z_0) = d(Tu_n, z_0) > 0, \ Q_5(u_n, z_0) = d(Tu_n, Tz_0) > 0.$$

Suppose by contradiction that

(posi-2) $b := d(z_0, Tz_0) > 0$ [whence, $Q_6(x_n, z_0) = b > 0, \forall n$].

We show that this is not compatible with Υ being npoint at b.

From the preceding observations, we have

$$(\forall n): B_3(u_n, z_0) = \min\{Q_1, \ldots, Q_6\}(u_n, z_0)\} > 0;$$
$$\text{whence, } \mathscr{Q}(u_n, z_0) \in \mathscr{Q}(B_3 > 0; [[X_0]]) \subseteq \Upsilon.$$

Let us evaluate the left part of this last relation. From the preceding facts

$$(t_1^n := d(u_n, Tu_n); n \geq 0), (t_2^n := d(u_n, z_0); n \geq 0),$$
$$(t_3^n := d(u_n, Tz_0); n \geq 0), \ (t_4^n := d(Tu_n, z_0); n \geq 0),$$

$$(t_5^n := d(Tu_n, Tz_0); n \geq 0), \text{ are sequences in } R_+^0 \text{ with}$$
$$(t_1^n, t_2^n, t_3^n, t_4^n, t_5^n) \to (0, 0, b, 0, b) \text{ as } n \to \infty.$$

At the same time,

$$(t_6^n := d(z, Tz) = b; n \geq 0) \text{ is a constant sequence (with } t_6^n \to b);$$

so, putting these together,

$$\text{(Rela-1)} \quad (t^n := (t_1^n, \ldots, t_6^n); n \geq 0) \text{ is point at } (0, 0, b, 0, b, b).$$

Finally, by simply replacing in the contractive condition above,

$$\text{(Rela-2)} \quad (\forall n) : \ t^n \in \mathscr{Q}(B_3 > 0; [[X_0]]) \subseteq \Upsilon.$$

This, via (Rela-1), contradicts the fact that $b \in \text{npoint}(\Upsilon; [[X_0]])$. Hence, the assumption $b > 0$ cannot be accepted; and then, $b = 0$; that is: $z_0 \in \text{Fix}(T)$.

Step 6. Take the points $z_1, z_2 \in \text{Fix}(T)$, according to

$$z_1 \mathscr{R} z_2; \text{ and (by contradiction) } z_1 \neq z_2 \text{ (hence, } a := d(z_1, z_2) > 0).$$

As $(z_1, z_2) \in \mathscr{R}$, the contractive condition applies to (z_1, z_2); and gives

$$\mathscr{Q}(z_1, z_2) \in \Upsilon; \text{ that is: } (0, a, a, a, a, 0) \in \Upsilon;$$

a contradiction with respect to the nnormal property of Υ. Hence, our working condition is not accepted; and the assertion follows.

Step 7. Take the points $z_1, z_2 \in \text{Fix}(T)$, according to

$$z_1 \neq z_2 \text{ (hence, } a := d(z_1, z_2) > 0).$$

This yields

$$B_0(z_1, z_2) = a > 0; \text{ whence, } (z_1, z_2) \in \mathscr{R};$$

if we note that $\mathscr{R}$ is B_0-admissible. Consequently, the contractive condition applies to (z_1, z_2); and gives

$$\mathscr{Q}(z_1, z_2) \in \Upsilon; \text{ that is: } (0, a, a, a, a, 0) \in \Upsilon;$$

a contradiction with respect to the nnormal property of Υ. Hence, our working condition is not accepted; and the assertion follows.

Note that, multivalued extensions of this result are possible, under the lines in Nadler [21]. On the other hand, an extended setting of these developments is possible, under the lines discussed in the 2001 PhD Thesis by Hitzler [13, Ch 1, Sect 1.2]; see also Pasicki [24]. Further aspects were delineated in Turinici [41].

6 Analytic Methods in RDC-Theory

As an application of the developments above, some analytic methods in the fixed point theory for Rhoades-Dutta-Choudhury contractions are being discussed.

Let (X, d) be a metric space. Further, let $T \in \mathscr{F}(X)$ be a selfmap of X. As precise, we are interested to determine sufficient conditions for (uniqueness and) existence of elements in $\mathrm{Fix}(T)$, via contractive type requirements involving iterative processes $X_0 = (x_n)$ starting from an element x_0 of X, and their associated sets

$$[X_0] = \{x_n; n \geq 0\}, [[X_0]] = \mathrm{cl}([X_0]).$$

The basic directions and regularity conditions (relative to X_0) under which the existence/uniqueness problem involving points of $\mathrm{Fix}(T)$ is to be solved were already listed; and the contractive type framework (involving X_0) was settled. As a by-product of these, we established our main result in this exposition, Theorem 4. It is our aim in the following to show that, starting from this principle, it is possible to describe the basic lines for a kind of analytical approach in treating fixed points of contractions over metric spaces of the type introduced by Rhoades [31] and refined by Dutta and Choudhury [12].

(A) Roughly speaking, the approach to be considered requires an appropriate description of contractive conditions to be used. Denote, for $x, y \in X$

$$Q_1(x, y) = d(x, Tx), \ Q_2(x, y) = d(x, y),$$
$$Q_3(x, y) = d(x, Ty), \ Q_4(x, y) = d(Tx, y),$$
$$Q_5(x, y) = d(Tx, Ty), \ Q_6(x, y) = d(y, Ty),$$
$$\mathscr{Q}(x, y) = (Q_1(x, y), Q_2(x, y), Q_3(x, y), Q_4(x, y), Q_5(x, y), Q_6(x, y)).$$

Then, let us construct the family of functions [for $x, y \in X$]

$$B_0(x, y) = \min\{Q_2, Q_5\}(x, y),$$
$$B_1(x, y) = \min\{Q_1, Q_2, Q_5, Q_6\}(x, y),$$
$$B_2(x, y) = \min\{Q_2, Q_3, Q_4, Q_5\}(x, y),$$
$$B_3(x, y) = \min\{Q_1, Q_2, Q_3, Q_4, Q_5, Q_6\}(x, y),$$

and put, for simplicity,

$$(B_i > 0) = \{(x, y) \in X \times X;\ B_i(x, y) > 0\},\ i \in \{0, 1, 2, 3\}.$$

Let (ψ, φ) be a pair of functions over $\mathscr{F}(R_+^0, R)$. Further, let (u, v, w) be a triple of functions over $\mathscr{F}(R_+^6, R_+)$, and define the couple of subsets

$$\Upsilon_0 = \{t \in R_+^6;\ u(t), v(t), w(t) > 0\},$$
$$\Upsilon = \{t \in \Upsilon_0;\ \psi(u(t)) \leq \psi(v(t)) - \varphi(w(t))\}.$$

Finally, define the couple of relations over X

$$R = Q^{-1}(\Upsilon_0),\ R^* = Q^{-1}(\Upsilon).$$

The obtained relations allow us defining two contractive conditions upon our data. The former of these is a functional contractive condition, written as:

(fct-contr) T is $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-*contractive*, if

$$\psi(u \circ \mathscr{Q}(x, y)) \leq \psi(v \circ \mathscr{Q}(x, y)) - \varphi(w \circ \mathscr{Q}(x, y)),\ \text{when } (x, y) \in \mathscr{R}.$$

The latter of these appears as a set contractive condition and writes

(set-contr) T is $(d, \mathscr{R}; \Upsilon)$-contractive:

$$\mathscr{Q}(x, y) \in \Upsilon,\ \text{if } (x, y) \in \mathscr{R};\ \text{that is } \mathscr{Q}(\mathscr{R}) \subseteq \Upsilon.$$

The connection between these is described in

Proposition 5 *Under the above developments, one has*

T is $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive iff T is $(d, \mathscr{R}; \Upsilon)$-contractive.

Proof Two steps must be passed.

Step 1. Suppose that the functional contractive property (fct-contr) is holding; and let the couple $(x, y) \in \mathscr{R}$ be arbitrary fixed. For the moment,

$$(x, y) \in \mathscr{R} \text{ implies } t := \mathscr{Q}(x, y) \in \Upsilon_0;\ \text{that is: } u(t) > 0, v(t) > 0, w(t) > 0.$$

Moreover, by the very definition of contractive property,

$$\psi(u(t)) \leq \psi(v(t)) - \psi(w(t));\ \text{that is: } t := \mathscr{Q}(x, y) \in \Upsilon.$$

This, by the arbitrariness of the couple $(x, y) \in \mathscr{R}$, shows that the set contractive property holds; and the left to right inclusion follows.

Step 2. Suppose that the set contractive property (set-contr) is holding; and let the couple $(x, y) \in \mathscr{R}$ be arbitrary fixed. For the moment,

$$(x, y) \in \mathscr{R} \text{ implies } t := \mathscr{Q}(x, y) \in \Upsilon_0;\ \text{that is: } u(t) > 0, v(t) > 0, w(t) > 0.$$

Moreover, by the very definition of Υ,

$$\psi(u(t)) \le \psi(v(t)) - \psi(w(t)); \text{ where } t := \mathscr{Q}(x, y).$$

This tells us that the couple (x, y) fulfills the functional contractive property; and proves the right to left inclusion.

In other words the functional contractive condition (fct-contr) is equivalent with the set-contractive condition (set-contr); to which the methods of the main result are applicable. Technically speaking, an effective application of the main result to the set contractive problem (set-contr) amounts to verifying the (sufficient) asymptotic, right, point, and normality properties upon Υ, in terms of its components; that is: the relation $\mathscr{R}$, the pair (ψ, φ), and the triple (u, v, w). Some preliminaries are needed.

For each function $\psi \in \mathscr{F}(R_+^0, R)$, and each $a > 0$, define

ψ is right regulated at a:
$\psi(a+0) = \lim_{t \to a+} \psi(t)$ (the right limit of ψ at a) exists (in R)
ψ is left regulated at a:
$\psi(a-0) = \lim_{t \to a-} \psi(t)$ (the left limit of ψ at a) exists (in R)
ψ is regulated at a: both $\psi(a+0)$ and $\psi(a-0)$ exist (in R)
ψ is continuous at a: ψ is regulated at a and $\psi(a+0) = \psi(a-0) = \psi(a)$.

The class of all such a will be indicated, respectively, as

rreg(ψ)= the right regulated domain of ψ,
lreg(ψ)= the left regulated domain of ψ,
reg(ψ)= the regulated domain of ψ,
cont(ψ)= the continuous domain of ψ.

In this case, we say that

(a-r-reg) ψ is *almost right regulated*, if $\Theta := \text{rreg}(\psi)$ is $(>)$-cofinal in R_+^0
(r-reg) ψ is *right regulated*, if $\Theta := \text{rreg}(\psi)$ is identical with R_+^0
(a-l-reg) ψ is *almost left regulated*, if $\Theta := \text{lreg}(\psi)$ is $(>)$-cofinal in R_+^0
(l-reg) ψ is *left regulated*, if $\Theta := \text{lreg}(\psi)$ is identical with R_+^0
(a-reg) ψ is *almost regulated*, if $\Theta := \text{reg}(\psi)$ is $(>)$-cofinal in R_+^0
(reg) ψ is *regulated*, if $\Theta := \text{reg}(\psi)$ is identical with R_+^0
(a-cont) ψ is *almost continuous*, if $\Theta := \text{cont}(\psi)$ is $(>)$-cofinal in R_+^0
(cont) ψ is *continuous*, if $\Theta := \text{cont}(\psi)$ is identical with R_+^0.

Remark 2 The class of almost right regulated functions is pretty large. This is shown in example below:

Let $(a_n; n \ge 0)$ be a strictly descending sequence in R_+^0 with

$$a_n \to 0 \text{ as } n \to \infty; \text{ whence: } \Theta := \{a_n; n \ge 0\} \text{ is } (>)\text{-cofinal in } R_+^0.$$

Then, let $\psi : R_+^0 \to R$ be a function with

$$\text{(r-lim)} \quad \psi(a_i + 0) \text{ exists, for all } i \in N.$$

In this case, ψ appears as an almost right regulated function. Note that, (r-lim) requires this behavior of ψ in the right neighborhoods of points in Θ; which means that, in the remaining points of R_+^0, the behavior of ψ is completely arbitrary.

In other words, the class of almost right regulated functions may contain many functions in $\mathscr{F}(R_+^0, R)$ with an arbitrary behavior over large subsets of R_+^0.

Remark 3 The class of almost continuous functions is pretty large too, in view of

(for each $\psi \in \mathscr{F}(R_+^0, R)$):
ψ is increasing or continuous implies ψ is almost continuous.

In fact, the affirmation concerning continuity is clear. On the other hand, the affirmation concerning increasing property is again clear, by a result in Natanson [23, Ch 8, Sect 1]. For, when ψ is increasing, there exists (by the quoted result) a denumerable part Δ of R_+^0, with

$$\Theta := R_+^0 \setminus \Delta \subseteq \text{cont}(\psi);$$

and this along with Θ being $(>)$-cofinal, proves the claim.

Remark 4 Concerning the algebraic properties of such functions, the class of right regulated functions is invariant with respect to linear combinations and products:

ψ_1, ψ_2=right regulated imply $\alpha_1\psi_1 + \alpha_2\psi_2$ (where $\alpha_1, \alpha_2 \in R$)
and $\psi_1\psi_2$ are right regulated too.

This property is no longer true for the class of almost right regulated functions, in view of immediate fact

the $(>)$-cofinal in R_+^0 property is not invariant to intersections: if Θ_1 and Θ_2 are $(>)$-cofinal in R_+^0, then $\Theta_1 \cap \theta_2$ need not be $(>)$-cofinal in R_+^0;

just let Θ_1 and Θ_2 be the class of all rationals and irrationals of R_+^0, respectively. However, when one of these functions is continuous, this happens, in the sense

$\psi_1 =$ almost right regulated and $\psi_2 =$ continuous, imply
$\alpha_1\psi_1 + \alpha_2\psi_2$ (where $\alpha_1, \alpha_2 \in R$) and $\psi_1\psi_2$ are almost right regulated.

Further aspects involving the concepts in question may be found in the 1960 monograph by Dieudonné [10, Ch VII, Sect 6].

Concerning these concepts, the following auxiliary fact is to be noted. Let $\psi : R_+^0 \to R$ be a regulated function. Denote, for each $b, c > 0$

(r-osc-bc) $\text{osc}(+)(\psi; b, c) = \psi(b+0) - \psi(c+0)$

(the *right oscillation* of ψ at (b, c))

(osc-bc) $\text{osc}(\psi; b, c) = \max\{\psi(u) - \psi(v); u \in \{b+0, b, b-0\},$

$v \in \{c+0, c, c-0\}\}$

(the *oscillation* of ψ at (b, c)).

Clearly, by the regulated hypothesis, these are finite real numbers. With an extra care, this definition may be extended to the case of $b = 0$. Precisely, assume that

(reg-ext) ψ is extended regulated: ψ is regulated, and $\psi(0+0)$ exists.

A strong variant of this is

ψ is strong extended regulated:
ψ is regulated, and *zero abrupt*: $\psi(0+0) = -\infty$.

For example, the extended regulated property holds under ψ=increasing. However, the strong extended regulated property is not assured, in this general way; so, to get it, we must impose this property in a mandatory way.

Having these precise, suppose that ψ is extended regulated. We may then define for each $c > 0$,

(r-osc-0c) $\text{osc}(+)(\psi; 0, c) = \psi(0+0) - \psi(c+0)$; the *right oscillation* of ψ at $(0, c)$

(osc-0c) $\text{osc}(\psi; 0, c) = \max\{\psi(0+0) - \psi(v); v \in \{c+0, c, c-0\}\}$; the *oscillation* of ψ at $(0, c)$.

In the same general context, given the function $\varphi : R_+^0 \to R$ and the point $a \geq 0$, let us introduce properties (where $b \geq 0$, $c > 0$):

(r-bd-osc-bca) (ψ, φ) is *right bounded oscillating at* $(b, c; a)$: for each sequence (t_n) in R_+^0 with $t_n \to a+$, we have $\limsup_n(\varphi(t_n)) > \text{osc}(+)(\psi; b, c)$

(bd-osc-bca) (ψ, φ) is *(bilateral) bounded oscillating at* $(b, c; a)$: for each sequence (t_n) in R_+^0 with $t_n \to a$, we have $\limsup_n(\varphi(t_n)) > \text{osc}(\psi; b, c)$.

Proposition 6 *Let (ψ, φ) be a couple of functions over $\mathcal{F}(R_+^0, R)$, with*

ψ is extended regulated

and (a, b, c) be a triple of points with $a, b \geq 0$, $c > 0$. The following are valid

***(62-1)** If the couple (ψ, φ) is right bounded oscillating at $(b, c; a)$, then there are no sequences (t_n), (s_n) and (r_n) in R_+^0, such that*

$$t_n \to a+, s_n \to b+, r_n \to c+ \text{ and } \varphi(t_n) \leq \psi(s_n) - \psi(r_n), \text{ for all } n$$

(62-2) *If the couple* (ψ, φ) *is bounded oscillating at* $(b, c; a)$ *where* $b > 0$*, then there are no sequences* (t_n)*,* (s_n) *and* (r_n) *in* R_+^0*, such that*

$$t_n \to a, s_n \to b, r_n \to c \text{ and } \varphi(t_n) \leq \psi(s_n) - \psi(r_n), \text{ for all } n$$

(62-3) *If the couple* (ψ, φ) *is bounded oscillating at* $(0, c; a)$*, then there are no sequences* (t_n)*,* (s_n) *and* (r_n) *in* R_+^0*, such that*

$$t_n \to a, s_n \to 0+, r_n \to c \text{ and } \varphi(t_n) \leq \psi(s_n) - \psi(r_n), \text{ for all } n.$$

Proof There are three parts to be passed.

Part 1. Suppose, by contradiction that there are sequences (t_n), (s_n) and (r_n) in R_+^0, such that

$$\text{(rela-1)} \quad t_n \to a+, s_n \to b+, r_n \to c+ \text{ and } \varphi(t_n) \leq \psi(s_n) - \psi(r_n), \forall n.$$

Passing to lim sup as $n \to \infty$ in this relation, gives

$\limsup_n \varphi(t_n) \leq \psi(b + 0) - \psi(c + 0)$;
whence (by the imposed conventions): $\limsup_n \varphi(t_n) \leq \text{osc}(+)(\psi; b, c)$;

absurd, by the oscillation type condition.

Part 2. Suppose, by contradiction that (under $b > 0$) there are sequences (t_n) (s_n) and (r_n) in R_+^0, such that

$$\text{(rela-2)} \quad t_n \to a, s_n \to b, r_n \to c \text{ and } \varphi(t_n) \leq \psi(s_n) - \psi(r_n), \text{ for all } n.$$

Let $\mathscr{E} \in \{(>), (=), (<)\}$ be some relation. By definition, the unique limit of the sequence $(\psi(c_n))$ when $c_n \to c$ and $(c_n \mathscr{E} c, \forall n)$ is denoted as $\psi(c\mathscr{E})$. Clearly,

$$\psi(c\mathscr{E}) = \psi(c + 0), \text{ when } \mathscr{E} \text{ is } (>),$$
$$\psi(c\mathscr{E}) = \psi(c), \text{ when } \mathscr{E} \text{ is } (=),$$
$$\psi(c\mathscr{E}) = \psi(c - 0), \text{ when } \mathscr{E} \text{ is } (<).$$

By the total property of the ordering in R, there exists a couple of relations $\mathscr{E}_1, \mathscr{E}_2 \in \{(>), (=), (<)\}$ with

$$H_1 := \{n \in N; s_n \mathscr{E}_1 b\}, H_2 := \{n \in N; r_n \mathscr{E}_2 c\} \text{ are infinite.}$$

As $H_1 := \{n \in N; s_n \mathscr{E}_1 b\}$ is infinite, there exists a strictly ascending sequence of ranks $(i(n))$ such that $s_{i(n)} \mathscr{E}_1 b$, $\forall n$. Without loss—passing to a subsequence if necessary—one may assume $(i(n) = n; n \geq 0)$; so that (by the above)

$$\text{(rela-3)} \quad t_n \to a, s_n \to b, (s_n \mathscr{E}_1 b, \forall n), r_n \to c$$
$$\text{and } \varphi(t_n) \leq \psi(s_n) - \psi(r_n), \text{ for all } n.$$

As $H_2 := \{n \in N; r_n \mathscr{E}_2 c\}$ is infinite, there exists a strictly ascending sequence of ranks $(j(n))$ such that $r_{j(n)} \mathscr{E}_2 c$, $\forall n$. Without loss—passing to a subsequence if necessary—one may assume $(j(n) = n; n \geq 0)$; so that (by the above)

$$\text{(rela-4)}\ \ t_n \to a, s_n \to b, (s_n \mathscr{E}_1 b, \forall n),$$
$$r_n \to c, (r_n \mathscr{E}_2 c, \forall n), \text{ and } \varphi(t_n) \leq \psi(s_n) - \psi(r_n), \text{ for all } n.$$

Passing to lim sup as $n \to \infty$ in this last relation, gives

$\limsup_n \varphi(t_n) \leq \psi(b\mathscr{E}_1) - \psi(c\mathscr{E}_2)$;
whence (by the imposed conventions): $\limsup_n \varphi(t_n) \leq \operatorname{osc}(\psi; b, c)$;

absurd, by the oscillation type condition.

Part 3. Suppose by contradiction that (under $b = 0$) there are sequences (t_n), (s_n) and (r_n) in R_+^0, such that

$$\text{(rela-5)}\ \ t_n \to a, s_n \to 0+, r_n \to c \text{ and } \varphi(t_n) \leq \psi(s_n) - \psi(r_n), \text{ for all } n.$$

Let the preceding notations be in force. By the total property of the ordering in R, there exists a relation $\mathscr{E} \in \{(>), (=), (<)\}$ with

$$H := \{n \in N; r_n \mathscr{E} c\} \text{ is infinite.}$$

There exists then a strictly ascending sequence of ranks $(i(n))$ such that $r_{i(n)} \mathscr{E} c$, $\forall n$. Without loss—passing to a subsequence if necessary—one may assume $(i(n) = n; n \geq 0)$; so that (by the above)

$$\text{(rela-6)}\ \ t_n \to a, s_n \to 0+, r_n \to c (r_n \mathscr{E} c, \forall n),$$
$$\text{and } \varphi(t_n) \leq \psi(s_n) - \psi(r_n), \text{ for all } n.$$

Passing to lim sup as $n \to \infty$ in this last relation, gives

$\limsup_n \varphi(t_n) \leq \psi(0+0) - \psi(c\mathscr{E})$;
whence (by the imposed conventions): $\limsup_n \varphi(t_n) \leq \operatorname{osc}(\psi; 0, c)$;

absurd, by the oscillation type condition.

The proof is complete.

(B) We may now pass to the specific objective of our developments: to determine the (sufficient) asymptotic, right, point, and normality properties upon Υ, in terms of its components; that is: the pair (ψ, φ) and the triple (u, v, w). Two more basic conventions are in order.

Let (ξ, η, ζ) be a triple of functions with $\xi \in \mathscr{F}(R_+^0)$, $\eta, \zeta \in \mathscr{F}(R_+)$; in particular, one may take $\xi = I$ (the identity function $(I(t) = t; t \in R_+^0)$); as well as $\eta = 0$ and/or $\zeta = 0$ (the identically zero function $(0(t) = 0; t \in R_+)$).

Conv-1) We say that (u, v, w) is *right asymptotic of type* (ξ, η, ζ), when

(r-asy) $\forall$ sequence $(t^n; n \geq 0)$ in $\mathscr{Q}(B_3 > 0; [X_0]) \cap \Upsilon$, and $\forall b > 0$: $(t^n; n \geq 0)$ has the right property at $(0, b, b, b, b, 0)$ implies $u(t^n) \to \xi(b)+$, $v(t^n) \to \eta(b)+$, $w(t^n) \to \zeta(b)+$.

Conv-2) We say that (u, v, w) is *point asymptotic of type* (ξ, η, ζ), when

(pt-asy) $\forall$ sequence $(t^n; n \geq 0)$ in $\mathscr{Q}(B_3 > 0; [[X_0]]) \cap \Upsilon$, and $\forall b > 0$:

$(t^n; n \geq 0)$ has the point property at $(0, 0, b, 0, b, b)$

implies $u(t^n) \to \xi(b)$, $v(t^n) \to \eta(b)$, $w(t^n) \to \zeta(b)$.

The following statement is an essential step towards the precise objective. Denote, for each couple $\alpha, \beta \in R$,

co(α, β)=the convex cover of $\{\alpha, \beta\}$;
that is: the interval $[\min\{\alpha, \beta\}, \max\{\alpha, \beta\}]$.

Theorem 5 *Let (ψ, φ) be a couple of functions over $\mathscr{F}(R_+^0, R)$, with*

(61-i) φ is strictly positive ($\varphi(R_+^0) \subseteq R_+^0$).

Further, let (u, v, w) be a triple of functions over $\mathscr{F}(R_+^6, R_+)$, such that (under the precise notations)

(61-ii) $\mathscr{R}$ is B_1-admissible: $(B_1 > 0) \subseteq \mathscr{R}$.

Then,

61-a) *Υ is descending asymptotic on $[X_0]$, provided*
(61-a-i) ψ is right regulated,

and there exists a continuous strictly increasing $\lambda \in \mathscr{F}(R_+^0)$, with

(61-a-ii) (u, v, w) is strongly λ-iterative, in the sense: for each sequence (r_n) in R_+^0 and each sequence (p_n) in R_+ with $(A_n := (r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \in \mathscr{Q}(B_1 > 0; [X_0]) \cap \Upsilon$, and $|p_n - r_n| \leq r_{n+1}$, for all n), we have $u^n := u(A_n) = r_{n+1}$, $v^n := v(A_n) = \max\{r_n, r_{n+1}\}$, $w^n := w(A_n) \in co(\lambda(r_n), \lambda(r_{n+1}))$, $\forall n$
(61-a-iii) (ψ, φ) is right bounded oscillating at $(b, b; \lambda(b))$, for each $b > 0$
61-b) *Υ is descending asymptotic on $[X_0]$, provided*
(61-b-i) ψ is increasing (hence right regulated),

and there exists a continuous strictly increasing $\lambda \in \mathscr{F}(R_+^0)$, with

(61-b-ii) (u, v, w) is weakly λ-iterative, in the sense: for each sequence (r_n) in R_+^0 and each sequence (p_n) in R_+ with $(A_n := (r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \in \mathscr{Q}(B_1 > 0; [X_0]) \cap \Upsilon$, and $|p_n - r_n| \leq r_{n+1}$, for all n), we have $u^n := u(A_n) = r_{n+1}$, $v^n := v(A_n) \leq \max\{r_n, r_{n+1}\}$, $w^n := w(A_n) \in co(\lambda(r_n), \lambda(r_{n+1}))$, $\forall n$
(61-b-iii) (ψ, φ) is right bounded oscillating at $(b, b; \lambda(b))$, for each $b > 0$
61-c) *Υ is almost nright on $[X_0]$, provided*
(61-c-i) ψ is almost right regulated

and there exists $\mu \in \mathscr{F}(R_+)$*, such that:*

(61-c-ii) (u, v, w) *is right asymptotic of type* (I, I, μ)
(61-c-iii) (ψ, φ) *is right bounded oscillating at* $(b, b; \mu(b))$*, for each* $b \in rreg(\psi)$
61-d) Υ *is npoint (hence, almost npoint) on* $[[X_0]]$*, provided*
(61-d-i) ψ *is extended regulated*

and there exists a triple (ξ, η, ζ) *with* $\xi \in \mathscr{F}(R_+^0)$*,* $\eta, \zeta \in \mathscr{F}(R_+)$*, such that:*

(61-d-ii) (u, v, w) *is point asymptotic of type* (ξ, η, ζ)
(61-d-iii) (ψ, φ) *is bounded oscillating at* $(\eta(b), \xi(b); \zeta(b))$*,* $\forall b > 0$
61-e) Υ *is nnormal, provided* (u, v, w) *is invariant, in the sense* $u(0, a, a, a, a, 0) = a$*,* $v(0, a, a, a, a, 0) = a$*,* $w(0, a, a, a, a, 0) = a$*,* $\forall a > 0$.

Proof The argument consists in a number of parts.

Part 1. [Υ is descending asymptotic on $[X_0]$ under the first lot of conditions].
Let the sequence (r_n) in R_+^0 and the sequence (p_n) in R_+ be such that

$$(\forall n):\ A_n := (r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \in \mathscr{Q}(B_1 > 0; [X_0]) \cap \Upsilon,$$
$$\text{and } |p_n - r_n| \le r_{n+1}.$$

As (u, v, w) is strongly λ-iterative,

$$(\forall n):\ u^n := u(A_n) = r_{n+1},\ v^n := v(A_n) = \max\{r_n, r_{n+1}\},$$
$$w^n := w(A_n) \in \mathrm{co}(\lambda(r_n), \lambda(r_{n+1})).$$

By the very representation of Υ, we must have

$$\text{(iter-1)}\ \ (\forall n):\ \psi(r_{n+1}) \le \psi(\max\{r_n, r_{n+1}\}) - \varphi(w^n),$$
$$\text{where } w^n \in \mathrm{co}(\lambda(r_n), \lambda(r_{n+1})).$$

If the alternative below holds

(alter-1) $r_n \le r_{n+1}$, for some n

then, by the above relation

$$\varphi(w^n) \le 0;\ \text{in contradiction with } \varphi = \text{strictly positive}.$$

Hence, necessarily,

$$\text{(alter-2)}\ \ r_n > r_{n+1}, \forall n;\ \text{that is: } (r_n) \text{ is strictly descending}.$$

As a first consequence of this, one has that (iter-1) becomes

$$\text{(iter-2)}\ \ (\forall n):\ \psi(r_{n+1}) \le \psi(r_n) - \varphi(w^n),\ \text{where } w^n \in [\lambda(r_{n+1}), \lambda(r_n)].$$

As a second consequence of this, $r := \lim_n(r_n)$ exists in R_+. Suppose by contradiction that $r > 0$. As (r_n) is strictly descending,

$$r_n \to r+;\ \text{hence } \lambda(r_n) \to \lambda(r)+,\ w^n \to \lambda(r)+$$

(cf. the choice of $\lambda(.)$ and (w^n)). By (iter-2), we get

$$\text{(iter-3)}\ \ (0 <)\varphi(w^n) \leq \psi(r_n) - \psi(r_{n+1}), \forall n.$$

Passing to lim sup as $n \to \infty$, one gets

$$0 \leq \limsup_n \varphi(w^n) \leq \psi(r+0) - \psi(r+0) = 0;\ \text{that is: } \lim_n \varphi(w^n) = 0;$$

in contradiction with (ψ, φ) being right bounded oscillating at $(r, r; \lambda(r))$. Hence, $r = 0$; and the assertion follows.

Part 2. [Υ is descending asymptotic on $[X_0]$ under the second lot of conditions]. Let the sequence (r_n) in R_+^0 and the sequence (p_n) in R_+ be such that

$$(\forall n):\ A_n := (r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \in \mathcal{Q}(B_1 > 0; [X_0]) \cap \Upsilon,$$

$$\text{and } |p_n - r_n| \leq r_{n+1}.$$

As (u, v, w) is λ-iterative,

$$(\forall n):\ u^n := u(A_n) = r_{n+1},\ v^n := v(A_n) \leq \max\{r_n, r_{n+1}\},$$

$$w^n := w(A_n) \in \text{co}(\lambda(r_n), \lambda(r_{n+1})),$$

By the very representation of Υ (and ψ=increasing)

$$\text{(iter-4)}\ \ (\forall n):\ \psi(r_{n+1}) \leq \psi(v^n) - \varphi(w^n) \leq \psi(\max\{r_n, r_{n+1}\}) - \varphi(w^n),$$

$$\text{where } w^n \in \text{co}(\lambda(r_n), \lambda(r_{n+1})).$$

By the strict positive condition upon φ and ψ=increasing, this yields

$$(\forall n):\ \psi(r_{n+1}) < \psi(\max\{r_n, r_{n+1}\});\ \text{whence } r_{n+1} < r_n;$$

which tells us that (r_n) is strictly descending. As a first consequence of this, one has that (iter-4) becomes

$$\text{(iter-5)}\ \ (\forall n):\ \psi(r_{n+1}) \leq \psi(r_n) - \varphi(w^n),\ \text{where } w^n \in [\lambda(r_{n+1}), \lambda(r_n)].$$

As a second consequence of this, $r := \lim_n(r_n)$ exists in R_+. Suppose by contradiction that $r > 0$. As (r_n) is strictly descending,

$$r_n \to r+;\ \text{hence } \lambda(r_n) \to \lambda(r)+,\ w^n \to \lambda(r)+$$

(cf. the choice of $\lambda(.)$ and (w^n)). By (iter-5), we get

$$\text{(iter-6)}\ \ (0<)\varphi(w^n) \le \psi(r_n) - \psi(r_{n+1}), \forall n.$$

Passing to lim sup as $n \to \infty$, one gets

$$0 \le \limsup_n \varphi(w^n) \le \psi(r+0) - \psi(r+0) = 0; \textit{that is} : \lim_n \varphi(w^n) = 0;$$

in contradiction with (ψ, φ) being right bounded oscillating at $(r, r; \lambda(r))$. Hence, $r = 0$; and the assertion follows.

Part 3. [Υ is almost nright on $[X_0]$, under the precise conditions].

Take some point $c = (c_1, \ldots, c_6)$ in R^6_+. We say that the sequence $(t^n := (t^n_1, \ldots, t^n_6); n \ge 0)$ in R^6_+ is *right* at c, if

$$\text{(r-c)}\ \ (t^n_i \to c_i, \forall i) \text{ and } (t^n_i \to c_i+, \text{ whenever } c_i > 0).$$

Given $b > 0$, let us say that Υ is *nright* at b on $[X_0]$, if

for each sequence $(t^n; n \ge 0)$ in $\mathscr{Q}(B_3 > 0; [X_0]) \cap \Upsilon$,
the right property at $(0, b, b, b, b, 0)$ is not true.

The class of all these $b > 0$ will be denoted as $\text{nright}(\Upsilon; [X_0])$. In this case, we say that Υ is

(a-n-r) *almost nright* on $[X_0]$, if $\Theta := \text{nright}(\Upsilon; [X_0])$ is $(>)$-cofinal in R^0_+ (for each $\varepsilon \in R^0_+$ there exists $\theta \in \Theta$ with $\varepsilon > \theta$)

(n-r) *nright* on $[X_0]$, if $\Theta := \text{nright}(\Upsilon; [X_0])$ is identical with R^0_+.

We have to establish that the former property is retainable for our data. To do this, we start by noting that

$$\psi = \text{almost right regulated implies } \Theta := \text{rreg}(\psi) \text{ is } (>)\text{-cofinal in } R^0_+.$$

We now claim that

$$\Theta \subseteq \text{nright}(\Upsilon; [X_0]);\ \text{wherefrom, nright}(\Upsilon; [X_0]) \text{ is } (>)\text{-cofinal in } R^0_+.$$

This amounts to establish that

for each sequence $(t^n := (t^n_1, \ldots, t^n_6); n \ge 0)$ in $\mathscr{Q}(B_3 > 0; [X_0]) \cap \Upsilon$,
and each $b \in \Theta$, the right property at $(0, b, b, b, b, 0)$ is not true.

Suppose by absurd that there exists a (vectorial) sequence $(t^n := (t^n_1, \ldots, t^n_6); n \ge 0)$ in R^6_+ and some $b \in \Theta$, with

(right-b-1) $(\forall n)$: $t^n := (t^n_1, \ldots, t^n_6) \in \mathscr{Q}(B_3 > 0; [X_0]) \cap \Upsilon$; whence, $\psi(u(t^n)) \le \psi(v(t^n)) - \varphi(w(t^n))$

(right-b-2) $(t^n := (t_1^n, \dots, t_6^n); n \geq 0)$ has the right property at $(0, b, b, b, b, 0)$; i.e.: $(t_i^n \to 0, \forall i \in \{1, 6\})$, and $(t_i^n \to b+, \forall i \in \{2, 3, 4, 5\})$.

By the former of these properties

$$(\forall n) : \ (0 <)\varphi(w(t^n)) \leq \psi(v(t^n)) - \psi(u(t^n)).$$

As (u, v, w) is right asymptotic of type (I, I, μ),

$$u(t^n) \to b+, v(t^n) \to b+, w(t^n) \to \mu(b) + .$$

Passing to lim sup as $n \to \infty$ in this relation gives (by φ=strictly positive and ψ=right regular at b)

$$0 \leq \limsup_n \varphi(w(t^n)) \leq \psi(b + 0) - \psi(b + 0) = 0; \text{ that is: } \lim_n \varphi(w(t^n)) = 0;$$

in contradiction with (ψ, φ)=right bounded oscillating at $(b, b; \mu(b))$. Hence, our working assumption cannot be true; and the assertion follows.

Part 4. [Υ is npoint on $[[X_0]]$ under the described assumptions].

Take some point $c = (c_1, \dots, c_6)$ in R_+^6. We say that the sequence $(t^n := (t_1^n, \dots, t_6^n); n \geq 0)$ in R_+^6 is *point* at c, if

$$\text{(pt-c)} \ \ (t_i^n \to c_i, \forall i) \text{ and } [t_6^n = c_6, \forall n].$$

Given $b > 0$, let us say that Υ is *npoint* at b on $[[X_0]]$, if

(npoint) for each sequence $(t^n; n \geq 0)$ in $\mathcal{Q}(B_3 > 0; [[X_0]]) \cap \Upsilon$, the 6-point property at $(0, 0, b, 0, b, b)$ is not true.

The class of all these $b > 0$ will be denoted as npoint$(\Upsilon; [[X_0]])$. In this case, we say that Υ is

(a-n-p) *almost npoint* on $[[X_0]]$, if $\Theta :=$ npoint$(\Upsilon; [[X_0]])$ is $(>)$-cofinal in R_+^0

(n-p) *npoint* on $[[X_0]]$, if $\Theta :=$ npoint$(\Upsilon; [[X_0]])$ is identical with R_+^0.

We have to establish that the latter of these properties holds; that is,

for each sequence $(t^n; n \geq 0)$ in R_+^6 and each $b > 0$, the point property at $(0, 0, b, 0, b, b)$ is not true.

Assume by contradiction that this assertion is false: there exists a (vectorial) sequence $(t^n = (t_1^n, \dots, t_6^n); n \geq 0)$ in R_+^6 and some $b > 0$, with

(point-b-1) $(\forall n)$: $t^n = (t_1^n, \dots, t_6^n) \in \mathcal{Q}(B_3 > 0; [[X_0]]) \cap \Upsilon$; whence, $\psi(u(t^n)) \leq \psi(v(t^n)) - \varphi(w(t^n))$

(point-b-2) $(t^n = (t_1^n, \dots, t_6^n); n \geq 0)$ is point at $(0, 0, b, 0, b, b)$; that is: $(t_i^n \to 0, \forall i \in \{1, 2, 4\})$, $(t_i^n \to b, \forall i \in \{3, 5, 6\})$, and $(t_6^n = b, \forall n)$.

By the former of these properties

$$(\forall n):\ \varphi(w(t^n)) \le \psi(v(t^n)) - \psi(u(t^n)).$$

As (u, v, w) is point asymptotic of type (ξ, η, ζ)

$$u(t^n) \to \xi(b),\ v(t^n) \to \eta(b),\ w(t^n) \to \zeta(b).$$

Taking an auxiliary fact into account, we have that the obtained relations are impossible via (ψ, φ)=bounded oscillating at $(\eta(b), \xi(b); \zeta(b))$. Consequently, the working assumption is not acceptable; and conclusion follows.

Part 5. [Υ is normal under the posed condition].

Remember that, this property means:

$$(0, a, a, a, a, 0) \in \Upsilon \text{ is impossible, for each } a > 0.$$

Suppose by contradiction that

$$\text{there exists } a > 0 \text{ with } (0, a, a, a, a, 0) \in \Upsilon.$$

In view of (u, v, w)=invariant,

$$u(0, a, a, a, a, 0) = a,\ v(0, a, a, a, a, 0) = a,\ \ w(0, a, a, a, a, 0) = a.$$

This firstly means

$$(0, a, a, a, a, 0) \in \Upsilon_0 \text{ (because } a > 0).$$

Secondly, by the complete definition of Υ, we derive

$$\psi(u(0, a, a, a, a, 0)) \le \psi(v(0, a, a, a, a, 0)) - \varphi(w(0, a, a, a, a, 0)).$$

So, again combining with (u, v, w)=invariant, we must have

$$\psi(a) \le \psi(a) - \varphi(a);\ \text{ that is: } \varphi(a) \le 0;$$

absurd, by φ=strictly positive; and our claim follows.

Now, by simply combining the obtained fact with Theorem 4, one gets the following couple of fixed point statements with a practical value. Let (ψ, φ) be a pair of functions over $\mathscr{F}(R_+^0, R)$. Further, let (u, v, w) be a triple of functions over $\mathscr{F}(R_+^6, R_+)$, and define the couple of subsets

$$\Upsilon_0 = \{t \in R_+^6;\ u(t), v(t), w(t) > 0\},$$
$$\Upsilon = \{t \in \Upsilon_0;\ \psi(u(t)) \le \psi(v(t)) - \varphi(w(t))\}.$$

Finally, define the couple of relations over X

$$\mathscr{R}_0 = \mathscr{Q}^{-1}(\Upsilon_0),\ \mathscr{R}^* = \mathscr{Q}^{-1}(\Upsilon).$$

The contractive property to be used here is

(fct-contr) T is $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-$contractive$, if

$$\psi(u \circ \mathscr{Q}(x, y)) \leq \psi(v \circ \mathscr{Q}(x, y)) - \varphi(w \circ \mathscr{Q}(x, y)),\ \text{when } (x, y) \in \mathscr{R}.$$

On the other hand, by the constructions above, one may consider the attached set contractive condition

(set-contr) T is $(d, \mathscr{R}; \Upsilon)$-contractive: $\mathscr{Q}(x, y) \in \Upsilon$, if $(x, y) \in \mathscr{R}$.

The connection between these conditions is expressed as

(equi) T is $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive iff T is $(d, \mathscr{R}; \Upsilon)$-contractive.

In other words: our initial contractive condition in terms of $\mathscr{R}$, (ψ, φ) and (u, v, w) is equivalent with a contractive condition in terms of $\mathscr{R}$ and Υ.

The former of these statements (referred to as: Rhoades-Dutta-Choudhury principle for regulated functions; in short: (RDC-reg)) is based on a regulated condition upon ψ; note that, in this case, ψ need not be increasing.

Theorem 6 *Let the selfmap T be $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive, where (ψ, φ) is a couple of functions over $\mathscr{F}(R_+^0, R)$, with*

(62-i) ψ is extended regulated and φ is strictly positive,

and (u, v, w) is a triple of functions over $\mathscr{F}(R_+^6, R_+)$, such that (under the notations we just proposed)

(62-ii) $\mathscr{R}$ is B_1-admissible: $(B_1 > 0) \subseteq \mathscr{R}$.

Further, let X be (Ba-o,d)-complete; and let $X_0 = (x_n)$ be a (Ba-o) sequence. Then,

(62-a) *$X_0 = (x_n)$ is semi-Picard (modulo $(d, \mathscr{R}; T)$), provided there exists a continuous strictly increasing $\lambda \in \mathscr{F}(R_+^0)$, with*

(62-a-i) (u, v, w) is strongly λ-iterative

(62-a-ii) (ψ, φ) is right bounded oscillating at $(b, b; \lambda(b))$, for each $b > 0$

(62-b) *$X_0 = (x_n)$ is Picard (modulo $(d, \mathscr{R}; T)$), provided (in addition to the above) there exists $\mu \in \mathscr{F}(R_+)$, such that*

(62-b-i) the associated triple (u, v, w) is right asymptotic of type (I, I, μ)

(62-b-ii) (ψ, φ) is right bounded oscillating at $(b, b; \mu(b))$, for each $b > 0$

(62-c) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever T is (Ba-o,d)-continuous*

(62-d) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever (in addition to the above) there exists a triple (ξ, η, ζ) with $\xi \in \mathscr{F}(R_+^0)$, $\eta, \zeta \in \mathscr{F}(R_+)$, such that*

(62-d-i) *the triple (u, v, w) is point asymptotic of type (ξ, η, ζ)*
(62-d-ii) *(ψ, φ) is bounded oscillating at $(\eta(b), \xi(b); \zeta(b))$, for each $b > 0$*
(62-e) *T is fix-$\mathscr{R}$-asingleton (hence, fix-$\mathscr{R}$-singleton, under any of these extra requirements) when, in addition to the conditions above, the triple (u, v, w) is invariant*

$$(u(0, a, a, a, a, 0) = v(0, a, a, a, a, 0) = w(0, a, a, a, a, 0) = a, \forall a > 0)$$

(62-f) *T is fix-asingleton (hence, fix-singleton, under any of these extra requirements) when, in addition to the conditions above,*

$$\mathscr{R} \text{ is } B_0\text{-admissible and the triple } (u, v, w) \text{ is invariant.}$$

The latter of these (referred to as: Rhoades-Dutta-Choudhury principle for increasing functions; in short: (RDC-incr)) is based on an increasing condition upon the ambient function ψ.

Theorem 7 *Let the selfmap T be $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive, where (ψ, φ) is a couple of functions over $\mathscr{F}(R_+^0, R)$, with*

(63-i) *ψ is increasing and φ is strictly positive*

and (u, v, w) is a triple of functions over $\mathscr{F}(R_+^6, R_+)$, such that (under the notations we just proposed)

(63-ii) *$\mathscr{R}$ is B_1-admissible: $(B_1 > 0) \subseteq \mathscr{R}$.*

Further, let X be (Ba-o,d)-complete; and let $X_0 = (x_n)$ be a (Ba-o) sequence. Then,

(63-a) *$X_0 = (x_n)$ is semi-Picard (modulo $(d, \mathscr{R}; T)$), provided there exists a continuous strictly increasing $\lambda \in \mathscr{F}(R_+^0)$, with*
(63-a-i) *(u, v, w) is weakly λ-iterative*
(63-a-ii) *(ψ, φ) is right bounded oscillating at $(b, b; \lambda(b))$, for each $b > 0$*
(63-b) *$X_0 = (x_n)$ is Picard (modulo $(d, \mathscr{R}; T)$), provided (in addition to the above) there exists $\mu \in \mathscr{F}(R_+)$, such that*
(63-b-i) *the associated triple (u, v, w) is right asymptotic of type (I, I, μ)*
(63-b-ii) *(ψ, φ) is right bounded oscillating at $(b, b; \mu(b))$, for each $b > 0$*
(63-c) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever T is (Ba-o,d)-continuous*
(63-d) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever (in addition to the above) there exists a triple (ξ, η, ζ) with $\xi \in \mathscr{F}(R_+^0)$, $\eta, \zeta \in \mathscr{F}(R_+)$, such that*
(63-d-i) *the triple (u, v, w) is point asymptotic of type (ξ, η, ζ)*
(63-d-ii) *(ψ, φ) is bounded oscillating at $(\eta(b), \xi(b); \zeta(b))$, for each $b > 0$*
(63-e) *T is fix-$\mathscr{R}$-asingleton (hence, fix-$\mathscr{R}$-singleton, under any of these extra requirements) when, in addition to the conditions above, the triple (u, v, w) is invariant*

$$(u(0, a, a, a, a, 0) = v(0, a, a, a, a, 0) = w(0, a, a, a, a, 0) = a, \forall a > 0)$$

(63-f) *T is fix-asingleton (hence, fix-singleton, under any of these extra requirements) when, in addition to the conditions above,*

$$\mathscr{R} \text{ is } B_0\text{-admissible and the triple } (u, v, w) \text{ is invariant.}$$

As already precise, an extended setting of these developments is possible, under the lines described in the 2001 PhD Thesis by Hitzler [13, Ch 1, Sect 1.2]; see also Turinici [41]. Further aspects will be discussed elsewhere.

7 Chandok-Choudhury Approach

Let (X, d) be a metric space. Further, let $T \in \mathscr{F}(X)$ be a selfmap of X. As precise, we are interested to determine sufficient conditions for (uniqueness and) existence of elements in $\mathrm{Fix}(T)$, via contractive type requirements involving iterative processes $X_0 = (x_n)$ starting from an element x_0 of X, and their associated sets

$$[X_0] = \{x_n; n \geq 0\}, [[X_0]] = \mathrm{cl}([X_0]).$$

Let (ψ, φ) be a pair of functions over $\mathscr{F}(R_+^0, R)$. Further, let (u, v, w) be a triple of functions over $\mathscr{F}(R_+^6, R_+)$, and define the couple of subsets

$$\Upsilon_0 = \{t \in R_+^6; u(t), v(t), w(t) > 0\},$$
$$\Upsilon = \{t \in \Upsilon_0; \psi(u(t)) \leq \psi(v(t)) - \varphi(w(t))\}.$$

Finally, define the couple of relations over X

$$\mathscr{R} = \mathscr{Q}^{-1}(\Upsilon_0), \ \mathscr{R}^* = \mathscr{Q}^{-1}(\Upsilon).$$

The starting functional contractive property to be considered is

(fct-contr) T is $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-$contractive$, if

$$\psi(u \circ \mathscr{Q}(x, y)) \leq \psi(v \circ \mathscr{Q}(x, y)) - \varphi(w \circ \mathscr{Q}(x, y)), \text{ when } (x, y) \in \mathscr{R}.$$

On the other hand, the attached set contractive condition is

(set-contr) T is $(d, \mathscr{R}; \Upsilon)$-contractive: $\mathscr{Q}(x, y) \in \Upsilon$, if $(x, y) \in \mathscr{R}$.

As precise, the connection between these conditions writes

(equi) T is $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive iff T is $(d, \mathscr{R}; \Upsilon)$-contractive.

For the set contractive condition, sufficient conditions were given upon $\mathscr{R}$ and Υ so that the main result be applicable; these, naturally, are ultimately expressed in terms of $\mathscr{R}$, (ψ, φ) and (u, v, w). A by-product of these developments is the couple of fixed point results we just exposed. It is our aim in the following to discuss a lot of particular cases of these, with a practical meaning.

(A) For the moment, some direct constructions will be proposed.

Constr-0) Define the auxiliary system of functions $(e_0, e_1, e_2, e_3, e_4)$ over the class $\mathscr{F}(R_+^6, R_+)$, as: for each $t = (t_1, \ldots, t_6) \in R_+^6$,

$$e_0(t) = t_6(1 + t_4)/(1 + t_3),$$
$$e_1(t) = t_6(1 + t_1)/(1 + t_2), e_2(t) = t_4(1 + t_3)/(1 + t_2),$$
$$e_3(t) = \max\{t_1, t_6\}, e_4(t) = (1/2)(t_3 + t_4).$$

Constr-1) Further, define the triple (u_1, u_2, u_3) over the class $\mathscr{F}(R_+^6, R_+)$, as: for each $t = (t_1, \ldots, t_6) \in R_+^6$,

$$\text{(u-def)}\ \ u_1(t) = \max\{t_5, e_0(t)\}, u_2(t) = \max\{t_5, t_6\}, u_3(t) = t_5.$$

Constr-2) Then, let us construct the triple of functions (v_1, v_2, v_3) over the class $\mathscr{F}(R_+^6, R_+)$, as: for each $t = (t_1, \ldots, t_6) \in R_+^6$,

$$\text{(v-def)}\ \ v_1(t) = \max\{t_2, e_1(t), e_2(t)\}, v_2(t) = \max\{t_2, e_3(t), e_4(t)\},$$
$$v_3(t) = t_2.$$

Constr-3) Finally, define the triple (w_1, w_2, w_3) of functions over $\mathscr{F}(R_+^6, R_+)$, as: for each $t = (t_1, \ldots, t_6) \in R_+^6$,

$$\text{(w-def)}\ \ w_1(t) = \max\{t_2, e_1(t)\}, w_2(t) = \max\{t_2, t_6\}, w_3(t) = t_2.$$

Remember that, for the above explained technical reasons, we are forced to accept a regularity condition like

$$\mathscr{R} \text{ is } B_1\text{-admissible: } (B_1 > 0) \subseteq \mathscr{R}.$$

It is our aim to discuss of to what extent are the triples (u, v, w), where

$$u \in \{u_1, u_2, u_3\}, v \in \{v_1, v_2, v_3\}, w \in \{w_1, w_2, w_3\}$$

compatible with the regularity conditions encountered in the couple of fixed point statements above. The response to this question is to be precise as below.

R-1) Given the sequence (r_n) in R_+^0 and the sequence (p_n) in R_+ with $(A_n := (r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \in \mathscr{Q}(B_1 > 0; [X_0]) \cap \Upsilon$, $|p_n - r_n| \leq r_{n+1}$, for all n), we have (by these definitions)

$$(\forall n):\ e_0(A_n) = r_{n+1}/(1+p_n)\ (\leq r_{n+1}),\ e_1(A_n) = r_{n+1},\ e_2(A_n) = 0,$$
$$e_3(A_n) = \max\{r_n, r_{n+1}\},\ e_4(A_n) = (1/2)p_n\ (\leq \max\{r_n, r_{n+1}\}).$$

This, along with our previous conventions, gives

$$\text{(u-iter)}\ \ u_1(A_n) = r_{n+1},\ u_2(A_n) = r_{n+1},\ u_3(A_n) = r_{n+1},$$
$$\text{(v-iter)}\ \ v_1(A_n) = \max\{r_n, r_{n+1}\},\ v_2(A_n) = \max\{r_n, r_{n+1}\},\ v_3(A_n) = r_n,$$
$$\text{(w-iter)}\ \ w_1(A_n) = \max\{r_n, r_{n+1}\},\ w_2(A_n) = \max\{r_n, r_{n+1}\},\ w_3(A_n) = r_n.$$

As a consequence of this, we have that

(s-iter) any triple (u, v, w), where $u \in \{u_1, u_2, u_3\}$, $v \in \{v_1, v_2\}$, $w \in \{w_1, w_2, w_3\}$ is strongly I-iterative

(iter) any triple (u, v_3, w), where $u \in \{u_1, u_2, u_3\}$, $w \in \{w_1, w_2, w_3\}$ is weakly I-iterative.

R-2) Let the sequence $(t^n = (t_1^n, \ldots, t_6^n); n \geq 0)$ in $\mathcal{Q}(B_3 > 0; [X_0]) \cap \Upsilon$, and the point $b > 0$ be such that

$$(t^n = (t_1^n, \ldots, t_6^n); n \geq 0) \text{ has the right property at } (0, b, b, b, b, 0);$$
$$\text{that is: } (t_i^n \to 0, \forall i \in \{1, 6\}), (t_j^n \to b+, \forall j \in \{2, 3, 4, 5\}).$$

Then, by definition,

$$e_0(t^n) = t_6^n(1+t_4^n)/(1+t_3^n) \to 0,$$
$$e_1(t^n) = t_6^n(1+t_1^n)/(1+t_2^n) \to 0,\ e_2(t^n) = t_4^n(1+t_3^n)/(1+t_2^n) \to b;$$
$$e_3(t^n) = \max\{t_1^n, t_6^n\} \to 0,\ e_4(t^n) = (1/2)[t_3^n + t_4^n] \to b.$$

This, along with our preceding conventions, yields

$$\text{(u-pos)}\ u_1(t^n) = \max\{t_5^n, e_0(t^n)\} \to b+,$$
$$u_2(t^n) = \max\{t_5^n, t_6^n\} \to b+,\ u_3(t^n) = t_5^n \to b+,$$
$$\text{(v-pos)}\ \ v_1(t^n) = \max\{t_2^n, e_1(t^n), e_2(t^n)\} \to b+,$$
$$v_2(t^n) = \max\{t_2^n, e_3(t^n), e_4(t^n)\} \to b+,\ v_3(t^n) = t_2^n \to b+,$$
$$\text{(w-pos)}\ \ w_1(t^n) = \max\{t_2^n, e_1(t^n)\} \to b+,$$
$$w_2(t^n) = \max\{t_2^n, t_6^n\} \to b+,\ w_3(t) = t_2^n \to b+.$$

Here, the convergence relations involving maximum type functions are obtained as

$$(u_1(t^n) \geq t_5^n, \forall n) \text{ and } u_1(t^n) \to b, t_5^n \to b+ \text{ imply } u_1(t^n) \to b+;$$

$$(u_2(t^n) \geq t_5^n, \forall n) \text{ and } u_2(t^n) \to b, t_5^n \to b+ \text{ imply } u_2(t^n) \to b+;$$
$$(v_1(t^n) \geq t_2^n, \forall n) \text{ and } v_1(t^n) \to b, t_2^n \to b+ \text{ imply } v_1(t^n) \to b+;$$
$$(v_2(t^n) \geq t_2^n, \forall n) \text{ and } v_2(t^n) \to b, t_2^n \to b+ \text{ imply } v_2(t^n) \to b+;$$
$$(w_1(t^n) \geq t_2^n, \forall n) \text{ and } w_1(t^n) \to b, t_2^n \to b+ \text{ imply } w_1(t^n) \to b+;$$
$$(w_2(t^n) \geq t_2^n, \forall n) \text{ and } w_2(t^n) \to b, t_2^n \to b+ \text{ imply } w_2(t^n) \to b+.$$

As a consequence of this, we have that

(r-I-I) any triple (u, v, w), where $u \in \{u_1, u_2, u_3\}$, $v \in \{v_1, v_2, v_3\}$, $w \in \{w_1, w_2, w_3\}$ is right asymptotic of type (I, I, I).

R-3) Let the sequence $(t^n = (t_1^n, \ldots, t_6^n); n \geq 0)$ in $\mathscr{Q}(B_3 > 0; [[X_0]]) \cap \Upsilon$, and the point $b > 0$ be such that

$$(t^n = (t_1^n, \ldots, t_6^n); n \geq 0) \text{ has the point property at } (0, 0, b, 0, b, b);$$
$$\text{that is: } (t_i^n \to 0, \forall i \in \{1, 2, 4\}), (t_j^n \to b, \forall j \in \{3, 5, 6\}), (t_6^n = b, \forall n).$$

Then, by definition,

$$e_0(t^n) = t_6^n(1 + t_4^n)/(1 + t_3^n) \to b/(1 + b),$$
$$e_1(t^n) = t_6^n(1 + t_1^n)/(1 + t_2^n) \to b,\ e_2(t^n) = t_4^n(1 + t_3^n)/(1 + t_2^n) \to 0,$$
$$e_3(t^n) = \max\{t_1^n, t_6^n\} \to b,\ e_4(t^n) = (1/2)[t_3^n + t_4^n] \to b/2.$$

This, along with our preceding conventions, yields

$$\text{(u-pt) } u_1(t^n) = \max\{t_5^n, e_0(t^n)\} \to b,$$
$$u_2(t^n) = \max\{t_5^n, t_6^n\} \to b,\ u_3(t^n) = t_5^n \to b,$$
$$\text{(v-pt) } v_1(t^n) = \max\{t_2^n, e_1(t^n), e_2(t^n)\} \to b,$$
$$v_2(t) = \max\{t_2^n, e_3(t^n), e_4(t^n)\} \to b,\ v_3(t^n) = t_2^n \to 0$$
$$\text{(w-pt) } w_1(t^n) = \max\{t_2^n, e_1(t^n)\} \to b,$$
$$w_2(t^n) = \max\{t_2^n, t_6^n\} \to b, w_3(t) = t_2^n \to 0.$$

As a consequence of this, we have that

(pt-I-I-I) any triple (u, v, w), where $u \in \{u_1, u_2, u_3\}$, $v \in \{v_1, v_2\}$, $w \in \{w_1, w_2\}$, is point asymptotic of type (I, I, I)

(pt-I-I-0) any triple (u, v, w_3), where $u \in \{u_1, u_2, u_3\}$, $v \in \{v_1, v_2\}$, is point asymptotic of type $(I, I, 0)$

(pt-I-0-I) any triple (u, v_3, w), where $u \in \{u_1, u_2, u_3\}$, $w \in \{w_1, w_2\}$, is point asymptotic of type $(I, 0, I)$

(pt-I-0-0) any triple (u, v_3, w_3), where $u \in \{u_1, u_2, u_3\}$, is point asymptotic of type $(I, 0, 0)$.

R-4) Let $a > 0$ be arbitrary fixed; and put $A = (0, a, a, a, a, 0)$. By the very definition of these functions,

$$e_0(A) = 0, e_1(A) = 0, e_2(A) = a,\ e_3(A) = 0, e_4(A) = a.$$

This, along with our preceding conventions, yields

$$\text{(u-inv)}\ \ u_1(A) = a, u_2(A) = a, u_3(A) = a,$$

$$\text{(v-inv)}\ \ v_1(A) = a, v_2(A) = a, v_3(A) = a$$

$$\text{(w-inv)}\ \ w_1(A) = a, w_2(A) = a, w_3(A) = a.$$

As a consequence of this, we have that

(inva) any triple (u, v, w), where $u \in \{u_1, u_2, u_3\}$, $v \in \{v_1, v_2, v_3\}$, $w \in \{w_1, w_2, w_3\}$ is invariant.

As a direct consequence of these facts and Rhoades-Dutta-Choudhury principles we just exposed, one may now derive a couple of fixed point statements with methodological value.

The former of these, starting from Rhoades-Dutta-Choudhury principle for regulated functions (RDC-reg) is based on a regulated condition upon ψ; note that, in this case, ψ need not be increasing.

Theorem 8 *Let the selfmap T be $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive, where (ψ, φ) is a couple of functions over $\mathscr{F}(R_+^0, R)$, with*

(71-i) *ψ is extended regulated and φ is strictly positive*

(71–ii) *(ψ, φ) is right bounded oscillating at $(b, b; b)$, for each $b > 0$,*

and (u, v, w) is a triple of functions over $\mathscr{F}(R_+^6, R_+)$, with

(71-iii) *$u \in \{u_1, u_2, u_3\}$, $v \in \{v_1, v_2\}$, $w \in \{w_1, w_2, w_3\}$,*

such that (under the notations we just proposed)

(71-iv) *$\mathscr{R}$ is B_1-admissible.*

Further, let X be (Ba-o,d)-complete; and let $X_0 = (x_n)$ be a (Ba-o) sequence. Then,

(71-a) *$X_0 = (x_n)$ is Picard (modulo $(d, \mathscr{R}; T)$)*

(71-b) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever T is (Ba-o,d)-continuous*

(71-c) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever (in addition to the above) one of the alternatives below is holding*

(71-c-i) $v \neq v_3$, $w \neq w_3$, (ψ, φ)=*bounded oscillating at* $(b, b; b)$, $\forall b > 0$

(71-c-ii) $v \neq v_3$, $w = w_3$, (ψ, φ)=*bounded oscillating at* $(b, b; 0)$, $\forall b > 0$

(71-c-iii) $v = v_3$, $w \neq w_3$, (ψ, φ)=*bounded oscillating at* $(0, b; b)$, $\forall b > 0$

(71-c-iv) $v = v_3$, $w = w_3$, (ψ, φ)=*bounded oscillating at* $(0, b; 0)$, $\forall b > 0$

(71-d) *T is fix-$\mathscr{R}$-asingleton (hence, necessarily, fix-$\mathscr{R}$-singleton, under any of these extra requirements)*

(71-e) *T is fix-asingleton (hence, fix-singleton, under any of these extra requirements) when, in addition to the conditions above,*

$$\mathscr{R} \text{ is } B_0\text{-admissible}.$$

The latter of these, starting from Rhoades-Dutta-Choudhury principle for increasing functions (RDC-incr) is based on an increasing assumption upon ψ. Note that, in this setting, we have

$(\forall i, j, k \in \{1, 2, 3\})$: T is $(d, \mathscr{R}; \psi, \varphi; u_i, v_j, w_k)$-contractive implies
T is $(d, \mathscr{R}; \psi, \varphi; u_3, v_j, w_k)$-contractive.

Theorem 9 *Let the selfmap T be $(d, \mathscr{R}; \psi, \varphi; u_3, v, w)$-contractive, where (ψ, φ) is a couple of functions over $\mathscr{F}(R_+^0, R)$, with*

(72-i) *ψ is increasing and φ is strictly positive*

(72-ii) *(ψ, φ) is right bounded oscillating at $(b, b; b)$, for each $b > 0$,*

and v, w are functions over $\mathscr{F}(R_+^6, R_+)$, with

(72-iii) $v \in \{v_1, v_2, v_3\}$, $w \in \{w_1, w_2, w_3\}$,

such that (under the notations we just proposed)

(72-iv) *$\mathscr{R}$ is B_1-admissible.*

Further, let X be (Ba-o,d)-complete; and let $X_0 = (x_n)$ be a (Ba-o) sequence. Then,

(72-a) *$X_0 = (x_n)$ is Picard (modulo $(d, \mathscr{R}; T)$)*

(72-b) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever T is (Ba-o,d)-continuous*

(72-c) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$) whenever (in addition to the above) one of the alternatives below is holding*

(72-c-i) $v \neq v_3$, $w \neq w_3$, (ψ, φ)=*bounded oscillating at* $(b, b; b)$, $\forall b > 0$

(72-c-ii) $v \neq v_3$, $w = w_3$, (ψ, φ)=*bounded oscillating at* $(b, b; 0)$, $\forall b > 0$

(72-c-iii) $v = v_3$, $w \in \{w_1, w_2, w_3\}$

(72-d) *T is fix-$\mathscr{R}$-asingleton (hence, necessarily, fix-$\mathscr{R}$-singleton, under any of these extra requirements)*

(72-e) *T is fix-asingleton (hence, fix-singleton, under any of these extra requirements) when, in addition to the conditions above,*

$$\mathscr{R} \text{ is } B_0\text{-admissible}.$$

Proof The only point to be clarified is that of the listed alternative related to $v = v_3$ being sufficient for deriving the strongly Picard property. To do this, the same reasonings as the ones in the main result will be applied.

From the Picard property, we have, as X is (Ba-o,d)-complete

$$X_0 = (x_n) \text{ is } d\text{-convergent: } x_n \xrightarrow{d} z_0 \text{ as } n \to \infty, \text{ for some } z_0 \in X.$$

We now claim that z_0 is a fixed point of T. Three alternatives are to be discussed.

Alter-1) Suppose that

$$\text{(Tz-rela-1)} \quad H_1 := \{n \in N; x_n = Tz_0\} \text{ is unbounded (in } N).$$

By the same procedure as the one in the main result, we derive $z_0 \in \text{Fix}(T)$.

Alter-2) Suppose that

$$\text{(Tz-rela-2)} \quad H_2 := \{n \in N; Tx_n = Tz_0\} \text{ is unbounded (in } N).$$

By the same procedure as the one in the main result, we derive $z_0 \in \text{Fix}(T)$.

Alter-3) Suppose that

both subsets H_1 and H_2 are bounded (in N).

This tells us that

$$\exists i = i(z_0) \in N, \text{ such that:}$$
$$n \geq i \text{ implies } Tx_n \neq Tz_0 \text{ (hence, } x_n \neq z_0) \text{ and } x_n \neq Tz_0.$$

Denote for simplicity $(u_n = x_{n+i}; n \geq 0)$; clearly, by the preceding relation,

$$\text{(non-id)} \quad (\forall n): \ Tu_n \neq Tz_0 \text{ (hence, } u_n \neq z_0) \text{ and } u_n \neq Tz_0.$$

Again combining with the (Ba-o) property of $X_0 = (x_n)$, one derives

$$\text{(posi-1) } (\forall n): \ Q_1(u_n, z_0) = d(u_n, Tu_n) > 0,$$
$$Q_2(u_n, z_0) = d(u_n, z_0) > 0, \ Q_3(u_n, z_0) = d(u_n, Tz_0) > 0,$$
$$Q_4(u_n, z_0) = d(Tu_n, z_0) > 0, \ Q_5(u_n, z_0) = d(Tu_n, Tz_0) > 0.$$

Suppose by contradiction that

$$\text{(posi-2)} \quad b := d(z_0, Tz_0) > 0 \text{ [whence, } Q_6(u_n, z_0) = b > 0, \forall n].$$

From the preceding observations, we have

$$(\forall n): \ B_3(u_n, z_0) = \min\{Q_1, \ldots, Q_6\}(u_n, z_0)\} > 0;$$
$$\text{so that, } (u_n, z_0) \in \mathscr{R}.$$

Consequence, the function contractive condition involving the systems (ψ, φ) and (u_3, v_3, w) applies; and gives

$$(\forall n): \ \psi(d(Tu_n, Tz_0)) \leq \psi(d(u_n, z_0)) - \varphi(w(\mathscr{Q}(u_n, z_0)).$$

By the strict positivity assumption upon φ, one derives

$$(\forall n)\psi(d(Tu_n, Tz_0)) < \psi(d(u_n, z_0));$$
$$\text{whence } d(Tu_n, Tz_0) < d(u_n, z_0);$$

if we remember that ψ is increasing. Passing to limit as $n \to \infty$, we get (by a metrical property of d)

$$0 < d(z_0, Tz_0) \leq 0; \text{ a contradiction.}$$

Hence, our working assumption relative to z_0 cannot be accepted; and then, we have $b = 0$; that is: $z_0 = Tz_0$.

Some remarks are in order.

Rem-1) An extension of these statements to quasi-ordered spaces is immediate, by the developments in Turinici [43]. Further aspects will be delineated elsewhere.

Rem-2) Under the formulation above, Theorem 8 contains 27 fixed point statements. Among these, we have

(part-1) the (quasi-order) variant (u_3, v_1, w_1) of Theorem 8 is identical with the fixed point statement in Chandok et al. [6]

(part-2) the (quasi-order) variant (u_3, v_2, w_2) of Theorem 8 is identical with the fixed point statement in Choudhury et al. [7].

In order words: the quoted results admit extensions to regulated functions. Some other variants of Theorem 8 include a number of related fixed point results described in Radenović et al. [28]. But, most of these seem to be new, at least from the perspective of regulated (modulo ψ) setting.

Rem-3) Under the formulation above, Theorem 9 contains 9 fixed point statements. The variant (u_3, v_3, w_3) of this statement includes the results in Wardowski [46] and Secelean [34]; see also Vujaković et al. [44]. For a different proof of it, we refer to the paper by Turinici [42]. The remaining variants (u_3, v_3, w) of the same where $w \in \{w_1, w_2\}$, seem to be new.

Finally, note that the proposed examples do not exhaust all possible variants described in the Rhoades-Dutta-Choudhury principles (RDC-reg) and (RDC-incr); for, e.g., the ones concerning strongly/weakly λ-iterative properties with $\lambda \in \mathscr{F}(R_+^0)$ fulfilling $\lambda \neq I$ are not present in this list of particular cases. Further technical aspects of this case will be delineated elsewhere.

8 Cosentino-Vetro Contractions

Let (X, d) be a metric space. Further, let $T \in \mathscr{F}(X)$ be a selfmap of X. As precise, we are interested to determine sufficient conditions for (uniqueness and) existence of elements in $\mathrm{Fix}(T)$, via contractive type requirements involving iterative processes $X_0 = (x_n)$ starting from an element x_0 of X, and their associated sets

$$[X_0] = \{x_n; n \geq 0\}, [[X_0]] = \mathrm{cl}([X_0]).$$

Let (ψ, φ) be a pair of functions over $\mathscr{F}(R_+^0, R)$. Further, let (u, v, w) be a triple of functions over $\mathscr{F}(R_+^6, R_+)$, and define the couple of subsets

$$\Upsilon_0 = \{t \in R_+^6; u(t), v(t), w(t) > 0\},$$
$$\Upsilon = \{t \in \Upsilon_0; \psi(u(t)) \leq \psi(v(t)) - \varphi(w(t))\}.$$

Finally, define the couple of relations over X

$$\mathscr{R} = \mathscr{Q}^{-1}(\Upsilon_0),\ \mathscr{R}^* = \mathscr{Q}^{-1}(\Upsilon).$$

The functional and set contractive properties to be used here are

(fct-contr) T is $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-*contractive*, if

$\psi(u \circ \mathscr{Q}(x, y)) \leq \psi(v \circ \mathscr{Q}(x, y)) - \varphi(w \circ \mathscr{Q}(x, y))$, when $(x, y) \in \mathscr{R}$

(set-contr) T is $(d, \mathscr{R}; \Upsilon)$-contractive: $\mathscr{Q}(x, y) \in \Upsilon$, if $(x, y) \in \mathscr{R}$.

The connection between these conditions is expressed as

(equi) T is $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive iff T is $(d, \mathscr{R}; \Upsilon)$-contractive.

For the set contractive property, sufficient conditions were given upon $\mathscr{R}$ and Υ so that the main result be applicable; these, as expected, were ultimately expressed in terms of $\mathscr{R}$, (ψ, φ) and (u, v, w). The by-product of these developments is the couple of Rhoades-Dutta-Choudhury fixed point results (RDC-reg) and (RDC-incr). Some particular cases of them were just exposed. It is our aim in the following to discuss some other particular cases of these, with a practical meaning.

(A) Let us say that $v \in \mathscr{F}(R_+^6, R_+)$ is a *Cosentino-Vetro function*, when

(CV-0) v is increasing and continuous in its variables
(CV-1) v is $(1, 2, 5, 6)$-positive: $v(a, a, 0, 0, a, a) > 0$, for each $a > 0$
(CV-2) v is I-iterative: $v(a, a, 2a, 0, a, a) \leq a$, for each $a > 0$
(CV-3) v is invariant: $v(0, b, b, b, b, 0) = b$, for each $b > 0$.

Fix in the following such a function; and define $\eta \in \mathscr{F}(R_+^0, R_+)$ as

$$\text{(E-def) } \eta(b) := v(0, 0, b, 0, b, b), b > 0.$$

Note that, by (v=increasing continuous) and the iterative condition:

(E-1) $\eta(.)$ is increasing and continuous (on R_+^0)
(E-2) $0 \le \eta(b) \le v(b, b, 2b, 0, b, b) \le b, \forall b > 0$; whence, $\eta(0+0) = 0$.

Unfortunately, the extremal inequalities above are not in general reducible to equalities. Then, define the couple (u, w) of functions over the class $\mathscr{F}(R_+^6, R_+)$, as: for each $t = (t_1, \ldots, t_6) \in R_+^6$,

$$\text{(uw-def)} \quad u(t) = t_5, w(t) = t_2;$$

in this case, (u, v, w) will be referred to as a *Cosentino-Vetro triple*.

Having this precise, let us introduce the class of contractive conditions

(C-V) T is Cosentino-Vetro $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive:

$$\psi(d(Tx, Ty)) \le \psi(v(Q_1(x, y), \ldots, Q_6(x, y)) - \varphi(d(x, y)), \ \forall (x, y) \in \mathscr{R}.$$

As usual, we assume that (under the notations we just proposed)

$$\text{(B1-adm)} \quad \mathscr{R} \text{ is } B_1\text{-admissible: } (B_1 > 0) \subseteq \mathscr{R}.$$

In this case, the following contractive property holds

(C-V-B1) T is Cosentino-Vetro $(d, (B_1 > 0); \psi, \varphi; u, v, w)$-contractive:

$$\psi(d(Tx, Ty)) \le \psi(v(Q_1(x, y), \ldots, Q_6(x, y)) - \varphi(d(x, y)), \ \forall (x, y) \in (B_1 > 0).$$

We are interested of to what extent is the triple (u, v, w) compatible with the regularity conditions encountered in Rhoades-Dutta-Choudhury principle for increasing functions (RDC-incr). The response to this question is contained in the developments below.

R-1) Given the sequence (r_n) in R_+^0 and the sequence (p_n) in R_+ with $(A_n := (r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \in \mathscr{Q}(B_1 > 0; [X_0]) \cap \Upsilon, |p_n - r_n| \le r_{n+1}$, for all $n)$, we have (by these definitions)

$$(\forall n): \ u(A_n) = r_{n+1}, w(A_n) = r_n.$$

On the other hand, under the convention

$$\alpha_n = \max\{r_n, r_{n+1}\}, n \ge 0$$

we have, by the iterative property

$$(\forall n): \ v(A_n) = v(r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \le \ v(\alpha_n, \alpha_n, 2\alpha_n, 0, \alpha_n, \alpha_n) \le \alpha_n.$$

This tells us that

the triple (u, v, w) is weakly I-iterative (see above);

and explains the choosing of Rhoades-Dutta-Choudhury principle for increasing functions (RDC-incr) as an appropriate tool for solving our problem.

R-2) Let the sequence $(t^n = (t_1^n, \ldots, t_6^n); n \geq 0)$ in $\mathcal{Q}(B_3 > 0; [X_0]) \cap \Upsilon$, and the point $b > 0$ be such that

$(t^n = (t_1^n, \ldots, t_6^n); n \geq 0)$ has the right property at $(0, b, b, b, b, 0)$; that is:
$(t_i^n \to 0, \forall i \in \{1, 6\})$, $(t_j^n \to b+, \forall j \in \{2, 3, 4, 5\})$.

For the moment, we have by definition

$$u(t^n) = t_5^n \to b+, w(t^n) = t_2^n \to b+ .$$

On the other hand, under the convention

$$\beta_n = \min\{t_2^n, t_3^n, t_4^n, t_5^n\}, n \geq 0 \text{ [hence, } \beta_n \to b+ \text{ as } n \to \infty],$$

we have (under the posed hypotheses upon v)

$$v(t^n) \geq v(0, \beta_n, \beta_n, \beta_n, \beta_n, 0) = \beta_n > b, \text{ for all } n,$$
$$\text{and } v(t^n) \to v(0, b, b, b, b, 0) = b \text{ as } n \to \infty;$$
$$\text{wherefrom } v(t^n) \to b+ \text{ as } n \to \infty.$$

Putting these together, it follows that

(r-asy) the triple (u, v, w) is right asymptotic of type (I, I, I).

R-3) Let the sequence $(t^n = (t_1^n, \ldots, t_6^n); n \geq 0)$ in $\mathcal{Q}(B_3 > 0; [[X_0]]) \cap \Upsilon$, and the point $b > 0$ be such that

$(t^n = (t_1^n, \ldots, t_6^n); n \geq 0)$ has the point property at $(0, 0, b, 0, b, b)$;
that is: $(t_i^n \to 0, \forall i \in \{1, 2, 4\})$, $(t_j^n \to b, \forall j \in \{3, 5, 6\})$, $(t_6^n = b, \forall n)$.

Then, by definition,

$$u(t^n) = t_5^n \to b, v(t^n) \to v(0, 0, b, 0, b, b) = \eta(b), \ w(t^n) = t_2^n \to 0.$$

As a consequence of this, we have that

the triple (u, v, w) is point asymptotic of type $(I, \eta, 0)$.

R-4) Let $a > 0$ be arbitrary fixed; and put $A = (0, a, a, a, a, 0)$. By the very definition of these functions,

$$u(A) = a, v(A) = v(0, a, a, a, a, 0) = a, w(A) = a.$$

And then, we have that

the triple (u, v, w) is invariant.

As a consequence of Rhoades-Dutta-Choudhury principle for increasing functions (RDC-incr), we derive the following fixed point statement (referred to as: *Cosentino-Vetro fixed point principle*; in short: (CV-fpp)).

Theorem 10 *Let the selfmap T be Cosentino-Vetro $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive, where (ψ, φ) is a couple of functions over $\mathscr{F}(R_+^0, R)$, with*

(81-i) ψ *is increasing and* φ *is strictly positive*
(81-ii) (ψ, φ) *is right bounded oscillating at* $(b, b; b)$*, for each* $b > 0$

and (u, v, w) is a Cosentino-Vetro triple over $\mathscr{F}(R_+^6, R_+)$ such that (under the notations we just proposed)

(81-iii) $\mathscr{R}$ *is* B_1*-admissible:* $(B_1 > 0) \subseteq \mathscr{R}$.

Further, let X be (Ba-o,d)-complete; and let $X_0 = (x_n)$ be a (Ba-o) sequence. Then,

(81-a) $X_0 = (x_n)$ *is Picard (modulo* $(d, \mathscr{R}; T)$*)*
(81-b) $X_0 = (x_n)$ *is strongly Picard (modulo* $(d, \mathscr{R}; T)$*), whenever T is (Ba-o,d)-continuous*
(81-c) $X_0 = (x_n)$ *is strongly Picard (modulo* $(d, \mathscr{R}; T)$*), whenever (in addition to the above)*

$$(\psi, \varphi) \text{ is bounded oscillating at } (\eta(b), b; 0), \text{ for all } b > 0$$

(81-d) *T is fix-$\mathscr{R}$-asingleton (hence, necessarily, fix-$\mathscr{R}$-singleton, under any of these extra requirements)*
(81-6) *T is fix-asingleton (hence, fix-singleton, under any of these extra requirements) when, in addition to the conditions above,*

$$\mathscr{R} \text{ is } B_0\text{-admissible: } (B_0 > 0) \subseteq \mathscr{R}.$$

A basic particular case of these developments corresponds to the choice

$$v(t) = \alpha_1 t_1 + \ldots + \alpha_6 t_6, t = (t_1, \ldots, t_6) \in R_+^6,$$

where $(\alpha_1, \ldots, \alpha_6)$ is a vector in R_+^6 with the properties

$$\text{(Pro-1)} \ \alpha_1 + \alpha_2 + \alpha_5 + \alpha_6 > 0,$$
$$\text{(Pro-2)} \ \alpha_1 + \alpha_2 + 2\alpha_3 + \alpha_5 + \alpha_6 \leq 1,$$
$$\text{(Pro-3)} \ \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 = 1.$$

Denote further

$$\eta := \alpha_3 + \alpha_5 + \alpha_6; \text{ hence, } 0 \leq \eta \leq 1.$$

The contractive condition upon T becomes a linear contraction like

(CV-lin) T is linear Cosentino-Vetro $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive:

$$\psi(d(Tx, Ty)) \leq \psi(\alpha_1 Q_1(x, y) + \ldots + \alpha_6 Q_6(x, y)) - \varphi(d(x, y)),\ \forall (x, y) \in \mathscr{R}.$$

Theorem 11 *Let T be linear Cosentino-Vetro $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive, where (ψ, φ) is a couple of functions over $\mathscr{F}(R_+^0, R)$, with*

(82-i) *ψ is increasing and φ is strictly positive*
(82-ii) *(ψ, φ) is right bounded oscillating at $(b, b; b)$ for each $b > 0$,*

and the constant vector $\alpha = (\alpha_1, \ldots, \alpha_6)$ in R_+^6 fulfilling (Pro-1)-(Pro-3), be such that (under the notations we just proposed)

(B1-adm) *$\mathscr{R}$ is B_1-admissible: $(B_1 > 0) \subseteq \mathscr{R}$.*

Further, let X be (Ba-o,d)-complete; and let $X_0 = (x_n)$ be a (Ba-o) sequence. Then,

(82-a) *$X_0 = (x_n)$ is Picard (modulo $(d, \mathscr{R}; T)$)*
(82-b) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever T is (Ba-o,d)-continuous*
(82-c) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$), whenever (in addition to the above)*

$$(\psi, \varphi) \text{ is bounded oscillating at } (\eta b, b; 0), \text{ for all } b > 0$$

(82-d) *T is fix-$\mathscr{R}$-asingleton (hence, necessarily, fix-$\mathscr{R}$-singleton, under any of these extra requirements)*
(82-e) *T is fix-asingleton (hence, fix-singleton, under any of these extra requirements) when, in addition to the conditions above,*

$$\mathscr{R} \text{ is } B_0\text{-admissible: } (B_0 > 0) \subseteq \mathscr{R}.$$

Some remarks are in order.

Rem-1) An extension of these statements to quasi-ordered spaces is immediate, by the developments in Turinici [43]. Further aspects will be delineated elsewhere.

Rem-2) The nonlinear fixed point statement expressed via Theorem 10 seems to be new.

Rem-3 The variant $(\alpha_2 = 1, \alpha_j = 0, j \neq 2)$ of Theorem 11 includes the statement in this area due to Wardowski [45, 46]; see also Secelean [34]. For a different proof of it, we refer to the paper by Turinici [42].

Rem-4) The remaining variants of Theorem 11 include a related statement due to Cosentino and Vetro [9], and its refinement in Vujaković et al. [44]. Further aspects may be found in Popescu and Stan [27].

9 Wardowski Type Contractions

Let (X, d) be a metric space. Further, let $T \in \mathscr{F}(X)$ be a selfmap of X. As precise, we are interested to determine sufficient conditions for (uniqueness and) existence of elements in $\mathrm{Fix}(T)$, via contractive type requirements involving iterative processes $X_0 = (x_n)$ starting from an element x_0 of X, and their associated sets

$$[X_0] = \{x_n; n \geq 0\}, [[X_0]] = \mathrm{cl}([X_0]).$$

Let (ψ, φ) be a pair of functions over $\mathscr{F}(R_+^0, R)$. Further, let (u, v, w) be a triple of functions over $\mathscr{F}(R_+^6, R_+)$, and define the couple of subsets

$$\Upsilon_0 = \{t \in R_+^6; u(t), v(t), w(t) > 0\},$$
$$\Upsilon = \{t \in \Upsilon_0; \psi(u(t)) \leq \psi(v(t)) - \varphi(w(t))\}.$$

Finally, define the couple of relations over X

$$\mathscr{R} = \mathscr{Q}^{-1}(\Upsilon_0),\ \mathscr{R}^* = \mathscr{Q}^{-1}(\Upsilon).$$

The functional and set contractive properties to be used here are

(fct-contr) T is $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-*contractive*, if $\psi(u \circ \mathscr{Q}(x, y)) \leq \psi(v \circ \mathscr{Q}(x, y)) - \varphi(w \circ \mathscr{Q}(x, y))$, when $(x, y) \in \mathscr{R}$

(set-contr) T is $(d, \mathscr{R}; \Upsilon)$-contractive: $\mathscr{Q}(x, y) \in \Upsilon$, if $(x, y) \in \mathscr{R}$.

And, the connection between these conditions is expressed as

(equi) T is $(d, \mathscr{R}; \psi, \varphi; u, v, w)$-contractive iff T is $(d, \mathscr{R}; \Upsilon)$-contractive.

For the set contraction, sufficient conditions were given upon $\mathscr{R}$ and Υ so that the main result be applicable; these, as expected, were ultimately expressed in terms of $\mathscr{R}$, (ψ, φ) and (u, v, w). The by-product of these developments is the couple of Rhoades-Dutta-Choudhury fixed point results (RDC-reg) and (RDC-incr). Some particular cases of them have been previously exposed.

As results from their proof, a common feature of all these is the fact that the only asymptotic property of Υ to be used there is the descending one; and not the general one. It is our aim in the following to show, by a standard example, that this general condition is ultimately applicable; and yields some interesting results.

To begin with, let (ψ, φ) be a pair of functions in $\mathscr{F}(R_+^0, R)$; and (u, v, w) be the triple of functions introduced as: for each $t = (t_1, \ldots, t_6) \in R_+^6$

$$u(t) = t_5, v(t) = w(t) = t_2.$$

We have, according to our conventions

$$\Upsilon_0 = \{t \in R_+^6; t_5, t_2 > 0\},\ \Upsilon = \{t \in \Upsilon; \psi(t_5) \leq \psi(t_2) - \varphi(t_2)\}.$$

In addition, the attached relations to be used here are (see above)

$$\mathscr{R} = \mathscr{Q}^{-1}(\Upsilon_0),\ \mathscr{R}^* = \mathscr{Q}^{-1}(\Upsilon).$$

The contractive property attached to these data writes, in a shorter way

(fct-contr) T is $(d, \mathscr{R}; \psi, \varphi)$-$contractive$, if
$\psi(d(Tx, Ty)) \leq \psi(d(x, y)) - \varphi(d(x, y)),$ when $(x, y) \in \mathscr{R}$.

As before, we impose a regularity condition like

$$\mathscr{R} \text{ is } B_1\text{-admissible: } (B_1 > 0) \subseteq \mathscr{R}.$$

This, combined with the functional contractive condition, yields

(B1-contr) T is $(d, (B_1 > 0); \psi, \varphi)$-$contractive$:
$\psi(d(Tx, Ty)) \leq \psi(d(x, y)) - \varphi(d(x, y)),$ when $(x, y) \in (B_1 > 0)$;

which is exactly the contractive condition we are dealing with. Note that under the stronger admissible assumption

$$\mathscr{R} \text{ is } B_0\text{-admissible:} (B_1 > 0) \subseteq \mathscr{R}$$

the underlying contractive condition is obtainable from the standard one

(B0-contr) T is $(d, (B_0 > 0); \psi, \varphi)$-$contractive$:
$\psi(d(Tx, Ty)) \leq \psi(d(x, y)) - \varphi(d(x, y)),$ when $(x, y) \in (B_0 > 0)$.

But, the reciprocal implication is not in general valid.

To get the announced result about our introduced class of contractions, a couple of specific conditions is needed:

(str-pos) φ is strictly positive: $\varphi(R_+^0) \subseteq R_+^0$
(r-as-pos) φ is *right asymptotic positive*: for each sequence $(t_n; n \geq 0)$ in R_+^0 and each $b > 0$ with $t_n \to b+$, we must have $\limsup_n(\varphi(t_n)) > 0$
(tele-admi) (ψ, φ) is *tele admissible*: each sequence (t_n) in R_+^0 with $(\psi(t_{n+1}) \leq \psi(t_n) - \varphi(t_n), \forall n)$ fulfills $\lim_n(t_n) = 0$
(zero-her) (ψ, φ) is *zero hereditary*: there are no couple of sequences (t_n) and (s_n) in R_+^0 and no points $b > 0$ with $(\psi(t_n) \leq \psi(s_n) - \varphi(s_n), \forall n)$ and $t_n \to b$, $s_n \to 0$.

Note that our asymptotic condition may be also written by means of the oscillation concepts we already introduced; but this is not important for us. Moreover, under the strict positive condition, this property is available under

(cont) φ is continuous (on R_+^0)

(loc-incr) φ is locally increasing: each $b \in R_+^0$ has a neighborhood V_b such that φ is increasing on V_b.

The following auxiliary statement is the basic step towards our objective.

Proposition 7 *Let (ψ, φ) be a couple of functions over $\mathscr{F}(R_+^0, R)$, with*

(91-I) *ψ is almost right regulated ($\Theta := rreg(\psi)$ is $(>)$-cofinal in R_+^0)*

(91-II) *φ is strictly positive and right asymptotic positive*

(91-III) *(ψ, φ) is tele admissible and zero hereditary.*

Then,

(91-1) *Υ is asymptotic on $[X_0]$*

(91-2) *Υ is nright on $[X_0]$*

(91-3) *Υ is npoint on $[[X_0]]$*

(91-4) *Υ is nnormal.*

Proof The argument consists in a number of parts.

Part 1. (Υ is asymptotic on $[X_0]$).

Let the sequence (r_n) in R_+^0 and the sequence (p_n) in R_+ be such that

$$(\forall n):\ (r_n, r_n, p_n, 0, r_{n+1}, r_{n+1}) \in \mathscr{Q}(B_1 > 0; [X_0]) \cap \Upsilon, \text{ and } |p_n - r_n| \le r_{n+1}.$$

By the very representation of Υ, we must have

$$(\forall n):\ \psi(r_{n+1}) \le \psi(r_n) - \varphi(r_n).$$

This, along with (ψ, φ) being tele admissible, gives $r_n \to 0$ (hence, $p_n \to 0$); and proves the desired assertion.

Part 2. (Υ is almost nright on $[X_0]$).

Take some point $c = (c_1, \ldots, c_6)$ in R_+^6. We say that the (vectorial) sequence $(t^n := (t_1^n, \ldots, t_6^n); n \ge 0)$ in R_+^6 is *right* at c, if

$$\text{(r-c)}\ \ (t_i^n \to c_i, \forall i) \text{ and } (t_i^n \to c_i +, \text{ whenever } c_i > 0).$$

Given $b > 0$, let us say that Υ is *nright* at b on $[X_0]$, if

(nright) for each sequence $(t^n; n \ge 0)$ in $\mathscr{Q}(B_3 > 0; [X_0]) \cap \Upsilon$, the right property at $(0, b, b, b, b, 0)$ is not true.

The class of all these $b > 0$ will be denoted as $\text{nright}(\Upsilon; [X_0])$. In this case, we say that Υ is

(a-n-r) *almost nright* on $[X_0]$, if $\Theta := \text{nright}(\Upsilon; [X_0])$ is $(>)$-cofinal in R_+^0 (for each $\varepsilon \in R_+^0$ there exists $\theta \in \Theta$ with $\varepsilon > \theta$)

(n-r) *nright* on $[X_0]$, if $\Theta := \text{nright}(\Upsilon; [X_0])$ is identical with R_+^0.

We have to establish that the former property is retainable for our data. To do this, we start by noting that

$$\psi = \text{almost right regulated implies } \Theta := \text{rreg}(\psi) \text{ is } (>)\text{-cofinal in } R^0_+.$$

We now claim that

$$\Theta \subseteq \text{nright}(\Upsilon; [X_0]);\ \text{wherefrom, nright}(\Upsilon; [X_0]) \text{ is } (>)\text{-cofinal in } R^0_+.$$

This amounts to establish that

for each sequence $(t^n := (t^n_1, \ldots, t^n_6); n \geq 0)$ in $\mathcal{Q}(B_3 > 0; [X_0]) \cap \Upsilon$,
and each $b \in \Theta$, the right property at $(0, b, b, b, b, 0)$ is not true.

Suppose—by *reductio ad absurdum*—that there exists a (vectorial) sequence $(t^n := (t^n_1, \ldots, t^n_6); n \geq 0)$ in R^6_+ and some $b \in \Theta$, with

(right-b-1) $(\forall n)$: $t^n := (t^n_1, \ldots, t^n_6) \in \mathcal{Q}(B_3 > 0; [X_0]) \cap \Upsilon$; whence, $\psi(t^n_5) \leq \psi(t^n_2) - \varphi(t^n_2)$

(right-b-2) $(t^n := (t^n_1, \ldots, t^n_6); n \geq 0)$ has the right property at $(0, b, b, b, b, 0)$; that is: $(t^n_i \to 0, \forall i \in \{1, 6\})$, and $(t^n_i \to b+, \forall i \in \{2, 3, 4, 5\})$.

By the former of these properties

$$(\forall n):\ (0 <)\varphi(t^n_2) \leq \psi(t^n_2) - \psi(t^n_5).$$

Passing to lim sup as $n \to \infty$ in this relation gives (via ψ=right regulated at b)

$$0 \leq \limsup_n \varphi(t^n_2) \leq \psi(b + 0) - \psi(b + 0) = 0;\ \text{that is: } \lim_n \varphi(t^n_2) = 0;$$

in contradiction with φ=right asymptotic positive. Hence, our working assumption cannot be true; and the assertion follows.

Part 3. (Υ is npoint on $[[X_0]]$).

Take some point $c = (c_1, \ldots, c_6)$ in R^6_+. We say that the sequence $(t^n := (t^n_1, \ldots, t^n_6); n \geq 0)$ in R^6_+ is *point* at c, if

$$\text{(pt-c)}\ \ (t^n_i \to c_i, \forall i) \text{ and } [t^n_6 = c_6, \forall n].$$

Given $b > 0$, let us say that Υ is *npoint* at b on $[[X_0]]$, if

(npoint) for each sequence $(t^n; n \geq 0)$ in $\mathcal{Q}(B_3 > 0; [[X_0]]) \cap \Upsilon$, the point property at $(0, 0, b, 0, b, b)$ is not true.

The class of all these $b > 0$ will be denoted as $\text{npoint}(\Upsilon; [[X_0]])$. In this case, we say that Υ is

(a-n-p) *almost npoint* on $[[X_0]]$, if $\Theta := \text{npoint}(\Upsilon; [[X_0]])$ is $(>)$-cofinal in R^0_+

(n-p) *npoint* on $[[X_0]]$, if $\Theta := \text{npoint}(\Upsilon; [[X_0]])$ is identical with R^0_+.

We have to establish that the latter of these properties holds; that is,

for each sequence $(t^n; n \geq 0)$ in R_+^6 and each $b > 0$, the point property at $(0, 0, b, 0, b, b)$ is not true.

Assume by contradiction that this assertion is false: there exists a sequence $(t^n = (t_1^n, \ldots, t_6^n); n \geq 0)$ in R_+^6 and some $b > 0$, with

(point-b-1) $(\forall n)$: $t^n = (t_1^n, \ldots, t_6^n) \in \mathscr{Q}(B_3 > 0; [[X_0]]) \cap \Upsilon$; whence, $\psi(t_5^n) \leq \psi(t_2^n) - \varphi(t_2^n)$

(point-b-2) $(t^n = (t_1^n, \ldots, t_6^n); n \geq 0)$ is 6-point at $(0, 0, b, 0, b, b)$; that is: $(t_i^n \to 0, \forall i \in \{1, 2, 4\})$, $(t_i^n \to b, \forall i \in \{3, 5, 6\})$, and $(t_6^n = b, \forall n)$.

The obtained inequality and convergence relations yield a contradiction with respect to (ψ, φ) being zero hereditary. Consequently, the working assumption is not acceptable; and conclusion follows.

Part 4. (Υ is nnormal).

Remember that, this property means:

$$(0, a, a, a, a, 0) \in \Upsilon \text{ is impossible, for each } a > 0.$$

Suppose by contradiction that

$$\text{there exists } a > 0 \text{ with } (0, a, a, a, a, 0) \in \Upsilon.$$

By the very definition of our triple,

$$u(0, a, a, a, a, 0) = a, v(0, a, a, a, a, 0) = a, \ w(0, a, a, a, a, 0) = a.$$

This, along with the definition of Υ, yields

$$\psi(a) \leq \psi(a) - \varphi(a); \ \text{ that is: } \varphi(a) \leq 0;$$

absurd, by φ=strictly positive; and our claim follows.

Now, by simply combining the obtained fact with our main result, one gets the following particular fixed point statement.

Theorem 12 *Assume that T is $(d, \mathscr{R}; \psi, \varphi)$-contractive, where (ψ, φ) is a couple over $\mathscr{F}(R_+^0, R)$ with*

(91-i) *ψ is almost right regulated ($\Theta := rreg(\psi)$ is $(>)$-cofinal in R_+^0)*

(91-ii) *φ is strictly positive and right asymptotic positive*

(91-iii) *(ψ, φ) is tele admissible and zero hereditary,*

and the associated relation $\mathscr{R}$ fulfills

(91-iv) *$\mathscr{R}$ is $(B_1 > 0)$-admissible: $(B_1 > 0) \subseteq \mathscr{R}$.*

Further, assume that X is (Ba-o,d)-complete; and let $X_0 = (x_n)$ be a (Ba-o) sequence. Then,

(91-a) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$)*
(91-b) *T is fix-$\mathscr{R}$-asingleton (hence, fix-$\mathscr{R}$-singleton)*
(91-c) *T is fix-asingleton (hence, fix-singleton) when, in addition to the conditions above*

$$\mathscr{R} \text{ is } (B_0 > 0)\text{-admissible: } (B_0 > 0) \subseteq \mathscr{R}.$$

A basic particular cases of these developments is the one of $\varphi \in \mathscr{F}(R_+^0, R)$ being a constant function. For both practical and theoretical reasons, it will be useful, in the following, to discuss its technical aspects.

Let $\psi \in \mathscr{F}(R_+^0, R)$ be a function; and $C > 0$ be a (real) constant. Then, let (u, v, w) be the triple of functions introduced as before: for each $t = (t_1, \ldots, t_6) \in R_+^6$

$$u(t) = t_5, v(t) = w(t) = t_2.$$

We have, according to our conventions

$$\Upsilon_0 = \{t \in R_+^6; t_5, t_2 > 0\},\ \Upsilon = \{t \in \Upsilon_0; \psi(t_5) \leq \psi(t_2) - C\}.$$

In addition, the attached relations to be used here are (see above)

$$\mathscr{R} = \mathscr{Q}^{-1}(\Upsilon_0),\ \mathscr{R}^* = \mathscr{Q}^{-1}(\Upsilon).$$

The contractive property attached to these data writes, in a shorter way

(fct-contr) T is $(d, \mathscr{R}; \psi, C)$-*contractive*, if
$\psi(d(Tx, Ty)) \leq \psi(d(x, y)) - C$, when $(x, y) \in \mathscr{R}$.

As before, we impose a regularity condition like

$$\mathscr{R} \text{ is } B_1\text{-admissible: } (B_1 > 0) \subseteq \mathscr{R}.$$

This, combined with the functional contractive condition, yields

(B1-contr) T is $(d, (B_1 > 0); \psi, C)$-$contractive$:
$\psi(d(Tx, Ty)) \leq \psi(d(x, y)) - C$, when $(x, y) \in (B_1 > 0)$;

which is exactly the contractive condition we are dealing with. Note that, under a stronger regularity condition like

$$\mathscr{R} \text{ is } B_0\text{-admissible: } (B_0 > 0) \subseteq \mathscr{R}.$$

the starting functional contractive condition yields

$$\text{(B0-contr)}\quad T \text{ is } (d, (B_0 > 0); \psi, C)\text{-}contractive:$$

$$\psi(d(Tx, Ty)) \leq \psi(d(x, y)) - C, \text{ when } (x, y) \in (B_0 > 0).$$

Moreover, in this B_0-admissible setting, the following inclusion holds

$$\text{(B0-B1)}\quad T \text{ is } (d, (B_0 > 0); \psi, C)\text{-contractive implies}$$

$$T \text{ is } (d, (B_1 > 0); \psi, C)\text{-contractive.}$$

But, the reverse inclusion does not hold, in general.

Formally, the conditions to be imposed upon the couple (ψ, C) are directly obtainable from the ones of Theorem 12, under the choice $(\varphi(t) = C; t > 0)$. In particular, the strictly positive and right asymptotic conditions upon φ are fulfilled here. Concerning the remaining ones, a basic instance when these hold writes

(s-zero-abr) ψ is *strongly zero abrupt*: for each sequence (t_n) in R_+^0, $(\psi(t_n) \to -\infty$ iff $t_n \to 0)$.

Putting these together, the following version of Theorem 12 is to be noted.

Theorem 13 *Assume that T is $(d, \mathscr{R}; \psi, C)$-contractive, where $C > 0$ is a constant, ψ is a function in $\mathscr{F}(R_+^0, R)$ with*

(92-i) ψ is almost right regulated and strongly zero abrupt, and the associated relation $\mathscr{R}$ fulfills
(92-ii) $\mathscr{R}$ is $(B_1 > 0)$-admissible: $(B_1 > 0) \subseteq \mathscr{R}$.

Further, assume that X is (Ba-o,d)-complete; and let $X_0 = (x_n)$ be a (Ba-o) sequence. Then,

(92-a) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$)*
(92-b) *T is fix-$\mathscr{R}$-asingleton (hence, fix-$\mathscr{R}$-singleton)*
(92-c) *T is fix-asingleton (hence, fix-singleton) when, in addition to the conditions above*

$$\mathscr{R} \text{ is } (B_0 > 0)\text{-admissible: } (B_0 > 0) \subseteq \mathscr{R}.$$

Proof As precise, $\varphi = C$ is strictly positive and right asymptotic positive. It will suffice establishing that the remaining conditions of Theorem 12 are working here to complete the argument.

Part 1. We claim that, necessarily,

(tele-admi) (ψ, C) is *tele admissible*: each sequence (t_n) in R_+^0 with $(\psi(t_{n+1}) \leq \psi(t_n) - C, \forall n)$ fulfills $\lim_n(t_n) = 0$

In fact, let (t_n) be as before. By a summation procedure, one arrives at

$$\psi(t_n) \leq \psi(t_0) - nC, \text{ for all } n.$$

Passing to limit as $n \to \infty$, yields

$$\psi(t_n) \to -\infty; \text{ hence, } t_n \to 0,$$

if we take the strongly zero abrupt property of ψ into account; hence the assertion.

Part 2. We claim that under the admitted hypotheses

(zero-her) (ψ, C) is *zero hereditary*: there are no couple of sequences (t_n) and (s_n) in R_+^0 and no points $b > 0$ with $(\psi(t_n) \leq \psi(s_n) - C, \forall n)$ and $t_n \to b$, $s_n \to 0$.

Suppose by contradiction that this is not true: there exists a couple of sequences (t_n) and (s_n) in R_+^0 and some point $b > 0$ with

(non-zh-1) $\psi(t_n) \leq \psi(s_n) - C$, for all $n \geq 0$

(non-zh-2) $t_n \to b$, $s_n \to 0$, as $n \to \infty$.

The second half of (non-zh-2) yields

$$\psi(s_n) \to -\infty \text{ (if we remember that } \psi\text{=strongly zero abrupt).}$$

But then, taking (non-zh-1) into account,

$$\psi(t_n) \to -\infty; \text{ hence, } t_n \to 0, \text{ in view of } \psi = \text{strongly zero abrupt.}$$

The obtained fact is in contradiction with the hypothesis $t_n \to b > 0$. Hence, our working assumption cannot be accepted; and this assertion is valid too.

Summing up, conditions of Theorem 12 are fulfilled by these data; and, from this, the conclusions in the statement follow.

A basic version of the obtained result is obtainable under the lines below. Let us say that $\psi : R_+^0 \to R$ is a *Wardowski function*, if

$$\psi \text{ is increasing and } \psi(0+0) = -\infty.$$

By a previous observation, we have that any such function is almost right regulated. On the other hand, by an auxiliary statement in Turinici [39], any Wardowski function is strongly zero abrupt. By Theorem 13 we then have the following fixed point statement with a methodological meaning.

Theorem 14 *Assume that T is $(d, \mathscr{R}; \psi, C)$-contractive, where $C > 0$ is a (real) constant, $\psi \in \mathscr{F}(R_+^0, R)$ is taken so as*

(93-i) *ψ is Wardowski (see above)*

and the associated relation $\mathscr{R}$ fulfills

(93-ii) *$\mathscr{R}$ is $(B_1 > 0)$-admissible: $(B_1 > 0) \subseteq \mathscr{R}$.*

Further, assume that X is (Ba-o,d)-complete; and let $X_0 = (x_n)$ be a (Ba-o) sequence. Then,

(93-a) *$X_0 = (x_n)$ is strongly Picard (modulo $(d, \mathscr{R}; T)$)*
(93-b) *T is fix-$\mathscr{R}$-asingleton (hence, fix-$\mathscr{R}$-singleton)*
(93-c) *T is fix-asingleton (hence, fix-singleton) when, in addition to the conditions above*

$$\mathscr{R} \text{ is } (B_0 > 0)\text{-admissible: } (B_0 > 0) \subseteq \mathscr{R}.$$

The obtained statement may be viewed as a technical improvement of the 2012 fixed point principle in Wardowski [45]; based, among others, on the extra condition

$$\psi \text{ is } subunitary\ power\ compatible\text{: } \lim_{t \to 0+} t^k \psi(t) = 0, \text{ for some } k \in]0, 1[.$$

This improvement is essentially related to

(impr-1) the $(B_1 > 0)$ setting of Theorem 14 is an effective extension of the $(B_0 > 0)$ setting of the quoted statement
(impr-2) the subunitary power compatible condition appearing in the quoted article is avoided in our statement.

For a different proof of Theorem 7 we refer to the paper by Turinici [40]. Some other partial aspects were discussed in Secelean [34].

References

1. M. Akkouchi, Common fixed points for weakly compatible maps satisfying implicit relations without continuity. Dem. Math. **44**, 151–158 (2011)
2. S. Banach, Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. Fund. Math. **3**, 133–181 (1922)
3. V. Berinde, F. Vetro, Common fixed points of mappings satisfying implicit contractive conditions. Fixed Point Theory Algorithms Sci. Eng. **2012**, Article ID 105 (2012)
4. P. Bernays, A system of axiomatic set theory: Part III. Infinity and enumerability analysis. J. Symb. Log. **7**, 65–89 (1942)
5. D.W. Boyd, J.S.W. Wong, On nonlinear contractions. Proc. Amer. Math. Soc. **20**, 458–464 (1969)
6. S. Chandok, B.S. Choudhury, N. Metiya, Fixed point results in ordered metric spaces for rational type expressions with auxiliary functions. J. Egyptian Math. Soc. **23**, 95–101 (2015)
7. B.S. Choudhury, P. Konar, B.E. Rhoades, N. Metiya, Fixed point theorems for generalized weakly contractive mapping. Nonlinear Anal. **74**, 2116–2126 (2011)
8. P.J. Cohen, *Set Theory and the Continuum Hypothesis* (Benjamin, New York, 1966)
9. M. Cosentino, P. Vetro, Fixed point results for F-contractive mappings of Hardy-Rogers type. Filomat **28**, 715–722 (2014)
10. J. Dieudonné, *Foundations of Modern Analysis* (Academic Press, New York, 1960)
11. N.V. Dung, V. I. Hang, A fixed point theorem for generalized F-contraction on complete metric space. Vietnam J. Math. **43**, 743–755 (2015)
12. P.N. Dutta, B.S. Choudhury, A generalisation of contraction principle in metric spaces. Fixed Point Theory Algorithms Sci. Eng. **2008**, Article ID 406368 (2008)
13. P. Hitzler, Generalized metrics and topology in logic programming semantics. PhD Thesis, National University Ireland, University College Cork (2001)

14. S. Kasahara, On some generalizations of the Banach contraction theorem. Publ. Res. Inst. Math. Sci. **12**, 427–437 (1976)
15. M.S. Khan, M. Swaleh, S. Sessa, Fixed point theorems by altering distances between the points. Bull. Aust. Math. Soc. **30**, 1–9 (1984)
16. S. Leader, Fixed points for general contractions in metric spaces. Math. Japon. **24**, 17–24 (1979)
17. J. Matkowski, *Integrable Solutions of Functional Equations*. Dissertationes Mathematicae, vol. 127 (Polish Sci. Publ., Warsaw, 1975)
18. A. Meir, E. Keeler, A theorem on contraction mappings. J. Math. Anal. Appl. **28**, 326–329 (1969)
19. G. H. Moore, *Zermelo's Axiom of Choice: its Origin, Development and Influence* (Springer, New York, 1982)
20. Y. Moskhovakis, *Notes on Set Theory* (Springer, New York, 2006)
21. S.B. Nadler Jr., Multi-valued contraction mappings. Pacific J. Math. **30**, 475–488 (1969)
22. H.K. Nashine, Z. Kadelburg, P. Kumam, Implicit-relation-type cyclic contractive mappings and applications to integral equations. Abstr. Appl. Anal. **2012**, Article ID 386253 (2012)
23. I.P. Natanson, *Theory of Functions of a Real Variable (Volume I)* (Frederick Ungar Publishing Co., New York, 1964)
24. L. Pasicki, A strong fixed point theorem Topology Appl. **282**, 107–130 (2020)
25. V. Popa, Some fixed point theorems for compatible mappings satisfying an implicit relation. Dem. Math. **32**, 157–163 (1999)
26. V. Popa, M. Mocanu, Altering distance and common fixed points under implicit relations. Hacettepe J. Math. Stat. **38**, 329–337 (2009)
27. O. Popescu, G. Stan, Two fixed point theorems concerning F-contraction in complete metric spaces. Symmetry **2020**, 12, 58 (2019)
28. S. Radenović, Z. Kadelburg, D. Jandrlić and A. Jandrlić, Some results on weakly contractive maps. Bull. Iranian Math. Soc. **38**, 625–645 (2012)
29. S. Reich, Fixed points of contractive functions. Boll. Un. Mat. Ital. **5**, 26–42 (1972)
30. B.E. Rhoades, A comparison of various definitions of contractive mappings. Trans. Amer. Math. Soc. **226**, 257–290 (1977)
31. B.E. Rhoades, Some theorems on weakly contractive maps. Nonlinear Anal. **47**, 2683–2693 (2001)
32. I.A. Rus, *Generalized Contractions and Applications* (Cluj University Press, Cluj-Napoca, 2001)
33. E. Schechter, *Handbook of Analysis and its Foundation* (Academic Press, New York, 1997)
34. N.-A. Secelean, Weak F-contractions and some fixed point results. Bull. Iran. Math. Soc. **42**, 779–798 (2016)
35. A. Tarski, Axiomatic and algebraic aspects of two theorems on sums of cardinals. Fund. Math. 35, 79–104 (1948)
36. M. Turinici, Fixed points of implicit contraction mappings. An. Şt. Univ. A. I. Cuza Iaşi (S I-a, Mat) **22**, 177–180 (1976)
37. M. Turinici, Fixed points of implicit contractions via Cantor's intersection theorem. Bul. Inst. Polit. Iaşi (Sect I: Mat., Mec. Teor., Fiz.) **26**(30), 65–68 (1980)
38. M. Turinici, Function pseudometric VP and applications. Bul. Inst. Polit. Iaşi (Sect.: Mat., Mec. Teor., Fiz.) **53**(57), 393–411 (2007)
39. M. Turinici, Wardowski implicit contractions in metric spaces. Arxiv, 1211–3164-v2 (2013)
40. M. Turinici, *Implicit Contractive Maps in Ordered Metric Spaces*. Topics in Mathematical Analysis and Applications, ed. by Th.M. Rassias, L. Toth (Springer International Publishing, Switzerland, 2014), pp. 715–746
41. M. Turinici, Contractive maps in locally transitive relational metric spaces. The Sci. World J. **2014**, Article ID 169358 (2014)
42. M. Turinici, Wardowski mappings and Meir-Keeler property, in *Modern Directions in Metrical Fixed Point Theory*, Paper 1–5 (Pim Editorial House, Iaşi, 2016)

43. M. Turinici, Rational implicit contractions in ordered metric spaces, in *Selected Topics in Metrical Fixed Point Theory* (Revised Ed.), Paper 3-1 (Pim Editorial House, Iaşi, 2017)
44. J. Vujaković, S. Mitrović, M. Pavlović, S. Radenović, On recent results concerning F-contraction in generalized metric spaces. Mathematics **8**, 767 (2020)
45. D. Wardowski, Fixed points of a new type of contractive mappings in complete metric spaces. Fixed Point Theory Algorithms Sci. Eng. **2012**, 94 (2012)
46. D. Wardowski, Solving existence problems via F-contractions. Proc. Amer. Math. Soc. **146**, 1585–1598 (2018)
47. E.S. Wolk, On the principle of dependent choices and some forms of Zorn's lemma. Canad. Math. Bull. **26**, 365–367 (1983)

Nonlinear Dynamics of the KdV-B Equation and Its Biomedical Applications

Michail A. Xenos and Anastasios C. Felias

Abstract In recent years there is an incremental degree of bridging open questions in biomechanics with the help of applied mathematics and nonlinear analysis. Recent advancements concerning the cardiac dynamics pose important questions about the cardiac waveform. A governing equation, namely the KdV-B equation (Korteweg–de Vries–Burgers),

$$\frac{\partial u}{\partial t} + \gamma u \frac{\partial u}{\partial x} - \alpha \frac{\partial^2 u}{\partial x^2} + \beta \frac{\partial^3 u}{\partial x^3} = 0, \quad u = u(t, x), \quad \alpha,\ \beta,\ \gamma \in \mathbb{R}, \tag{1}$$

is a partial differential equation utilized to answer several of those questions. The cardiac dynamics mathematical model features both solitary and shock wave characteristics due to the dispersion and dissipation terms, as occurring in the arterial tree. In this chapter a focus is given on describing cardiac dynamics. It is customarily difficult to solve nonlinear problems, especially by analytical techniques. Therefore, seeking suitable solving methods, exact, approximate or numerical, is an active task in branches of applied mathematics. The phase plane of the KdV–B equation is analyzed and its qualitative behavior is derived. An asymptotic expansion is presented and traveling wave solutions under both shock and solitary profiles are sought. Numerical solutions are obtained for the equation, by means of the Spectral Fourier analysis and are evolved in time by the Runge–Kutta method. This whole analysis provides vital information about the KdV–B equation and its connection to cardiac hemodynamics. The applications of KdV–B, presented in this chapter, highlight its essence to human hemodynamics.

Mathematics Subject Classification (2020) 35Q35, 35Q53

M. A. Xenos (✉) · A. C. Felias
Department of Mathematics, University of Ioannina, Ioannina, Greece
e-mail: mxenos@uoi.gr

Th. M. Rassias (ed.), *Nonlinear Analysis, Differential Equations, and Applications*,
Springer Optimization and Its Applications 173,
https://doi.org/10.1007/978-3-030-72563-1_26

1 Introduction

1.1 Background Information for KdV–B

In the last few decades, much attention from a rather diverse group of scientists such as physicists, engineers and applied mathematicians has been attracted to two contrasting themes: (a) the theory of dynamical systems, most popularly associated with the study of chaos, and (b) the theory of integrable (or nonintegrable) systems associated, among other things, with the study of solitary waves.

It is common knowledge that many physical phenomena, such as nonlinear shallow-water waves and wave motion in plasma, can be described by the Korteweg–de Vries (KdV) equation [29]. It is well known that solitons and solitary waves are the class of special solutions of the KdV equation. In order to study propagation of undular bores in shallow water [6, 27], liquid flow containing gas bubbles [54], fluid flow in elastic tubes [28], crystal lattice theory, nonlinear circuit theory and turbulence [20, 30, 51], the governing equation can be reduced to the so-called Kortweg–de Vries–Burgers equation (KdV–B) as follows [10],

$$\frac{\partial u}{\partial t} + \gamma u \frac{\partial u}{\partial x} - \alpha \frac{\partial^2 u}{\partial x^2} + \beta \frac{\partial^3 u}{\partial x^3} = 0, \quad u = u(t, x), \quad \alpha,\ \beta,\ \gamma \neq 0. \tag{2}$$

This is a nonintegrable equation in the sense that its spectral problem is nonexistent [19]. Multiplying t, x and u, by constants can be used to make the coefficients of any of the above four terms equal to any given nonzero constant. Therefore, we focus on the case where $\alpha \geq 0$, $\beta > 0$ and $\gamma \neq 0$.

This equation is equivalent to the KdV equation with the addition of a viscous dissipation term ($\alpha \frac{\partial^2 u}{\partial x^2}$). The studies of the KdV equation [29] ($\alpha = 0$) and the Burgers equation [9] ($\beta = 0$) have been undertaken, but the exact solution for the general case of equation (2) ($\alpha \geq 0$, $\beta > 0$, $\gamma \neq 0$) has still not been completed.

1.2 Biomechanical Applications

Solitons are mathematical entities appearing as solutions of nonlinear wave equations [8]. They are waves of stable and steady form, although internal oscillations may occur, exhibiting unique characteristics when colliding with other solitary waves as described by Ablowitz and Segur [1]. During the last decade, soliton profiles are found when studying nonlinear optics, condensed matter Physics and quantum theory of matter and gravity [43]. Lately, an increasing number of studies focuses on describing the cardiac pulse as a soliton, due to the features those two seem to share. The pulsatility synchronization of the smooth arterial muscle allows the consideration of solitary profiles in cardiac hemodynamics [34].

Theoretical investigation for the blood waves have been developed by many researchers through the use of weakly nonlinear theories. The theoretical investigation of pulse wave propagation in human arteries has a long history starting from ancient times until today. Over the past decade, the scientific efforts have been concentrated on theoretical investigations of nonlinear wave propagation in arteries with a variable radius. The question "How local imperfections appeared in the artery can disturb the arterial wall deformation?" is important for understanding the nature and main features of various cardiovascular diseases, such as stenoses and aneurysms. Rowlands (1982) reported some extraordinary features of the cardiac pulse, leading to his conception of the arterial flows as a solitary motion [44]. A few years later, Otwinowski and collaborators presented a nonlinear differential equation whose solutions exhibited similar characteristics with those reported by Rowlands [38].

Based on those evolutionary theories, adding the inertial behavior of blood vessel in an one-dimensional cardiovascular model, researchers concluded that the KdV equation is a seemingly reliable tool in modelling cardiac dynamics. It was supported that the solitary wave formulation fits much better in describing the arterial pulse wave experimental results than the wave equation proposed by the majority of researchers [57]. An additional reason to support the above formulation is the peaking and steepening features of the pressure pulse, which coincide with the structure of soliton profiles of KdV [11].

The majority of studies on the wave propagation in blood flow is mainly based on linear waves. The linearized theories proposed by Resal, Witzig, Womersly, McDonald and others, consider the vessel as a straight, infinite, circular elastic tube filled with an isotropic and Newtonian fluid, blood [55]. Blood is studied as an incompressible fluid, a characterization justified by its compressibility being rather insignificant, compared to the dilation of the blood vessels. In 1958, Lambert based on the Euler equations of fluid motion, proposed the Method of Characteristics for the calculations concerning the nonlinear blood flow. All theories presented to model nonlinear blood flow are one-dimensional, meaning that both pressure and flow velocity are seen as functions of the axial distance along the vessel in time. Contributions in nonlinear modulation were done by Rudinger, Skalak, Rockwell, Hawley and Anliker. The suggested equations are basically the equations of continuity and motion coupled with an extra equation to describe the vessel wall distensibility [3, 43, 45, 49]. Sakanishi and Hasegawa proposed a soliton profile pulsatile wave modeling, based on the nonlinear elasticity of the vessel wall [46]. Yomosa and collaborators proposed a theory describing solitons in long arteries, where the viscous effects, the reflective effects caused by the arterial branch as well as the effects of the peripheral resistance are neglectible. For the above reasons, the latter modulation is unable to describe the pressure drop caused when moving away from the heart. Nevertheless, it points out that it does make some sense to attribute the special features of the pulsatile wave, including the "sudden steepening" and the change in the phase velocity, to the solitary profile [57]. While the pulsatile wave travels to arteries with smaller radius, viscosity seems to play a vital role in both the flow decay and the widening of the wave width [57].

Antar and Demiray studied the propagation of weak nonlinear waves in a thin elastic tube, under an initial stress distribution, due to the flow of an incompressible viscous fluid [4]. The propagation of pressure pulses in dilatable tubes has been studied by various researchers [22, 41]. Most of those studies, consider waves of small width, neglecting the nonlinear characteristics and focusing on their dispersive character [5, 12, 42]. It is widely accepted that a long-term evolution of weak nonlinear waves of either dispersion or dissipation, can be modeled by nonlinear dispersive equations. Two classical simplified and indicative examples are the Burgers equation and the KdV equation, exhibiting balance between nonlinearity and dissipation and nonlinearity and dispersion, respectively. On the other hand, when a balance is exhibited among nonlinearity, dissipation and dispersion, the simplest and most representative dispersive equation is the KdV–B equation, combining the KdV and Burgers equations. Via asymptotic methods, the propagation of small, but with finite width, waves in dilatable tubes has been studied sufficiently [4].

Hashizume and Yomosa showed that propagation, in the case of weak nonlinear waves in a thin and nonlinear elastic tube for incompressible flow, is determined by the KdV equation [57]. Erbay and collaborators, examining the propagation of weak nonlinear waves in a thin viscoelastic tube filled with fluid, were lead to the Burgers, KdV and KdV–B equations, depending on the parameters considered [15]. Demiray studied the propagation of slightly nonlinear waves in thin elastic and viscoelastic tubes for an incompressible fluid and finally concluded to the KdV and KdV–B equations, respectively. In all the above studies, an inviscid fluid was considered and the axial movement of the tube wall was neglected. However, regarding biological applications, blood is an incompressible and viscous fluid. So, Antar and Demiray formulated their mathematical model toward this direction [4].

In this chapter, an emphasis is given to the theoretical and numerical analysis of the KdV–B equation and its applications, providing vital information about the KdV–B equation and its connection to cardiac hemodynamics. More precisely, in the next section the phase plane of the KdV–B equation is analyzed and its qualitative behavior is derived. Furthermore, an asymptotic expansion is presented and traveling wave solutions under both shock and solitary profiles are derived. Finally, numerical solutions are obtained for the KdV–B equation, by means of spectral Fourier analysis and are evolved in time by the well known explicit 4th order Runge–Kutta method.

2 Phase Plane Analysis of KdV–B

In this section, the phase plane of the KdV–B equation is analyzed and its qualitative behavior is derived and further described. The wave variable ζ is introduced as [25, 47],

$$\zeta = x - \lambda t, \tag{3}$$

with λ being the wave velocity. Then equation (2) using (3), can be written as,

$$(\gamma u - \lambda)\frac{du}{d\zeta} - \alpha\frac{d^2u}{d\zeta^2} + \beta\frac{d^3u}{d\zeta^3} = 0, \quad u = u(t,x) = u(x - \lambda t) = u(\zeta)\,. \tag{4}$$

The so-called traveling–wave solution, $u = u(\zeta)$, shall be considered here. By integrating equation (4) with respect to ζ, a nonlinear differential equation can be obtained as follows,

$$\frac{d^2u}{d\zeta^2} + c_1\frac{du}{d\zeta} + c_2u^2 + c_3u = c_0, \tag{5}$$

where $c_1 = -\dfrac{\alpha}{\beta}$, $c_2 = \dfrac{\gamma}{2\beta}$, $c_3 = -\dfrac{\lambda}{\beta}$ and the integral constant $c_0 > -\dfrac{\lambda^2}{2\beta}$.

In the case where $c_0 \neq 0$, a simple translation transformation,

$$u = u' + c_0', \quad c_0' = \frac{-c_3 \pm \sqrt{c_3^2 + 4c_0c_2}}{2c_2},$$

can be made, with u' satisfying the following equation,

$$\frac{d^2u'}{d\zeta^2} + c_1\frac{du'}{d\zeta} + c_2u'^2 + (c_3 + 2c_2c_0')u' = 0.$$

Without loss of generality, we shall confine ourselves to the consideration of $c_0 = 0$ alone from now on. It can be further assumed that $\lambda \geq 0$, because the discussion on $\lambda' = -\lambda$ can be made in the same manner for $\lambda < 0$.

Equation (5) can be written as an autonomous system of first-order equations,

$$\begin{cases} \dfrac{du}{d\zeta} = v, \\ \dfrac{dv}{d\zeta} = -\dfrac{u}{\beta}(\gamma\dfrac{u}{2} - \lambda) + \dfrac{\alpha}{\beta}v. \end{cases}$$

Now, we study the above system according to the qualitative theory of ordinary differential equations. Initially, we find the system's singular points, setting,

$$f_1(u,v) = v, \quad f_2(u,v) = -\frac{u}{\beta}(\gamma\frac{u}{2} - \lambda) + \frac{\alpha}{\beta}v.$$

The following conditions should be met,

$$f_1(u,v) = f_2(u,v) = 0$$

$$\Leftrightarrow \begin{cases} v = 0 \\ -\frac{u}{\beta}(\gamma\frac{u}{2} - \lambda) + \frac{\alpha}{\beta}v = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} v = 0 \\ u = 0, \quad u = \frac{2\lambda}{\gamma}. \end{cases}$$

Therefore, the singular points are,

$$\begin{cases} P_1 = (0, 0), \\ P_2 = (\frac{2\lambda}{\gamma}, 0). \end{cases}$$

Next, we are to find the eigenvalues of the linearization matrices, defined for our singular points, as follows,

$$A(P_1) = \begin{bmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} \end{bmatrix}, \ (P_1) = \begin{bmatrix} 0 & 1 \\ \frac{\lambda}{\beta} & \frac{\alpha}{\beta} \end{bmatrix},$$

so, for its eigenvalues we get,

$$det(A_{P_1} - sI_2) = 0 \Leftrightarrow s^2 - \frac{\alpha}{\beta}s - \frac{\lambda}{\beta} = 0$$

$$\Leftrightarrow \begin{cases} s_1 = \dfrac{\frac{\alpha}{\beta} + \frac{1}{\beta}\sqrt{\alpha^2 + 4\lambda\beta}}{2} > 0 \\ s_2 = \dfrac{\frac{\alpha}{\beta} - \frac{1}{\beta}\sqrt{\alpha^2 + 4\lambda\beta}}{2} < 0 \end{cases} .$$

$$A(P_2) = \begin{bmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} \end{bmatrix}, \ (P_2) = \begin{bmatrix} 0 & 1 \\ -\frac{\lambda}{\beta} & \frac{\alpha}{\beta} \end{bmatrix},$$

so, for its eigenvalues we get,

$$det(A_{P_2} - sI_2) = 0 \Leftrightarrow s^2 - \frac{\alpha}{\beta}s + \frac{\lambda}{\beta} = 0$$

$$\Leftrightarrow \begin{cases} s_1 = \dfrac{\frac{\alpha}{\beta} + \frac{1}{\beta}\sqrt{\alpha^2 - 4\lambda\beta}}{2} > 0 \\ s_2 = \dfrac{\frac{\alpha}{\beta} - \frac{1}{\beta}\sqrt{\alpha^2 - 4\lambda\beta}}{2} > 0 \end{cases} , \quad \alpha \geq 2\sqrt{\lambda\beta} .$$

$$\begin{cases} s_1 = \dfrac{\frac{\alpha}{\beta} + i\frac{1}{\beta}\sqrt{4\lambda\beta - \alpha^2}}{2} \\ s_2 = \dfrac{\frac{\alpha}{\beta} - i\frac{1}{\beta}\sqrt{4\lambda\beta - \alpha^2}}{2} \end{cases} , \quad \alpha \in (0, 2\sqrt{\lambda\beta}) .$$

$$\begin{cases} s_1 = i\sqrt{\dfrac{\lambda}{\beta}} \\ s_2 = -i\sqrt{\dfrac{\lambda}{\beta}} \end{cases} \quad \alpha = 0 .$$

We conclude that $(0, 0)$ is invariably a saddle point, whereas $(\frac{2\lambda}{\gamma}, 0)$ has three cases depending on the values of α, β, λ [37],

A. a source for $\alpha \geq 2\sqrt{\lambda\beta}$,
B. a spiral source for $\alpha \in (0, 2\sqrt{\lambda\beta})$,
C. a central point for $\alpha = 0$ (KdV).

Regarding the geometric nature of the above characterizations, we have the following [37],

1. $(0, 0)$ being a saddle point, means that it's an unstable node and phase trajectories tend to move around it in hyperbolas, defined by the separatrices (i.e. straight lines directed along the two eigenvectors of the linearization matrix).
2. $((\frac{2\lambda}{\gamma}, 0) : \alpha \geq 2\sqrt{\lambda\beta})$ being a source, means that it's an unstable node from where phase trajectories diverge away without any (or relatively little) rotation.
3. $((\frac{2\lambda}{\gamma}, 0) : \alpha \in (0, 2\sqrt{\lambda\beta}))$ being a spiral source, means that it's an unstable focus where phase trajectories tend to spiral around before eventually diverge away from it.
4. $((\frac{2\lambda}{\gamma}, 0) : \alpha = 0)$ being a central point, means that the phase trajectories tend to move in ellipses around the point, describing periodic motion of a point in the phase space.

The phase plots of Fig. 1 depict our three cases, where we have set for convenience $\gamma = 1$, since it is not related to any of α, β, λ in effecting the stabillity statuses.

An essential tool in studying the phase portrait of nonlinear autonomous systems, like the above, is the Hartman–Grobman Theorem [22, 23, 37],

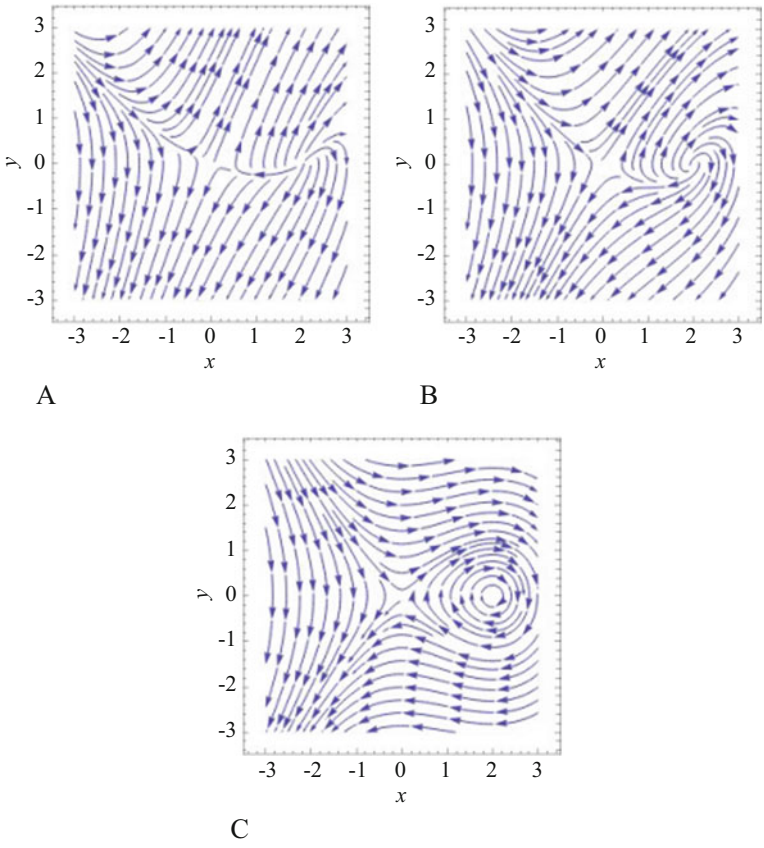


Fig. 1 The point $(0, 0)$ is invariably a saddle point whereas $(2, 0)$ is a source point in **a**, a spiral source point in **b** and a central point in **c**. (**a**) ($\alpha = 2, \beta = \lambda = 1$). (**b**) ($\alpha = \beta = \lambda = 1$). (**c**) ($\alpha = 0, \beta = \lambda = 1$)

Theorem 1 (Hartman-Grobman) *Consider a two-dimensional nonlinear autonomous system with a continuously differentiable field* $\bar{f}$*,*

$$\bar{x}' = \bar{f}(\bar{x})$$

and consider its linearization at a hyperbolic critical point $\bar{x}_0$ *(that is the Jacobian matrix has eigenvalues with non-zero real part),*

$$\bar{u}' = (Df_0)(\bar{u}).$$

Then there is a neighborhood of the hyperbolic critical point where all the solutions of the linear system can be transformed into solutions of the nonlinear system by a continuous, invertible transformation.

Remark 1 The above theorem implies that the phase portrait of the linear system in a neighborhood of the hyperbolic critical point can be transformed to the phase portrait of the nonlinear system by a continuous, invertible transformation. When that happens, we say that the two phase portraits are topologically equivalent.

Additional information about the phase plane of KdV–B equation can be found in [14].

3 Asymptotic Expansion for KdV–B

In the study of ordinary differential equations and their applications, an asymptotic expansion is of high importance. It would be very useful to understand thoroughly the property of the solution to the KdV–B equation. The asymptotic expansion would provide a reliable basis for estimating the advantages and disadvantages when seeking and applying numerical methods to our equation.

Here, by means of variable transformation and the qualitative theory of ordinary differential equations, the asymptotic behavior of the traveling wave solutions to the KdV–B equation is presented. The asymptotic expansion is real and continuous, if the argument is greater than a certain value.

The following variable transformation can be made [36, 48],

$$u = -e^{-\frac{c_1(1-k)\zeta}{2}} \frac{c_1 k^2}{c_2} y(\xi), \quad \xi = e^{-c_1 k \zeta}, \quad k = \sqrt{1 - \frac{4c_3}{c_1^2}} = \sqrt{1 + \frac{4\beta\lambda}{\alpha^2}} \geq 1. \tag{6}$$

Equation (5) ($c_0 = 0$), can be reduced to the Emden–Fowler equation [35],

$$\frac{d^2 y}{d\xi^2} = \xi^\sigma y^2, \quad \sigma = \frac{1 - 5k}{2k}. \tag{7}$$

It is obvious that,

$$\begin{cases} \sigma = -2, & \lambda = 0, \\ \sigma \in [\frac{-5}{2}, -2), & \lambda > 0. \end{cases} \tag{8}$$

Some characteristics of the KdV–B equation can be derived from equation (7).

Next we demonstrate some essential results that will help us in deriving the asymptotic expansion of KdV–B.

First, we show that the KdV–B equation, Equation (5) has finite isolated zero points only. Since y satisfies (7), y'' does not change sign for $\xi \in (0, \infty)$, so Equation (7) has finite zero points only, except that it identically vanishes for some intervals. Equation (7) has finite zero points only. This indicates that the solution of

KdV–B is consistently positive negative or zero for large arguments, which depends upon the condition of infinite point.

Our next important tool is the Integral Rule of asymptotic formulae [31, 36, 48].

Lemma 1 (Integral Rule of Asymptotic Formulae) *Let,*

$$\phi(t) \sim f(t),$$

where $f \neq 0$ *and* f *does not change in sign. Then,*

$$\begin{cases} \int_{t_0}^{t} \phi(t)dt \sim \int_{t_0}^{t} f(t)dt, & if \quad \int_{t_0}^{\infty} |f(t)|dt = \infty, \\ \int_{t}^{\infty} \phi(t)dt \sim \int_{t}^{\infty} f(t)dt, & if \quad \int_{t_0}^{\infty} |f(t)|dt < \infty. \end{cases}$$

Following the above result, we demonstrate the character of asymptotic expansion [31, 36, 48].

Lemma 2 (Character of Asymptotic Expansion) *If* $f(t) > 0$ *and* f' *is continuous and non-negative as* $t \geq t_0$*, then,*

$$f' \leq f^{1+\epsilon}$$

for any $t \geq t_0$ *and for any* $\epsilon > 0$*, except perhaps in a set of intervals of finite total length, which depends upon* ϵ*.*

The final necessary result will be Hardy's Theorem [31, 36, 48].

Theorem 2 (Hardy) *Any solution of an equation,*

$$\frac{df}{dt} = \frac{P(f,t)}{Q(f,t)},$$

which is continuous for $t \geq t_0$*, is ultimately monotonic, together with all of its derivatives, and satisfies one of the following relations,*

$$f \sim at^b e^{E(t)},$$

or

$$f \sim at^b (\ln t)^c,$$

where $E(t)$ *is a polynomial in time and* a, b, c *are constants.*

Now, all the above three results can be adopted to derive the asymptotic expansion of KdV–B, for the different values of λ.

Claim (Shu [48]) Let $\lambda > 0$. The negative asymptotic expansion of KdV–B has the following form,

$$u=-\frac{2k^2v^2U_\infty}{\delta}e^{-\frac{(k-1)v\zeta}{2\delta}}-\frac{8k^4v^2U_\infty^2}{(k-1)(3k-1)\delta}e^{-\frac{(k-1)v\zeta}{\delta}}[1+O(1)],\quad \zeta\to\infty, \tag{9}$$

where $k=\sqrt{1+\frac{4\lambda\delta}{v^2}}$ and $U_\infty>0$ is a constant.

In order to prove that, we consider $\lambda>0$ for which $\sigma\in(-\frac{5}{2},-2)$. If u has a negative asymptotic expansion then y has a positive asymptotic expansion. Since,

$$\frac{d^2y}{d\xi^2}=\xi^\sigma y^2>0,\quad \xi>0,$$

y' must be strictly monotonically increasing for $\xi>0$ and y must be a monotone function for large ξ. Thus y' has three possible cases as $\xi\to\infty$,

1. $y'\to 0$
2. $y'\to y_0'=const>0$
3. $y'\to\infty$

Let us show that case (2) cannot hold.
If

$$y'\to y_0'=const>0,$$

then,

$$y\sim y_0'\xi$$

and from equation (7),

$$y''=y^2\xi^\sigma\sim y_0'^2\xi^{\sigma+2}>\frac{1}{2}y_0'^2\xi^{\sigma+2},$$

whose integration yields,

$$y'>\frac{y_0'^2}{2(\sigma+3)}\xi^{\sigma+3}\to\infty,$$

for large ξ, which leads to a contradiction. Then it will be shown that case (3) leads to a contradiction as well.
If

$$y'\to\infty,$$

then,

$$y' > M,$$

for large ξ and some $M > 0$, and hence,

$$y > M\xi\ .$$

Reverting to equation (7),

$$y'' = \xi^{\sigma} y^2 > M^2 \xi^{\sigma+2},$$

so,

$$y > \frac{M^2}{(\sigma+3)(\sigma+4)} \xi^{\sigma+4},$$

for large ξ. Continuing in this fashion,

$$y > y_0\, \xi^5,$$

can be obtained for large ξ and the constant y_0. Hence, from equation (7),

$$y'' = \xi^{\sigma} y^2 > \sqrt{y_0}\, y^{\frac{3}{2}},$$

for large ξ. Since y' is positive,

$$y' y'' > \sqrt{y_0}\, y^{\frac{3}{2}} y',$$

whose integration yields,

$$y' > \frac{2 y_0^{\frac{1}{4}}}{\sqrt{5}} y^{\frac{3}{4}},$$

which is impossible due to Lemma (2).

Consequently, we are left with case (1). Since $y' < 0$ is strictly monotone increasing for $\xi > 0$, and y is strictly monotone decreasing for $\xi > 0$. Since $y > 0$ for large ξ, y has a finite limit $U_\infty \geq 0$ as $\xi \to \infty$. Now, let us show that $U_\infty \neq 0$. If $U_\infty = 0$, $y(\xi_0) = \delta > 0$ is set to be small. Since y is strictly monotone decreasing,

$$\delta = y(\xi_0) = \int_{\xi_0}^{\infty} \left(\int_t^{\infty} \tau^{\sigma} y^2 d\tau \right) dt < \delta^2 \int_{\xi_0}^{\infty} \left(\int_t^{\infty} \tau^{\sigma} d\tau \right) dt$$

or

$$\delta > \frac{(\sigma+1)(\sigma+2)}{\xi_0^{\sigma+2}},$$

which leads to the contradiction for δ sufficiently small.

Then let,

$$y(\infty) = U_\infty > 0, \quad y(\xi) = U_\infty + O(1), \quad \xi \to \infty\,.$$

Then

$$y'(\xi) = -\int_\xi^\infty y''dt = -\int_\xi^\infty t^\sigma y^2 dt = \frac{U_\infty^2}{\sigma+1}\xi^{\sigma+1}[1+O(1)]$$

and thus,

$$y(\xi) = U_\infty - \int_\xi^\infty y'dt = U_\infty + \frac{U_\infty^2}{(\sigma+1)(\sigma+2)}\xi^{\sigma+2}[1+O(1)]\,.$$

The latter proves our claim.

Claim (Shu [48]) Let $\lambda = 0$. The negative asymptotic expansion of the KdV–B equation has the following form,

$$u = -\frac{2kv}{\zeta}e^{\frac{-(k-1)v\zeta}{2\delta}}, \quad \zeta \to \infty, \tag{10}$$

where $k = \sqrt{1+\frac{4\lambda\delta}{v^2}}$.

For the proof, we consider $\lambda = 0$, which gives us $\sigma = -2$. If u has a negative asymptotic expansion, y has a positive asymptotic expansion. Let $\xi = e^s$, obtaining from equation (7),

$$\frac{d^2y}{ds^2} - \frac{dy}{ds} - y^2 = 0. \tag{11}$$

If $\frac{dy}{ds} = 0$ at s_0, then,

$$\frac{d^2y}{ds^2} = y^2 > 0$$

and y can only have a minimum at s_0. Hence, y is a monotone function for large ξ. Thus y has three possible cases as $s \to \infty$:

1. $y \to 0$

2. $y \to y_0 = const > 0$
3. $y \to \infty$

Let us show that case (2) cannot hold.
If

$$y \to y_0 = const > 0,$$

then

$$\frac{d^2y}{ds^2} - \frac{dy}{ds} \sim y^2 .$$

Integrating, we get,

$$\frac{dy}{ds} - y \sim y_0^2 s .$$

Since $y \to y_0$, this implies,

$$\frac{dy}{ds} \sim y_0^2 s$$

from which,

$$y \sim \frac{1}{2} y_0^2 s^2,$$

which contradicts $y \to y_0$. Next, we will show that case (3) is impossible.
If $y \to \infty$, let,

$$p = \frac{dy}{ds}.$$

Then equation (11) becomes,

$$p\frac{dp}{dy} - p - y^2 = 0. \tag{12}$$

Since $y \to y_0$, we have,

$$p\frac{dp}{dy} > 0 .$$

Now, Theorem (2) indicates that p has two possible cases for large y,

1. $p \sim ay^b e^{E(y)}$,
2. $p \sim ay^b (\ln y)^c$,

where $E(y)$ is a polynomial in y and $a > 0$, b, c are constants. We also show that case (1) is impossible.

If $E(y) \to -\infty$, then,

$$p \to 0, \quad \frac{dp}{dy} \to 0$$

which leads to a contradiction by referring to equation (12).

If $E(y) \to \infty$, then,

$$p > y^2,$$

for large y, which contradicts Lemma (2). Hence,

$$E(y) = const\ .$$

If $b > 1$, then,

$$p > y^{\frac{b+1}{2}},$$

for large y, which is impossible due to Lemma (2).
If $b \leq 1$, then,

$$p\frac{dp}{dy} \sim y^2,$$

is obtained from equation (12). By integration,

$$\frac{1}{2}p^2 \sim \frac{1}{3}y^3,$$

is obtained, so that $b = \dfrac{3}{2} > 1$, which leads to a contradiction. Let us now show that case (2) is also impossible.

If $b > 1$, then,

$$p > y^{\frac{b+1}{2}},$$

for large y, which is impossible due to Lemma (2).

If $b \leq 1$, then,

$$\frac{1}{2}p^2 \sim \frac{1}{3}y^3,$$

is obtained, so that $b = \dfrac{3}{2} > 1$, which leads to a contradiction.

Consequently, we are left with case (1), where $y \to 0$. Let $v = \frac{1}{y}$ and $w = \frac{dv}{ds}$, obtaining from equation (12),

$$w\frac{dw}{dv} - \frac{2w^2}{v} - w + 1 = 0. \tag{13}$$

Since $y \to 0$, $v \to \infty$ and $\frac{dv}{ds} < 0$, we have,

$$w = \frac{dv}{ds} = -\frac{1}{y^2}\frac{dy}{ds} > 0,$$

is obtained. Theorem (2) indicates that w has two possible cases for large v,

1. $w \sim av^b e^{E(v)}$,
2. $w \sim av^b(\ln v)^c$,

where $E(v)$ is a polynomial in v and $a > 0$, b, c are constants. It is now shown that if case (3) is satisfied, $E(v) = const$ and $b = 0$. Similar to above, $E(v) = const$ and $b \leq 1$.

If $b = 1$, then,

$$\frac{dw}{dv} \sim a > 0.$$

From equation (13), $a = -1$ is obtained, which leads to a contradiction.

If $b \in (0, 1)$, then,

$$\frac{dw}{ds} \sim 1,$$

is obtained from equation (13). By integrating, we get,

$$w \sim v,$$

so that $b = 1$, which also leads to a contradiction.

If $b < 0$, then,

$$w\frac{dw}{dv} \sim -1,$$

is obtained from equation (13). By integrating, we get,

$$\frac{1}{2}w^2 \sim -v,$$

which leads to a contradiction.

Let us now show that if case (2) is satisfied, then $b = c = 0$. Similar to above, either $b = 1, c \neq 0$ or $b = 0$.
If $b = 1, c < 0$ or $c > 0$, then,

$$\frac{dw}{dv} \sim 1 \quad or \quad \frac{dw}{dv} \sim \frac{2w}{v},$$

is obtained from equation (13). By integrating, we get,

$$w \sim v \quad or \quad w \sim v^2,$$

is obtained, so that $c = 0$, which leads to a contradiction. Hence $b = 0$.
If $c < 0$, then,

$$w\frac{dw}{dv} \sim -1,$$

is obtained from equation (13). By integrating, we get,

$$\frac{1}{2}w^2 \sim -v,$$

is obtained, which leads to a contradiction.

If $c > 0$, then,

$$\frac{dw}{dv} \sim 1,$$

is obtained from equation (13). By integrating, we get,

$$w \sim v,$$

so that $c = 0$, which leads to a contradiction.

Summing up and from equation (13), we get,

$$w \sim 1,$$

so that,

$$\frac{dv}{ds} \sim 1, \quad s \to \infty .$$

By integrating, we finally obtain,

$$v \sim s,$$

as $s \to \infty$, so that,

$$y \sim \frac{1}{\ln \xi}, \quad \xi \to \infty .$$

That proves our claim.

Claim (Shu [48])
Let $\lambda > 0$. The negative asymptotic expansion of KdV–B can be written in the form (see Fig. 2),

$$u = -\frac{2k^2v^2U_\infty}{\delta} e^{\frac{-(k-1)v\zeta}{2\delta}} - \frac{2k^4v^2}{\delta} \sum_{i=1}^{\infty} \frac{(2U_\infty)^{i+1} e^{-\frac{(i+1)(k-1)v\zeta}{2\delta}}}{\prod_{j=1}^{i} [j(k-1)+2k]\, j(k-1)}, \quad \zeta \to \infty, \tag{14}$$

where $k = \sqrt{1 + \frac{4\lambda\delta}{v^2}}$ and $U_\infty > 0$ is a constant.

To prove the latter, we first notice that since,

$$e^{-\frac{(i+1)(k-1)v\zeta}{2\delta}}$$

exists, the infinite series converges. Let,

$$u_m = -\frac{2k^2v^2U_\infty}{\delta} e^{\frac{-(k-1)v\zeta}{2\delta}} - \frac{2k^4v^2}{\delta} \sum_{i=1}^{\infty} \frac{(2U_\infty)^{i+1} e^{-\frac{(i+1)(k-1)v\zeta}{2\delta}}}{\prod_{j=1}^{i} [j(k-1)+2k]\, j(k-1)}$$

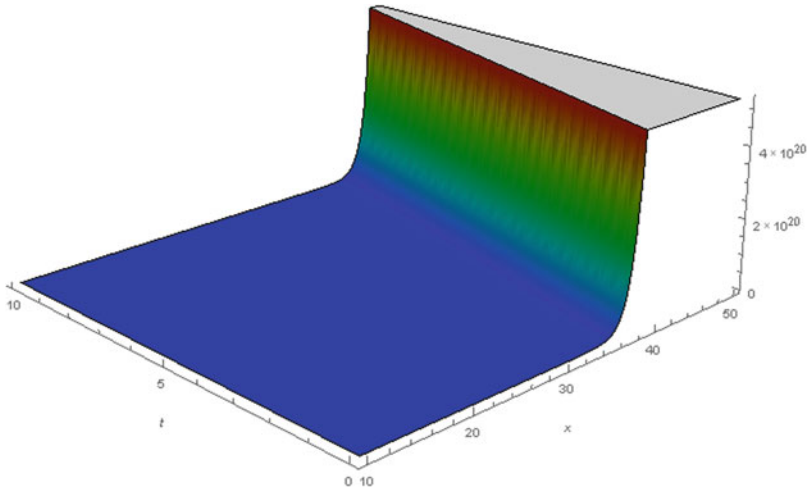


Fig. 2 The graph shows the asymptotic expansion of the KdV–B equation (see Equation (14)) for the parameters, $\alpha = 0.1$, $\beta = 0.7$, $\gamma = 1$, $\lambda = 1$, $U_\infty = 1$ and $\sigma = -2.25$. The shock wave characteristics can be observed

and

$$y_m = U_\infty + \sum_{i=1}^{m} \frac{2^{i-1} u_\infty^{1+i} \xi^{i(\sigma+2}}{\prod_{j=1}^{i} [j(\sigma+2)-1]\, j(\sigma+2)}\,.$$

Then,

$$y_{m+1}[1+O(1)] = U_\infty + \int_\xi^\infty \left(\int_t^\infty \tau^\sigma y_m^2 [1+O(1)]^2 d\tau \right) dt$$

can be obtained for an arbitrary integer m. Since $u_m \to u$ as $m \to \infty$, we get,

$$y_m \to y_\infty, \quad m \to \infty,$$

so that,

$$y_\infty = U_\infty + \int_\xi^\infty \left(\int_t^\infty \tau^\sigma y_\infty^2 d\tau \right) dt$$

and y_∞ is the positive asymptotic expansion of equation (7).

4 Hyperbolic Methods for Traveling Wave Solutions of KdV–B

Since the late 1980s, various methods for seeking explicit exact solutions to the KdV–B equation have been independently proposed by many mathematicians, engineers and physicists. The first analytical traveling wave solution to the Burgers-KdV equation was obtained by Xiong [56] in 1989. Two different methods for the construction of exact solutions to the KdV–B equation were proposed by Jeffrey and Mohamad [26]. Wang [52] applied the homogeneous balance method to the study of exact solutions of the compound KdV–B equation. Demiray [13] proposed a so-called "hyperbolic tangent approach" for finding the exact solution to the KdV–B equation, which is actually the Parkes and Duffy's automated method [39, 40]. Recently, Feng [16–18] introduced the first-integral method to study the exact solution of KdV–B, which is based on the ring theory of commutative algebra. The Cauchy problem for the KdV–B equation was investigated by Bona and Schonbek [7]. They proved the existence and uniqueness of bounded traveling wave solutions which tend to constant states at plus and minus infinity.

We focus on deriving traveling wave solutions for KdV–B, using the Tanh and Sech methods [21, 24, 32, 33, 53]. We start with the Tanh method, considering,

$$-\lambda u + \gamma \frac{u^2}{2} - \alpha \frac{du}{d\zeta} + \beta \frac{d^2u}{d\zeta^2} = 0, \tag{15}$$

where λ is the wave velocity, assuming that both our solution and its spatial derivatives vanish at either plus or minus infinity.

The Tanh method uses a finite series,

$$u(x,t) = u(\mu\zeta) = s(y) = \sum_{m=0}^{M} a_m y^m, \tag{16}$$

where μ is the wave number, inversely proportional to the width of the wave, and M is a positive integer, in most cases, that will be determined. However if M is not an integer, a transformation formula is usually used. Substituting equation (16) into equation (15) yields an equation in powers of y.

To determine the parameter M, we usually balance the linear terms of highest order in the resulting equation with the highest order nonlinear terms. With M determined, we collect all coefficients of powers of y in the resulting equation where these coefficients have to vanish.

This will give a system of algebraic equations involving the parameters a_m, $m = 0, \ldots, M$, μ and λ. Having determined these parameters, knowing that M is a positive integer in most cases, and using equation (16), we obtain an analytic solution in a closed form. We introduce,

$$y = tanh(\mu\zeta), \tag{17}$$

that leads to the change of derivatives,

$$\begin{cases} \dfrac{d}{d\zeta} = \dfrac{d}{dy}\dfrac{dy}{dz} = \mu(1-y^2)\dfrac{d}{dy}, \\ \dfrac{d^2}{d\zeta^2} = \dfrac{d}{d\zeta}\dfrac{d}{d\zeta} = \mu^2(1-y^2)\left(-2y\dfrac{d}{dy} + (1-y^2)\dfrac{d^2}{dy^2}\right). \end{cases} \tag{18}$$

Therefore, by replacing equation (16) in equation (15) and using equation (18), we derive an equation with respect to u as follows,

$$-\lambda\left(\sum_{m=0}^{M} a_m y^m\right) + \frac{\gamma}{2}\left(\sum_{m=0}^{M} a_m y^m\right)^2 - \alpha\mu(1-y^2)\frac{d}{dy}\left(\sum_{m=0}^{M} a_m y^m\right)$$
$$+ \beta\mu^2(1-y^2)\left(-2y\frac{d}{dy}\left(\sum_{m=0}^{M} a_m y^m\right) + (1-y^2)\frac{d^2}{dy^2}\left(\sum_{m=0}^{M} a_m y^m\right)\right) = 0\,.$$

To determine M, we follow the procedure described above to get $M = 2$. This gives the solution in the form,

$$s = \sum_{m=0}^{M} a_m y^m = a_0 + a_1 y + a_2 y^2, \quad a_2 \neq 0. \tag{19}$$

Substituting equation (19) into equation (15), we get,

$$-\lambda(a_0 + a_1 y + a_2 y^2) + \frac{\gamma}{2}(a_0 + a_1 y + a_2 y^2)^2 - \alpha\mu(1 - y^2)(a_1 + 2a_2 y) + \beta\mu^2(1 - y^2)\Big(-2y(a_1 + 2a_2 y) + 2a_2(1 - y^2)\Big) = 0.$$

Collecting the coefficients of different powers of y, gives the following system of algebraic equations for λ μ, a_0, a_1 and a_2,

$$\begin{cases} a_2(\gamma a_2 + 12\beta\mu^2) = 0 \\ \gamma a_1 a_2 + 2\beta\mu^2 a_1 + 2\alpha\mu a_2 = 0 \\ -\lambda a_2 - 8\beta\mu^2 a_2 + \gamma a_0 a_2 + \frac{\gamma}{2} a_1^2 + \alpha\mu a_1 = 0 \\ -\lambda a_1 - 2\beta\mu^2 a_1 + \gamma a_0 a_1 - 2\alpha\mu a_2 = 0 \\ -\lambda a_0 + \frac{\gamma}{2} a_0^2 - \alpha\mu a_1 + 2\beta\mu^2 a_2 = 0 \end{cases}$$

with solution,

$$\begin{cases} \lambda = \pm\frac{6}{25}\frac{\alpha^2}{\beta}, \\ \mu = \pm\frac{\alpha}{10\beta}, \\ a_0 = \frac{\lambda}{\gamma} + 12\frac{\beta\mu^2}{\gamma}, \\ a_1 = -\frac{12}{5}\frac{\alpha\mu}{\gamma}, \\ a_2 = -12\frac{\beta\mu^2}{\gamma}. \end{cases} \tag{20}$$

Using the trigonometric identities,

$$\begin{cases} tanh^2(\theta) = 1 - sech^2(\theta) \\ tanh(-\theta) = -tanh(\theta) \end{cases}, \quad \theta \in \mathbb{R}$$

and requiring for both our solution and its spatial derivatives to vanish at plus infinity, we get the following traveling wave solution,

$$u_{1\,\infty}(\zeta) = \frac{3}{25}\frac{\alpha^2}{\beta\gamma}\Big(sech^2(\mu\zeta) - 2tanh(\mu\zeta) + 2\Big), \quad \mu, \lambda > 0 \tag{21}$$

Requiring for both our solution and its spatial derivatives to vanish at minus infinity, we get the following traveling wave solution,

$$u_{2\,-\infty}(\zeta) = \frac{3}{25}\frac{\alpha^2}{\beta\gamma}\Big(sech^2(\mu\zeta) - 2tanh(\mu\zeta) - 2\Big), \quad \mu > 0, \quad \lambda < 0 \tag{22}$$

Remark 2 A notable result is that our traveling wave solutions are expressed as a composition of a bell-profile solitary wave (KdV) and a kink-profile solitary wave (Burgers') with velocity $\lambda = \pm\frac{6}{25}\frac{\alpha^2}{\beta}$. The shock profile is dominant here. All those exact solutions, and others mentioned in literature, can be proved to be algebraically equivalent to each other [18]. That is, essentially only one explicit traveling solitary wave solution to the KdV–B equation is known which can be expressed as a composition of a bell-profile solitary wave and a kink-profile solitary wave. In other words, a feature of this solution is that is a linear combination of particular solutions of the KdV equation and the Burgers equation [18, 26].

By following similar steps as with the Tanh method for traveling wave solutions of KdV–B, the Sech method uses the variable transformation [21, 24, 53],

$$\begin{cases} y = sech(\mu\zeta), \quad \mu \neq 0, \\ \dfrac{d}{d\zeta} = -\mu y\sqrt{1-y^2}\dfrac{d}{dy}, \\ \dfrac{d^2}{d\zeta^2} = \mu^2 y[(1-2y^2)\dfrac{d}{dy} + (y-y^3)\dfrac{d^2}{dy^2}] \end{cases} \tag{23}$$

and the ansatz,

$$u(x,t) = u(\mu\zeta) = s(y) = \sum_{m=0}^{2} a_m y^m. \tag{24}$$

Then by replacing equation (24) in equation (15) and using equation (23), we derive an equation with respect to u as follows,

$$\begin{aligned} [a_2(\gamma\frac{a_2}{2} - \mu\alpha - 6\beta\mu^2)]y^4 + [a_1(\gamma a_2 - \frac{\alpha\mu}{2} - 2\beta\mu^2)]y^3 \\ + [-\lambda a_2 + \frac{\gamma}{2}a_1^2 + \gamma a_0 a_2 + 2\alpha\mu a_2 \\ + 4\beta\mu^2 a_2]y^2 + [a_1(-\lambda + \gamma a_0 + \alpha\mu + \beta\mu^2)]y \\ + \frac{\gamma}{2}a_0^2 - \lambda a_0 = 0. \end{aligned}$$

Above, we used the first order Taylor approximation,

$$\sqrt{1-y^2} \sim 1 - \frac{y^2}{2}.$$

Collecting the coefficients of different powers of y, gives the following system of algebraic equations for λ μ, a_0, a_1 and a_2,

$$\begin{cases} a_2(\gamma\dfrac{a_2}{2} - \mu\alpha - 6\beta\mu^2) = 0 \\ a_1(\gamma a_2 - \dfrac{\alpha\mu}{2} - 2\beta\mu^2) = 0 \\ -\lambda a_2 + \dfrac{\gamma}{2}a_1^2 + \gamma a_0 a_2 + 2\alpha\mu a_2 + 4\beta\mu^2 a_2 \\ a_1(-\lambda + \gamma a_0 + \alpha\mu + \beta\mu^2) = 0 \\ \dfrac{\gamma}{2}a_0^2 - \lambda a_0 = 0 \end{cases}$$

with a solution being,

$$\begin{cases} \lambda = 2\mu(\alpha + 2\beta\mu) \\ a_0 = \dfrac{4\mu}{\gamma}(\alpha + 2\beta\mu) \\ a_1 = 0 \\ a_2 = \dfrac{2\mu}{\gamma}(\alpha + 6\beta\mu) \end{cases}, \quad \mu \neq 0 \tag{25}$$

giving a solitary profile traveling wave solution,

$$\frac{2\mu}{\gamma}\left(2(\alpha + 2\beta\mu) + (\alpha + 6\beta\mu)sech^2(\mu\zeta)\right), \quad \mu \neq 0. \tag{26}$$

Remark 3 A notable result is that the Sech method can give "purely" solitary profile traveling wave solutions.
The following graph, Fig. 3, depicts the solutions studied in this section.

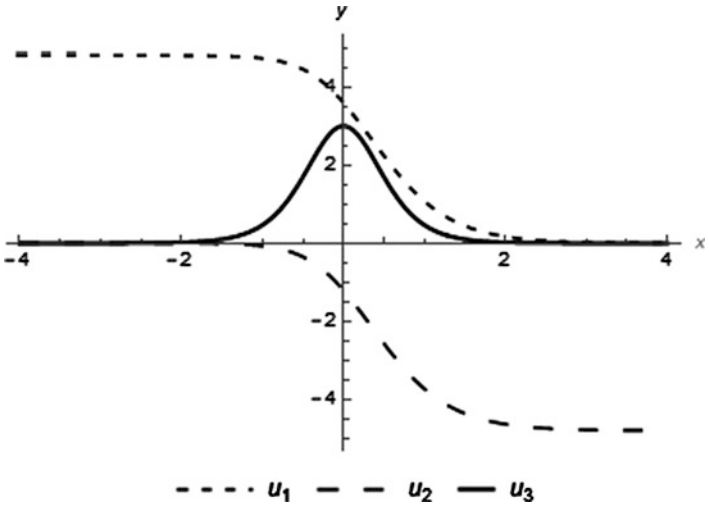


Fig. 3 u_1 and u_2 are two shock profile traveling waves of the KdV–B equation vanishing at plus and minus infinity, respectively, whereas u_3 is a solitary profile traveling wave solution of the KdV equation, for the parameters, $\alpha = 1$, $\beta = 0.1$, $\gamma = 1$, $\lambda = 1$

5 Spectral Fourier Analysis for the Numerical Solution of KdV–B

We define the Fourier and Inverse Fourier Transform of a function, say f, in the sense that the following symbols make sense, to be [2, 31],

$$\begin{cases} F[f(x)] = \hat{f}(k) = \int_{-\infty}^{\infty} e^{-ikx} f(x) dx, \\ F^{-1}[\hat{f}(k)] = f(x) = \dfrac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} \hat{f}(k) dk. \end{cases} \tag{27}$$

It is easy to see, integrating by parts, that regarding the nth derivative of f and its Fourier Transform, the following results hold,

$$\begin{cases} \dfrac{d^n f}{dx^n} = (i^n) F^{-1}\left[k^n \hat{f}(k)\right], \\ \dfrac{d^n \hat{f}}{dx^n} = (-i)^n F\left[x^n f(x)\right]. \end{cases} \tag{28}$$

Now consider the KdV–B equation,

$$\frac{\partial u}{\partial t} + \gamma u \frac{\partial u}{\partial x} - \alpha \frac{\partial^2 u}{\partial x^2} + \beta \frac{\partial^3 u}{\partial x^3} = 0, \quad u = u(t, x). \tag{29}$$

Rearranging the terms of equation (29), we get,

$$\frac{\partial u}{\partial t} = -\gamma u \frac{\partial u}{\partial x} + \alpha \frac{\partial^2 u}{\partial x^2} - \beta \frac{\partial^3 u}{\partial x^3}. \tag{30}$$

By means of the Inverse Fourier Transform, F^{-1}, equation (30) can be written in the form [2, 50],

$$\frac{\partial u}{\partial t} = f(t, u), \tag{31}$$

where we have substituted the x-partial derivatives with,

$$\begin{cases} \dfrac{\partial u}{\partial x} = i F^{-1}(\kappa \hat{u}), \\ \dfrac{\partial^2 u}{\partial x^2} = -F^{-1}(\kappa^2 \hat{u}), \\ \dfrac{\partial^3 u}{\partial x^3} = -i F^{-1}(\kappa^3 \hat{u}). \end{cases}$$

Now, equation (31) is suitable for applying the 4th order explicit Runge–Kutta method, giving us the following,

$$\begin{cases} u_{n+1} = u_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ t_{n+1} = t_n + h \\ k_1 = f(t_n, u_n) \\ k_2 = f(t_n + \frac{h}{2}, u_n + \frac{hk_1}{2}) \\ k_3 = f(t_n + \frac{h}{2}, u_n + \frac{hk_2}{2}) \\ k_4 = f(t_n + h, u_n + hk_3) \end{cases}, \quad n = 0, 1, \ldots \tag{32}$$

For $n = 0$, we may choose either a soliton of the KdV equation or a similarity solution of the viscous Burgers equation or a traveling wave solution of KdV–B.

Below we exhibit the obtained numerical results for each case separately. In Fig. 4, the evolution of an initial solitary profile solution of the KdV equation is depicted, where diffusive effects are absent. It can be observed that the solitary waveform is retained. In Fig. 5, the evolution of a solitary profile solution of the

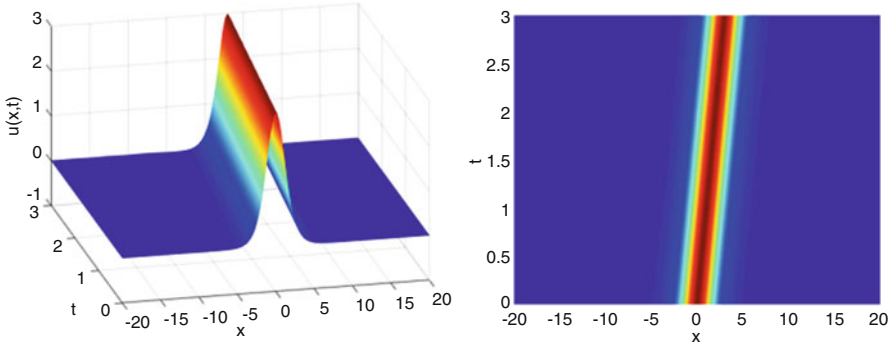


Fig. 4 A solution of the KdV–B equation, evolving a solitary profile solution of the KdV equation, where diffusive effects are absent (KdV case), for the parameters, $\lambda = 1$, $\alpha = 0$, $\beta = 0.7$, $\gamma = 1$

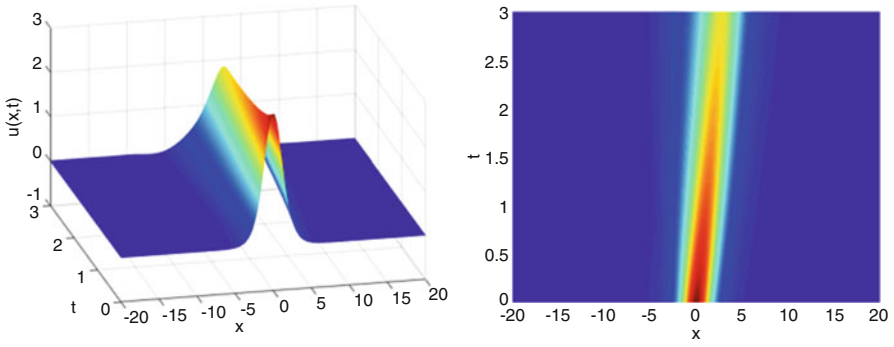


Fig. 5 A solution of the KdV–B equation, evolving a solitary profile solution of the KdV equation, where both diffusive and dispersive effects coexist, for the parameters, $\lambda = 1$, $\alpha = 0.5$, $\beta = 0.7$, $\gamma = 1$

KdV equation is presented, where both diffusive and dispersive effects coexist. It is observed that the profile loses energy and reduces in amplitude drastically. Additionally, in Fig. 6 we present the evolution of a similarity shock profile solution of the viscous Burgers equation, where both diffusive and dispersive effects coexist. In this case a wavefront can be observed revealing a shock-like behavior. Finally, in Fig. 7, the evolution of a traveling shock wave profile solution of the KdV–B equation is presented, where both diffusive and dispersive effects coexist. These numerical solutions clearly reveal both solitary and shock wave features of the KdV–B equation, revealing its connection to cardiac hemodynamics where all these phenomena, such as convection, diffusion and dispersion, can be observed.

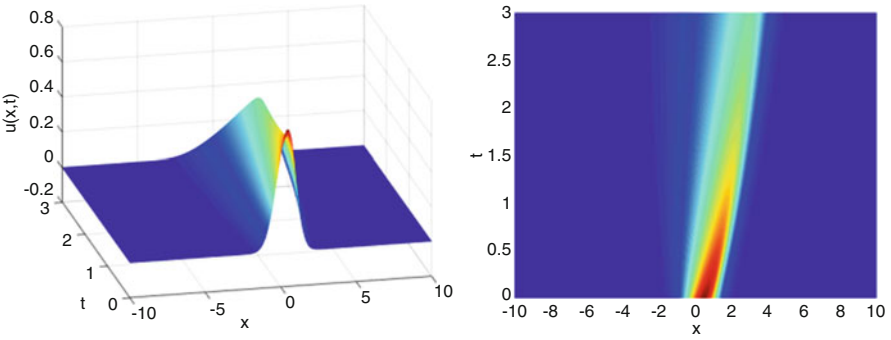


Fig. 6 A solution of the KdV–B equation, evolving a similarity shock profile solution of the viscous Burgers equation, where both diffusive and dispersive effects coexist, for the parameters, $\lambda = 1.8, \alpha = 0.19, \beta = 0.01, \gamma = 3.4$

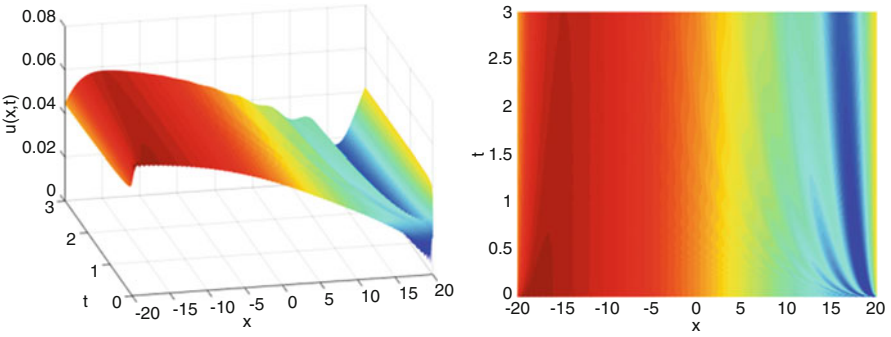


Fig. 7 A solution of the KdV–B equation, evolving a traveling wave shock profile solution of the KdV–B equation, where both diffusive and dispersive effects coexist, for the parameters, $\alpha = 0.3, \beta = 0.7, \gamma = 1$

6 Conclusions

Recent advancements concerning cardiac dynamics pose important questions about the cardiac waveform. A governing equation, namely the KdV–B equation (Korteweg–de Vries–Burgers), which is a partial differential equation can be utilized to answer several of those questions. The KdV–B equation features both solitary and shock wave characteristics due to the dispersion and dissipation terms, as also occurring in the arterial tree. This study focuses on describing cardiac dynamics with the applications of mathematics and nonlinear analysis. It is customarily difficult to solve nonlinear problems, especially by analytical techniques. Therefore, seeking suitable solving methods, such as, exact, approximate or numerical methods, is an active task in branches of applied mathematics and nonlinear analysis.

In this chapter, the phase plane of the KdV–B equation is analyzed and its qualitative behavior is derived, depicting the stability states of the equation contributing to the decisions made for further analytical and numerical consideration. The analysis reveals a saddle point (0, 0), and an additional one that could be a source point or a spiral source point or a central point depending on the equation's parameters, α, β and λ.

Furthermore, an asymptotic expansion is presented, providing a reliable basis for estimating the advantages and disadvantages when seeking and applying numerical methods to KdV–B equation. Furthermore, traveling wave solutions under both solitary and shock profiles are obtained from the hyperbolic methods, whose strength is their ease of use to find which solitary wave structures and/or shock-wave (kinks) profiles satisfy nonlinear wave and evolution equations. These techniques allow to develop algorithms for symbolic software packages, so that nonlinear partial differential equations and difference equations, can be studied automatically whether (or not) they possess traveling wave solutions. Additionally, numerical solutions are obtained for the equation, by means of the Spectral Fourier analysis. Both these solutions and the latter traveling wave solutions are evolved in time by the Runge–Kutta method. These solutions clearly depict both solitary and shock wave characteristics of the KdV–B equation. This analysis provides vital information about the equation and its connection to cardiac hemodynamics.

References

1. M.J. Ablowitz, H. Segur, *Solitons and the Inverse Scattering Transform*. (SIAM, Philadelphia, 1981)
2. M.J. Ablowitz, D.J. Kaup, A.C. Newell, H. Segur, The inverse scattering transform-fourier analysis for nonlinear problems. Stud. Appl. Math. **53**(4), 249–315 (1974)
3. M. Anliker, M.B. Histand, E. Ogden, Dispersion and attenuation of small artificial pressure waves in the canine aorta. Circ. Res. **23**(4), 539–551 (1968)
4. N. Antar, H. Demiray, Weakly nonlinear waves in a prestressed thin elastic tube containing a viscous fluid. Int. J. Eng. Sci. **37**(14), 1859–1876 (1999)

5. H. Atabek, Wave propagation through a viscous fluid contained in a tethered, initially stressed, orthotropic elastic tube. Biophys. J. **8**(5), 626–649 (1968)
6. D. Benney, Long waves on liquid films. J. Math. Phys. **45**(1–4), 150–155 (1966)
7. J. Bona, M. Schonbek, Travelling-wave solutions to the Korteweg-de Vries-Burgers equation. Proc. Roy. Soc. Edinburgh Sect. A **101**, 207–226 (1985)
8. R.K. Bullough, P.J. Caudrey, *Solitons*, vol. 17 (Springer Science & Business Media, Berlin, 2013)
9. J. Burgers, *Correlation Problems in a One-dimensional Model of Turbulence II* (North-Holland Publishing, Amsterdam, 1950)
10. J. Canosa, J. Gazdag, The korteweg-de vries-burgers equation. J. Comput. Phys. **23**(4), 393–403 (1977)
11. E. Crépeau, M. Sorine, A reduced model of pulsatile flow in an arterial compartment. Chaos Solitons Fractals **34**(2), 594–605 (2007)
12. H. Demiray, Wave propagation through a viscous fluid contained in a prestressed thin elastic tube. Int. J. Eng. Sci. **30**(11), 1607–1620 (1992)
13. H. Demiray, A note on the exact travelling wave solution to the KdV–burgers equation. Wave motion **38**(4), 367–369 (2003)
14. G.A. El, M.A.Hoefer, Dispersive shock waves and modulation theory. Phys. D Nonlin. Phenom. **333**, 11–65 (2016)
15. H. Erbay, S. Erbay, S. Dost, Wave propagation in fluid filled nonlinear viscoelastic tubes. Acta Mech. **95**(1–4), 87–102 (1992)
16. Z. Feng, The first-integral method to study the Burgers–Korteweg–de Vries equation. J. Phys. A Math. Gen. **35**(2), 343 (2002)
17. Z. Feng, On explicit exact solutions to the compound Burgers–KdV equation. Phys. Lett. A **293**(1–2), 57–66 (2002)
18. Z. Feng, Qualitative analysis and exact solutions to the Burgers-Korteweg-de vries equation. Dyn. Contin. Discret. Impul. Syst. Ser. A Math. Anal. **9**(4), 563–580 (2002)
19. F. Feudel, H. Steudel, Non-existence of prolongation structure for the Korteweg-de Vries-Burgers equation. Phys. Lett. A **107**(1), 5–8 (1985)
20. G. Gao, A theory of interaction between dissipation and dispersion of turbulence. Sci. Sin. (Ser. A) Math. Phys. Astronom. Tech. Sci. **28**, 616–627 (1985)
21. Y.T. Gao, B. Tian, Generalized hyperbolic-function method with computerized symbolic computation to construct the solitonic solutions to nonlinear equations of mathematical physics. Comput. Phys. Commun. **133**(2-3), 158–164 (2001)
22. D.M. Grobman, Homeomorphism of systems of differential equations. Dokl. Akad. Nauk SSSR **128**(5), 880–881 (1959)
23. P. Hartman, A lemma in the theory of structural stability of differential equations. Proc. Am. Math. Society **11**(4), 610–620 (1960)
24. R. Hirota, Direct method of finding exact solutions of nonlinear evolution equations, in *Bäcklund Transformations, the Inverse Scattering Method, Solitons, and Their Applications* (Springer, Berlin, 1976), pp. 40–68
25. A. Jeffrey, T. Kakutani, Weak nonlinear dispersive waves: a discussion centered around the Korteweg–de Vries equation. SIAM Rev. **14**(4), 582–643 (1972)
26. A. Jeffrey, M. Mohamad, Exact solutions to the kdv-burgers' equation. Wave Motion **14**(4), 369–375 (1991)
27. R. Johnson, Shallow water waves on a viscous fluid the undular bore. Phys. Fluid. **15**(10), 1693–1699 (1972)
28. T. Kawahara, Weak nonlinear magneto-acoustic waves in a cold plasma in the presence of effective electron-ion collisions. J. Phys. Soc. Jpn. **28**(5), 1321–1329 (1970)
29. D.J. Korteweg, G. De Vries, Xli. on the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. London, Edinburgh, Dublin Philos. Mag. J. Sci. **39**(240), 422–443 (1895)
30. S.D. Liu, S.K. Liu, KdV-burgers equation modelling of turbulence. Sci. China Ser. A Math. Phys. Astron. Technol. Sci. **35**(5), 576–586 (1992)

31. J.D. Logan, *Applied Mathematics* (John Wiley & Sons, Hoboken, 2013)
32. W. Malfliet, Solitary wave solutions of nonlinear wave equations. Am. J. Phys. **60**(7), 650–654 (1992)
33. W. Malfliet, W. Hereman, The tanh method: I exact solutions of nonlinear evolution and wave equations. Phys. Scripta **54**, 563–568 (1996)
34. A. Mangel, M. Fahim, C. Van Breemen, Control of vascular contractility by the cardiac pacemaker. Science **215**(4540), 1627–1629 (1982)
35. B. Mehta, R. Aris, A note on a form of the emden-fowler equation. J. Math. Anal. Appl. **36**(3), 611–621 (1971)
36. J.D. Murray, *Asymptotic Analysis*, vol. 48 (Springer Science & Business Media, Berlin, 2012)
37. G. Nagy, Ordinary Differential Equations. Online notes (2020)
38. M. Otwinowski, R. Paul, J. Tuszynski, An answer to the question: is the arterial pulse a soliton? J. Biol. Phys. **14**(2), 43–48 (1986)
39. E. Parkes, B, Duffy, An automated tanh-function method for finding solitary wave solutions to non-linear evolution equations. Comput. Phys. Commun. **98**(3), 288–300 (1996)
40. E. Parkes, B. Duffy, Travelling solitary wave solutions to a compound KdV-Burgers equation. Phys. Lett. A **229**(4), 217–220 (1997)
41. T.J. Pedley, Y. Lx, *Fluid Mechanics of Large Blood Vessels* (Shaanxi People's Press, Beijing, 1995)
42. A.I. Rachev, Effects of transmural pressure and muscular activity on pulse waves in arteries. J. Biomech. Eng., 119–123 (1980)
43. R.L. Rockwell, *Nonlinear Analysis of Pressure and Shock Waves in Blood Vessels. Ph. D* (Stanford Univ. CA, US, NASA Archives, Thesis, 1969)
44. S. Rowlands, Is the arterial pulse a soliton? J. Biol. Phys. **10**(4), 199–200 (1982)
45. G. Rudinger, Review of current mathematical methods for the analysis of blood flow, in *Biomedical Fluid Mechanics Symposium* (ASME, New York, 1966), pp. 1–33
46. Y. Shi, P. Lawford, R. Hose, Review of zero-d and 1-d models of blood flow in the cardiovascular system. Biomed. Eng. Online **10**(1), 33 (2011)
47. J.J. Shu, Exact n-envelope-soliton solutions of the hirota equation. Preprint. arXiv:1403.3645 (2014)
48. J.J. Shu, The proper analytical solution of the korteweg-de vries-burgers equation. Preprint. arXiv:1403.3636 (2014)
49. R. Skalak, Wave propagation in blood flow, in *Biomechanics Symposium* (ASME, New York, 1966), pp. 20–46
50. L.N. Trefethen, *Spectral Methods in Matlab, Volume 10 of Software, Environments, and Tools*, vol. 24 (Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2000)
51. M. Wadati, Wave propagation in nonlinear lattice. I. J. Phys. Society Jpn. **38**(3), 673–680 (1975)
52. M. Wang, Exact solutions for a compound KdV-Burgers equation. Phys. Lett. A **213**(5-6), 279–287 (1996)
53. A.M. Wazwaz, The tanh–coth and the sech methods for exact solutions of the jaulent–miodek equation. Phys. Lett. A **366**(1–2), 85–90 (2007)
54. L.V. Wijngaarden, One-dimensional flow of liquids containing small gas bubbles. Ann. Rev. Fluid Mech. **4**(1), 369–396 (1972)
55. J.R. Womersley, WADC technical report tr 56-614 and phil. Mag **46**, 199–221 (1955)
56. S. Xiong, An analytic solution of Burgers-KdV equation. Chin. Sci. Bull. **34**(14), 1158–1162 (1989)
57. S. Yomosa, Solitary waves in large blood vessels. J. Phys. Society Jpn. **56**(2), 506–520 (1987)